Aquaculture Reports 21 (2021) 100877

Contents lists available at ScienceDirect

Aquaculture Reports

journal homepage: www.elsevier.com/locate/aqrep

Association of wastewater determinants with fish hematological and plasma biochemical responses: Multivariate analysis approach

Natalija Topić Popović [a,b,*], Jasenka Gajdoš Kljusurić [c], Ivančica Strunjak-Perović [a,b], Josip Barišić [d], Dušan Palić [e], Blanka Beer Ljubić [f], Sanja Babić [a,b], Rozelindra Čož-Rakovac [a,b]

[a] Laboratory for Aquaculture Biotechnology, Ruđer Bošković Institute, Bijenička 54, Zagreb, Croatia
[b] Scientific Centre of Excellence for Marine Bioprospecting-BioProCro, Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, Croatia
[c] Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, Zagreb, Croatia
[d] School of Health and Life Sciences, University of the West of Scotland, Paisley PA1 2BE, Scotland, United Kingdom
[e] Faculty of Veterinary Medicine, Ludwig-Maximilians University Munich, Geschwister-Scholl-Platz 1, Munich, Germany
[f] Faculty of Veterinary Medicine, University of Zagreb, Heinzelova 55, Zagreb, Croatia

ARTICLE INFO

Keywords:
WWTP
Fish
MVA
Hematology
Plasma biochemistry

ABSTRACT

The aim of the study was to assess the value and applicability of multivariate tools for hematological and plasma biochemical responses of fish living in treated wastewater. Physicochemical water properties and heavy metals concentration of water in spring and fall were used as determinants of multiple fish stressors. Three methods of data analysis (Agglomerative Hierarchical Clustering, Factor Analysis, Principal Component Analysis) and one method of data modeling (Partial Least Square Regression) were applied. These methods enabled identification of clustering based on observed parameters, identification of significant variables in the observed data set, and correlation of observed variables with samples collected in different places and at different seasons. Prediction of total leukocytes, lymphocytes, granulocytes, hematocrit, glucose, alanine aminotransferase, triglycerides and cholesterol from fish blood ($R^2 > 0.9$) was better for fall than for spring variables, regardless the sampling site ($R^2 > 0.98$). For hematocrit and glucose (determination coefficient over 0.99), prediction was successful regardless the season and the sampling site. The effectiveness of prediction models was also evaluated using ratio of standard error of performance to standard deviation (RPD), and range error ratio (RER). High applicability of these models was found for multiple purposes (RPD > 8 and RER > 15), including prediction of parameters from fish blood with regard to water quality.

1. Introduction

Anthropogenic activities, particularly environmental pressures caused by pollution from excess nutrient loading deriving from wastewater, agricultural and urban runoff, are the main factors for degradation of environments. Primary and secondary treatments of wastewater treatment plants (WWTPs) remove a good portion of fecal bacteria and reduce biological oxygen demand, suspended solids, nitrogen and phosphorus. However, quality of effluents is often insufficient for well-being of aquatic animals living in such waters (Topić Popović et al., 2015). Fish living in treated wastewaters are thus exposed to a variety of stressors. Although free-living fish react with avoidance of polluted waters when in sublethal threat, they nevertheless tend to adapt to altered environmental parameters, allowing them to survive in unfavorable conditions (Vosyliene et al., 2003). Cyprinid fish have the ability to live in waters of diminished quality and compensate for environmental changes (Topić Popović et al., 2016). The insight of their responses to various environmental factors can be obtained through hematological and plasma biochemical variables.

Blood exhibits pathological changes before the occurrence of other clinical symptoms, and blood withdrawal is a non-lethal method (Bani and Vayghan, 2011). Fish blood variables are thus evermore used in toxicological research and as indicators of environmental stress (Li et al., 2011). Specific plasma biochemistry indicators are reflecting acute exposure to certain stressors, or damage to a specific tissue. However, an array of non-specific indicators may better demonstrate the outcome of chronic exposure to a pollutant (Folmar, 1993). Hematologic disorders, on the other hand, are marked by aberrations in numbers, structure and

* Corresponding author at: Laboratory for Aquaculture Biotechnology, Ruđer Bošković Institute, Bijenička 54, Zagreb, Croatia.
E-mail address: ntopic@irb.hr (N. Topić Popović).

https://doi.org/10.1016/j.aqrep.2021.100877
Received 5 March 2021; Received in revised form 10 September 2021; Accepted 19 September 2021
Available online 23 September 2021

Table 1
Intra- and inter-assay coefficients of variation (CV) for biochemical parameters in fish serum pools.

Assay	Comparison	Pool	Mean	SD	CV (%)
Urea (mmol/L)	intra	low	0.67	0.06	8.66
		high	1.19	0.04	3.20
	inter		1.22	0.07	5.68
Creatinin (μmol/L)	intra	low	9.33	1.53	16.37
		high	17.83	0.75	4.22
	inter		18.28	1.18	6.47
Total proteins (g/L)	intra	low	16.00	0.73	10.83
		high	29.00	0.87	2.99
	inter		28.97	0.29	1.01
Albumin (g/L)	intra	low	5.50	0.58	10.50
		high	10.11	0.33	3.30
	inter		10.11	0,19	1.90
Glucose (mmol/L)	intra	low	6.80	1.26	10.91
		high	12.27	0.34	2.79
	inter		11.57	0.35	5.94
Alkaline phosphatase (U/L)	intra	low	5.75	0.96	16.65
		high	9.00	0.89	9.94
	inter		8,67	1.04	12.01
Alanine aminotransferase (U/L)	intra	low	11.50	1.91	16.65
		high	18.00	1.79	9.94
	inter		18.00	2.00	11.11
Triglycerides (mmol/L)	intra	low	1.13	0.07	6.54
		high	2.23	0.24	10.80
	inter		2.27	0.31	13.48
Cholesterol (mmol/L)	intra	low	3.60	0.34	9.32
		high	6.44	0.13	2.01
	inter		5.90	0.97	16.51

function of the blood cells (Clauss et al., 2008). Obtained hematological and plasma biochemical data therefore require adequate statistical evaluation.

Since conventional statistical techniques cannot analyze large numbers of interrelated variables, a statistical approach with the power to extract principal components from a multitude of data needs to be applied (Sun et al., 2011). Application of mathematical and statistical methods used in order to process large data sets in an optimal way is termed multivariate data analysis (Kurtanjek and Gajdoš Kljusurić, 2014). Multivariate analysis was proven as a powerful tool in data analysis (Hassoun and Karoui, 2017) including tasks as data reduction, regression modeling, and classification of observed samples (Cheng et al., 2014; Dai et al., 2014). Multivariate tools used for reduction of variables, classification and grouping are Factor Analysis (FA) and Principal Component Analysis (PCA). They can give an overview of complex multivariate data and reveal relations between observed samples, allowing detection of significant variables as well as outliers (Bro and Smilde, 2014).

The aim of the study thus was to assess the value and applicability of multivariate tools (MVA) for hematological and plasma biochemical responses of fish living in the treated wastewater over two seasons (spring and fall), using physicochemical water properties and heavy metals concentration of water as determinants of multiple fish stressors. Important parameters were extracted from 1. physicochemical water properties, 2. heavy metals from water, 3. leukogram with PCV values, and 4. plasma biochemistry from Prussian carp (*Carassius gibelio*) blood over two seasons and two locations. After extracting primary parameters, the Partial Least Squares Regression (PLSR) was used to investigate the possibility for direct and rapid prediction of expected hematological and plasma biochemistry responses based on the water quality.

2. Materials and methods

2.1. Fish and sampling sites

Water and free-living fish were sampled during the operation of a municipal wastewater treatment plant (WWTP in NE Croatia). Representative seasons for this work were spring (S) and fall (F). Representative sampling sites for this work were: 1. canal receiving the outflow of treated wastewater (polluted), and 2. WWTP-unrelated natural creek in vicinity of the plant (unpolluted). From both locations, fish were netted, angled and manipulated in accordance with the Bioethical Committee approval (No. BEP-274/2-2012), and European Council Directive 86/609/EEC for animal experiments. Fish belonging to several species were captured in spring and fall, but only data from Prussian carp (*Carassius gibelio*) were taken into consideration, since it represented the most numerous fish group ($n = 24$ and 45, respectively). Fish were randomly sampled, kept in aerated tanks and anesthetized with tricaine methanesulfonate (MS-222, Sigma, St. Louis, Missouri, USA) as described in Topić Popović et al. (2012). Blood was withdrawn by caudal vein puncture.

2.2. Hematology and plasma biochemistry

Blood was centrifuged for 90 s at 12,000 × *g* (Statspin VT (Idexx, USA)). Separated plasma was pipetted and frozen (−80 °C). Concentrations of plasma metabolites and enzymes were determined by Beckman Coulter commercial kits (Olympus Life and Material Science Europe, Ireland) on the Olympus AU 640 biochemistry analyzer (Olympus, Japan) as concentrations of glucose (GLU), total proteins (TP), albumin (ALB), cholesterol (CHOL), triglyceride (TRIG), urea (URE), creatinine (CRE), activity of alanine aminotransferase (ALT), and alkaline phosphatase (ALP). All analyses were conducted in duplicates. All reagents were calibrated on at least two calibration points each, and they were controlled on three levels. For analytical performance of URE, CRE, TP, ALB, GLU, ALP, TRIG and CHOL, assay precision tests were performed (Westgard, 2020; Flatland et al., 2010). For intra-assay precision, two pools of samples with different concentrations of analytes (low and high) were prepared from fish plasma. Intra-assay coefficient of variation (CV) was calculated after analysis of the low and high pools, five times in a single assay run (Table 1). For inter-assay precision, one pool with high concentration of analytes was divided into aliquots and stored at −20 °C until analysis. All samples used for repetitive analyses were frozen to avoid possible changes caused by repetitive thawing and freezing. Inter-assay CV was calculated by analyzing the same sample in separate runs performed on five consecutive days (Table 1).

Blood smears were prepared in triplicate, air dried and stained with Diff-quick stain. Leukocyte morphology was evaluated under magnification of 400× and later 1000× with immersion. Differential count was performed on 200 cells minimum (excluding platelets), under 400× in minimum 10 visual fields. The standardized total leukocyte count per μL of blood and their ratio as % was calculated adapted from Fudge (2000). In short, the total leukocyte count was estimated using the following formula: total slide leukocyte count/number of high power fields (400×, 10 fields minimum) × 2000 = Total white blood cells. Total white blood cells number was then used to determine numbers of each leukocyte type with the following formula: differential count (in %) of counted leukocyte type/100× total number of white blood cells = number of leukocyte type.

The percentage of PCV (packed cells to total volume) was determined by direct measurement on StatSpin microhaematocrit capillary tube reader after centrifuging heparinized microhaematocrit capillaries at 12,000*g* for 120 s. Data are presented in Supplementary material.

2.3. Heavy metals concentration and physicochemical water properties

Determination of water quality parameters was performed according to the following protocols: ISO 15586:2003 and ISO 15586:2004 Trace elements using atomic absorption spectrometry with graphite furnace; ISO 12846:2012 Mercury – Method using atomic absorption spectrometry (AAS) with and without enrichment; ISO 11885:2007 Selected elements by inductively coupled plasma optical emission spectrometry

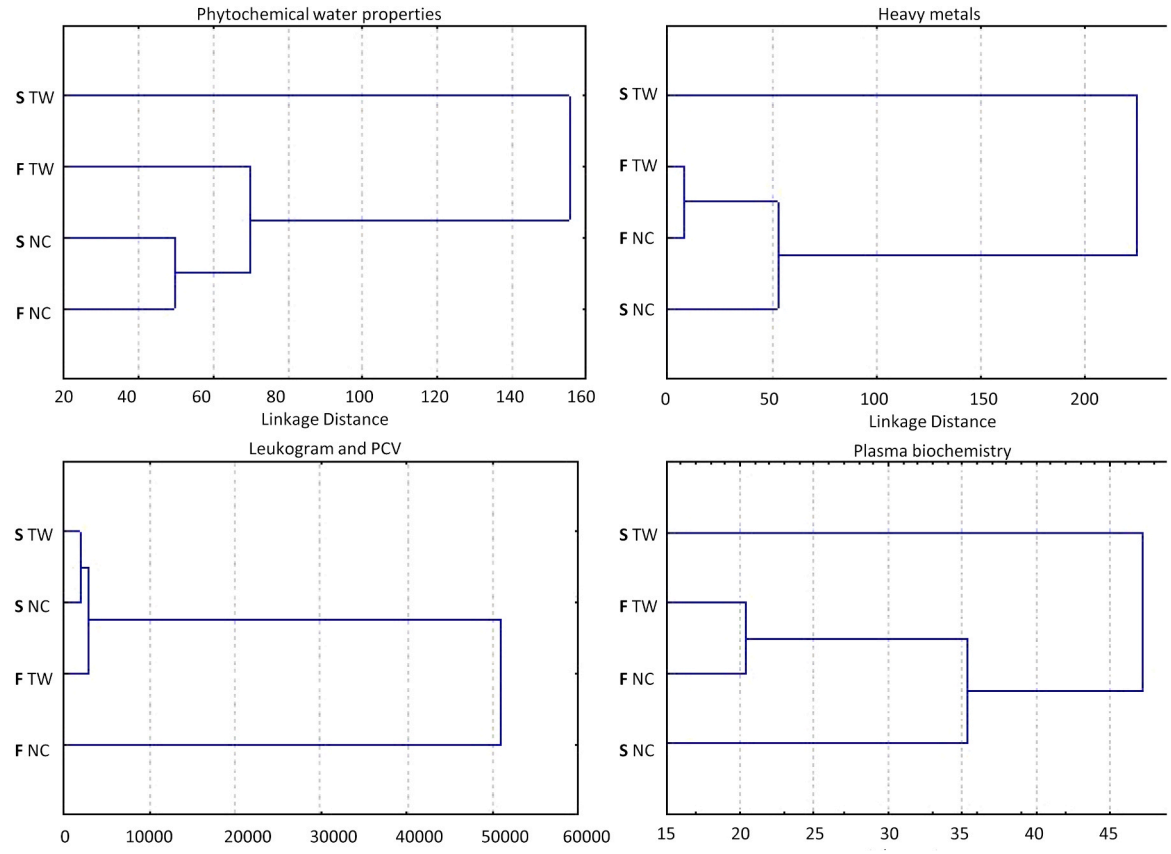

Fig. 1. Dendrogram of the agglomerative hierarchical clustering (AHC) for physicochemical water properties, heavy metals in water, leukogram with PCV, and plasma biochemistry, containing data for two seasons (spring S/fall F) and two sampling sites (treated wastewater TW/natural creek NC).

(ICP-OES); ISO 872:2005 Suspended solids – Method by filtration through glass fiber filters; ISO 6878:2004 Phosphorus – Ammonium molybdate spectrometric method; ISO 10523:2008 pH; ISO 7888:1985 Electrical conductivity; ISO 8467:1993 Permanganate index; ISO 5813:1983 Dissolved oxygen – Iodometric method; ISO 5815:1989 Biochemical oxygen demand after 5 days (BOD 5) – Dilution and seeding method; ISO 7150-1:1984 Ammonium – Part 1: Manual spectrometric method; Standard Methods: 4500-NO3-E Nitrate in water after cadmium reduction; ISO 6777:1984 Nitrite – Molecular absorption spectrometric method; ISO 15705:2002 Chemical oxygen demand index (ST-COD) – Small-scale sealed-tube method. All measurements were conducted in duplicates from three water samples in each season. Data are presented in Supplementary material.

2.4. Data analysis

Processing data contained (i) observations of water per two seasons, which included physicochemical properties (14 parameters), heavy metals (7 parameters); (ii) observations of Prussian carp blood per two seasons: leukogram and PCV (8 parameters), and plasma biochemistry (9 parameters). The total data matrix consisted of 38 rows (of the previously mentioned parameters) for which triplicates were measured per two seasons and two sampling sites. Therefore, the initial data matrix for processing contained 456 complex data. For all statistical analyses, Software Statistica v.10 (StatSoft, Tulsa, OK, USA) was used. Four different multivariate tools were used; three of them were Analysis of Data tools and one was a Modeling Data tool. Analysis of Data tools: Agglomerative Hierarchical Clustering (AHC), Factor Analysis (FA) and Principal Component Analysis (PCA). Modeling Data tool: Partial Least Square Regression (PLSR).

2.4.1. Agglomerative hierarchical clustering (AHC)

Cluster analysis (CA) was used to classify samples of observed parameters based on the physicochemical water properties, heavy metals in water, leukogram, PCV and plasma biochemistry, containing data for two seasons (spring (S) and fall (F)), and two sampling sites (treated wastewater (TW) and natural creek (NC)). CA forms groups based on their similarity or difference. From the analysis, it is possible to determine the qualitative reasons for such grouping, although CA is not a method that can differentiate between relevant and irrelevant variables. This method depends on the measurement units of the observed parameters and therefore, it is important to standardize them. Different initial clusterings can lead to different final clusterings (as our clusters S & F) and it is advisable to run the procedure several times with different (random) initial clusterings.

2.4.2. Factor analysis (FA)

In order to determine which variables were important in the observed set of data, the FA was applied. FA was used to test the relationships between factor loadings and to test the strength and relationship between each common factor. First step was to determine the suitability of the input data size, and to create a correlation matrix for testing of adequacy. Second step was the extraction of factors minding to determine the number of factors to retain. For better interpretation of the factor structure, the factor rotation was used (Varimax rotation) to simplify the expression of a particular sub-space (Kurtanjek and Gajdoš Kljusurić, 2014). FA resulted with the appropriateness for reducing a number of factors, examining relationships between categories, and evaluating the construct validity of a measurement scale (as in the case of the leukogram and PCV values). FA analysis thus helped to indicate important variables (Ruscio and Roche, 2012; Schmitt, 2011). The outcome of applying this method will largely depend on the ability to

Table 2
Factor analysis showing the distribution of factor patterns for the first two factors.

Observations in water per two seasons	Variables	F1	F2
Physicochemical properties	Temperature	**0.9098**	-0.3978
	Dissolved oxygen	**-0.7940**	-0.6079
	Oxygen saturation	**-0.8342**	-0.5493
	pH	**-0.9401**	0.1034
	Suspended solids	**0.8612**	-0.5082
	COD	0.5591	**-0.6598**
	COD-Mn	**0.9688**	-0.1271
	BODn	**0.9315**	-0.3578
	Ammonium	0.4743	**0.8758**
	Nitrite	0.3288	**0.9343**
	Nitrate	0.1059	**0.9914**
	Total nitrogen	**0.8831**	-0.4692
	Phosphate	**0.9749**	0.2214
	Total phosphorus	**0.9875**	0.1574
Heavy metals	Cadmium	**0.9603**	0.0462
	Chromium	**0.7652**	0.1339
	Mercury	0.6727	**-0.7399**
	Lead	0.6807	**-0.7326**
	Nickel	**0.9631**	0.2690
	Zinc	**0.9417**	0.3364
	Copper	**0.9604**	0.2786
Observations in Prussian carp blood per two seasons	**Variables**	**F1**	**F2**
Leukogram and PCV	Total Leukocytes	**0.9792**	0.2027
	Lymphocytes	**0.8825**	-0.2215
	Lymphocytes	**-0.9624**	-0.2528
	Granulocytes	**0.9639**	0.2484
	Granulocytes	**0.9389**	0.3360
	Monocytes	**0.8947**	-0.4466
	Monocytes	0.0670	**-0.9839**
	PCV	0.3604	**-0.9300**
Plasma biochemistry	URE	-0.5398	**-0.8418**
	CRE	**-0.9310**	0.3651
	TP	**0.9892**	0.0278
	ALB	**0.9671**	0.2016
	GLU	**-0.8068**	0.2821
	ALT	**-0.8303**	-0.4735
	ALP	0.1067	**-0.9943**
	TRIG	**0.6685**	-0.3778
	CHOL	**0.9742**	-0.1451

develop a complete and accurate selection of important variables; we have decided to use the value 0.7 as the value of the significance of the contribution of the variable.

2.4.3. Principal component analysis (PCA)

For the factor extraction, PCA method can also be used, particularly when the aim is to reduce the number of variables while retaining as much of the original variance as possible. PCA resolves the input data matrix into product of scores and loading matrix. PCA was thus used to emphasize variations and to bring out strong patterns in the observed data set. The data were observed with reduced variables and parameters (Conway and Huffcutt, 2003; Jennrich and Bentler, 2011). By use of score and loading plots (observed were parameters in the rows and variables in the columns), PCA helped to explore and visualize the data easily. Scores matrix gave the position of samples in the PCA quadrant with two axes: the principal components (PCs) explaining the majority of data variability (Kurtanjek and Gajdoš Kljusurić, 2014), and the loadings presenting contributions of the observed variables to the PCs. This method is affected by scale, so the data standardization is a mandatory process before applying the analysis itself.

2.4.4. Partial least square regression (PLSR)

In order to examine the possibility of prediction and quantitative determination of certain parameters (expected values form the blood of Prussian carp) based on input parameters (water quality), PLSR models were used. PLSR models were evaluated regarding following parameters: coefficient of determination (R^2), ratio of performance to deviation (RPD) and Range Error Ratio (RER). More acceptable models have higher values of R^2, RPD and RER. The models rating according the R^2 and RPD are typically interpreted according the guidelines published by Williams (2004) and Williams et al. (2012). The interpretation and utility of the calibration model is categorized as follows: R^2 in the range 0.50–0.64 presents a model usable for rough screening; R^2 ranging from 0.66 to 0.81 presents a model usable for screening and some "approximate" calibrations. Good model will have R^2 in the range 0.83–0.90 presenting a usable model that should be used with caution. A usable model in most applications will have R^2 in the range of 0.92–0.96 while the models usable in any application will result with $R^2 > 0.98$. When predicting accuracy of a model based on the RPD, its value tends to rise with the prediction ability of the model. Thus values from 3.1 to 4.9 refer to models with fair prediction ability. Values greater than 8.1 indicate a suitability of the model for predictions (Fearn, 2001; Williams, 2001). RER values for quantification should be greater than 15 (American Association Cereal Chemists International: Approved Method 39-00.01, 1999). The main limitation of PLSR is the need of a large number of samples required for accurate calibration. There are also a number of parameters that assess the efficiency of the model (R^2; RPD, SME etc.), and each of them has its shortcomings due to the way their amount is calculated.

3. Results and discussion

Multivariate tools were used to investigate potential relationships between variables in an overarching way. They were also used to qualify, and if possible, to quantify the relationship between the observed variables. In a system that seeks to examine the impact of quality on the observed parameters of fish inhabiting different (vulnerable) areas, MVA seems to be a good choice because these tools determine the links between the independent and dependent variables and specify the conditions under which the association takes place. MVA controls association between variables by using correspondences, partial correlation and multiple regressions, which emphasizes the main advantage of multivariate analysis: giving a more realistic assessment than if using a single variable. Results of this work explain which of the multivariate tools used were the most suitable for data sets divided in four different groups (1. physicochemical water properties, 2. heavy metals in water, 3. leukogram with PCV, and 4. plasma biochemistry) containing data for two seasons (spring S/fall F) and two sampling sites (treated wastewater TW/natural creek NC).

3.1. Agglomerative hierarchical clustering (AHC)

According to AHC, concentration of heavy metals had the same grouping pattern as plasma biochemistry (Fig. 1). The characteristic of AHC is that based on tables and charts it allows determining the rationale for such grouping. Clustering allowed grouping of observed data sets based on similarity of the observed cases, i.e. based on the season and the site of sampling (Fig. 1). Clustering regarding fish plasma biochemistry showed an increase of linkage distance (> 45) for fish caught in spring in TW vs. NC. The same trend, but with larger linkage distance (> 200) was noted when the observed variables in the clustering process were heavy metals.

Physicochemical water properties are grouped based on the sampling site and the main parameters (dissolved oxygen, saturated oxygen, suspended solids, COD, COD-Mn and BODn). These parameters were significantly responsible for such grouping, and the polluted site (TW) showed differences based on the season, i.e. they were significantly different in spring.

Heavy metals are grouped significantly primarily depending on the

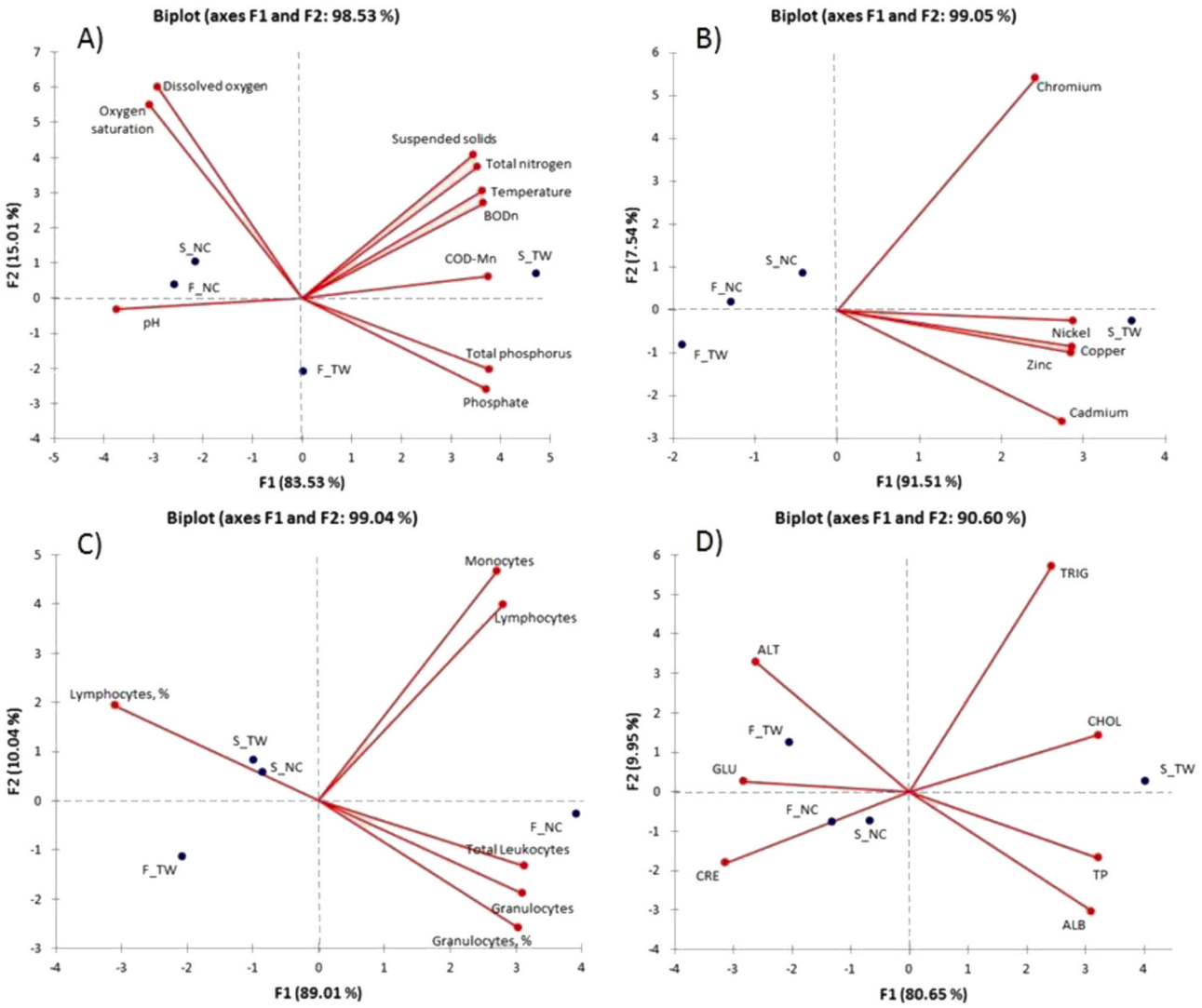

Fig. 2. Biplots of principal component analysis (PCA) for A) physicochemical water properties, B) heavy metals in water, C) leukogram with PCV, and D) plasma biochemistry, containing data for two seasons (spring S/fall F) and two sampling sites (treated wastewater TW/natural creek NC).

season. Significant differences were caused by concentrations of Ni, Zn and Cu. Groupings of plasma biochemistry had no significant changes for URE, TRIG and CHOL, while all other parameters differed based on the season and the sampling site. This multivariate analyzing data method thus proved to be a useful qualitative tool in differentiation of observed data based on physicochemical water properties, heavy metals in water and fish plasma biochemistry.

3.2. Factor analysis (FA)

As previously demonstrated, cluster analysis groups data based on the similarity of cases. Factor analysis, as an explorative analysis, groups similar variables into dimensions (Schmitt, 2011). As an explorative analysis, FA does not distinguish between independent and dependent variables. It reduces the information in a model by reducing dimensions of observations (Jennrich and Bentler, 2011), and is often used to reduce the number of variables (Schmitt, 2011).

In this work, FA was conducted to determine a potential reduction of input parameters for further analysis, for each set of data (Table 2). It was used to eliminate variables which were not significant enough. The stringent criteria i.e. cut-off was set at 0.7. The rotated factor loading is presented in Table 2.

All parameters with the absolute value of loading greater than the cut-off were used in the subsequent multivariate analysis, Principal Component Analysis (PCA). The first factor (F1) of FA explained the largest percentage of variance in the observed set of data (Williams et al., 2012). Therefore, from the data set of physicochemical properties some data were reduced from further observation, thus excluding COD, ammonium, nitrite and nitrate from further calculations. In the data set of heavy metals, Hg and Pb were excluded from further calculations without any significant impact. Monocytes and PCV values were also excluded from the data set of leukogram and PCV, as well as ALP from the data set of plasma biochemistry.

3.3. Principal component analysis (PCA)

PCA analysis was conducted to determine similarities and differences in the observed sets of data depending on sampling sites and seasons (Fig. 2A–D). Analyses were conducted on the reduced data, using only significant F1 data from Table 2.

PCA confirmed the results of AHC which showed clear groupings (S/F and TW/NC). It also demonstrated which parameters were dominant in which season and on which sampling site. Unlike in the AHC method, there were quantitative differences for the observed parameters. In Fig. 2A it is visible that total phosphorus and phosphate have the highest values in treated wastewater (F_TW and S_TW). Heavy metals were the

Table 3
PLS regression model parameters (coefficient of determination, R^2 and the ratio of standard error of prediction to sample standard deviation, RPD). The input data for prediction: physicochemical water properties (PWP) or heavy metals from water (HMW), over two seasons (spring (S) and fall (F)), in treated wastewater (TW) and a natural creek (NC).

			Total Leukocytes	Lymphocytes	Granulocytes	PCV	GLU	ALT	TRIG	CHOL
S_TW	model PWP	R^2	0.9465	0.9637	0.9983	0.9980	0.9988	0.9423	0.9990	0.9973
		RPD	5.3	6.4	8.6	8.6	8.6	5.0	8.6	8.5
		RER	10.6	12.8	17.1	17.1	17.2	10.1	17.2	17.0
	model HMW	R^2	0.9465	0.9637	0.9983	0.9980	0.9988	0.9423	0.7500	0.9973
		RPD	5.3	6.4	8.6	8.6	8.6	5.0	2.4	8.5
		RER	10.6	12.8	17.1	17.1	17.2	10.1	6.1	17.0
F_TW	model PWP	R^2	0.9909	0.9843	0.9866	0.9997	0.9895	0.9995	0.9991	0.9868
		RPD	8.1	7.7	7.8	8.7	8.0	8.6	8.6	7.8
		RER	16.2	15.4	15.7	17.3	16.0	17.3	17.2	15.7
	model HMW	R^2	0.9909	0.9843	0.9866	0.9997	0.9895	0.9995	0.9356	0.9868
		RPD	8.1	7.7	7.8	8.7	8.0	8.6	4.6	7.8
		RER	16.2	15.4	15.7	17.3	16.0	17.3	9.2	15.7
S_NC	model PWP	R^2	0.9948	0.9981	0.9616	0.9957	0.9998	0.8246	0.9900	0.9176
		RPD	8.4	8.6	6.2	8.4	8.7	-2.4	8.0	3.5
		RER	16.7	17.1	12.5	16.8	17.3	-4.8	16.1	6.9
	model HMW	R^2	0.9948	0.9981	0.9616	0.9957	0.9998	0.8246	0.9891	0.9176
		RPD	8.4	8.6	6.2	8.4	8.7	-2.4	8.0	3.5
		RER	16.7	17.1	12.5	16.8	17.3	-4.8	16.0	6.9
F_NC	model PWP	R^2	0.9997	0.9960	0.9794	0.9996	0.9990	0.9316	0.9990	0.9996
		RPD	8.7	8.4	7.4	8.7	8.6	4.4	8.6	8.7
		RER	17.3	16.9	14.7	17.3	17.2	8.7	17.2	17.3
	model HMW	R^2	0.9997	0.9960	0.9794	0.9996	0.9990	0.9316	0.9990	0.9996
		RPD	8.7	8.4	7.4	8.7	8.6	4.4	8.6	8.7
		RER	17.3	16.9	14.7	17.3	17.2	8.7	17.2	17.3

highest in treated wastewater in spring (S_TW), positioned in the right side of the biplot, in the fourth quadrant.

Reduced data set for leukogram and PCV demonstrated that unpolluted natural creek during fall (F_NC) was positioned inversely proportional to lymphocytes (%). Similar positioning was in polluted waters (F_TW) for the lowest blood values of monocytes and lymphocytes.

Biplot of plasma biochemistry parameters brought out a pattern for the lowest values of CRE in spring fish (S_TW), and the lowest values of TP and ALB in fall fish (F_TW) in polluted waters.

PCA thus demonstrated an objective way of calculating indices that reduce dimensionality of the data by their linear combination, generating latent variables uncorrelated to each other, and concisely accounted for variability of the data (Bhuiyan et al., 2016). It also demonstrated a general picture of trends and groupings of the individual variables, as well as pinpointed particular determinants influencing the variability of data (Li et al., 2011). Qualitative determination ending with quantitative determination was reliable.

3.4. Partial least square regression (PLSR)

PLSR was used to investigate a possibility for direct and rapid prediction of expected hematological and plasma biochemistry responses to water quality (Table 3). It explained and qualitatively evaluated data with three different parameters: coefficient of determination (R^2), ratio of standard error of performance to standard deviation (RPD), and range error ratio (RER). Each parameter described the PLSR model in a particular way. The R^2 described how well the data points fitted the model, with values ranging from 0 to 1, and accuracy of the model expressed by values closer to 1. Higher RPD values also suggested models with increasing accuracy. RER values greater than 10 indicated the model as acceptable for quality control while values over 15 indicated a possibility of applying the model for quantification (American Association Cereal Chemists International: Approved Method 39-00.01, 1999).

Based on the results from FA (Table 2), significant values were chosen for leukogram and PCV data set. From plasma biochemistry parameters, four (GLU, ALT, TRIG and CHOL) were chosen as parameters related to water pollution. The lowest model indicators (R^2, RPD and RER; 0.75; 2.4; 6.1, respectively) were established for TRIG when the content of heavy metals in water (HMW model) was used as the input data of the PLSR model. However, for TRIG in S_TW the standard deviation was equal to zero. All other parameters were evaluated as suitable for quality control (RPD $\geq$ 5 and/or RER $\geq$10), for good calibration for quantification (RER $\geq$ 15), and for applied research (RPD $\geq$ 8) based on the American Association Cereal Chemists International: Approved Method 39-00.01 (1999). This is an additional confirmation for the soundness of conducted methods of multivariate analyses, as well as the applicability of the model for prediction of expected parameters in the blood of Prussian carp based on water pollution in different seasons.

Application of a statistical approach on a certain data set often carries dilemmas. Multivariate approach on environmental data is complex and the input data set needs to be informative enough (without missing data) to enable qualitative and quantitative analyses. Qualitative analysis observed similarities and/or differences in the computed data through different seasons and sampling sites, and enabled quantitative analyses. These presented the prediction potential of expected hematological and plasma biochemistry responses to water quality.

3.5. Conclusions

The results corroborate the concept that multivariate tools are valuable in determination of biological responses of fish to multiple fish stressors, particularly the PCA analysis. PCA confirmed clustering groups and determined dominant parameters. In a biological sense, leucopenia and lymphopenia, namely mononuclear leukocytes (lymphocytes and monocytes) grouped for F_TW fish, indicate a reaction to stressor(s) (Clauss et al., 2008). These stressors were likely elevated ammonium, nitrite, nitrate and phosphorus, which were excessively high in that period in treated wastewater. Such an elevation could be attributed to industrial waters generated from the sugar beet processing (Topić Popović et al., 2016). Water quality diminished to such an extent might stimulate the algae overgrowth and hence reduce the oxygen level needed for fish breathing (Sahu and Chaudhari, 2015). Although the FA reduced the data set of physicochemical properties excluding ammonium, nitrite and nitrate, it included the phosphorus variable as a representative of the effect. Granulocytes decreased in S_TW fish and were correctly grouped with the leukocytes variables, as they contribute a significant portion of white blood cells in the form of

polymorphonuclears (Folmar, 1993). Their decrease might be connected to immunosuppression resulting from the poor quality of wastewater, particularly high concentrations of all tested heavy metals in that period (Dunier, 1996). Although AHC positioned leukogram variables in a small linkage distance, PCA highlighted the granulocyte parameter as dominant. Elevated levels of ALB and TP in spring TW fish were also interrelated with the increased levels of heavy metals in that period, providing valuable information on fish physiological status and stress impact on blood biochemistry (Topić Popović et al., 2016). Overall, PCA gave a strong definition of factors relative to the immune suppression of fish captured from treated wastewater in both seasons.

Multivariate tools are powerful and versatile methods that have shown their capacity for a detailed insight in complex multivariate data sets. Depending on the set hypothesis, different multivariate tools for analyses and modeling should be used. Complex data sets in this work contained data of physicochemical properties and heavy metals in water over two seasons measured in two sampling sites as well as data of leukogram, PCV and plasma biochemistry from Prussian carp blood. By use of multivariate tools, it was possible to provide a comprehensible scrutiny of complex multivariate data, such as relations between variables and relations between samples (clustering). It was also possible to detect which parameters could be reduced based on their (in)significance and quantifying patterns, as well as to predict significant trends of variables to be expected in Prussian carp blood based on the water quality and season.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Scientific Centre of Excellence for Marine Bioprospecting – BioProCro, a project co-financed by the Croatian Government and the European Union through the European Regional Development Fund – the Competitiveness and Cohesion Operational Programme (KK.01.1.1.01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no conflict of interest. The authors gratefully acknowledge Dr. Slavko Kepec from Virkom d.o.o., Public Water Supply and Wastewater Services, Virovitica, Croatia for enabling sample acquisition and water analyses.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.aqrep.2021.100877.

References

American Association Cereal Chemists International: Approved Method 39-00.01, 1999. Near-Infrared Methods – Guidelines for Model Development and Maintenance. First Approval 3 November, pp. 1–5.

Bani, A., Vayghan, A.H., 2011. Temporal variations in haematological and biochemical indices of the Caspian kutum, *Rutilus frisii kutum*. Ichthyol. Res. 58, 126–133.

Bhuiyan, M.A.H., Bodrud-Doza, M., Towfiqul, I.A.R.M., Rakib, M.A., Rahman, M.A., Ramanathan, A.L., 2016. Assessment of groundwater quality of Lakshimpur district of Bangladesh using water quality indices, geostatistical methods, and multivariate analysis. Environ. Earth Sci. 75, 1020. https://doi.org/10.1007/s12665-016-5823-y.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6, 2812–2831. https://doi.org/10.1039/C3AY41907J.

Cheng, J.-H., Qu, J.-H., Sun, D.-W., Zeng, X.-A., 2014. Texture and structure measurements and analyses for evaluation of fish and fillet freshness quality: a review. Compr. Rev. Food Sci. Food Saf. 9, 52–61.

Clauss, T.M., Dove, A.D.M., Arnold, J.E., 2008. Hematologic disorders of fish. Vet. Clin. Exot. Anim. 11, 445–462.

Conway, J.M., Huffcutt, A.I., 2003. A review and evaluation of exploratory factor analysis practices in organizational research. Organ. Res. Methods 6, 147–168.

Dai, Q., Sun, D.-W., Xiong, Z., Cheng, J.-H., Zeng, X.-A., 2014. Recent advances in data mining techniques and their applications in hyperspectral image processing for the food industry. Compr. Rev. Food Sci. Food Saf. 13, 891–905.

Dunier, M., 1996. Water pollution and immunosuppression of freshwater fish. Ital. J. Zool. 63 (4), 303–309.

Fearn, T., 2001. Assessing calibrations: SEP, RPD, RER and R2. NIR News 13, 12–14.

Flatland, B., Freeman, K.P., Friedrichs, K.R., Vap, L.M., Getzy, K.M., Evans, E.W., Harr, K. E., 2010. ASVCP quality assurance guidelines: control of general analytical factors in veterinary laboratories. Vet. Clin. Pathol. 39 (3), 264–277. https://doi.org/10.1111/j.1939-165X.2010.00251.x.

Folmar, L.C., 1993. Effects of chemical contaminants on blood chemistry of teleost fish: a bibliography and synopsis of selected effects. Environ. Toxicol. Chem. 12, 337–375.

Fudge, A.M., 2000. Laboratory Medicine for Avian and Exotic Pets. WB Saunders, Philadelphia.

Hassoun, A., Karoui, R., 2017. Quality evaluation of fish and other seafood by traditional and nondestructive instrumental methods: advantages and limitations. Crit. Rev. Food Sci. Nutr. 57 (9), 1976–1998.

Jennrich, R.I., Bentler, P.M., 2011. Exploratory bi-factor analysis. Psychometrika 76, 537–549.

Kurtanjek, Ž., Gajdoš Kljusurić, J., 2014. Statistical modelling of anthropometric characteristics evaluated on nutritional status. In: Granato, D., Ares, G. (Eds.), Mathematical and Statistical Methods in Food Science and Technology. John Wiley and Sons, Oxford, UK, pp. 285–302 (str).

Li, Z.-H., Velisek, J., Grabic, R., Li, P., Kolarova, J., Randak, T., 2011. Use of hematological and plasma biochemical parameters to assess the chronic effects of a fungicide propiconazole on a freshwater teleost. Chemosphere 83, 572–578.

Ruscio, J., Roche, B., 2012. Determining the number of factors to retain in exploratory factor analysis using comparison data of known factorial structure. Psychol. Assess. 24 (2), 282–292.

Sahu, O.P., Chaudhari, P.K., 2015. The characteristics, effects, and treatment of wastewater in sugarcane industry. Water Qual. Expo. Health 7, 435–444.

Schmitt, T.A., 2011. Current methodological considerations in exploratory and confirmatory factor analysis. J. Psychoeduc. Assess. 29 (4), 304–321.

Sun, C.C., Wang, Y.S., WU, M.L., Dong, J.D., Wang, Y.T., Sun, F.L., Zhang, Y.Y., 2011. Seasonal variation of water quality and phytoplankton response patterns in Daya Bay, China. Int. J. Environ. Res. Public Health 8, 2951–2966.

Topić Popović, N., Strunjak-Perović, I., Čož-Rakovac, R., Barišić, J., Jadan, M., Peršin Beraković, A., Sauerborn Klobučar, R., 2012. Tricaine methanesulfonate (MS-222) application in fish anaesthesia. J. Appl. Ichthyol. 28, 553–564.

Topić Popović, N., Strunjak-Perović, I., Sauerborn Klobučar, R., Barišić, J., Babić, S., Jadan, M., Kepec, S., Kazazić, S., Matijatko, V., Beer Ljubić, B., Car, I., Repec, S., Stipaničev, D., Klobučar, G., Čož-Rakovac, R., 2015. Impact of treated wastewater on organismic biosensors at various levels of biological organization. Sci. Total Environ. 538, 23–37.

Topić Popović, N., Strunjak-Perović, I., Barišić, J., Kepec, S., Jadan, M., Babić, S., Jadan, M., Beer Ljubić, B., Matijatko, V., Palić, D., Klobučar, G., Babić, S., Gajdoš Kljusurić, J., Čož-Rakovac, R., 2016. Native Prussian carp (*Carassius gibelio*) health status, biochemical and histological responses to treated wastewaters. Environ. Pollut. 218, 689–701.

Vosyliene, M.Z., Kazlauskiene, N., Svecevičius, G., 2003. Effect of a heavy metal model mixture on biological parameters of rainbow trout *Oncorhynchus mykiss*. Environ. Sci. Pollut. Res. 10 (2), 103–107.

Westgard, J.O., 2020. Basic Method Validation and Verification, 4th edition. Training in Statistical Quality Control for Medical Laboratories. ⟨www.westgard.com⟩.

Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P., Norris, K. (Eds.), Near Infrared Technology in the Agricultural and Food Industries, 2nd ed. American Association of Cereal Chemists, St. Paul, MN, USA.

Williams, P.C., 2004. Near-infrared technology: getting the best out of light. In: Short, A. (Ed.), Course in the Practical Implementation of Near Infrared Spectroscopy for the User. PDK Projects Inc., Nanaimo, BC, Canada, p. 109.

Williams, P.J., Geladi, P., Britz, T.J., Manley, M., 2012. Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus Fusarium. Anal. Bioanal. Chem. 404, 1759–1769.