The Application of Multivariate Statistical Methods in Ecotoxicology and Environmental Biochemistry

Halina Falfushynska[1], Oleg Lushchak[2] and Eduard Siemens[3]
[1]*Department of physical rehabilitation and vital sequrity, Ternopil V. Hnatiuk National Pedagogical University, 2 Kryvonosa Str., Ternopil, Ukraine*
[2]*Department of biochemistry and biotechnology, Vasyl Stefanyk Precarpathian National University, 57 Shevchenko Str., Ivano-Frankivsk, Ukraine*
[3]*Anhalt University of Applied Sciences, 57 Bernburger Str., Köthen, Germany*
falfushynska@tnpu.edu.ua, olehl@pu.if.ua, eduard.siemens@hs-anhalt.de

Keywords: Multivariate Statistical Analysis, Principal Component Analysis, Linear Discriminant Analysis, Classification and Regression Tree Analysis, Pesticide Pollution, Ecotoxicology and Environmental Biochemistry.

Abstract: Pesticide pollution of surface- and groundwater are a subject of national importance indeed. However, far too little attention has been paid to find out suitable protocols and algorithms for ecotoxicological data analysis and generalisation. The aim of the present study was to implement Multivariate statistical analysis techniques for prediction of toxicity level of widely-used organophosphate pesticides to living organisms and find out the most appropriate statistical technique out of implemented to integrate biological data. The generalization of the results of biochemical and physiological measurements in zebrafish, Daphnia and Drosophila exposed to widely-used pesticides namely chlorpyrifos, roundup, and malathion have been done using principal component analysis, linear discriminant analysis and classification and regression tree analysis. All of three applied multivariate statistical techniques claimed chlorpyrifos to be the most toxic pesticides out of tested based on responses of living organisms. The importance of battery of biomarkers for risk assessment when compare to individual indices was proved using classification and regression tree analysis and discriminant analysis and daphnia's protein carbonyls level and zebrafish's lactate dehydrogenase activity pertain to the most sensitive indices for group distinguishing. We propose to combine the most widely used in life sciences Principal Component Analysis with classification and regression tree analysis and discriminant analysis to better highlight the important biological entities and reveal insightful patterns in the data.

1 INTRODUCTION

Being rapidly increasing and causing 14,000 deaths of people daily [1], water pollution has to be the matter of global concern. Sure enough that pesticides make a considerable contribution in water pollution. For example, more than 1 billion pounds of pesticides are used annually in the USA and these substances and/or their metabolites are commonly detected in both surface and groundwater bodies[1]. Likewise, 81% of small streams in Germany were accused in pollution by several pesticides in amounts that exceeded permissible concentration, while in 18% of them, these concentrations were simultaneously breached by more than 10 different pesticides [2]. Also, high concentrations of carbendazim, malathion, and diuron were found in the basin of the Llobregat River, Catalonia, Spain [3]. Nevertheless, pesticide contamination become the global threat both for animal and human because its proven toxicity against living organisms, particularly non-targets, far too little attention has been paid to find out suitable protocols and algorithms for ecotoxicological data analysis, integration and generalisation that allow researchers to conjoin efforts regarding water quality monitoring across the world and then implement standardized water quality criteria.

Multivariate statistical analysis is a quickly developed field of classical statistics which can help data processing and analysing associations between the different variables measured[2]. Principal

[1]https://biologicaldiversity.org

[2]https://web.stanford.edu/class/bios221/book/Chap-Multivariate.html

component analysis (PCA), factor analysis, and discriminant analysis belong to the most popular modules of Multivariate statistical analysis and all of them should be very helpful for data generalization, interpretation, and making meaningful conclusions and prediction. It is a valuable tool for identifying factors and sources that may affect quality of surface and groundwater systems [4].

PCA has been widely used to estimate the potential effects of pollutants on water animals and environment. In particular, data processing related to water quality conditions, spatial–temporal changes, and the driving factors in pond and cage aquaculture areas of Zhuanghe area using PCA showed that the most relevant factors of water quality in marine ecosystems are salinity, dissolved oxygen, and antibiotic resistance genes. For the cage aquaculture area chlorophyll a was determined as the additional qualification parameter [5]. Likewise, PCA emphasized that nutrient factor (39.2%), sewage and feacal contamination (29.3%), physicochemical sources of variability (6.2%) and waste water pollution from industrial and organic load (5.8%) affected water quality in the Ganges River [6]. Our prior studies also have noted the importance of PCA for risk assessment and toxic level prediction based on not only water chemical parameters but also biochemical and physiological parameters of living organisms [7]. Moreover, more profound findings come to hand when data analysis using PCA combines with other module of Multivariate statistical analysis namely discriminant analysis or data mining tools [7]. However, a few studies that used different modules of Multivariate statistical analysis and/or data mining in ecotoxicological purpose could be found in literature. Therefore, the aim of the present study was to implement Multivariate statistical analysis techniques for prediction of toxicity level of widely-used organophosphate pesticides to living organisms and find out the most appropriate statistical technique out of implemented to integrate biological data.

2 MATERIALS AND METHODS

For experimental research, we have chosen three species that differ in the level of origin (invertebrate/vertebrate) and ecological needs. We have used the cyprinid fish *Danio rerio*, as a conventional biological model for mechanistic and toxicological studies, which demonstrates universal to vertebrates, responses to stress and toxicity. *Daphnia magna* is widely spread in nature, easy in cultivation, has high sensitivity to various xenobiotics, which makes it very convenient objects for bioindication. The fruit fly *Drosophila melanogaster* was used as an extremely convenient, popular and relatively cheap model object.

Fish were treated with organophosphate pesticides roundup (commercial form of glyphosate), malathion and chlorpyrifos in two concentrations, low and high which could be designated as environmentally relevant due to average concentration of correspondent chemicals in water [8, 9]. Higher tested concentrations correspond to pesticides levels in waste waters and heavily polluted water bodies. In particular, experimental fish were exposed to roundup (RL, 15 $\mu g \cdot L^{-1}$ and RH, 500 $\mu g \cdot L^{-1}$), malathion (ML, 5 μg L^{-1} and MH, 50 μg L^{-1}) and to chlorpyrifos (CPL, 0.1 $\mu g \cdot L^{-1}$ and CPH, 3.0 $\mu g \cdot L^{-1}$) for 14 days. Daphnia were exposed to the same organophosphate pesticides in the following concentration, Roundup (0.1; 0.5; 1 and 5 μg L^{-1}), chlorpyrifos at concentrations of 0.001; 0.005 and 0.01 μg L^{-1}, and malathion (0.1; 0, 5; 1 μg L^{-1}).

Complex study of zebrafish responses included next parameters: antioxidant defense (total antioxidant capacity, catalase, glutathione total, glutathione transferase); oxidative damage (formation of reactive oxygen and nitrogen species, protein carbonyls and lipid peroxidation); cytotoxicity (stability of lysosomal membranes, lactate dehydrogenase activity in the blood); neurotoxicity (cholinesterase activity); apoptosis (expression of caspase-3, cathepsin D level), DNA damage and repair (DNA fragmentation level, expression of GADD45); immune status (IgM); endocrine disruption (expression of vitellogenin in male, cortisol and triiodothyronine). The impact of pesticides on *Daphnia* was evaluated using parameters of antioxidant defense (superoxide dismutase, catalase, glutathione total); oxidative damage (lipid peroxidation and protein carbonyls). Fruit fly response was evaluated based on the activity of antioxidants (SOD, catalase, aconitase, low molecular weight thiols) and markers of oxidative stress (protein carbonyls, lipid peroxides). All abovementioned parameters were measured according to protocols as described in detail in [10, 11].

Data were tested for normality and homogeneity of variance by using Kolmogorov–Smirnoff and Levene’s tests, respectively. Whenever possible, data were normalized by Box–Cox transforming method. For the data that were not normally distributed, non-parametric tests (Kruskall–Wallis

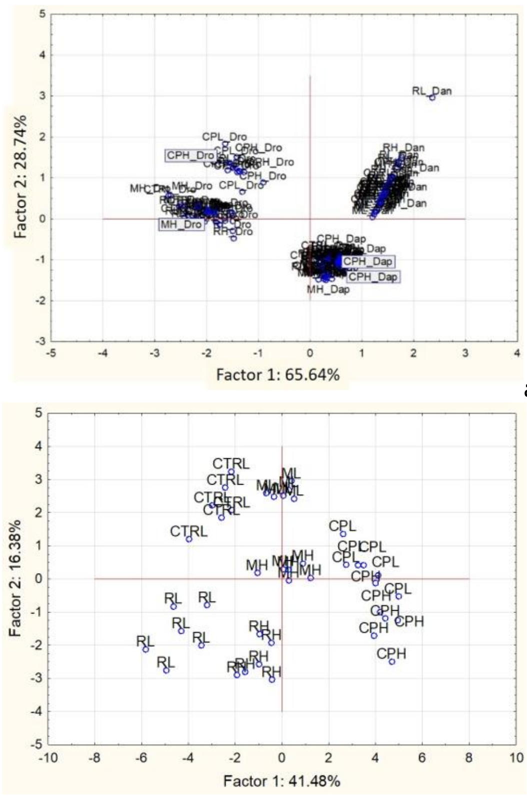
ANOVA and Mann–Whitney *U*-test) were performed. Normalized, Box–Cox transformed data were subjected to principal component analysis (PCA) to differentiate individual specimens by the set of their indices. PCA is the standard tool for extracting components and to visualize the similarities between the biological samples. It serves to find a low dimensional representation of the data which captures most of the variance. In the analysis of animal responses to adverse effects have been considered only variables with correlation >0.60 with the first two dimensions of the PCA (Factor Loadings).

The classification tree in terms of determination the complex interactions among variables and sensitive criterions that distinguish evaluated groups based was built using Classification and Regression Tree (CART)-style exhaustive search for univariate splits using non-transformed studied traits. Canonical discriminant analysis was applied to extract variables that were able to maximize the between-group separation. Scores for each CVA and for each individual were calculated and then plotted in the canonical space. Differences between groups were specified based on the Mahalanobis distance, which reflects the distance between the centroids of each group.

All statistical calculations were performed with Statistica v. 12.0. For all traits and experimental treatment groups, sample size was 8.

3 RESULTS

The generalization of the results of biochemical and physiological measurements in zebrafish, daphnia and Drosophila exposed to organophosphate pesticides by Multivariate statistical analysis confirms the peculiarities of the response of stress-responsive and detoxifying systems of Drosophila on the one hand as a target organism to the effects of insecticides and daphnia / zebrafish as non-target organisms on the other. The first two PCs account for 94 % of the total variation in the dataset. All of the factor loadings in the first PC for biological traits of zebrafish and daphnia have the same sign, so it is a weighted average of all variables, representing 'non-targets' mode of reaction'. Contrary, all groups of fruit fly which is target organism to insecticides and non-target to herbicides have negative factor loadings in the first PC. In general, in depends on the mode of response and adaptation strategy to adverse effects, insects are on the left and cyprinid fish and crustacean which are not intentionally selected for control by pesticides, but which may suffer damage because of exposure to it on the right. The purest intergroup distribution is characteristic for fruit fly, whose specimens based on the oxidative stress markers are on one side of PC 1, and daphnia / zebrafish, whose specimens are located opposite the axis of PC 1 (Figure 1 a). The joint localization of daphnia and zebrafish in the plane of plot area of factor analysis confirms the realization of common strategies of adaptation in non-target organisms to pesticides, regardless of phylogenetic origin.



a) b)

Figure 1: Principal component analysis biplot based on all measured biomarkers of zebrafish *Danio rerio* (Dan), Daphnia (Dap), and Drosophila (Dro) exposed to roundup, chlorpyrifos, and malathion. CTRL – control, RL - roundup low concentration, RH – roundup high concentration, ML – malathion low concentration, malathion high concentration, CPL – chlorpyrifos low concentration, CPH – chlorpyrifos high concentration.

On the other hand, grouping data according to the type of pesticide and its concentration shows a significant dependence of the adaptation strategy on the nature of the acting factor and the depth of its impact (Figure 1 b). Tangentially, groups separation along the axis of Factor2 occurred mainly due to parameters of zebrafish (Table 1) when for Factor1 responses of all three organisms were important.

Chlorpyrifos, is an organophosphate insecticide used to control foliage and soil-borne insect pests, becomes apparent as the most toxic pesticide to living organisms based on the sum of the biochemical and physiological markers of three studied organisms, and roundup (in environmentally relevant concentrations) and malathion manifested to be the least toxic. The specific response to the effects of pesticides in high tested concentrations corresponds to the general oppression of the health status of the organism. However, even in the case of minimal toxicity, signs of oxidative stress, cytotoxicity, inhibition of detoxification processes were observed in zebrafish and daphnia, which can probably reduce the organism's tolerance to stress when organism exposed to additional stressors and lead to irreversible changes in molecular and cellular levels that can potentially appeared themselves over time at both the organismal and population-species levels, remotely affecting biodiversity loss.

Table 1: Factor coordinates of the variables, based on correlations.

Variable	Factor 1	Factor 2
Low-weight molecular thiols (Dro)	0.842451	-0.050171
High-weight molecular thiols (Dro)	0.772227	0.022814
LOOH (Dro)	0.679562	0.200390
CAT (Dro)	-0.857893	-0.023008
Aconitase (Dro)	-0.872867	-0.329750
SOD (Dap)	-0.198962	-0.160684
CAT (Dap)	-0.730681	-0.334843
LOOH (Dap)	-0.724947	0.072561
Low-weight molecular thiols (Dap)	0.341826	-0.417394
High-weight molecular thiols (Dap)	-0.636850	0.483128
Protein Carbonyls (Dap)	-0.762287	-0.502726
ROS (Dan)	0.779434	-0.171246
TAC (Dan)	-0.673789	-0.175300
Glutathione (Dan)	-0.423798	-0.561634
RNS (Dan)	0.408960	-0.611201
GST (Dan)	-0.657437	-0.035853
CAT (Dan)	-0.723644	-0.316562
TBARS (Dan)	0.501207	-0.742588
Protein Carbonyls (Dan)	0.486866	-0.222567
Lactate dehydrogenase (Dan)	0.506237	-0.800050
Caspase 3 (Dan)	-0.693161	-0.365407
Vitellogenin (Dan)	0.260017	-0.650203
Eigenvalues	9.13	3.61

Results of the linear discriminant analysis stated that only 13 out of 22 parameters analysed of zebrafish, daphnia and drosophila were proven important in discriminating studied groups according to pesticide toxicity and adaptation strategy ($F_{(132,89)}$ =22.507 p<0.0000) (Figure 2). The first two discriminant functions accounted for 99.3% of the variation in group separation, being significant, with high Chi–Square value. The distinguishing of chlorpyrifos-treated groups from other groups were primarily due to more prominent signs of cytotoxicity and rate of lipid and protein peroxidation of animals regardless their evolutionary level. These groups locate farthest away from others according to Mahalanobis distance (> 2970) and may well designate as "the most impacted" that corresponds to "High toxicity". On the oppose, animals have been treated with roundup and malathion locate closest to control (<721) and highly likely belong to "the less impacted" groups that corresponds to "Low toxicity". The ratio of observed to predicted values of the biological parameters equalled to 100%.

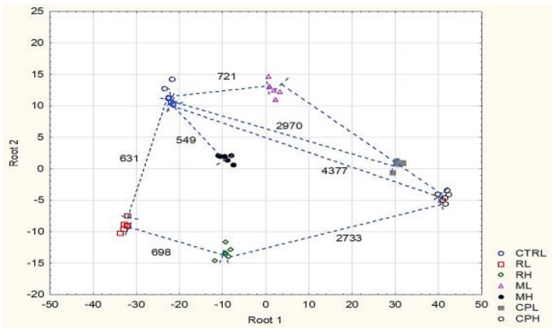


Figure 2: Discriminant analysis biplot of the physiological and biochemical traits of zebrafish *D. rerio*, Daphnia, and Drosophila exposed to roundup, chlorpyrifos, and malathion. Numbers next to the lines indicate the squared Mahalanobis distance between the respective groups. CTRL – control, RL - roundup low concentration, RH – roundup high concentration, ML – malathion low concentration, malathion high concentration, CPL – chlorpyrifos low concentration, CPH – chlorpyrifos high concentration.

It has been recently shown that discriminant analysis could be very helpful to integrate environmental data and establish a set of variables for significant discrimination between affected groups [12]. Several attempts disclosed the power of discriminant analysis to explain the differences between control and polluted vicinities based on water parameters or set of biochemical/ histopathological indices of bioindicator organism [13]. On the other hand, discriminant analysis has been implemented successfully to disclose prepotent biomarkers that allow to separate localities or conditions significantly. In particular, results of the stepwise discriminant analysis selected nine indices that distinguished studied sites in Hong Kong due to benthic infaunal structure, physical and chemical characteristics of sediment samples, toxicity data and metal accumulation and rate of survival belonged to the most sensitive parameters [14]. We have tried to explain the differences in the mechanisms of toxic mode of pesticides action on the basis of molecular descriptors of exposed organisms which differ in ecological demands and biological organisation. It let us not only to evaluate the biohazards of widely-used pesticides, but also digitize the value of toxicity using Mahalanobis square distance. The PCA, which is more popular in biostatistics, doesn't allow that. Indeed, the combination of different statistical approach for data integration may well optimize a procedure of parameters selection for risk assessment protocols and determined the depths of biohazards. Therefore, despite the small dataset, our results depict the applicability and usefulness of discriminant analysis with close relation with other tools of multivariate statistical analysis in integrating ecotoxicological and biochemical data in terms of water quality monitoring and risk assessment. The predictive capability of the model might take advantage from an increase in sampling parameters, exposure conditions, number of chemical parameters and/or biomarker.

The data analysis using CART algorithm allowed us to identify the valuable parameters that most likely distinguish the groups of interests. The built tree consists of seven terminal leaf nodes and accounts six splits (Figure 3). The set of biomarkers includes lactate dehydrogenase activity of zebrafish blood, which primarily separates the least impacted groups (control and malathion in environmentally relevant concentration), the total antioxidant capacity and caspase 3 activity of zebrafish, as well as protein carbonyls level and catalase activity of daphnia. The most obvious finding to emerge from this study is that none of the Drosophila parameters was included in the classification tree, which clearly indicates the impossibility of using an organism that is targeted to a certain factor as a bioindicative for assessing toxicity and biohazards based on a set of biomarkers. The importance of battery of biomarkers for risk assessment when compare to individual indices was also proved by discriminant analysis and daphnia's protein carbonyls level (F-remove $_{(6,14)}$ = 25.85, $p<0.0001$) and zebrafish's lactate dehydrogenase activity (F-remove $_{(6,14)}$=13.9, $p<0.0001$) pertain to the most sensitive indices for group distinguishing.

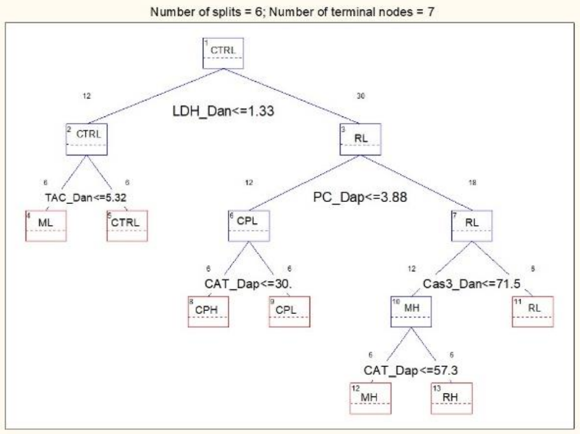


Figure 3: Classification tree of the studied biological traits of zebrafish (Dan), Daphnia (Dap), and Drosophila (Dro) exposed to roundup, chlorpyrifos, and malathion. CTRL – control, RL - roundup low concentration, RH – roundup high concentration, ML – malathion low concentration, malathion high concentration, CPL – chlorpyrifos low concentration, CPH – chlorpyrifos high concentration.

3 CONCLUSIONS

Data analysis is important in processing and generalization data sets in ecotoxicology and life sciences. We have described an approach that uses multivariate statistical techniques namely PCA, discriminant analysis and CART to find variables that are correlated across the samples, separate studied groups of specimens, localities and/or conditions, as well as discover the most sensitive indices/biomarkers that might be helpful to predict outcomes and effects while being treated. Being processed with PCA, discriminant analysis and CART, biochemical and physiological indices of zebrafish, daphnia and drosophila claimed chlorpyrifos to be the most toxic pesticides out of roundup, chlorpyrifos, and malathion. The adaptation strategy of living organisms to adverse effects depends significantly on the nature of the acting factor and the depth of its impact. The importance of battery of biomarkers for risk assessment when compare to individual indices was proved using CART and discriminant analysis and daphnia's protein carbonyls level and zebrafish's lactate dehydrogenase activity pertain to the most sensitive indices for group distinguishing. We propose to combine the most widely used in life sciences Principal Component Analysis with CART and discriminant analysis to better highlight the important biological entities and reveal insightful patterns in the data.

ACKNOWLEDGMENTS

The work was supported by the National Research Foundation of Ukraine (#2020.02/0270), Alexander von Humboldt Stiftung (H.F.), DAAD (DigIn.Net2), and Ministry of Education and Science of Ukraine (#MV-2 and BF/2-2021).

REFERENCES

[1] A. Agrawal, R. Pandey, and B. Sharma, "Water Pollution with Special Reference to Pesticide Contamination in India", Journal of Water Resource and Protection, Vol. 2, 2010, pp. 432-448.

[2] M. Liess, L. Liebmann, P. Vormeier, O. Weisner, R. Altenburger, D. Borchardt, W. Brack, A. Chatzinotas, B. Escher, K. Foit, R. Gunold, S. Henz, K. L. Hitzfeld, M. Schmitt-Jansen, N. Kamjunke, O. Kaske, S. Knillmann, M. Krauss, E. Küster, M. Link, M. Lück, M. Möder, A. Müller, A. Paschke, R. B. Schäfer, A. Schneeweiss, V. C. Schreiner, T. Schulze, G. Schüürmann, W. von Tümpling, M. Weitere, J. Wogram, and T. Reemtsma, "Pesticides are the dominant stressors for vulnerable insects in lowland streams", Water Research, Vol. 201, 2021, pp. 117262.

[3] A. Masiá, J. Campo, A. Navarro-Ortega, D. Barceló, and Y. Picó, "Pesticide monitoring in the basin of Llobregat River (Catalonia, Spain) and comparison with historical data", Sci. Total Environ., vol. 503-504, 2015, pp. 58-68.

[4] J. Liu, D. Zhang, Q. Tang, H. Xu, S. Huang, D. Shang, and R. Liu, "Water quality assessment and source identification of the Shuangji River (China) using multivariate statistical methods", PLoS ONE, vol. 16, 2021, e0245525.

[5] X. Zhang, Y. Zhang, Q. Zhang, P. Liu, R. Guo, S. Jin, J. Liu, L. Chen, Z. Ma, and Y. Liu, "Evaluation and Analysis of Water Quality of Marine Aquaculture Area", Int. J. Environ. Res. Public Health, vol. 17, 2020, pp. 1446.

[6] A. Mishra, "Assessment of water quality using principal component analysis: A case study of the river Ganges", J. Water Chem. Technol, vol. 32, 2010, pp. 227-234.

[7] H. I. Falfushynska, L. L. Gnatyshyna, and O. B. Stoliar, "In situ exposure history modulates the molecular responses to carbamate fungicide Tattoo in bivalve mollusk", Ecotoxicology, vol. 22, 2013, pp. 433-445.

[8] K. L. Newhart, "Environmental fate of malathion. California Environmental, Protection Agency Department of Pesticide Regulation", Environmental Monitoring Branch, 2006, [Online]. Available: http://www.bio-nica.info/Biblioteca/Newhart2006 Malathion.pdf.

[9] A. Ccanccapa, A. Masia, A. Navarro-Ortega, Y. Pico, and D. Barcelo, "Pesticides in the Ebro River basin: occurrence and risk assessment", Environ. Pollut, vol. 211, 2016, pp. 414-424.

[10] B. Rovenko, O. Kubrak, D. Gospodaryov, I. Yurkevych, A. Sanz, O. Lushchak, and V. Lushchak, "Restriction in glucose and fructose causes mild oxidative stress independently of mitochondrial activity and reactive oxygen species in Drosophila melanogaster", Comp. Biochem. Physiol., vol. 187, 2015, pp. 27-39.

[11] O. Bodnar, O. Horyn, I. Khatib, and H. Falfushynska, "Multibiomarker assessment in zebrafish Danio rerio after the effects of malathion and chlorpyrifos", Toxicol. Environ. Health Sci., vol. 13, 2021, pp. 165-174.

[12] P. J.Van den Brink, N. W. Van den Brink, and C. J. F. Ter Braak, "Multivariate analysis of ecotoxicological data using ordination: demonstrations of utility on the basis of various examples", Australasian journal of ecotoxicology, vol. 9, 2003, pp. 141-156.

[13] N. C. Ghisi, E. C. Oliveira, I. C. Guiloski, S. B. de Lima, H. C. Silva de Assis, S. J. Longhi, and A. J. Prioli, "Multivariate and integrative approach to analyze multiple biomarkers in ecotoxicology: A field study in Neotropical region", Sci. Total Environ., Vol. 609, 2017, pp. 1208-1218.

[14] P. K. S Shin and K. Y. S Fong, "Multiple Discriminant Analysis of Marine Sediment Data", Mar. Pollut. Bul., Vol. 39, 1999, pp. 285-294.