1-1-2011

Development of a Library of Responses for an Early-Warning Biomonitoring System to Detect and Identify Various Aquatic Contaminants

Aryo Maradona
Ryerson University

DEVELOPMENT OF A LIBRARY OF RESPONSES FOR AN EARLY-WARNING BIOMONITORING SYSTEM TO DETECT AND IDENTIFY VARIOUS AQUATIC CONTAMINANTS

By

ARYO MARADONA

B.Eng.Biosci in Chemical & Biological Engineering

(McMaster University, 2007)

A thesis

Presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the Program of

Environmental Applied Science and Management

Toronto, Ontario, Canada, 2011

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

_____

Aryo Maradona

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

_____

Aryo Maradona

ACKNOWLEDGMENTS

Starting this graduate thesis was akin to embarking on an exciting journey to the unknown, and such exploration would not have been fruitful without the assistance and guidance of many benevolent and talented individuals along the way.

First and foremost, I would like to thank my supervisors, Dr. Mehrab Mehrvar and Dr. Ron Pushchak, for constantly supporting my research work throughout these past two and a half years. My sincerest gratitude is also offered to other principals from our project team: Dr. Lynda McCarthy, Dr. Andrew Laursen, Dr. Vadim Bostan and Dr. Kimberley Gilbride. All six project principals had supplied me with immense amount of knowledge pertinent to their field, and I am truly grateful for the opportunity to work with them as a research assistant and graduate student. A very special appreciation is also extended to our laboratory technician, Karen Puddephatt, for her kind assistance in preparing the algal bioassay.

My family, friends, and fellow students were also a remarkable source of inspiration for this research. I am thankful to my mother for her unending support, patience and encouragement to believe in my own abilities, and to my father for his creative advice, wisdom and humour to keep my spirit high even during the most difficult times. I also extend my deepest gratitude to Stephen Jones for sharing his friendship and his mathematical expertise, as well as continuously motivating me throughout my studies. In addition, the staff at the International Student Services office, especially Esther Yip and Diana Ning, relentlessly offered me their guidance and fostered a very warm and welcoming community at Ryerson. Last but certainly not the least, I offer many thanks to my graduating cohort from the Environmental Applied Science & Management program for always enlightening me with their brilliant work and enthusiasm.

This research was financially supported by the Natural Science and Engineering Research Council of Canada (NSERC) through the provision of an NSERC Strategic Research Grant and Ryerson University.

This thesis is dedicated to my parents, Anita Hyacintha and Rachmad Pandia.

Your humble spirit and perseverance have always been my inspiration,

and to make both of you proud would be my greatest accomplishment.

ABSTRACT

Development of a Library of Responses for an Early-Warning Biomonitoring System to Detect and Identify Various Aquatic Contaminants

By:

Aryo Maradona
Master of Applied Science,
Environmental Applied Science and Management
Ryerson University
2011

Biomonitors can be implemented in aquatic ecosystems to continuously assess water quality, but existing monitors are still reliant on a single species and unable to identify any stressor. A library of responses could potentially address these drawbacks by stereotyping the responses of several aquatic species to different contaminants. A model for the library was developed by conducting a bioassay on *Pseudokirchneriella subcapitata* and collecting the response data of *Daphnia magna, Hyalella azteca* and *Lumbriculus variegatus* from published ecotoxicological studies. Multivariate statistical tools were then employed to process the response data set and evaluate the ability of the model to distinguish contaminations by atrazine and tributyltin. Based on preliminary tests, the library was able to detect and identify each contaminant within 4 hours with an accuracy of 97%. These findings supported the integration of a library of responses in a biomonitoring system to provide a more comprehensive water quality assessment.

TABLE OF CONTENTS

AUTHOR'S DECLARATION ii

ACKNOWLEDGMENTS iii

LIST OF TABLES x

LIST OF FIGURES xi

NOMENCLATURE xiii

CHAPTER 1 – INTRODUCTION 1

1.1 Objectives 5

1.2 Expectations 7

CHAPTER 2 – LITERATURE REVIEW 9

2.1 Introduction 9

2.2 Evaluation of Current Biomonitoring Technologies 9

2.2.1 Fish Biomonitors 9

2.2.2 Daphnid Biomonitors 12

2.2.3 Bivalve Biomonitors 14

2.2.4 Algal Biomonitors 17

2.2.5 ECOTOX 18

2.2.6 Multispecies Freshwater Biomonitor 21

2.2.7 Remarks on Current Biomonitoring Technologies 24

2.3 Current Progress of the NSERC Project 25

2.3.1 Background on Test Contaminants in Previous and Current Studies 28

2.3.2 Analysis of Bioassays by Marshall (2009) 31

2.3.3 Analysis of Bioassays by Pearce (2009) 35

2.3.4 Supplementary Algal Bioassay 38

2.4 Multivariate Statistical Tools 41

2.4.1 Principal Component Analysis 41

2.4.2 Cluster Analysis 48

2.4.3 Discriminant Analysis 53

2.5 Concluding Remarks 54

CHAPTER 3 – MATERIALS AND METHODS 57

3.1 Introduction 57

3.2 Experimental Setup for Online Algal Fluorometer 58

3.2.1 Procedures for Cleaning and Disinfection 58

3.2.2 Procedures for Algal Culturing 58

3.2.3 Procedures for Algal Bioassay 61

3.3 Distinct Conditions Established for the Present study 66

3.4 Methodology for Creating the Library of Responses 68

3.5 Methodology for Testing the Library of Responses 72

CHAPTER 4 – RESULTS AND DISCUSSION 73

4.1 Introduction 73

4.2 Results and Evaluation of the Algal Bioassay 73

4.2.1 Construction of the Algal Growth Curve 73

4.2.2 Influence of Atrazine and Tributyltin on the Effective Photosynthetic Yield of *Pseudokirchneriella subcapitata* 76

4.2.3 Performance Evaluation of the Algae Online Monitor™ 2800 83

4.3 Application of Multivariate Statistical Tools on the Endpoint Data Set 86

4.3.1 Results of Principal Component Analysis 86

4.3.2 Further Analysis of Endpoints Using a Biplot 96

4.3.3 Results of Cluster Analysis 100

4.3.4 Results of Discriminant Analysis 104

4.4 Potential Implementation of the Library of Responses 111

CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS 113

5.1 Conclusions 113

5.2 Recommendations for Future Research 115

REFERENCES 119

APPENDICES 129

APPENDIX A: Raw Data for Algal Bioassay 130

A.1 Sample Calculations for Preparing Test Solutions 130

A.2 Sample Calculations and Raw Data for Algal Growth Curve 133

A.3 Endpoint Data for Effective Photosynthetic Yield Experiments 137

APPENDIX B: Raw Data for Invertebrates Bioassay 139

B.1 *Daphnia magna* at t = 6 h 139

B.2 *Hyalella azteca* at t = 6 h 140

B.3 *Lumbriculus variegatus* at t = 6 h 141

APPENDIX C: MATLAB Codes 143

APPENDIX D: Intermediate Calculations in PCA 146

APPENDIX E: Raw Data for Discriminant Analysis 153

APPENDIX F: Raw Data for Average Silhouette Plot 156

LIST OF TABLES

Table 2.1. Chemical properties of atrazine, tributyltin and dimethyl sulfoxide.....29

Table 2.2. A series of endpoints measured and recommended by Marshall (2009) to evaluate the effects of atrazine and tributyltin on the selected aquatic organisms. 32

Table 2.3. A series of endpoints measured and recommended by Pearce (2009) to evaluate the effects of atrazine and tributyltin on the selected aquatic organisms. 36

Table 2.4. Summary of all endpoints from the present study and previous work by Marshall (2009) for inclusion in the library of responses. 40

Table 3.1. Stock nutrient solutions as recommended by Environment Canada (2007) for growing *Pseudokirchneriella subcapitata.* 59

Table 3.2. Three distinct conditions established for the library of responses 67

Table 4.1. Transformed non-normalized matrix of responses at t = 6 h. 87

Table 4.2. Transformed normalized matrix of responses at t = 6 h. 89

Table 4.3. Eigenvector coefficients which governed the contribution of each endpoint towards the construction of the transformed data set. 95

Table 4.4. Classification of the endpoint measurements at t = 2 h using the current library of responses. 106

Table 4.5. Classification of endpoint measurements at t = 4 h using the current library of responses. 109

LIST OF FIGURES

Figure 2.1. The application of Principal Component Analysis where data points in a three-dimensional space are described using a two-dimensional plane. ...43

Figure 2.2. Flowchart of the data-reduction procedure using PCA 46

Figure 2.3. Some possible patterns found in a data set after simplification using principal component analysis (Adapted from Shaw, 2003).....49

Figure 2.4. Algorithms for K-means clustering (Adapted from Govaert, 2009).....52

Figure 2.5. Graphical illustration of the discriminant analysis technique (adapted from Dillon and Goldstein, 1984).....55

Figure 3.1. Experimental setup and chemical composition for the test solutions prepared for the algal bioassay.....62

Figure 3.2. Instrumental setup for the algal bioassay. 64

Figure 4.1. Growth curve of *Pseudokirchneriella subcapitata* used for the algal bioassay in the present study.....75

Figure 4.2. Effective photosynthetic yields of *Pseudokirchneriella subcapitata* measured over a period of 6 h in dechlorinated tap water with and without 0.1% v/v dimethyl sulfoxide. 77

Figure 4.3. Effective photosynthetic yields of *Pseudokirchneriella subcapitata* measured over a period of 6 h in dechlorinated tap water and in various solutions of atrazine.....79

Figure 4.4. Effective photosynthetic yields of *Pseudokirchneriella subcapitata* measured over a period of 6 h in dechlorinated tap water and in various solutions of tributyltin. 81

Figure 4.5. A plot of the transformed non-normalized matrix of responses containing endpoints at t = 6 h. 88

Figure 4.6. A plot of the transformed normalized matrix of responses containing endpoints at t = 6 h. 90

Figure 4.7. Biplot of the unit vectors formed by the coefficients in Table 4.3 and data points scaled by a factor of 1/200 from Table 4.1..... 98

Figure 4.8. Classification of data points in Figure 4.5 using K-means clustering. ... 101

Figure 4.9. Average Silhouette Plot describing the cluster quality for data points in Figure 4.8. 103

Figure 4.10. Plot of the endpoint measurements at t = 2 h against the library of responses. 105

Figure 4.11. Plot of the endpoint measurements at t = 4 h against the library of responses. 108

Figure 4.12. Illustration of the early-warning biomonitoring system in action where (a) depicts the overall system configuration, and (b) displays the MATLAB program to process the endpoint data. 112

NOMENCLATURE

Photosynthetic Parameters

F	Fluorescence
F_m	Maximum fluorescence (dark-adapted)
F_m'	Maximum fluorescence (light-adapted)
H	Non-radiative heat dissipation
P	Photochemical conversion
Y	Photosynthetic yield

Principal Component Analysis

a	Eigenvector coefficient
i	Component number
j	Observation number
k	Number of variables in original data set
x	Magnitude of original measured endpoint
z	Coordinate in the principal component

Cluster Analysis

m_i	Average distance for point i to all other points in the same cluster
n_i	Average distance for point i to all other points in the nearest cluster

S_C	Silhouette coefficient

Abbreviations

AOM	Algae Online Monitor
CA	Cluster Analysis
DA	Discriminant Analysis
DMSO	Dimethyl sulfoxide
EWS	Early-Warning System
EWBS	Early-Warning Biomonitoring System
iABS	Intelligent Aquatic Biomonitor System™
IPF	*In-vivo* Prompt Fluorescence
MFB	Multispecies Freshwater Biomonitor
NSERC	Natural Sciences and Engineering Research Council of Canada
PAM	Pulse-amplitude modulation
PCA	Principal Component Analysis
PC1	First Principal Component Axis
PC2	Second Principal Component Axis
PSII	Photosystem II
US EPA	United States Environmental Protection Agency
USACEHR	United States Army Center for Environmental Health Research

CHAPTER 1

INTRODUCTION

Freshwater is an integral component of human lives, yet the availability of clean and potable water is constantly threatened by a variety of factors. Traditionally, many human settlements have been concentrated near a lake or a river due to the convenience of having a nearby freshwater supply. Close proximity to a water body also increased the accessibility of the settlements and facilitated the transport of goods and raw materials through shipping. As the populations of settlements grew and more water-side communities were established, the risk of contamination due to anthropogenic activities also increased.

According to a survey conducted in 1999 among 153 water providers in the United States and Canada, the most common causes of water contamination can be attributed to transportation accidents, pipeline and storage tank leaks, pesticides from agricultural runoff, and pathogenic microbes from untreated sewage (Gullick *et al.*, 2003). Water contamination also occurred frequently according to the 2009 Spill Action Report published by the Ontario Ministry of Environment which recorded 1,162 spills into water bodies in Ontario that year, with 595 of these cases confirmed to have caused a significant environmental impact (Ontario Ministry of the Environment, 2010). Although far more unlikely, the United States Environmental Protection Agency (US EPA) also recently evaluated the heightened concerns of terrorist attacks through the release of harmful chemical, microbial or radioactive materials into the national water infrastructure

and suggested the installation of an Early-Warning System (EWS) to mitigate these risks (US EPA, 2005). Based on the various threats highlighted above, it is evident that a strategy must be developed to better protect and preserve freshwater sources as well as drinking water reservoirs and distribution networks.

The installation of an EWS in water-monitoring stations is one of the recommended strategies for water quality protection. An EWS typically comprises sensors to detect contaminants, a computer unit to process incoming data, an alarm system to notify the operators, and protocols to facilitate decision-making and emergency responses (US EPA, 2005). A number of criteria has been proposed for the development of a robust and reliable EWS (US EPA, 2005; ILSI, 1999; Grayman, 2004; Hasan, 2004), and the system must be:

- Accurately and rapidly responsive to changing water quality with minimal false-positives and false-negatives;
- Capable of providing real-time, continuous measurements and detection for a wide range of potential contaminants;
- Mostly automated with little need for human supervision;
- Inexpensive to install, maintain and upgrade;
- Easy to operate by low-to-moderately skilled technicians; and
- Capable of identifying the source, type and concentration of the contaminant.

A technological gap exists because no system currently satisfies all of the above recommendations (US EPA, 2005). Water-monitoring stations typically only measure

common physico-chemical parameters such as temperature, dissolved oxygen content, conductivity, pH, and turbidity as well as other target chemicals which vary among different jurisdictions (Evans *et al.*, 1986; Roig *et al.,* 2007; Dort, 2010). Physico-chemical parameters alone do not describe the composition of chemicals which may be present in the water and are not adequate for measuring toxicity. Some real-time contaminant monitors have been installed in certain rivers to detect a number of predetermined target compounds (*e.g.*, volatile organic compounds), but the range of detection for these monitors is very narrow and thus renders the monitors inefficient (Dort, 2010; Calder, 2010).

When more thorough analyses are required, spot-sampling must be conducted where a sample of water is sent to an off-site laboratory for chemical or microbial analysis (Roig *et al*., 2007). A variety of analytical instruments, such as spectrophotometers, gas chromatographs, high-performance liquid chromatographs or mass spectrometers, are often employed to identify the chemical species in the sample. These instruments generally require highly-trained professionals to operate, involve large capital and maintenance costs, and are unable to detect all known chemicals (Van der Schalie *et al.*, 2004; Gerhardt *et al*., 2006). Furthermore, there is a delay between sample collection and the conclusion of the analysis which can vary from hours to days (Miller *et al.*, 1985). This delay presents an additional risk as a contaminant may potentially travel undetected downstream and reach a drinking water intake.

Another approach to water quality monitoring is the use of biological organisms as components of an online sensor system to detect contaminants. Similar to a miner's canary, a variety of aquatic organisms display rapid changes in physiology and behaviour upon exposure to acute, fast-acting toxins (Streb *et al.*, 2002). A water safety operator can thus indirectly assess the water quality by monitoring a variety of endpoints exhibited by these aquatic organisms. When a deviation occurs in these endpoints, an alarm is raised and appropriate mitigative actions can then be taken.

A number of bioassays have evaluated the use of fish (Cairns *et al.*, 1970; Roig *et al.,* 2003; Van der Schalie *et al.*, 2004; Van der Schalie *et al.*, 2006; bbe Moldaenke, n.d.b), cladocerans (Knie, 1985; Green *et al.*, 2003; Jeon *et al.,* 2008), bivalves (Gunkel and Streit, 1980; Kramer and Foekema, 2001; White *et al.*, 2001; Borcheding, 2006), and algae (Pandard *et al.*, 1993; Osbild *et al.*, 1995; Durrieu *et al.*, 2006; Fai *et al.*, 2007) in biomonitoring systems. A major drawback to the biomonitors in these studies is their reliance on a single species which may lead to an overestimation or underestimation of risks posed by a certain contaminant. Employing only a single species will severely compromise the integrity of the EWS as many organisms exhibit varying levels of sensitivities to different contaminants (Bunn, 1995). To address this issue, a number of water monitoring stations in Germany and the Netherlands have installed several biomonitoring systems including multiple species of various trophic levels (Diehl *et al.*, 2006). Despite the system redundancy, there has been no published work to characterize the responses of these biomonitors to different contaminants, and as a result, current

biomonitoring systems are only capable of indicating the occurrence of a pollutant without identifying the actual contaminating agent.

1.1 Objectives

The present study is a component of a large-scale project funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) – Strategic Research Grants Program to develop a holistic, real-time, multi-organism Early-Warning Biomonitoring System (EWBS). A novel feature of this biomonitoring system is its ability to more accurately detect and identify contaminants in a sample of water. The proposed EWBS could potentially bridge the aforementioned technological gap in water quality monitoring as the system satisfies most of the previously-mentioned criteria for a robust and reliable early-warning system.

A number of studies have been previously conducted for the development of the EWBS. Marshall (2009) examined the changes in the behavioural responses (movements) and respiration rates of *Daphnia magna, Hyalella azteca,* and *Lumbriculus variegatus* when exposed to varying concentrations of atrazine and tributyltin over a brief time period. Similarly, Pearce (2009) monitored various endpoints of *Lemna minor*, *Pseudokirchneriella subcapitata, Euglena gracilis,* and *Anodonta grandis* to different levels of atrazine and tributyltin. The objectives of these studies were to evaluate the suitability of the selected aquatic organisms for use in the EWBS and to determine which endpoints were particularly sensitive to the two contaminants. By examining the findings

from the previous studies and supplementing the collected data with another experiment using an online algal fluorometer, the present study intended to investigate whether the responses exhibited by a suite of aquatic organisms to different contaminants could be characterized. Ultimately, the present study aimed to create a model for a library of responses which can be incorporated into the design of the EWBS to provide a more accurate detection and identification of contaminants.

In summary, the objectives of the present study are:

- To analyze all previous bioassays and to evaluate the suitability of each endpoint for inclusion in the library of responses for the EWBS.
- To supplement the endpoint data collected from the previous studies by conducting another bioassay involving the green algae *Pseudokirchneriella subcapitata.* An online algal monitor was used to measure the effective photosynthetic yield of the algae, and the performance of this instrument was also evaluated.
- To characterize the behavioural and physiological responses collected from the previous and current studies according to test contaminants. A number of multivariate statistical tools were employed to reduce the variable interdependence and to highlight the greatest variance within the data set.
- To establish a preliminary library of responses and to provide suggestions for the future expansion of this library.

1.2 Expectations

In theory, contaminants with distinct modes of action should elicit different sets of reactions in a test organism. Furthermore, different test organisms may also exhibit varying sensitivities to a particular contaminant, resulting in various species-specific dose-response relationships. Analyzing this network of interactions between various species and doses of contaminants would be cumbersome, and thus a series of data-simplification tools were applied to facilitate data analysis. Overall, it was expected that the aquatic organisms selected in the present study would exhibit distinct patterns of responses to each test contaminant, and this finding could be incorporated in the construction of the library of responses.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The overall objective of this NSERC project is to develop and implement a real-time early-warning system for drinking water facilities that is capable of detecting chemical contaminants using a suite of biomonitoring organisms. Prior to achieving this objective, further theoretical background information on current biomonitoring organisms and technologies must be explored. The present chapter also examines the strengths and caveats of many commercially-available biomonitors. A few of these biomonitors were in fact obtained for this NSERC project and evaluated by past researchers (Fleet, 2010; Netto, 2010). This chapter then provides a progress summary of the NSERC project based on the findings of past researchers and outlines the strategies to accomplish the objectives of the present study. Finally, some of the challenges in developing the library of responses for the EWBS are the large numbers of variables and data observations, and thus a series of multivariate statistical tools must be employed to facilitate data analysis. These multivariate statistical tools are elaborated as well in this chapter.

2.2 Evaluation of Current Biomonitoring Technologies

2.2.1 Fish Biomonitors

The research for an automated biomonitoring system started as early as the beginning of the 1970s. In a study by Cairns *et al.* (1970), a goldfish was placed in an aquarium equipped with several light beams and photoreceptors at various levels to

provide a rudimentary method of monitoring the swimming pattern of the fish. By counting the number of light beam interruptions, the swimming activity of the fish was assessed. Simple polygraph kits were also set up to measure the heart rate and breathing pattern of the fish. The results of this study supported the use of fish as an indicator species in a biomonitoring system.

Recent refinements in fish biomonitoring systems have introduced video-based technologies to track other endpoints such as swimming speed, turning rate and swarm formation. The bbe Fish Toximeter (bbe Moldaenke GmbH, Kiel, Germany) consists of a tank where a number of fish are observed by a camera, and the recording is analyzed using an image-analysis software (bbe Moldaenke, n.d.b). The unit also minimizes the probability of a false-alarm by simultaneously monitoring several parameters such as swimming speed, swimming behaviour (*e.g.* height, the number of turns and circular motions), size of each fish, number of active fish, and the location of the fish in the aquarium. A drawback to many video-based biomonitoring systems, however, is the need to pre-treat the water because high turbidity levels can affect the ability of the camera to track the movement of the animals (Gerhardt *et al.*, 2006). Alternative technologies employing non-visual endpoint measurements have since been developed.

The United States Army Center for Environmental Health Research (USACEHR) has conducted extensive research to develop a portable online drinking water monitoring system which is capable of rapidly detecting a broad range of chemicals such as cyanide, malathion, pentachlorophenol, phenol, tetrachloroethane, tricaine methanesulfonate, and

zinc (United States Army Center for Environmental Health Research, 2004; Van Der Schalie, 2004). This system was designed to identify any potential threat of bioterrorism, and the equipment was dubbed the Intelligent Aquatic Biomonitor System™ (iABS) (IAC 1090, Honeywell Corporation, USA). Preliminary experiments measured the ventilatory rate, ventilatory depth, cough rate and movement of bluegills (*Lepomis macrochirus* Rafinesque) upon exposure to a variety of toxicants at high concentrations approaching the 96-h LC_{50} concentrations of the species (Van Der Schalie *et al.,* 2001; Van Der Schalie *et al.*, 2004). The bluegills were placed in individual chambers with carbon block electrodes placed at the top and bottom of each chamber, and as the fish ventilated their gills, the electrodes received electrical signals from the muscular contractions and recorded the frequency. The study found that the responses by the fish to the toxicants were rapid and consistent with the high exposure levels, and that different toxicants even resulted in unique patterns of responses.

Despite the reported success with the instrument, there are still very few peer-reviewed publications evaluating the performance of the iABS, and thus it is difficult to objectively assess the efficacy of the system. The study by Van Der Schalie *et al.* (2004) utilized toxicants at very high concentrations, but such high levels are unlikely to be found in the environment. Roig *et al.* (2003) cited the relatively low chemical sensitivity of fish as bioindicator organisms, and thus higher concentrations of contaminants are typically required in fish bioassays. Furthermore, the chambers for the fish in the iABS are very small and this limited containment may induce additional stress on the fish. Gerhardt *et al.* (2006) also raised some ethical concerns with placing higher organisms

such as fish in a very confined space as the organisms would be constantly under stress due to the lack of space for manoeuvring. Due to the limited number of reliable fish-based biomonitoring systems as well as the relative insensitivity of fish to contaminants at environmentally-relevant concentrations, the current NSERC project does not recommend the use of fish in the proposed early-warning biomonitoring system.

2.2.2 Daphnid Biomonitors

Daphnids have been studied extensively in various environmental bioassays, and as a result, there is a large amount of ecotoxicological data for the organism (Kieu et al., 2001; Kiss et al., 2003; Schmidt et al.,2005; Gerhardt *et al.,* 2006; Ren et al., 2009). One of the earliest implementations of daphnids in a biomonitoring system was the Dynamic Daphnia Test developed by Knie (1982). Similar to the fish test by Cairns *et al.* (1970), this test attempted to monitor the swimming activity of several daphnids in an aquarium by placing several infrared light beams at various heights and subsequently counting the number of beam interruptions. This setup was partly unsuccessful because of several technical interferences. One hyperactive animal crossing through several beams would yield a response pattern that was similar to numerous less active animals passing through fewer beams. In order to address this counting error, a number of technologies have integrated digital image analysis to track the location and record the swimming activity of individual organisms.

The bbe Daphnia Toximeter II (bbe Moldaenke GmbH, Kiel, Germany) was developed based on the Dynamic Daphnia Test (bbe Moldaenke, n.d.a). The instrument

obtains a live image of each daphnid using an onboard surveillance camera, and the endpoints are then calculated using digital image-analysis software. The Daphnia Toximeter tracks a variety of endpoints, and an alarm is raised when a statistically significant departure is exhibited by two or more endpoints. The bbe Daphnia Toximeter II was deployed during the Olympic Games of 2002 in Salt Lake City, Utah, where the instrument monitored the water supply and distribution system for any deliberate contamination (Green *et al*., 2003).

A study by Green *et al*. (2003) evaluated the performance of the bbe Daphnia Toximeter in detecting chemical warfare agents in a water sample. When exposed to varying concentrations of sarin, soman, tabun and cyclosarin, the daphnids exhibited relatively rapid changes in behaviour, indicating the reaction of the organisms to the toxicants. The study, however, noted that the daphnids were very sensitive to a change in pH as well as high levels of chlorine in the water. Since water distribution systems often involve chlorination, dechlorination is required if the toximeter is to be employed to monitor the water. Gerhardt *et al.* (2006) also noted the hypersensitivity of daphnids to a variety of physico-chemical parameters of their environment such as pH, temperature and dissolved oxygen level, and this extreme sensitivity presents a significant drawback in using daphnids as biomonitors.

Despite reported hypersensitivities of daphnids in various ecotoxicological studies, the current study still recommended the use of this organism due to its acute sensing ability. To avoid false readings and minimize any experimental noise, a number

of design factors must be carefully controlled. First, the water used during experiments must be pre-treated to eliminate any disruptive particles. Furthermore, since daphnids require continuous feeding by a stable algal culture, a consistent procedure for growing the algae must be established to prevent any experimental uncertainty (Gerhardt *et al.*, 2006). Last, multiple trials are recommended to ensure that experimental results are replicable and not affected by other unaccounted factors. By following the above recommendations, daphnids could be implemented as a sensing organism in a biomonitoring system.

2.2.3 Bivalve Biomonitors

Bivalves are another type of aquatic organism which has been extensively evaluated for use in a biomonitoring system (Gunkel and Streit, 1980; Kramer and Foekema, 2001; White *et al.*, 2001; Borcheding, 2006). Gunkel and Streit (1980) found that bivalves filtered not only large volumes of water but also suspended particulates, and therefore these organisms would be an excellent indicator for hydrophobic pollutants. A typical measured endpoint in a bivalve monitoring system is the opening and closing of the shell which is related to the health of the organism (Kramer and Foekema, 2001).

An example of a bivalve-based biomonitor is the Dreissena-Monitor. In 1989, the 'German Commission for the Protection of the Rhine Against Pollution' (DKRR) initiated a research project to develop, test, and install biomonitors along the River Rhine and its tributaries (Bund/Länder-Projektgruppe 'Wirkungstests Rhein', 1994). The research project resulted in the installation of the Dreissena-Monitor in 1991 in the city of

Bergheim to monitor the River Erft, which is one of the tributaries of the Rhine (Borcherding, 2006). Presently, a number of other Dreissena-Monitors have also been installed in other municipalities such as Hattingen, Frödenberg and Amsberg to monitor the River Ruhr; and Jochenstein and Bad Abbach to monitor the River Donau. A complete diagram showing all operational sites of the Dreissena-Monitor in Germany is published by Borcherding (2006).

The Dreissena-Monitor employs eighty-four zebra mussels (*Dreissena polymorpha*) which are distributed equally into two parallel channels, and a reed switch and magnet which are attached to the top valve of each mussel to record the number of open and closed valves (Borcherding and Volpers, 1994). The percentage of open mussels and the number of valve movements per mussel per hour are computed as running averages for each channel. Dynamic limits are also calculated for each channel such that, if the latest measurement plus the three-fold standard deviation is lower than the previous measurement, an alarm is signalled.

While the measurement of valve movements in the Dreissena-Monitor is binary (*e.g.*, 1 = open, 0 = closed), a newer bivalve biomonitoring technology developed in the Netherlands measures shell opening as a continuous variable. Dubbed as the Mossel Monitor™, this instrument attaches an emitter and receiver on each shell of the mussel in order to record the width of the shell opening (White *et al.*, 2002). Due to this modification, the Mossel Monitor™ is capable of providing more extensive information on the mussel activities.

The biggest drawback in using bivalves in a biomonitoring system is the difficulty in distinguishing whether any behavioural changes in the mussels are caused by the presence of a contaminant, or if the drift is simply a natural variation of the behaviour of the organisms. Based on the data collected over the past ten years with the Dreissena-Monitor, the valve movements and the percentage of open mussels change regularly, and thus the baseline parameters must be adjusted accordingly to prevent false alarms (Borcherding, 2006). An analysis of the mussel behaviour throughout the year also reveals a direct relationship between temperature and the number of valve movements. As a result, during summer times when the water is warmer, there is > 3-fold increase in valve movement activity (Borcherding, 2006).

Because of the large variation in the activity of the bivalves as well as the difficulties in characterizing the general behaviour of the organism, implementing bivalves in a biomonitoring requires extensive preliminary tests. A preliminary study on bivalves for the current NSERC project was conducted by Pearce (2009) where the percentage of open valves and the respiration rates of a group of *Anodonta grandis* were measured. The study reported some difficulties in maintaining the culture as well as a number of inconsistencies among the percentages of open valves and the concentration of toxicants. Due to the poor results in the preliminary tests, bivalves are not recommended at the moment for inclusion in the present study. Additional experiments, however, may be considered in the future to further assess the implementation of this organism as a biological sensor.

2.2.4 Algal Biomonitors

The use of aquatic plants and algae in biomonitoring systems has been extensively studied due to the sensitivity of the photosynthetic activity of aquatic plants to various contaminants (Pandard *et al.*, 1993; Osbild *et al.*, 1995; Durrieu *et al.*, 2006; Fai *et al.*, 2007). One of the earliest algal biomonitor experiments was conducted by Pandard *et al.* (1993) where photosynthesis rates were quantified using an oxygen electrode. Measuring only oxygen production, however, may lead to inaccuracies in the detection of contaminants because other factors such as nutrients and ambient conditions may also affect the rates of oxygen production and release. In order to improve the accuracy, subsequent algal biomonitors have incorporated other methods to measure the photosynthetic activities such as by quantifying the amount of *in-vivo* prompt fluorescence (IPF). Other plant endpoints which have been studied are cell growth, delayed fluorescence and sometimes motility (for example, in the case of the flagellate *Euglena gracilis*) (Osbild *et al.,* 1995; Tahedl and Häder, 1999; Tahedl and Häder, 2001; Gerhardt *et al.,* 2006).

IPF is a short illumination pulse which determines the biomass and the physiological state of the organism (Osbild *et al.,* 1995). When chlorophyll is irradiated by a short pulse of light, the energy is either absorbed to drive photosynthesis or re-emitted as radiation or fluorescence (Strasser *et al.,* 2000). When a plant is affected by the presence of a contaminant, the ability of the plant to photosynthesize could be compromised, and thus the ratio of light absorbed versus light re-emitted, or the photosynthetic yield, would change (Berden-Zrimec *et al.,* 2007). Effective

photosynthetic yield can be measured using a pulse-amplitude modulation (PAM) fluorometer where a suspension of algae is irradiated using a short pulse of light of a specific wavelength for a brief period of time. A study was conducted by Fai *et al.* (2007) where the changes in the IPF and photosynthetic yield of *Pseudokirchneriella subcapitata* were analyzed after exposing the algae to different types and concentrations of herbicides. The study found the algal fluorescence to be a quick indicator of stress in plants and could be used in the screening of environmental samples (Fai *et al.*, 2007).

Based on published studies, aquatic plants and algae could potentially be employed as a rapid sensing organism in a biomonitoring system. For the current NSERC project, preliminary algal bioassays were conducted by Pearce (2009) where the respiratory and physiological parameters of *Lemna minor, Pseudokirchneriella subcapitata,* and *Euglena gracilis* were measured in response to varying levels of atrazine and tributyltin. Pearce (2009) found a number of endpoints to be useful in providing rapid water quality assessment and recommended the inclusion of aquatic plants in the EWBS. The present study aimed to verify these recommendations by conducting a similar algal bioassay.

2.2.5 ECOTOX

The ECOTOX is a commercially-available automated biotest system which records different movement parameters of the motile, unicellular flagellate *Euglena gracilis*. *Euglena gracilis* possesses a single flagellum at its front end which controls its movement and orientation according to external factors such as light and gravity (Tahedl

and Häder, 1998). In the absence of light, the organisms exhibit negative gravitaxis where they swim upwards toward the water surface. At a low level of light exposure, however, the organisms tend to travel toward any light source until a certain irradiance threshold is reached, and after which they start to exhibit negative phototaxis. In their natural habitat, *Euglena gracilis* maintain their orientation in the water column through an antagonistic balance between negative gravitaxis and negative phototaxis (Netto, 2010).

Experiments with ECOTOX have shown that the gravitactic orientation is the most sensitive movement parameter with significant deviations recorded in as little as two minutes after the addition of different toxins (Tahedl and Häder, 1998). This behaviour was attributed to the influence of toxins on the stretch-sensitive ion channels located asymmetrically on the cell membrane which act as gravireceptors (Lebert and Häder, 1996). Streb *et al.* (2002) also found movement and orientation responses of the organism to be very susceptible to the presence of heavy metals and organic compounds.

By monitoring the orientation of *Euglena gracilis*, the ECOTOX can detect contaminants in a water sample (Netto, 2010). The instrument consists of three stock chambers which contain the test organisms in water, a distilled water reservoir for dilution and rinsing, and the test solution to be analyzed. The three stocks are pumped into the mixing chamber according to the ratio set by the operator, and a small portion of this mixture is transferred to a cuvette made of an opaque stainless steel frame to exclude phototaxis. To visualize the organisms, an infrared diode with a wavelength of 875 nm is

used as a light source, and the image is recorded using a combination of a miniaturized microscope and camera. A system with a similar setup can also be used to measure the behavioural parameters of daphnids, and this modified instrument is dubbed the DaphniaTox (Netto, 2010).

The ECOTOX can be used as a rapid biotest in an early-warning system due to the very short time required for a complete measurement (Tahedl and Häder, 1999; Tahedl and Häder, 2001). The efficacy of the instrument for longer-term bioassays had also been reported where the motility and photosynthetic parameters of *Euglena gracilis* were monitored after twenty-four hours of exposure to various concentrations of nickel (Ahmed and Häder, 2010). Other claimed major advantages were the small size of the instrument, the reliability of the image analysis, the ability to simultaneously monitor several endpoints, and the automation of the instrument. The majority of published studies to date, however, were authored or co-authored by at least one of the inventors of the instrument (*i.e.* Harald Tahedl and Donat-P. Häder). This strong bias may have hindered any objective assessment of the performance of the ECOTOX based on literature reviews alone.

Netto (2010) examined the usability of the ECOTOX and DaphniaTox in monitoring the behavioural parameters of selected aquatic organisms to varying levels of atrazine, tributyltin, and copper. Visual observations revealed some contaminant-specific sensitivities of the organism where the cell shape and motility were significantly different from the reference condition at higher levels of exposure to copper but no significant

difference when exposed to atrazine and tributyltin. Based on this finding, *Euglena gracilis* may be an excellent indicator to differentiate contamination by metals and organic contaminants, but further research must be conducted to investigate the discriminating ability of the organism for the two types of contaminants (Netto, 2010).

Despite the claimed advantages of the ECOTOX and DaphniaTox, Netto (2010) reported some major hardware and software issues which hindered the performance of both instruments. The image analysis software in the ECOTOX was not able to adequately capture the movements of the *Euglena gracilis*, and the study on the sensitivity of the ECOTOX was thus considered inconclusive. Similarly, the DaphniaTox system was not functioning at all due to a major instability in the associated image analysis software. Based on the above findings, Netto (2010) recommended some major modifications for both the ECOTOX and the DaphniaTox before the systems could be implemented in a water quality testing facility.

2.2.6 Multispecies Freshwater Biomonitor

All other previously-discussed biomonitoring systems have been single-species systems where only one type of aquatic organism could be employed at a time or a signal disturbance would occur. The Multispecies Freshwater Biomonitor (MFB), however, is capable of measuring the behavioural responses of several different species at the same time (Gerhardt *et al.*, 2006). Despite the claim that the MFB was capable of monitoring the behavioural endpoints of several species simultaneously, however, there has been no

research documenting such an experimental setup. Currently, there are only published studies documenting the performance of the MFB in single-species experiments.

The instrument comprises a computer, a processing unit, and several individual test chambers in which different aquatic organisms could be placed and monitored. Two pairs of electrodes are installed on the walls of each test chamber where the first pair generates a high-frequency alternating current of 100 kHz, and the non-current carrying second pair senses any change in impedance due to the movement of the organisms. By changing the band of the instrument, different types of movements such as locomotion and ventilation can be tracked and recorded (Kirkpatrick *et al.,* 2006). Because the MFB measurements are non-optical, there is no need to pre-treat the incoming water for sediments or other impurities (Gerhardt *et al.*, 2006). This advantage allows the MFB to provide a realistic representation of both dissolved and particle-bound pollutants, as well as their synergistic effects. The impedance measurement also enables the MFB to monitor the movements of organisms buried in sediments because the measurements are unaffected by non-living materials (Gerhardt *et al.,* 2003).

In-situ tests have been reportedly conducted in monitoring stations along the Rhine River, France; Aller River, Germany; and Meuse River, the Netherlands using *Gammarus pulex* (Gerhardt *et al.,* 2003; Gerhardt *et al.,* 2007). Other single-species experiments examined the movement patterns of *Caenorhabditis elegans* (nematoda) to study their behavioural parameters in sediments (Gerhardt *et al.*, 2002), as well as the assessment of behavioural changes of *Corophium volutator* in response to toxicant

exposure in sediment (Kirkpatrick *et al.*, 2006). Interestingly, successful non-aquatic application of the MFB has also been reported where Bednarska *et al.* (2010) examined the locomotor activity of the ground beetle *Pterostichus oblongopunctatus* to determine any links between the behavioural and physiological activities of the beetle.

In all aforementioned studies, the respective organisms were sensitive to the tested toxicants, and the use of each organism was recommended in future MFB experiments. Similar to published studies involving ECOTOX, however, the majority of publications advocating the use of the MFB are also authored or co-authored by the inventor of the biomonitor (Dr. Almut Gerhardt). The objectivity of these studies may therefore be questioned, and the findings must be carefully considered.

Fleet (2010) examined the use of the MFB to assess the behavioural change of *Daphnia magna* and *Hyalella azteca* upon exposure to varying levels of atrazine and tributyltin. The behavioural endpoints of both organisms were previously studied in bioassays containing atrazine and tributyltin, and the organisms were found to be excellent indicators due to their sensitivities, even at low, environmentally-relevant concentrations (Marshall, 2009). The MFB, however, was unable to detect any behavioural changes from the *Daphnia magna* and *Hyalella azteca*, even at high contaminant concentrations. The organisms also exhibited large variations in behaviour when studied using the MFB, resulting in major difficulties to reproduce and replicate experiments. An attempt by the researcher to contact Dr. Almut Gerhardt to verify the

conflicting results was also unsuccessful, and as a result, the MFB was deemed unsuitable for use in a multi-species early-warning biomonitoring system (Fleet, 2010).

2.2.7 Remarks on Current Biomonitoring Technologies

Sections 2.2.2 – 2.2.6 have outlined a number of published studies involving some of the latest commercially-available biomonitoring technologies. When compared against the recommendations listed in Chapter 1 for a robust and reliable early-warning system, none of these biomonitors meets all of the requirements. For example, the ECOTOX and the DaphniaTox are not flow-through systems, and thus these instruments would require frequent sampling or a major modification to allow flow-through modes. Many of the biomonitors are optical-based, and thus some pre-treatment of the incoming water might be necessary to reduce turbidity and remove excess sediment.

In addition, the selected organism must be sensitive enough to low and environmentally-relevant levels of contamination, but the exhibited stress response must be consistent and replicable. The fish biomonitoring system is not sensitive enough to low levels of contaminants, and thus such a system is only useful for monitoring sudden, large spikes of contamination in the water supply. Conversely, bivalves are quite sensitive to low concentrations of contaminants, but their behavioural parameters are reported to be largely variable according to season and other water parameters. More studies where the water parameters are carefully controlled must therefore be conducted to fully characterize the behaviours of bivalves before employing these organisms in a water monitoring facility. Extreme caution must also be exercised when selecting a

particular technology, as published results may be biased, and objective evaluations of the technology must be obtained if available.

As mentioned earlier, these biomonitoring systems rely mostly on the endpoints of a single aquatic species, and thus despite some reported success in detecting contaminants, there are still issues associated with species-specific sensitivities. Single-species biomonitoring systems are also only capable of sensing the presence of a contaminant but not identifying the source, type and concentration of the contaminant. While it is impossible to test for every single contaminant and potential threat to aquatic systems, many chemicals can be grouped together according to their physico-chemical, structural and functional classifications. By exploiting species-specific sensitivities, it may be possible to create a fingerprint of responses and characterize the responses of the organisms to different contaminants. This hypothesis forms the basis of the current research, which is to develop a library of responses of aquatic organisms to multiple contaminants and to analyze for any specific patterns of responses which would help distinguish different types of contaminants and potentially estimate the concentration of the contaminant.

2.3 Current Progress of the NSERC Project

As previously stated, the large-scale NSERC project aims to develop a multi-species early-warning biomonitoring system (EWBS) that would detect and identify different chemical contaminants and pathogens. This EWBS is aimed to be fully

implemented in a water treatment facility in the Niagara Region within the next few years. In order to achieve this long-term goal, the project is divided into a few smaller objectives:

1. To measure responses in aquatic plants and invertebrates when exposed to pathogens and chemical stressors at environmentally-relevant concentrations;
2. To develop a microarray-based test to directly detect the presence of pathogens as well as a UV-based system to treat the pathogens;
3. To build and test a biomonitoring system for measuring stress-response in real-time; and
4. To develop a profile of stereotyped responses for the whole suite of biomonitoring organisms to chemical contaminants and pathogens.

A number of studies and experiments have been conducted by previous graduate students to meet the objectives listed above. Marshall (2009), Pearce (2009), and Dort (2010) accomplished a portion of the first objective by measuring the effects of tributyltin, atrazine and copper at various concentrations on a variety of aquatic plants and invertebrates. Upon the completion of their studies, Marshall (2009) and Pearce (2009) recommended a set of endpoints for a subset of the species tested which were sensitive to the selected levels of atrazine and tributyltin, and some of their suggestions were incorporated into the experimental designs of Netto (2010) and Fleet (2010). As elaborated in Sections 2.2.5 and 2.2.6 of this thesis, Netto (2010) and Fleet (2010) evaluated the use of several automated biomonitoring systems, and the results of their

studies concluded that their respective instruments were inadequate in providing real-time stress-response measurements as currently configured.

For the second objective, Barrera (2011) examined different photochemical treatments for pathogens, while Clark (2010) designed and validated oligonucleotide primers for detecting waterborne bacterial pathogens, resulting in primer sets with high specificity and sensitivity for *Escherichia coli* O157:H7, *Salmonella Typhimurium*, *Campylobacter jejuni*, *Pseudomonas aeruginosa*, and *Shigella flexneri* (Clark, 2011). Other studies relevant to the project are the analysis of land use and potential sources of contaminants surrounding the Welland Canal (Labbaf, 2010), the investigation of drinking water frameworks across Canada (Dort, 2010), and the study of benthic invertebrates to some toxicants by a current graduate student, Jason Solnik.

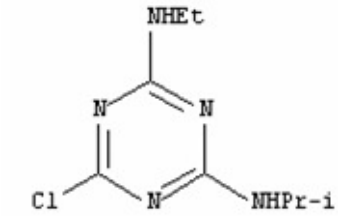
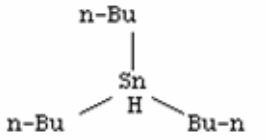
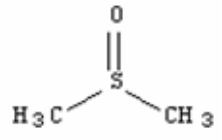
This thesis aims to contribute to the fourth objective by creating a library which describes and characterizes the responses of several aquatic organisms to some contaminants. Marshall (2009) and Pearce (2009) had studied the effects of atrazine and tributyltin on a suite of organisms, and their results were analyzed for inclusion in the library of responses. To supplement the results by Marshall (2009) and Pearce (2009), a bioassay measuring the effects of atrazine and tributyltin on the effective photosynthetic yield of *Pseudokirchneriella subcapitata* was also conducted in the present study. Finally, a number of multivariate statistical tools were employed to create a model for the library of responses.

2.3.1 Background on Test Contaminants in Previous and Current Studies

In the previous and current studies, three different chemicals were tested on the selected aquatic organisms. The first two chemicals, atrazine and tributyltin, are some of most commonly found contaminants in aquatic systems. The third chemical, dimethyl sulfoxide (DMSO), is not classified as a common waterway pollutant, but this chemical was used during experiments as a solvent for the delivery of atrazine and tributyltin. A summary of the chemical properties of the three compounds is listed in Table 2.1.

Atrazine is a chloro-N-diakyl triazine compound commonly used as herbicide. Other similar herbicides include simeton, simazine, propazine, terbumeton, terbuthylazine, cyprazine, simetryn, prometryn, terbutryn, methoprotryne, and hexazinone (Zhang *et al.*, 2006). In North America, between 70,000 and 90,000 tonnes of atrazine are applied annually to control the growth of weeds in croplands (Graymore *et al.*, 2001). Due to leaching and run-off, atrazine may enter the aquatic ecosystem, and trace levels of the compound have been found in surface and well waters across Canada and the United States (Health Canada, 1993). Various studies have classified atrazine as a potentially carcinogenic and endocrine-disrupting compound (Donna *et al.*, 1984; Hoar *et al.*, 1988; Health Canada, 1993). An extensive re-evaluation of the effects of atrazine on human health is currently underway by the United States Environmental Protection Agency (US EPA, 2011) which may lead to further restrictions in the use of the chemical. Currently, however, atrazine is still widely applied and thus remains a threat to various water bodies.

Table 2.1. Chemical properties of atrazine, tributyltin and dimethyl sulfoxide.

	[a] **Atrazine**	[b] **Tributyltin hydride**	[c] **Dimethyl sulfoxide**
Chemical Structure			
Molecular Formula	$C_8H_{14}ClN_5$	$SnC_{12}H_{26}$	C_2H_6OS
Density	1.187 g/cm^3 at 20 ºC	1.17 g/ cm^3 at 20 ºC	1.1004 g/ cm^3 at 20 ºC
Melting Point	175 – 177 ºC	<45 ºC	19 ºC
Solubility in Water at 20 ºC	0.030 g/litre	0.020 g/litre	miscible
Toxicity (Oral LD_{50})	1869 to 3080 mg/kg body weight (rat)	672 to 3000 mg/kg body weight (rat)	17400 to 28300 mg/kg body weight (rat)

[a] World Health Organization (1996)
[b] United States Environmental Protection Agency (1997)
[c] Gaylord Chemical Company (2007)

The second contaminant was tributyltin, a tri-substituted organotin compound commonly used as a biocide and an anti-fouling agent in the paints of boats and cargo ships, oil rigs, fish cages and other floating structures (Alzieu, 1998; Konstantinou and Albanis, 2004). Other similar biocides include fentin hydroxide, cyhexatin, azocyclotin, and fenbutatin oxide (Ma, 2005). The teratogenic effects of tributyltin in several crustacean species have been reported, and an international ban on the use of tributyltin was proposed in 2003 by the International Maritime Organization (Oberdöster *et al.,* 1998; Alzieu, 1998). Many goods-exporting countries, however, have not agreed to this convention, and thus trace levels of tributyltin can still be found in heavily-travelled waterways (Konstantinou and Albanis, 2004).

Atrazine and tributyltin uniquely affect aquatic organisms due to their distinct structures and properties as listed in Table 2.1. Their toxicities are similar as indicated by their relatively close oral LD_{50} values in rats. Atrazine strongly exerts its toxicities by inhibiting the electron transport process in the photosystem II (PSII) complex in plants (Solomon *et al.,* 1996). Conversely, tributyltin causes malformations of the mitochondrial membranes in aquatic invertebrates and decreases their metabolic outputs by inhibiting the conversion of ATP to ADP (Alzieu, 1998; Fent, 1996).

Another chemical was also used in the current and previous studies in conjunction with atrazine and tributyltin. The solubilities of atrazine and tributyltin in water are low, and thus organic solvents such as DMSO are necessary to help evenly distribute the contaminants in the test solutions (Stratton, 1985; Haap *et al.*, 2007). To prevent the

organic solvent from interfering with the bioassays, a concentration of 0.1% v/v should not be exceeded for acute toxicity tests (LeBlanc and Surprenant, 1983). Studies involving 0.1% v/v DMSO found that the solvent had a minimal impact on a number of aquatic organisms (Martins *et al.*, 2007; Ren *et al.*, 2008; Ren *et al.*, 2009). Marshall (2009) and Pearce (2009) also confirmed the negligible impact of 0.1% v/v DMSO on the organisms used in the previous and current studies by comparing the endpoints of the organisms in dechlorinated tap water with and without DMSO using a one-way analysis of variance.

2.3.2 Analysis of Bioassays by Marshall (2009)

Marshall (2009) examined the changes in many behavioural and respiratory endpoints of *Daphnia magna, Hyalella azteca,* and *Lumbriculus variegatus* when exposed to various concentrations of tributyltin and atrazine. For the majority of the bioassays, the prepared concentrations were 0.010, 0.050, and 0.100 mg/L for tributyltin and 0.005, 0.050, and 0.100 mg/L for atrazine. When creating solutions of atrazine or tributyltin, 0.1% v/v DMSO was also added to increase the solubility of atrazine and tributyltin in dechlorinated tap water. Sixteen endpoints were recommended by Marshall (2009) due to the sensitivities of the organisms to atrazine and tributyltin. For example, the percentages of *Daphnia magna* exhibiting a change in their swimming height was drastically affected by the additions of tributyltin and atrazine, thus making this endpoint an excellent indicator for the two contaminants and for inclusion in the EWBS. The present study followed most of the recommendations by Marshall (2009) except for several cases as outlined in Table 2.2.

Table 2.2. A series of endpoints measured and recommended by Marshall (2009) to evaluate the effects of atrazine and tributyltin on the selected aquatic organisms.

Measured Endpoints (% Organism)	Recommended by Marshall (2009)	Included in the present study
Daphnia magna		
Changing swimming height	√	√
Spinning	√	√
Changing body orientation		√
Immobilized	√	√
Using secondary antennae	√	√
Changing swimming style	√	√
Hyalella azteca		
Changing swimming height		√
Crawling on substrate	√	
Immobilized	√	√
Burrowing	√	√
Changing grouping behaviour	√	√
Shortening body length	√	√
Changing body orientation		√
Lumbriculus variegatus		
Burrowing	√	
Changing grouping behaviour		
Swimming in the middle of the beaker		
Displaying abnormal behaviour	√	√
Immobilized	√	√
Shortening body length	√	√
Changing body orientation	√	√
Moving within groups	√	√
Total Number of Endpoints	16	17

Several endpoints were previously not recommended for inclusion in the EWBS, but the present study found these endpoints to be potentially useful and should be included in the library of responses. For example, Marshall (2009) did not recommend monitoring the percentage of *Daphnia magna* displaying a change in their body orientation due to the lack of observable trends between the endpoint and the tested concentrations of atrazine or tributyltin. It was noted, however, that despite the difficulty in discriminating between concentrations, there was still a statistically significant difference between all tributyltin and atrazine treatments versus the reference condition. Based on this finding, the present study included this endpoint for the development of the EWBS library as the endpoint would be very useful in discriminating between the reference condition and a condition where either one of the two contaminants is present.

In addition, Marshall (2009) did not recommend measuring the percentages of *Hyalella azteca* displaying a change in the swimming behaviour because all of the organisms were immobilized after 6 hours of exposure to all concentrations of tributyltin, while exposure to atrazine did not affect the swimming behaviour to such a severe extent. Similarly, the percentages of *Hyalella azteca* displaying a change in body orientation was also not recommended for measurement because the exposure to atrazine did not affect the organisms except at higher concentrations after a longer incubation time, while exposure to tributyltin affected the organisms within one hour even at lower concentrations. Such species-specific sensitivities, however, would be crucial in distinguishing cases of contamination by the two compounds. As a result, the present

study recommended the inclusion of the above two endpoints in the library of responses due to their potential to discriminate atrazine and tributyltin contaminations.

Conversely, a number of endpoints recommended by Marshall (2009) were not included in the present study for several reasons. First, Marshall (2009) suggested measuring the percentage of *Hyalella azteca* crawling on substrate due to the sensitivity of the parameter to the tested contaminants. Upon analysis of the data, however, it was found that the percentage of organisms crawling was highly related to the percentage of organisms immobilized (*i.e.* it was impossible for the organism to crawl and be immobilized at the same time). Measuring both endpoints would therefore be unnecessarily redundant, and only one of the two endpoints should be selected for the EWBS library. From the standpoint of automation, measuring the percentage of immobilized organisms would be easier as monitoring the percentage of crawling organisms requires a more detailed image analysis, and the percentage of immobilized *H. azteca* was then selected for inclusion in the current library.

Additionally, Marshall (2009) recommended measuring the percentage of *Lumbriculus variegatus* burrowing, but the concentrations of tributyltin used in this bioassay (0.0001, 0.001 and 0.010 mg/L) were completely different than the rest of the bioassays (0.010, 0.050 and 0.100 mg/L, as stated previously). As a result, it was impossible to incorporate the measured endpoint data into the EWBS library as the tested concentrations must be consistent at 0.010, 0.050 and 0.100 mg/L for tributyltin, and 0.005, 0.050 and 0.100 mg/L for atrazine.

Last, Marshall (2009) also measured the effects of tributyltin and atrazine on the respiration rates of *Daphnia magna*, *Hyalella azteca* and *Lumbriculus variegatus*. While the contaminants exerted some effects on the respiratory rates of each species, the tests employed to measure the dissolved oxygen content were inconsistent. For example, three adult *Hyalella azteca* were employed in solutions of tributyltin, but five *Hyalella azteca* were used to measure the effects of atrazine. Furthermore, despite the respiratory rates being normalized in relation to the total average body weight of the organisms, replicates of the reference condition yielded very large difference in magnitude. Based on these findings, the respiration data collected for the three organisms should not be used for the development of the EWBS library.

2.3.3 Analysis of Bioassays by Pearce (2009)

Pearce (2009) monitored various endpoints of *Lemna minor*, *Pseudokirchneriella subcapitata, Euglena gracilis,* and *Anodonta grandis* in response to different levels of tributyltin and atrazine. The concentrations used in this study ranged from 0.001 to 0.100 mg/L for tributyltin, and 0.005 to 0.500 for atrazine. Similar to the experimental design of Marshall (2009), DMSO at a concentration of 0.1% v/v was added to increase the solubility of atrazine and tributyltin in dechlorinated tap water. Table 2.3 summarizes the endpoints measured and recommended by Pearce (2009) for inclusion in the EWBS library, and a total of five endpoints were suggested for use in the proposed early-warning biomonitoring system. Despite some success with the studied organisms, however, the results from the study by Pearce (2009) could not be included in the present study due to a number of factors.

Table 2.3. A series of endpoints measured and recommended by Pearce (2009) to evaluate the effects of atrazine and tributyltin on the selected aquatic organisms.

Measured Endpoints	Recommended by Pearce (2009)	Included in the present study
Lemna minor		
Population growth	√	
Pseudokirchneriella subcapitata		
Photosynthesis rate	√	
Respiration rate		
Population growth	√	
Euglena gracilis		
Respiration under light-saturated conditions		
Respiration in dark conditions		
Population growth		
Percentage of cysts formed under light-saturated conditions	√	
Percentage of cysts formed under dark conditions		
Anodonta grandis		
Percentage of opened valves		
Respiration rate	√	
Total Number of Endpoints	5	0

The first recommended endpoint was the population growth of *Lemna minor*. While this parameter showed a statistically-significant difference between the reference condition and higher levels of tributyltin (0.100 mg/L) and atrazine (0.500 mg/L), the growth rate measurement was conducted over a period of four days (96 hours). Since the EWBS aims for a rapid detection of contaminants, this measurement period is considered too long and therefore unsuitable for use in a rapid detection, early-warning system.

The largest hindrance in incorporating the endpoint data by Pearce (2009) into the EWBS library was the inconsistent doses of contaminants used in different bioassays which did not match the concentrations used in the previous study by Marshall (2009). For example, Pearce (2009) conducted experiments involving *Pseudokirchneriella subcapitata* and *Euglena gracilis* in tributyltin solutions of 0.010 and 0.100 mg/L, and atrazine solutions of 0.050 and 0.500 mg/L, whereas Marshall (2009) worked with concentrations of 0.010, 0.050 and 0.100 mg/L for tributyltin, and 0.005, 0.050 and 0.100 mg/L for atrazine. The concentration levels tested on the *Anodonta grandis* were even more significantly different at 0.0001, 0.001 and 0.01 mg/L of tributyltin, and 0.0005, 0.005 and 0.05 mg/L of atrazine. Furthermore, large standard deviations were often found for different replicates, and in most cases no clear trend was observed between the measured endpoint and the concentration. Due to the non-linear dose-response relationship, data interpolation or extrapolation methods could not be employed to match the concentrations by Marshall (2009) as more uncertainties would have resulted. The recommendations by Pearce (2009) were therefore not included for the development of

the EWBS library and could only serve as guidance for future experiments involving aquatic plants.

2.3.4 Supplementary Algal Bioassay

Despite the inability to incorporate experimental data by Pearce (2009), the present study recognized the importance of including photosynthetic organisms among the suite of organisms employed for biomonitoring. Since atrazine is a herbicide which directly affects the photosynthetic activity of aquatic plants, measuring the responses from photosynthetic organisms may help differentiate between contaminations by herbicidal and non-herbicidal compounds. Based on the recommendations by Pearce (2009), the present study re-measured the photosynthesis rate of *Pseudokirchneriella subcapitata* using an online algal fluorometer.

As elaborated in Section 2.2.4, a fluorometer can measure the *in-vivo* prompt fluorescence (IPF) and the photosynthetic yield of an algal solution. When chlorophyll is irradiated by a short pulse of light, a portion of the energy is used to perform photosynthesis while the remainder is re-emitted as heat and fluorescence (Strasser *et al.*, 2000). Thus, a relationship can be inferred such that (Balan *et al.*, 2006):

$$F + P + H = 1 \qquad (1)$$

where

F = Fluorescence,

P = Photochemical conversion, and

H = Non-radiative heat dissipation.

If the plant is dark-adapted, the photochemical conversion approaches zero and heat dissipation can be assumed constant. Equation (1) can thus be rearranged as:

$$F_m = 1 - H \quad (2)$$

where

F_m = Maximum fluorescence.

The photosynthetic yield is the ratio between the absorbed energy and the re-emitted fluorescence. This parameter can be expressed by the following equation:

$$Y = (F_m - F)/ F_m \quad (3)$$

where

Y = Photosynthetic yield, and

$(F_m - F)$ = P.

When the sample is not dark-adapted and experiments are conducted on light-adapted plants, a different notation is used to describe the maximum fluorescence (F_m') and the photosynthetic yield is termed the *effective* photosynthetic yield. The presence of a herbicide may damage the structure of chlorophyll, resulting in lower photochemical conversion by the plant (P), higher re-emitted fluorescence (F), and thus less effective photosynthetic yield (Y). By including the change in effective photosynthetic yield in the library of responses, the effects of atrazine and tributyltin on an algal population can be examined, and a summary of all included endpoints is listed in Table 2.4.

Table 2.4. Summary of all endpoints from the present study and previous work by Marshall (2009) for inclusion in the library of responses.

Measured Endpoints	
Daphnia magna	
1	Changing swimming height
2	Spinning
3	Changing body orientation
4	Immobilized
5	Using secondary antennae
6	Changing swimming style
Hyalella azteca	
7	Changing swimming height
8	Immobilized
9	Burrowing
10	Changing grouping behaviour
11	Shortening body length
12	Changing body orientation
Lumbriculus variegatus	
13	Displaying abnormal behaviour
14	Immobilized
15	Shortening body length
16	Changing body orientation
17	Moving within groups
Pseudokirchneriella subcapitata	
18	Effective photosynthetic yield
Total Number of Endpoints = 18	

2.4 Multivariate Statistical Tools

Multivariate analysis is the process of examining three or more variables which may or may not be intercorrelated to each other (Shaw, 2003). By analyzing a large number of variables, one is able to obtain a more comprehensive depiction of the analyzed subject. As outlined in Table 2.4 and Section 2.3.4, the present study would incorporate a total of 18 endpoints, where 17 endpoints were extracted from Marshall (2009) and 1 endpoint was obtained from the algal bioassay. Analyzing these 18 endpoints simultaneously would be very difficult as they contained a large number of data points, and complex interdependencies could also exist within the data set. A number of multivariate statistical tools must therefore be utilized to facilitate the analysis.

2.4.1 Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical tool that reduces a large number of variables into a smaller set of independent variables with minimum loss of the original information (Reimann *et al.*, 2008). First derived by Pearson (1901) and refined later by Hotelling (1933), PCA was widely used in many ecological studies in the 1950s and remains one of the more popular ordination techniques due to its ease of calculations and non-dependence on the normality of the data (Shaw, 2003).

Figures 2.1(a) and 2.1(b) provide a step-by-step example of data-reduction using PCA. In Figure 2.1(a), data points which are distributed in a three-dimensional data space are to be expressed in a two-dimensional plane. In order to create this plane, two new independent axes, or principal components, must be constructed. The first principal

component (PC1) can be obtained by drawing an imaginary line through the cloud of data points, similar to the line-of-best-fit in a linear regression. This imaginary line must also pass through an imaginary point that represents the overall mean of the original data set, and thus the first principal component describes the greatest degree of data variation. After constructing the first principal component, the procedure can be repeated to obtain the second principal component. A second line-of-best-fit can be fitted similarly by passing through the overall mean, but this secondary line must be orthogonal to the first principal component as shown in Figure 2.1(a).

The dynamics between the two principal components can be compared to a wheel, with the first axis being the axle and the second axis lying on the plane of the wheel. The second axis is able to rotate in a complete circle until this axis reaches a position that describes the second largest variance among the data points. Because both axes pass through the overall mean of the original data set, the new transformed data set is said to be mean-centered where the average is located on the origin. PC1 and PC2 then form a plane describing the relative positions of the data points in a two-dimensional space as shown by the top view of the plane in Figure 2.1(b). This procedure thus accomplished data reduction from three to two dimensions.

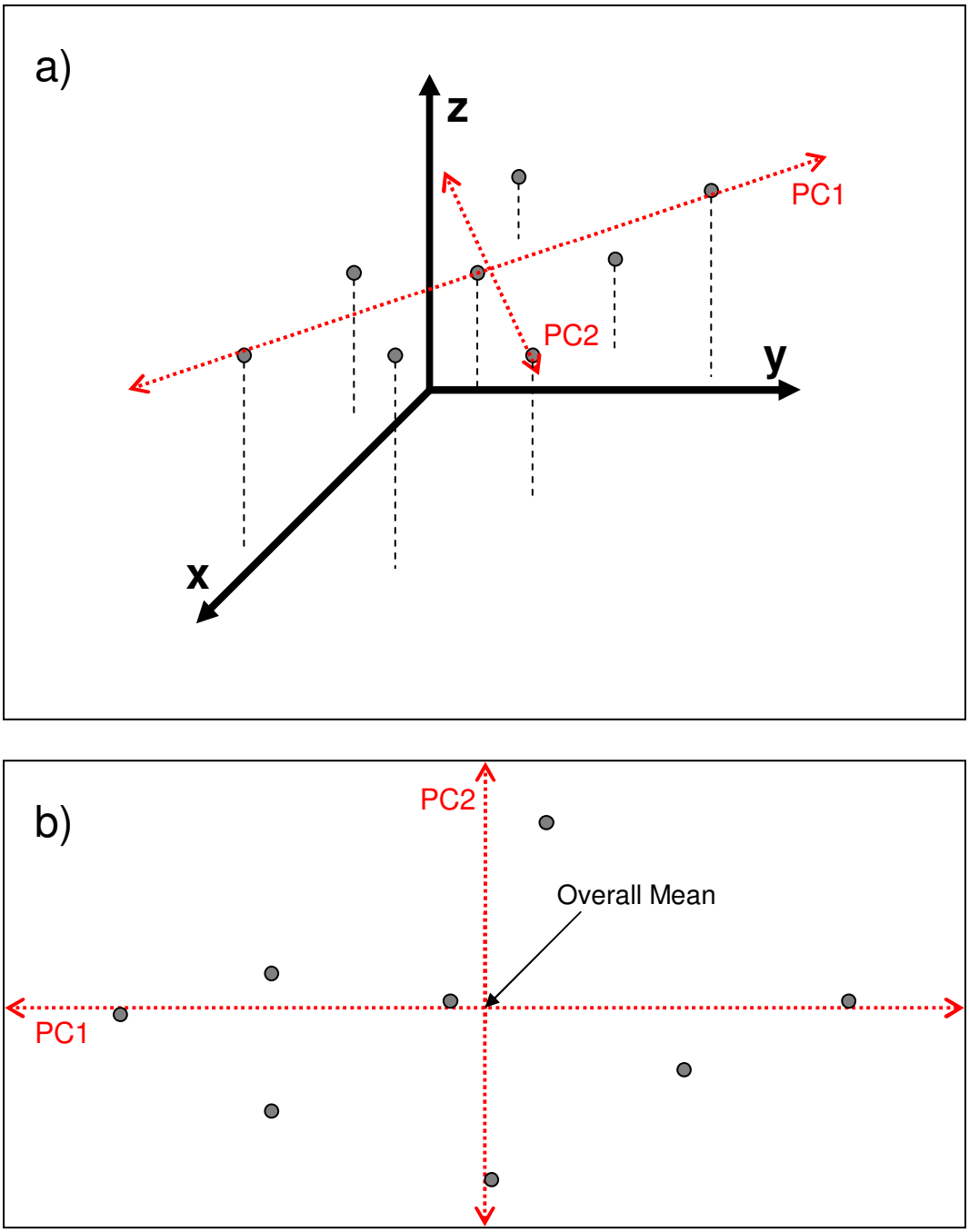


Figure 2.1. The application of Principal Component Analysis where data points in a three-dimensional space are described using a two-dimensional plane.

When performing data-reduction from a large number of variables, more than two principal components are sometimes required. In this case, it is harder to visualize the original data set as the human depth perception is limited to three dimensions, but the process of data reduction is very similar to the previous example. After constructing the first two principal components, the third axis is defined by default as this axis must be orthogonal to both the first and second principal components. It is possible to construct more than three principal components, but the analysis of such multi-dimensional data set would be cumbersome, and such practice is impractical when applied to interpret environmental data (Shaw, 2003).

The following steps describe the general mathematical algorithms involved in PCA. Figure 2.2 also provides a graphical depiction of these steps.

1) *Tabulate data points*

   Research findings are collected and presented in an *j*-by-*k* matrix where the columns (*k*) correspond to the variables and the rows (*j*) list the observations for each variable.

2) *Pre-condition certain variables*

   The variables in the matrix may differ significantly from each other in unit and magnitude. For example, a researcher may examine the mass of different fish and the dissolved oxygen content in the water. In order to prevent strong bias by variables with larger magnitudes, some data transformation may be necessary.

One of the most commonly applied transformations is mean-centering, in which the mean across each dimension is subtracted from each data point, resulting in a modified matrix with a mean of zero for each variable. Other transformations may include additive, centred or isometric log transformations, as well as normalization of each variable. Extreme caution must be applied when selecting methods for data transformation so as to avoid highly-skewed data sets. After applying appropriate transformations, a modified matrix is created.

3) *Calculate the covariance matrix based on the modified matrix*

The covariance matrix is a square matrix displaying all possible interrelations among the variables. The size of the matrix (*k*-by-*k*) corresponds to the total number of variables in the experiment. This matrix can be easily constructed using most statistical software.

4) *Calculate the eigenvectors and eigenvalues of the covariance matrix*

From the covariance matrix, eigenvectors and eigenvalues can be calculated. A detailed explanation of eigenvectors and eigenvalues can be found in most textbooks on matrix algebra (Anton, 2003). In simple terms, the eigenvectors of a matrix are a series of vectors that, upon multiplication with the matrix, only change the scale and not the direction. The eigenvalue is then a measure of the variance described by each eigenvector. For a square (*k*-by-*k*) covariance matrix, *k* eigenvectors are generated, and these eigenvectors are arranged in order of decreasing eigenvalues.

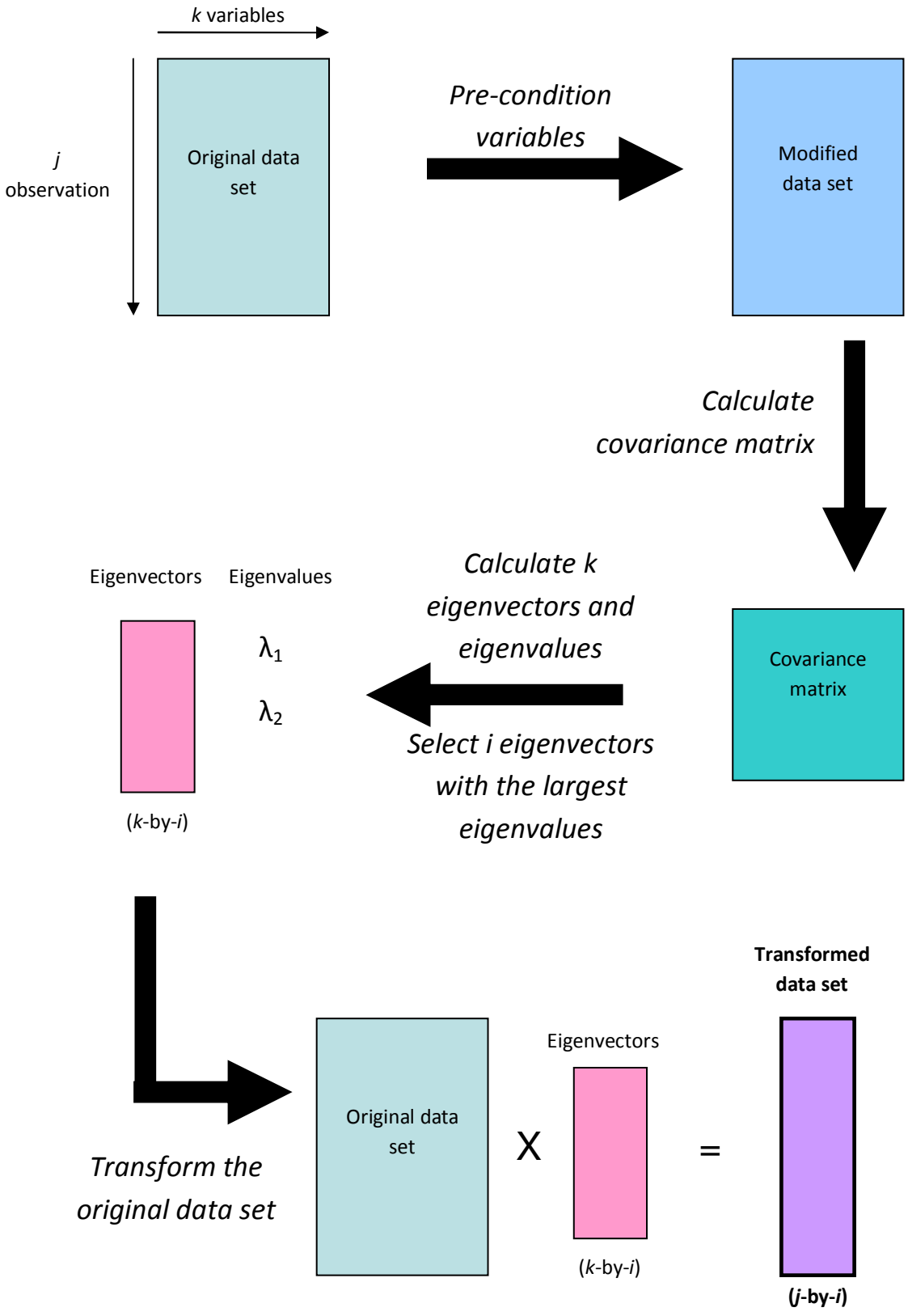


Figure 2.2. Flowchart of the data-reduction procedure using PCA

5) *Transform the original data set*

The original data set, which is composed of k possibly correlated variables, can be transformed into a new data set containing fewer completely independent variables (also known as principal components). The number of principal components selected must be specified according to the intent of the researcher. If the original data set is to be reduced into i principal components, then i eigenvectors with the largest eigenvalues are selected, and matrix multiplication is carried out between the original data set matrix and these i eigenvectors. The resulting matrix is a transformed data set which can be expressed as (Helena *et. al,* 2000; Shrestha and Kazama, 2007):

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \ldots + a_{ik}x_{kj} \quad (4)$$

where

z = the coordinate in the new principal component,

a = the eigenvector coefficient,

x = the original measured value,

i = the component number,

j = the observation number, and

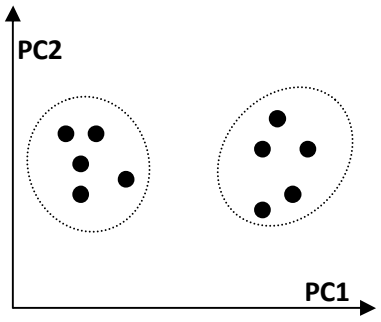
k = the number of variables in the original data set

A number of published ecological studies have incorporated PCA in their data treatment and analysis process. Shrestha and Kazama (2007) employed PCA, along with

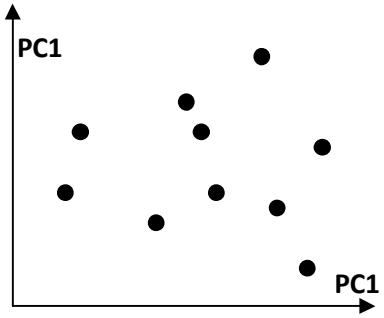
other multivariate statistical tools such as cluster analysis, discriminant analysis and factor analysis, to assess the surface water quality of the Fuji River, Japan. In this study, a large data matrix containing water quality parameters collected during an eight-year monitoring program was constructed. The objectives of the study were to determine any similarities or dissimilarities between sampling sites and to identify factors which might cause spatial and temporal variations in water quality. By applying PCA, the study was able to extract and identify the factors affecting the water quality of the river at three different sampling sites. Akbal *et al.* (2011) also conducted a similar study using PCA and other multivariate statistical tools, where the causes of variation in surface water quality at the mid-Black Sea coast of Turkey were determined. These studies demonstrated the capability of PCA to handle and resolve patterns within a complex data set, and this powerful multivariate statistical tool was thus employed in the current study.

2.4.2 Cluster Analysis

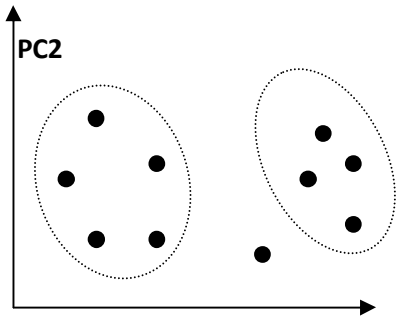
After applying PCA on a data set, some patterns among the data points may be observed where a number of points appear to separate into discrete clusters. PCA, however, is an ordination technique which simply displays a representation of a data set and does not perform actual data set separation (Shaw, 2003). Since no formal rules exist in defining a cluster, a subjective decision must often be made by the researcher regarding the clustering patterns within the data (Rao, 1952). Figure 2.3 displays some examples of possible clustering patterns typically found in an environmental data set.



(a) Clear separation into two clusters



(b) No strong evidence of clustering



(c) Some evidence of clustering

Figure 2.3. Some possible patterns found in a data set after simplification using principal component analysis (Adapted from Shaw, 2003).

Figure 2.3(a) presents an ideal situation where the data points are clearly partitioned into two different clusters. This condition is known as ball clusters, a condition where the distance between the clusters is greater than the longest distance between any two data points within each cluster (Shaw, 2003). Environmental data sets, however, may not always resolve into ball clusters, and some overlaps are often observed. In Figure 2.3(b), the data points appear as a continuum across the two principal components with clear separation among data points, and there is no strong indication of a subpopulation within the data set. Environmental data points are more likely to be distributed as shown in Figure 2.3(c) (Shaw, 2003). In this case, the data points show some evidence of clustering, with a point lying in between the two groups.

In order to assess whether a data set could be objectively partitioned into several clusters, Cluster Analysis (CA) could be employed. A number of CA techniques currently exist, and these techniques can be categorized as hierarchical or non-hierarchical. In hierarchical techniques, each cluster is designated as a subset of a higher-order cluster, and a branching diagram can be constructed. An example of a hierarchical separation is the biological taxonomic classification which separates all living organisms according to kingdom, phylum/division, class, order, family, genus, and species (Gingerich, 1987). Conversely, non-hierarchical techniques subdivide data points into different clusters without exploring how different clusters may relate to each other. For the present study, non-hierarchical techniques would be more appropriate since the data points were to be separated according to the different water conditions, and each condition was not a subset of any higher-order group.

An example of a non-hierarchical classification technique is the K-means clustering. In this classification technique, a series of iterative algorithms are employed to separate *n* data points into *k* clusters according to the cluster means. An example of the application of K-means clustering is shown in Figure 2.4, where the algorithm is employed to separate 10 data points into 2 clusters. By randomly selecting 2 data points, 2 clusters are created. Then, new means for these 2 clusters are identified, and 2 new clusters are re-calculated based on these means. This step is repeated several times until a convergence is reached. By applying K-means clustering to the data set in the present study, the number of correctly assigned data points could be identified, and the results would describe whether the clusters in the data set are clearly defined and well-separated.

Kaufman and Rousseeuw (1990) devised a technique to further assess cluster quality called the Average Silhouette Width Plot. In this technique, a data point from a cluster is selected, and the average distance of this data point to all other data points in the same cluster is calculated. Then, the average distance of this data point to all data points in the next closest cluster is also determined. By comparing the two averages, a silhouette coefficient can be calculated as:

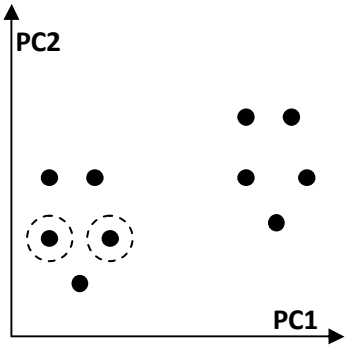
$$S_C = \frac{n_i - m_i}{\max(m_i, n_i)} \tag{5}$$

where

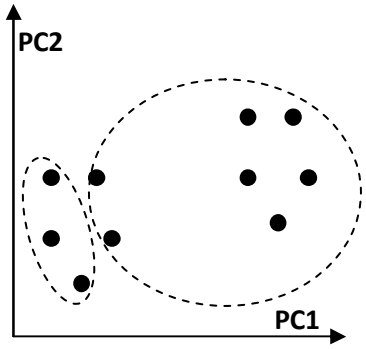
m_i = the average distance for point *i* to all other points in the same cluster

n_i = the average distance for point *i* to all other points in the nearest cluster, and

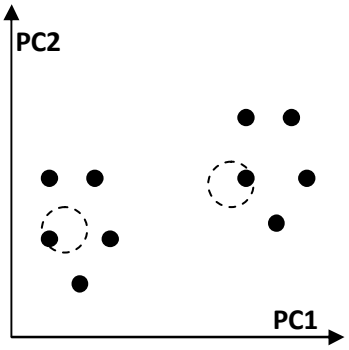
S_C = the silhouette coefficient.



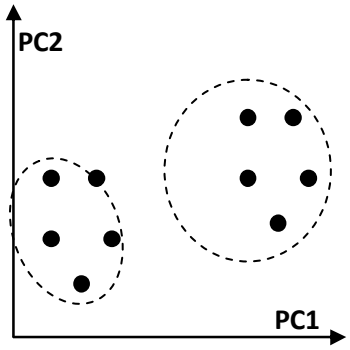
Step 1: If the data points are to be separated into 2 clusters, 2 random data points are selected as cluster means



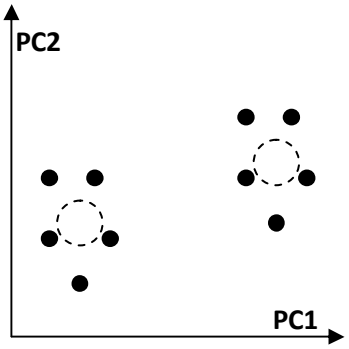
Step 2: Clusters are then created where data points are grouped according to their distance to the cluster means in Step 1



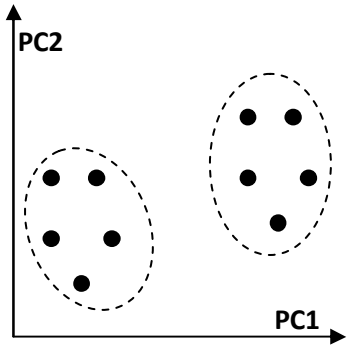
Step 3: New means for the clusters constructed in Step 2 are identified



Step 4: Based on the means in Step 3, 2 new clusters are re-calculated



Step 5: New means for the clusters constructed in Step 4 are identified



Step 6: New clusters are re-calculated again based on the previous step. This process is repeated until a convergence is reached.

Figure 2.4. Algorithms for K-means clustering (Adapted from Govaert, 2009).

The silhouette coefficient in Equation (5) ranges from -1 to 1, and this number is a measure of how well each data point is represented in the cluster (Reimann, 2008). A data point with a coefficient of -1 denotes that the point is assigned to the wrong cluster because average distances of that point to other points in the same cluster is greater than the average distances of that point to other points in the nearest cluster. Conversely, a coefficient of 1 indicates a very well-clustered object. In general, a cluster exhibits strong cluster qualities if the average coefficient for that cluster is higher than 0.7, while an average coefficient of less than 0.25 indicates the absence of a cluster structure (Kaufman and Rousseeuw, 1990)

2.4.3 Discriminant Analysis

The proposed EWBS aims to not only provide a more sensitive and accurate detection of contaminants but also the means to identify different contaminants by comparing the responses of the organisms to a set of stereotyped responses. When a series of responses are observed by the EWBS, this incoming reading is assessed against the library of responses and matched according to established conditions in order to identify the offending toxicant. To perform this type of data classification, a statistical tool such as the Discriminant Analysis (DA) method could be employed.

Discriminant analysis uses *a priori* knowledge on existing group memberships to develop a function which will predict the group membership of new observations (Fisher, 1936). A variety of different methods exist for performing DA, and one of the simplest techniques is linear discriminant analysis where group covariances are assumed to be

equal, and linear functions are used to separate the groups (Huberty, 1994; Johnson and Wichern, 2002). An example of linear discriminant analysis is depicted in Figure 2.5.

Figure 2.5(a) illustrates two independent variables, X and Y, as well as two clusters of data points, group A and group B. A discriminant function Z is also projected in the sample space of X and Y, and data points in groups A and B can subsequently be projected onto this discriminant function to form two distribution curves. When a new data point is introduced, the membership of this point is determined based on a combination of functions that minimizes the probability of misclassifications (Figure 2.5(b)). The calculations to perform DA techniques can be complicated, but statistical packages such as MATLAB can easily process data sets and calculate simple linear cases.

2.5 Concluding Remarks

This chapter provides an overview of some recently-developed biomonitoring technologies, their strengths, and their limitations. As previously elaborated, current biomonitoring technologies do not sufficiently meet the criteria outlined in Chapter 1 for a robust and reliable early-warning system. A large-scale project funded by NSERC was initiated to develop an early-warning biomonitoring system which would more accurately detect and identify contaminants, and studies have been conducted by previous graduate students to complete some of the objectives of the project. The present study aimed to accomplish the last objective by creating a model for a library of responses which can be incorporated into the design of the EWBS.

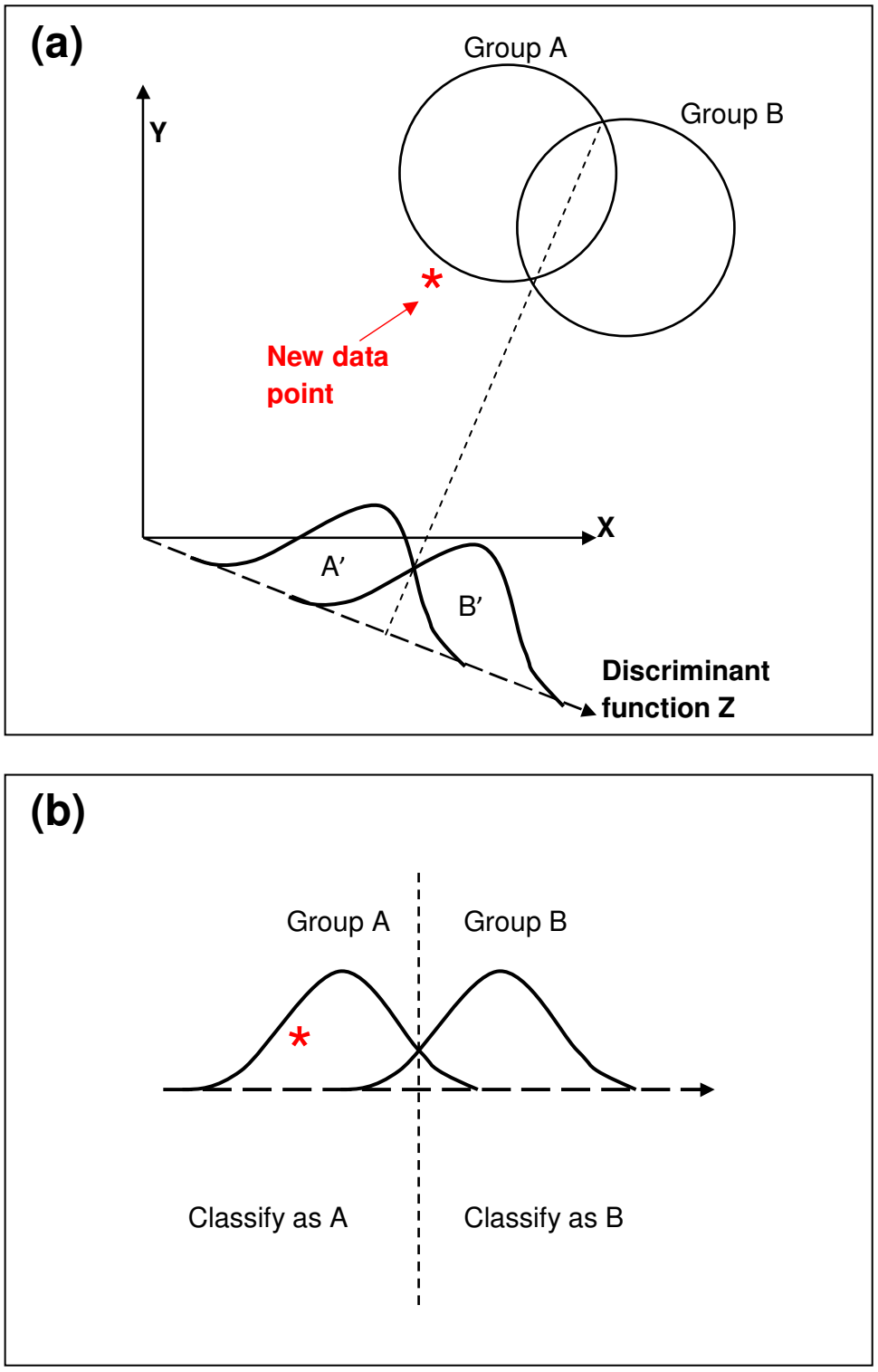


Figure 2.5. Graphical illustration of the discriminant analysis technique (adapted from Dillon and Goldstein, 1984).

CHAPTER 3

MATERIALS AND METHODS

3.1 Introduction

A key objective of the current work is to establish a model for the library of responses which could be incorporated into an early-warning biomonitoring system. As outlined in Table 2.4 in the previous chapter, a total of 18 endpoints were used in the current study to distinguish atrazine and tributyltin contaminations from the reference condition. To establish this library of responses, a number of endpoint data sets were obtained from a previous study by Marshall (2009), and an algal bioassay was conducted in the present study to supplement these data sets. Following the recommendations by Pearce (2009), the effective photosynthetic yield of the green alga *Pseudokirchneriella subcapitata* was measured in the present study using an online algal fluorometer.

This chapter starts by elaborating on the procedures for growing and testing the algal cultures. After conducting the algal bioassay and combining all endpoint data from previous and present studies, this chapter then elaborates on the statistical treatment of these endpoint data and the design parameters chosen to perform the multivariate statistical analysis. By applying principal component analysis on the endpoint data, a preliminary library of responses was created. A series of cluster analysis methods were then applied on the preliminary library to objectively assess the cluster quality. Finally, the chapter describes the discriminant analysis method employed to evaluate the potential ability of the library in distinguishing different contaminants.

3.2 Experimental Setup for Online Algal Fluorometer

3.2.1 Procedures for Cleaning and Disinfection

Prior to any experiment, each glassware and equipment was thoroughly cleaned to avoid any bacterial contamination and remove any potential residues which could affect the growth of the green algae. To perform sterilization, glassware was autoclaved at 120 ºC and 15 psi for an hour and cooled prior to use. Before and after performing experiments, each glassware and equipment was washed using acetone and distilled water three times to remove any residues which might have adhered to the glass. After rinsing with acetone, each glassware and equipment was washed using the Extran organic decontaminating soap (VWR Cat#: CAEX0995-1) to remove any residual organics and then rinsed with distilled water. Finally, each glassware and equipment was washed using 10% v/v hydrochloric acid, rinsed with distilled water, and placed in the drying rack in an inverted position to dry.

3.2.2 Procedures for Algal Culturing

For the algal bioassay, *Pseudokirchneriella subcapitata* was grown according to the protocols outlined by Environment Canada (2007) with a few modifications. A starter culture of the algae was obtained from an agar slant (Wards™ Natural Science, item# 86V0620), and a group of cells from this agar slant was aseptically transferred using a sterile loop and resuspended into several 250-mL Erlenmeyer flasks, each containing 50 mL of sterile growth medium as prepared in Table 3.1.

Table 3.1. Stock nutrient solutions as recommended by Environment Canada (2007) for growing *Pseudokirchneriella subcapitata.*

Stock Nutrient Solution	Compound	Amount per 500 mL of dechlorinated tap water
1	$NaNO_3$	12.75 g
2	$MgCl_2 \cdot 6H_2O$	6.08 g
	$CaCl_2 \cdot 2H_2O$	2.1919 g
	H_3BO_3	92.8 mg
	$MnCl_2 \cdot 4H_2O$	207.2 mg
	$ZnCl_2$	1.64 mg
	$FeCl_3 \cdot 6H_2O$	79.9 mg
	$CoCl_2 \cdot 6H_2O$	0.714 mg
	$NaMoO_4 \cdot 2H_2O$	3.63 mg
	$CuCl_2 \cdot 2H_2O$	0.006 mg
	$Na_2EDTA \cdot 2H_2O$	150.1 g
3	$MgSO_4 \cdot 7H_2O$	7.35 g
4	K_2HPO_4	0.522 g
5	$NaHCO_3$	7.5 g

Note:

- To prepare the '**original growth medium'** as recommended by Environment Canada (2007), 1 mL of each stock solution was added to approximately 995 mL of autoclaved dechlorinated tap water, yielding ~1000-mL of growth medium.
- To prepare the '**modified growth medium'** as recommended by the present study, 4 mL of each stock solution was added to approximately 980 mL of autoclaved dechlorinated tap water, yielding ~1000-mL of growth medium.

The original protocol by Environment Canada (2007) suggested sterilizing each stock nutrient solution by filtering though a sterile 0.2 μm membrane and recommended against autoclaving as the heat could destroy the nutrients in the stock solutions and result in a reduced algal growth. Following this recommendation, filter-sterilization was initially performed in preliminary trials, but the filter-sterilized stock nutrient solutions were found to result in very low algal growth rates. Conversely, the autoclaved stock nutrient solutions were found to yield faster algal growth rates while still eliminating any unwanted microorganisms. Based on these findings, autoclaving appeared to be a more advantageous sterilization technique and was therefore employed in the present study.

The procedures for culturing the algae and preparing the algal growth medium were also modified from the protocols of Environment Canada (2007). For culturing, Environment Canada (2007) suggested aseptically transferring approximately 1 mL of the first generation algal culture into a 250-mL Erlenmeyer flask containing 50 mL of the 'original growth medium' as outlined in Table 3.1. This procedure, however, did not result in sufficient algal growth as the concentration of the algae peaked at only ~1×10^5 cells/mL. Such concentration was too low to be detected by the online algal fluorometer, and the procedure must be modified to yield more algal growth.

In order to achieve more growth, a larger amount of the first generation algal culture was transferred, and a more nutrient-rich 'modified growth medium' was prepared as outlined in Table 3.1. An optimal growth was achieved when approximately 20 mL of the first generation algal culture was transferred into a 500-mL Erlenmeyer

flask containing 130 mL of the modified growth medium. The modified growth medium was prepared by adding 4 mL of each stock nutrient solution into 980 mL of autoclaved dechlorinated tap water, yielding 1000 mL of solution containing four-times the amount of nutrients originally specified by Environment Canada (2007). By following this adjusted recipe, an algal cell concentration of up to 1.3×10^7 cells/mL was obtained.

After inoculating the algal cells from the agar slant into several flasks containing the modified growth medium, the flasks were incubated at 21.8 ºC under continuous 'cool white' fluorescent light with a light quantal flux of 56 – 66 μmol/m^2-s. To ensure complete mixing, the flasks were also placed on a continuous shaker at 100 rotations per minute. Environment Canada (2007) did not recommend testing on the first generation algal cultures, and thus the algae were sub-cultured at least once. Between 3 – 7 days, the cultures reached a logarithmic growth phase and were sub-cultured. The algal cells were also enumerated daily using a haemocytometer, and a growth curve was constructed to verify that the cells were at the logarithmic growth phase. The algal bioassay must be conducted during the logarithmic growth phase because the cells are actively dividing at this stage and would produce the highest effective photosynthetic yield.

3.2.3 Procedures for Algal Bioassay

As previously elaborated, *Pseudokirchneriella subcapitata* was cultured for testing using the fluorometer. The setup for the algal bioassay is shown in Figures 3.1 and 3.2.

Replicate	Reference (dechlorinated tap water)	DMSO At 0.1% v/v	Atrazine at 0.005 mg/L	Atrazine at 0.050 mg/L	Atrazine at 0.100 mg/L	Tributyltin at 0.010 mg/L	Tributyltin at 0.0510 mg/L	Tributyltin at 0.100 mg/L
1								
2								
3								

Compositions:

- Reference: 50 mL algae
- DMSO at 0.1% v/v: 50 mL algae + 50.0 µL DMSO
- Atrazine at 0.005 mg/L: 50 mL algae + 47.5 µL DMSO + 2.50 µL atrazine stock at 100 mg/L
- Atrazine at 0.050 mg/L: 50 mL algae + 25.0 µL DMSO + 25.0 µL atrazine stock at 100 mg/L
- Atrazine at 0.100 mg/L: 50 mL algae + 50.0 µL atrazine stock at 100 mg/L
- Tributyltin at 0.010 mg/L: 50 mL algae + 45.0 µL DMSO + 5.00 µL tributyltin stock at 100 mg/L
- Tributyltin at 0.050 mg/L: 50 mL algae + 25.0 µL DMSO + 25.0 µL tributyltin stock at 100 mg/L
- Tributyltin at 0.100 mg/L: 50 mL algae + 50.0 µL tributyltin stock at 100 mg/L

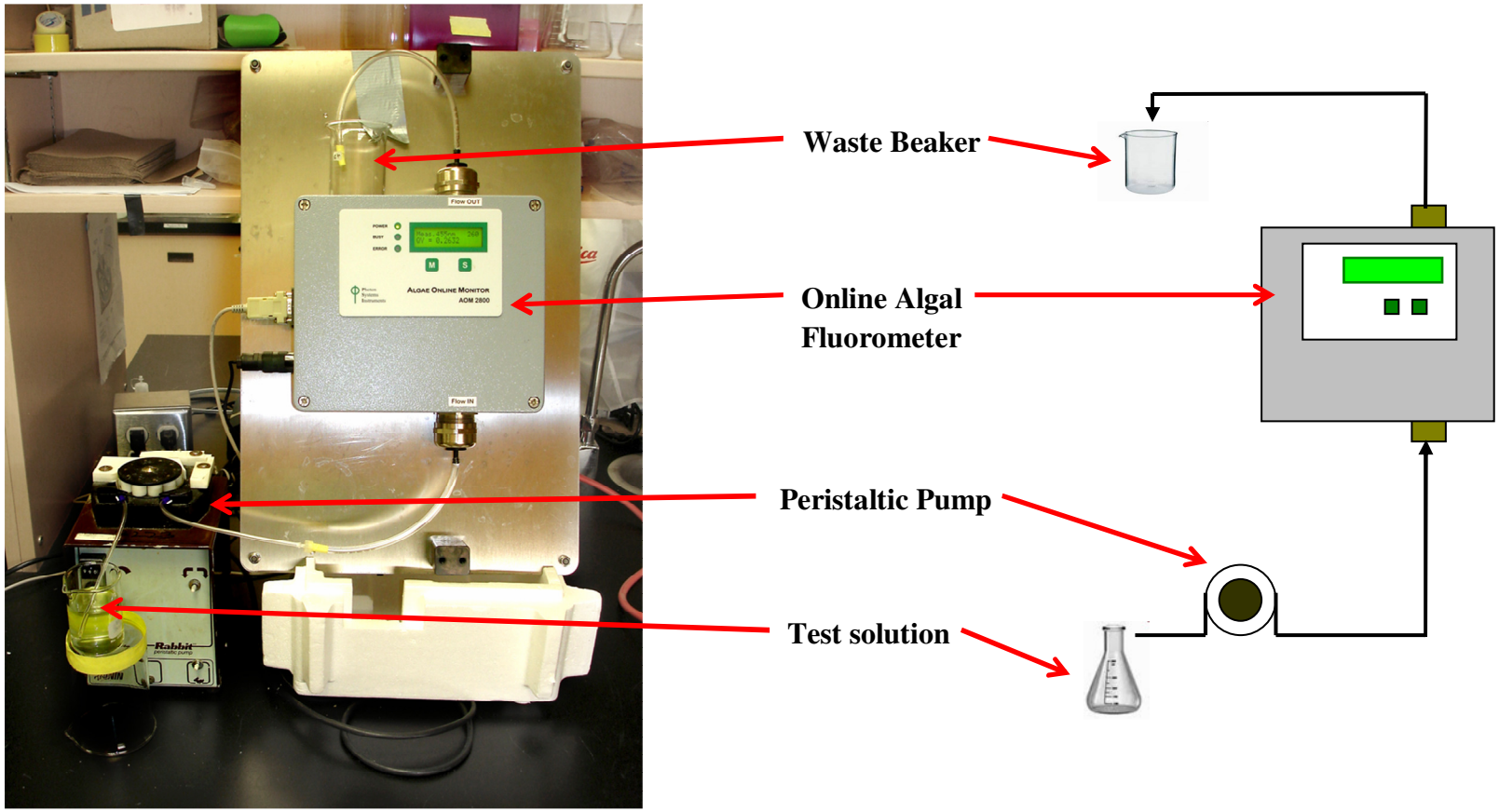
Note: The atrazine and tributyltin stocks contained 100 mg of the respective contaminant per litre of DMSO

Figure 3.1. Experimental setup and chemical composition for the test solutions prepared for the algal bioassay.

As shown in Figure 3.1, test solutions were prepared at the same concentrations used in the previous study by Marshall (2009) so that the results from the current and previous studies could be integrated into one data set. To create the test solutions, the algal suspensions were distributed into twenty-four 250-mL Erlenmeyer flasks in 50-mL aliquots. As elaborated in Section 3.2.2, the algae used in the bioassay must be at their logarithmic growth phase to ensure maximum photosynthetic activity. The algal suspension also should not be too dilute (~1×10^5 cells/mL) or too concentrated (greater than 2×10^6 cells/mL), otherwise the fluorometer would signal an error in measurements. If the suspension was too concentrated, the algae solution was diluted using the modified growth medium to ensure that the cells were not starved of nutrients. An ideal algal concentration for the bioassay was approximately $1 – 1.5 \times 10^6$ cells/mL.

After aliquoting the algal suspension into each Erlenmeyer flask, a concentrated stock of atrazine or tributyltin was added to some flasks to obtain the desired concentration of contaminants. The concentrated stock solutions of atrazine and tributyltin were prepared at a concentration of 100 milligrams contaminant per litre DMSO. A stock of pure DMSO was also prepared to be added into some flask to achieve a ratio of 0.1% v/v DMSO. Sample calculations for determining the chemical composition of each test solution are available in Appendix A.1. Test solutions were prepared in triplicates for:

- Reference condition with and without 0.1% v/v dimethyl sulfoxide;
- Atrazine at concentrations of 0.005, 0.050, and 0.100 mg/L; and
- Tributyltin at concentrations of 0.010, 0.050, and 0.100 mg/L.



Note: The test solution contained a suspension of *Pseudokirchneriella subcapitata* at a concentration of approximately 1.25×10^6 cells/mL. For each reading, a small portion of the algal suspension was transferred using a peristaltic pump to the fluorometer where the effective photosynthetic yield of the algae was determined.

Figure 3.2. Instrumental setup for the algal bioassay.

Figure 3.2 illustrates the setup for the equipments used in the algal bioassay. Each of the test solution in Figure 3.1 was shaken to ensure thorough mixing, and a small portion of the test solution was extracted using a peristaltic pump (Rabbit Peristaltic Pump No: 53890, Rainin Instrument Co, Boston, MA) at a flow rate of 6.7 mL/minute. The small extract was then fed into the fluorometer (Algae Online Monitor™ – AOM 2800, Photon Systems Instruments, Czech Republic) where the effective photosynthetic yield of the algal solution was determined. Photosynthetic yield is a highly-sensitive parameter which can be affected by changing light conditions, and therefore it was important to calibrate the amount of light present during the study. During the algal bioassay, the test solutions were placed on a shaker at 100 rpm and under a continuous 'cool white' fluorescent light. Ambient light condition was measured using a light meter (Solar Electric Quantum Meter, Item#: 3415FSE, Spectrum Technologies Inc.), and flasks were positioned to maintain a light intensity between 56 – 66 μmol/m^2-s of light quantal flux. The flasks were also randomized in between trials to remove any bias due to the position of each flask relative to the light source.

After preparing and mixing all the chemical constituents as outlined in Figure 3.1, the effective photosynthetic yields of the test solutions were immediately measured using the AOM. This measurement at t = 0 formed a baseline condition to which later measurements were compared, and measurements were conducted every 2 hours for 6 hours starting from the measurement at t = 0.

3.3 Distinct Conditions Established for the Present study

Based on the experimental setup outlined in Section 3.2 for the algal bioassay and the experimental setup developed in the previous study by Marshall (2009), a number of distinct conditions could formally be defined to establish groups of data points. First, 'Reference Condition' contained experiments conducted in dechlorinated tap water. Since the effects of DMSO were considered to be negligible on the behavioural and physiological endpoints of the aquatic organisms, test solutions containing dechlorinated tap water with 0.1% v/v could be categorized under reference condition as well. The reference condition thus represented a condition where no contaminant was present or was present at a concentration much lower than the allowable concentration limit.

Another distinct condition established by the experimental setup was 'Contamination by Atrazine', and this condition simulated a potential atrazine runoff into the aquatic systems by preparing solutions containing atrazine at various environmentally-relevant concentrations. The last distinct condition established by the experimental setup was 'Contamination by Tributyltin', where solutions containing various doses of tributyltin were prepared to represent a realistic spill of this contaminant into the waterways. The three conditions above were expected to form distinct clusters of data sets in PCA and subsequent Cluster Analysis, and the descriptions for each condition are summarized in Table 3.2.

Table 3.2. Three distinct conditions established for the library of responses

Condition	Basis for Establishing Condition
'Reference Condition'	Endpoints measured in dechlorinated tap water only Endpoints measured in dechlorinated tap water with the addition of 0.1% v/v dimethyl sulfoxide
'Contamination by Tributyltin'	Endpoints measured in tributyltin solutions with the addition of 0.1% v/v dimethyl sulfoxide. Test solutions were prepared at concentrations of 0.010, 0.050 and 0.100 mg tributyltin per litre of dechlorinated tap water.
'Contamination by Atrazine'	Endpoints measured in atrazine solutions with the addition of 0.1% v/v dimethyl sulfoxide. Test solutions were prepared at concentrations of 0.005, 0.050 and 0.100 mg atrazine per litre of dechlorinated tap water.

3.4 Methodology for Creating the Library of Responses

After measuring the endpoint in the algal bioassay and extracting endpoint data from a study by Marshall (2009), a total of 18 parameters were proposed for inclusion in the library of responses. As outlined in Table 2.2, 17 endpoints were extracted from the study by Marshall (2009) which monitored the parameters for up to 24 h in solutions containing various levels of tributyltin and atrazine with the addition of 0.1% v/v DMSO. The goal of the EWBS, however, was to provide rapid, early-warning detection, and thus the present study only required endpoint measurements up to 6 h in order to capture the immediate, sub-acute responses of the organisms. The endpoint data from Marshall (2009) at t = 6 h for *Daphnia magna, Hyalella azteca* and *Lumbriculus variegatus* are attached in Appendices B.1 – B.3.

After conducting the algal bioassay, the responses collected from all four aquatic organisms were then arranged into a matrix of responses and analyzed for any contaminant-specific patterns. As previously elaborated, Principal Component Analysis and Cluster Analysis were applied in the current research to simplify the data set of responses and objectively identify the presence of any clusters within the data set. The clusters could then be labelled according to the three established conditions as outlined in Table 3.2. The step-by-step methodologies for PCA were established earlier in Section 2.4.1 and Figure 2.4. All statistical computations were conducted using the MATLAB software (Version 7.7.0.471 (R2008b), MathWorks). The detailed computer codes to perform each operation are included in Appendix C.

The following steps were taken in the present work to perform PCA:

1) *Data points were tabulated*

Measurements of the 18 endpoints were arranged into an 18-by-30 matrix. Each column represented an endpoint, and 30 observations were collected for each endpoint where 6 replicates were conducted in dechlorinated tap water, 6 replicates in dechlorinated tap water with 0.1% v/v DMSO, 9 replicates in various concentrations of atrazine, and 9 replicates in various concentrations of tributyltin.

2) *Variables were pre-conditioned*

To minimize any bias exhibited by variables of larger magnitudes, mean-centering was applied by subtracting the mean from each data point across each dimension. This procedure resulted in a modified matrix with a mean of zero for each endpoint. Two options were then applied to further pre-condition the mean-centered matrix. In the first option, data normalization was not performed, while in the second option, data normalization was applied to further reduce any bias by larger variables. By conducting the two options separately, the effects of normalization could be compared and analyzed.

3) *The covariance matrix was calculated from the pre-conditioned matrix*

After mean-centering, a covariance matrix was calculated from the modified matrix in Step 2. The covariance matrix described the linear relationships among all 18 variables.

4) *The eigenvectors and eigenvalues of the covariance matrix were calculated*

After the covariance matrix was constructed, a total of 18 unit eigenvectors and eigenvalues could be calculated. Each eigenvector was composed of coefficients which would transform the original data set, and the eigenvalues corresponded to the amount of variance described by the eigenvectors.

5) *The original data set was transformed*

In order to minimize the data complexity and facilitate pattern analysis, it was decided to reduce the number of variables from 18 to 2 principal components. To accomplish this reduction in dimension, matrix multiplication was applied between the original data set and 2 eigenvectors with the largest eigenvalues, resulting in a final matrix of 30 observations and 2 principal components.

After transforming the original data set using PCA, the new matrix was then graphed and examined for any contaminant-specific pattern. Non-hierarchical cluster analysis tools, such as K-means clustering, was applied to objectively separate the data points into the 3 clusters as defined in Section 3.3, and the number of correctly assigned data points would describe how clearly defined and well-separated the clusters were. Calculations for CA were performed using the MATLAB software, and the codes are listed in Appendix C.

The following steps described the K-means clustering algorithm employed in the present study:

1) *Three random data points were selected*

Since the data set was to be separated into the 3 conditions as described in Section 3.3, 3 random data points were selected from the transformed matrix from Step 5 of the PCA algorithm.

2) *Clusters were generated based on the 3 chosen points*

Using the 3 random points chosen in Step 1, 3 clusters were established where the remaining data points would be assigned to the most similar cluster. The distance between each remaining data point and each 3 chosen points were calculated, and the data point would be assigned to the cluster with the shortest distance.

3) *Cluster means were re-calculated*

After grouping data points into 3 clusters, a new mean was calculated for each cluster. All objects are then re-evaluated and re-assigned into the most similar cluster.

4) *Convergence was reached*

Step 3 was repeated several times until no data point was re-assigned to a different cluster and no changes were made in between iterations. A full convergence was then reached, and 3 clusters were established.

After applying K-means clustering, the cluster quality was further assessed using the Average Silhouette Width Plot technique as described in Section 2.4.2. By calculating the silhouette coefficient for each cluster, the validity of each group of data points was evaluated. If distinct clusters were indeed formed by the data points, this finding would then demonstrate the capability of the suite of organisms to resolve different classes of contaminants as well as illustrate the versatility of multivariate statistics in revealing hidden patterns among large data sets. The transformed matrix of responses could then be incorporated as a preliminary library of responses for the EWBS.

3.5 Methodology for Testing the Library of Responses

After constructing the preliminary library of responses, the capability of this library to resolve different contaminants was evaluated. As discussed in Section 3.3, the library of responses was constructed using endpoints at t = 6 h from the present study and a previous study by Marshall (2009). The endpoints at t = 6 h thus served as the standard to which subsequent readings could be compared. Since no blind experiments had been conducted for the current NSERC project, there was no actual reading data yet to test the library. To simulate real *in-situ* reading data, two matrices of responses of 18 endpoints and 30 observations (12 observations for reference condition and 9 for each contaminant) were created using experimental data at t = 2 h and t = 4 h from the present algal bioassay and previous experiments by Marshall (2009). Raw data for these two matrices are provided in Appendix E. The resolving capability of the library was then determined according to the number of correctly classified data points.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

As discussed in Chapter 3, a bioassay was developed to measure the effective photosynthetic yield of *Pseudokirchneriella subcapitata* over a period of six hours in solutions containing different doses of atrazine and tributyltin. The results from the bioassay were analyzed in this chapter, and the suitability of the Algae Online Monitor™ for inclusion in the early-warning biomonitoring system was evaluated. Data from the algal bioassay was then combined with the measurements collected by Marshall (2009), and the combined data set was transformed using principal component analysis. Any distinct patterns pertaining to each contaminant was identified, as well as the general trend and behaviour associated with the data points. Then, a preliminary library of responses was constructed, and the ability of this library to resolve the two contaminants was evaluated. Last, the present study ended with a discussion of how the library of responses could be incorporated into an early-warning biomonitoring system and applied in an automated system to identify contaminants in a water sample.

4.2 Results and Evaluation of the Algal Bioassay

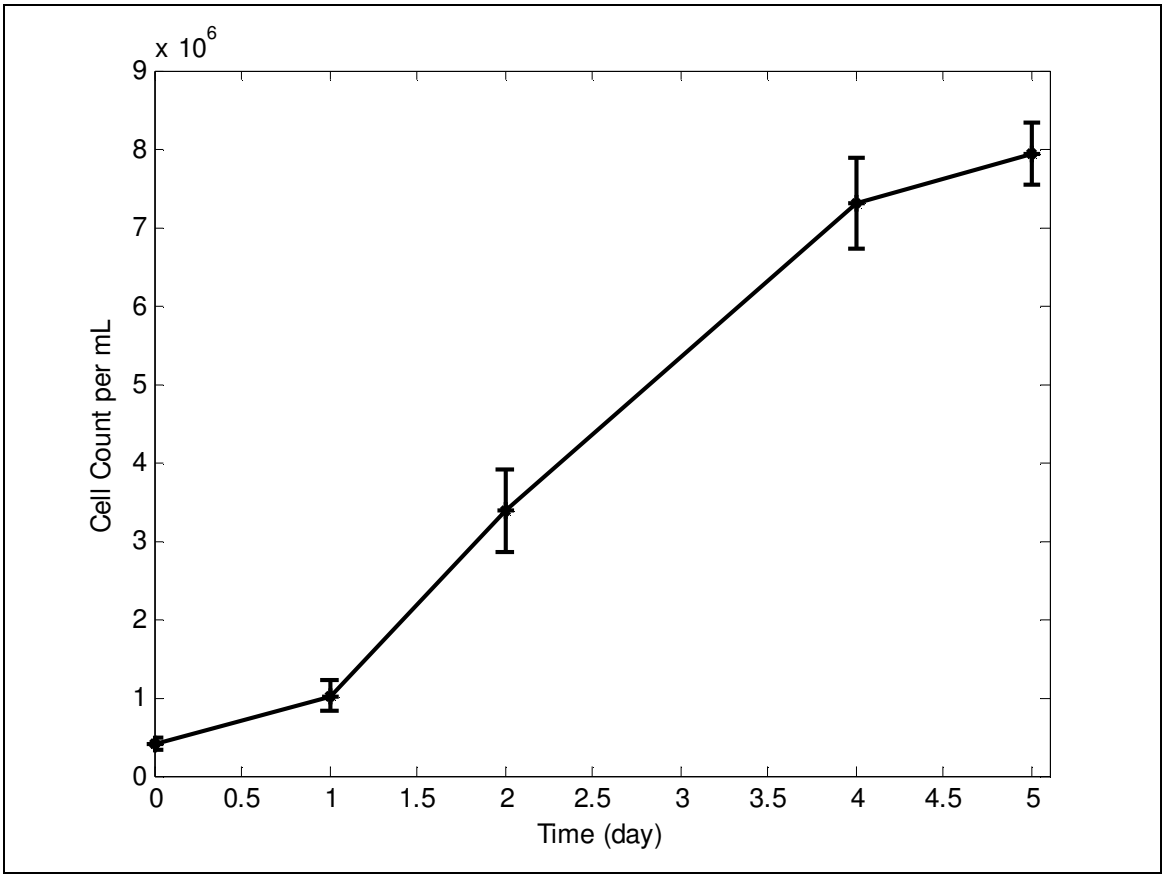
4.2.1 Construction of the Algal Growth Curve

Pseudokirchneriella subcapitata cultures were grown for use in the bioassay according to the protocols modified from Environment Canada (2007). As elaborated in Section 3.2.2, the protocols for preparing the growth medium were modified, resulting in

a rich medium containing four times more nutrients than the original protocol. More starting culture was also used to subculture the alga, and the algal concentration was measured every day using a haemocytometer to monitor for the exponential growth phase. The resulting growth curve is shown in Figure 4.1, and the raw data and sample calculations are included in Appendix A.2.

The growth curve for the algae in the present study was comparable to the 8 to 10-day growth curve of *P. subcapitata* published by Environment Canada (2007). As seen in Figure 4.1, the algal culture started at a concentration of $4.0 \pm 0.8 \times 10^5$ cells/mL and grew exponentially before finally reaching a stationary phase at t = 5 d. In a similar study by Environment Canada (2007), the algal culture reached the exponential growing phase at t = 3 d before finally arriving at the stationary phase around t = 8 d. This finding confirmed that the algal culture used in the present study was healthy since the culture exhibited an exponential growth within 7 days, and this culture was therefore suitable for use in bioassays (Environment Canada, 2007).

Constructing the growth curve also helped determine the suitability of the culture for use in a bioassay by identifying the growth stage of the algae. For example, conducting a bioassay using a senescing algal culture would introduce a large bias in the measurements of effective photosynthetic yield as the culture would be undergoing a natural decline (Behera *et al.*, 2003). By utilizing an algal culture that is actively growing, the bioassay would be able to capture the maximum effective photosynthetic yield of the aquatic plant.



Note: The bioassay was conducted at t = 4 d, when the concentration of the algal suspension was ~7.3 ± 0.58 × 10^6 cells/mL.

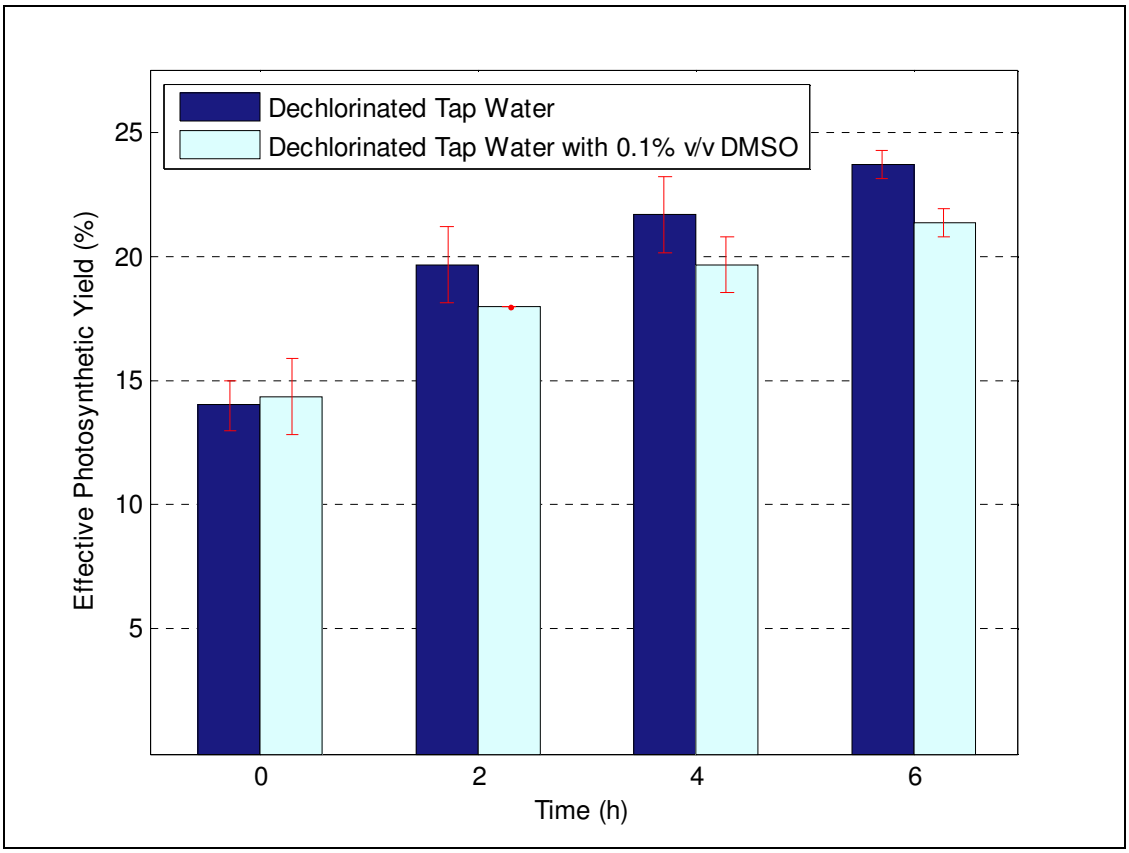
Figure 4.1. Growth curve of *Pseudokirchneriella subcapitata* used for the algal bioassay in the present study.

At t = 4 d, the concentration of the algal culture reached 7.3 ± 0.58 × 10^6 cells/mL, and this culture was ideal for use in a bioassay. As noted in Section 3.2.3, however, the algal concentration used in the bioassay should be around 1 × 10^6 cells/mL to prevent any error in measurements by the fluorometer. The algal culture was then diluted using the enriched media until the concentration was 1.25 ± 0.23 × 10^6 cells/mL. The effects of atrazine and tributyltin on the effective photosynthetic yield of this diluted algal culture were subsequently measured.

4.2.2 Influence of Atrazine and Tributyltin on the Effective Photosynthetic Yield of Pseudokirchneriella subcapitata

The effective photosynthetic yield can be defined as a measure of the performance of the photosystem II complex in plants (Strasser *et al.*, 2000). Photosystem II is a protein complex which contributes to photosynthetic reactions by transferring the energy absorbed from photons of light into electrons and subsequently driving photosynthesis through a series of cascade reactions. Electron transport is crucial in photosynthesis, and thus any inhibition on the activity of the photosystem II complex may affect the physiological state of the plant (Juneau and Popovic, 1999).

To examine the effects of atrazine and tributyltin on the effective photosynthetic yield of *P. subcapitata,* a series of test solutions were prepared according to the compositions listed in Figure 3.1, and the effective photosynthetic yield of each of these test solutions was measured. Figure 4.2 depicted the effects of adding DMSO on the photosynthetic yield of the algae.

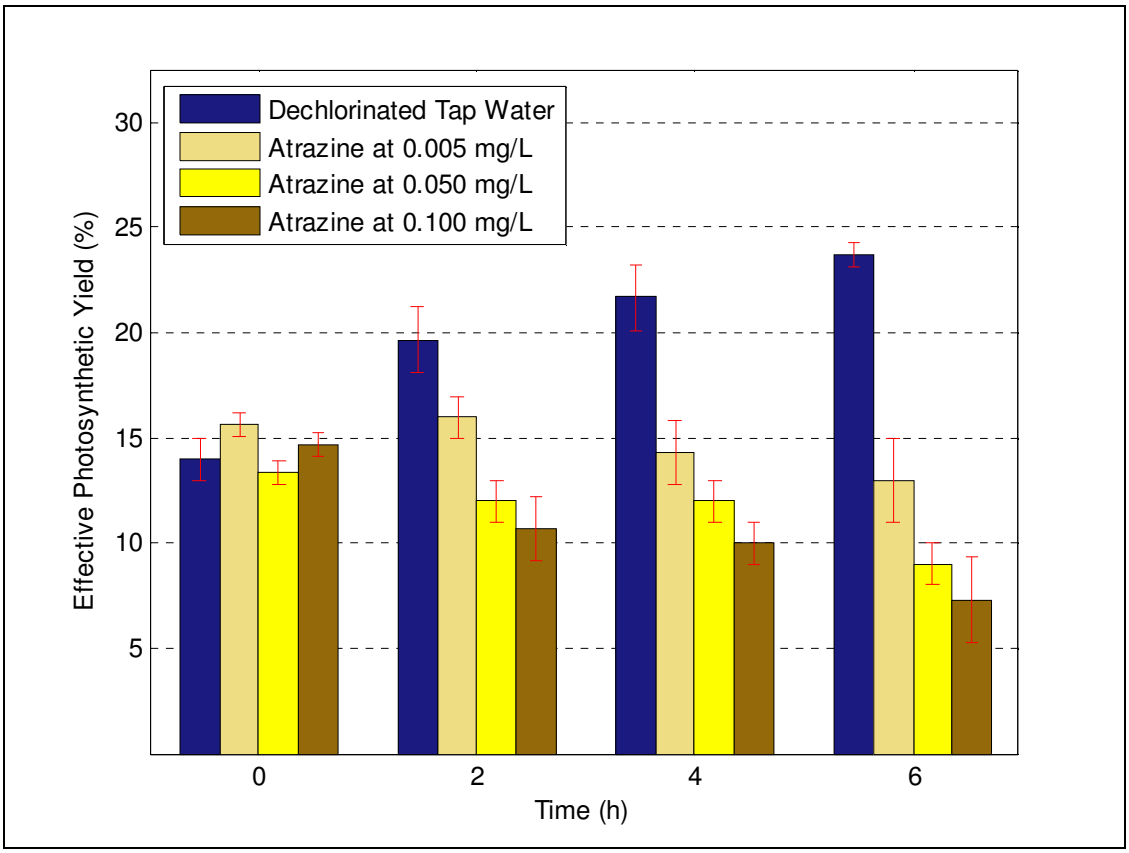


Note: **Error bars were drawn at ± 1 standard deviation.**

Figure 4.2. Effective photosynthetic yields of *Pseudokirchneriella subcapitata* measured over a period of 6 h in dechlorinated tap water with and without 0.1% v/v dimethyl sulfoxide.

From Figure 4.2, the effective photosynthetic yields for the two treatments showed a tendency to increase as the experiment progressed. Since the amount of fluorescence was proportional to the amount of absorbing pigments (Strasser *et al.*, 2000), this small increase in photosynthetic yield could potentially be attributed to the fact that the test algae were at the exponential growth phase and thus were actively growing. The addition of DMSO, however, appeared to have slightly influenced the effective photosynthetic yield, resulting in lower values at t = 6 h. A one-way analysis of variance (ANOVA) comparing the photosynthetic yields between reference and DMSO treatments at t = 6 h revealed a p-value of 0.0078, signifying a statistically-significant difference between the two treatments at 95% confidence. This finding did not agree with some published studies as DMSO is a commonly used solvent for bioassays involving herbicides and pesticides (El Jay, 1996; Pearce, 2009; Marshall, 2009). Further testing was therefore recommended to ensure that the DMSO did not significantly influence the endpoint measurements in the present study and was suitable for use in future bioassays.

In the next analysis, the effects of various doses of atrazine on the effective photosynthetic yield of algae were studied, and the results for the atrazine experiments are shown in Figure 4.3. As previously noted, dimethyl sulfoxide at a concentration of 0.1% v/v was added to all atrazine solutions to increase the solubility of the compound in the dechlorinated tap water.



Note: **Error bars were drawn at ± 1 standard deviation.**

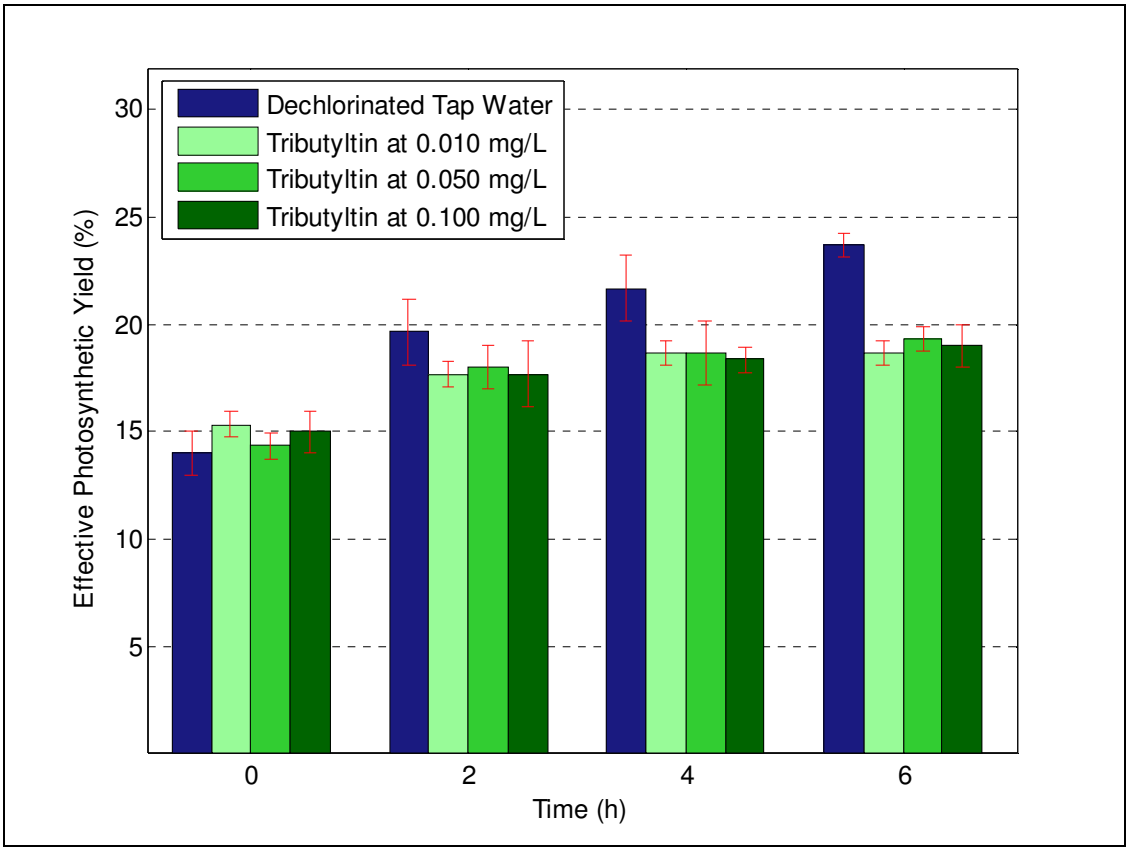
Figure 4.3. Effective photosynthetic yields of *Pseudokirchneriella subcapitata* measured over a period of 6 h in dechlorinated tap water and in various solutions of atrazine.

Since atrazine is a photosystem II inhibitor, the addition of this contaminant was expected to strongly affect the effective photosynthetic yield of *P. subcapitata* as the photosynthetic yield is a direct measurement of the photosystem II efficiency (Fai *et al.*, 2007). Atrazine exerts its phytotoxic action on aquatic plants by interfering with the electron transport process in photosystem II and binding specifically to the quinone site within the thylakoid membrane of the protein (Solomon *et al.,* 1996). As described by Equation (1) in Section 2.3.4, if the ability of the plant to absorb and process energy from photons is reduced, more of the exerted energy would be dissipated as fluorescence, resulting in a lower photosynthetic yield.

From Figure 4.3, exposure to atrazine resulted in a decrease in the effective photosynthetic yield of the algae with a response time of at least 2 h. A similar study by Fai *et al.* (2007) also noted the inhibitory effects of atrazine on the photosynthetic yields of *P. subcapitata* which was observed after 1 h of incubation. In the present study, higher atrazine concentrations also resulted in a further reduction of the endpoint. A one-way ANOVA revealed statistically-significant differences at 95% confidence level between:

- 0.005 mg /L atrazine and dechlorinated tap water at t = 6 h (p-value = 0.0009);
- 0.050 mg/L atrazine and dechlorinated tap water at t = 6 h (p-value = 2.53×10^{-5}); and
- 0.100 mg/L atrazine and dechlorinated tap water at t = 6 h (p-value = 0.0002).

Last, the effects of various doses of tributyltin on the effective photosynthetic yield of the algae were investigated, and the results are displayed in Figure 4.4.



Note: **Error bars were drawn at ± 1 standard deviation.**

Figure 4.4. Effective photosynthetic yields of *Pseudokirchneriella subcapitata* measured over a period of 6 h in dechlorinated tap water and in various solutions of tributyltin.

Unlike the experiments with atrazine, the various doses of tributyltin did not as strongly influence the effective photosynthetic yield of *P. subcapitata,* but a significant decrease in the endpoint was still observed after 2 h of exposure. Increasing the concentration of tributyltin also did not seem to more strongly affect the endpoint as shown in Figure 4.4, and the effective photosynthetic yields for all tributyltin experiments remained very similar throughout the entire bioassay. Overall, the effective photosynthetic yield of the algae increased over a period of 6 h, but the rates of increase were reduced by at least 40% from the reference condition. A one-way ANOVA revealed statistically significant differences at 95% confidence level between:

- 0.010 mg /L tributyltin and dechlorinated tap water at t = 6 h (p-value = 0. 0004);
- 0.050 mg /L tributyltin and dechlorinated tap water at t = 6 h (p-value = 0.0008); and
- 0.100 mg/L tributyltin and dechlorinated tap water at t = 6 h (p-value = 0.022).

Very few studies have actually examined the effects of tributyltin on the photosynthetic parameters of plants. Jensen *et al.* (2004) studied the effects of tributyltin-contaminated soil on the seagrass *Ruppia maritima.* The performance of the plants was monitored over a period of 3-4 weeks, and the photosynthetic activity was found to be reduced by up to 60%. Fargasová (1998) also examined the inhibitory effects of various tributyltin compounds on the freshwater planktonic alga *Scenedesmus quadricauda*, where a reduction in growth, photosynthetic activities, and chlorophyll content was observed after 12 days of cultivation. Despite some noted inhibitory effects, the above studies were long-term bioassays which lasted for several weeks and thus could not be

incorporated into an early-warning biomonitoring system. It was interesting to note, however, that the Algae Online Monitor™ used in the present study was able to detect a change in the physiological state of the algae upon exposure to tributyltin within only 2 h. Similar rapid detections were also reported by Pearce (2009) where tributyltin at 0.100 mg/L was found to significantly inhibit the photosynthetic rate of *P. subcapitata* after 6 h.

Overall, the algal bioassay illustrated the importance of employing multiple aquatic organisms from different trophic levels in a biomonitoring system, as each organism may exhibit distinct sensitivities to a particular contaminant. In this case, the alga *Pseudokirchneriella subcapitata* was susceptible to herbicides, and the effective photosynthetic yield of the alga was strongly affected by the presence of a PSII-inhibitor such as atrazine. By measuring the photosynthetic yield, the bioassay was able to detect the presence of atrazine as rapid as within 2 h of exposure. Exposure to tributyltin also caused a change in the effective photosynthetic yield, but the deviation was not as significant as when exposed to atrazine. This difference in response intensity would be a useful discriminating parameter to help resolve contamination cases by atrazine and tributyltin. Due to the sensitivity of the effective photosynthetic yield, this endpoint was recommended for inclusion in the library of responses for the early-warning biomonitoring system.

4.2.3 Performance Evaluation of the Algae Online Monitor™ 2800

A schematic of the Algae Online Monitor™ (model: AOM 2800) was shown in Figure 3.2 in Section 3.2.3. According to the specifications from the manufacturer, the

device is capable of online detection and continuous monitoring of several photosynthetic parameters in a field or laboratory setting. Since one of the contaminants tested in the present study was a known PSII-inhibitor, the effective photosynthetic rate was deemed the most suitable parameter for monitoring. Besides photosynthetic yield, the device can also measure the instantaneous fluorescence (F_T) and construct a curve for the chlorophyll fluorescence induction kinetics (0JIP curve). Experiments measuring the 0JIP curve are typically conducted on dark-adapted plants because dark-adaptation results in a reduced photochemical activity of the plant which then allows for the measurement of the maximum algal fluorescence as described by Equation (2) in Section 2.3.4. The shape of the 0JIP curve can then provide rich information on the physiological state of the sample (Strasser *et al.*, 2000). For the current study, measuring the effective photosynthetic rates was already adequate in characterizing the effects of atrazine and tributyltin on *Pseudokirchneriella subcapitata*, but if more studies on the physiology of the algal cells are required, experiments measuring the behaviour of dark-adapted algal cultures as well as the change in the 0JIP curve could be considered.

A recent literature search did not find any published study employing this particular model of fluorometer, and thus an objective assessment of the AOM 2800 could not be obtained. A published study involving the same device would have been very useful as some of the features of the AOM 2800 were not clearly detailed in the operation manual. As a result, it was necessary to perform a number of trial-and-error experiments to obtain empirical settings for several experimental parameters such as flow rate and concentration. Despite the claimed detection limit of 10 algal cells/mL, the

fluorometer was not able to measure the photosynthetic yield of algal cultures with a concentration of less than 1×10^5 cells/mL. When the concentration exceeded ~2×10^6 cells/mL, an 'overflow' error message was displayed by the fluorometer. Several preliminary experiments found an optimal concentration of ~1×10^6 cells/mL when used in conjunction with a pump flow rate of 6.7 mL/minute.

Preliminary experiments also found that by reducing the algal concentration, the AOM yielded lower effective photosynthetic yield measurements even though the physiological state of the algal cells was unchanged. Since the current bioassay was conducted as a semi-batch experiment, where aliquots of different test solutions were fed into the fluorometer at different times, maintaining roughly the same concentration throughout the experiment was easily accomplished. When applied in a continuous flow-through setting, however, an instrument such as a turbidostat must be employed to sustain the concentration of the algal cells and prevent a false assessment of the water condition.

Overall, the present study found the performance of the AOM 2800 to be satisfactory. Despite some initial difficulties encountered in the experimental setup, the bioassay was able to rapidly detect the change in the effective photosynthetic yields of *P. subcapitata* when exposed to atrazine and tributyltin. Upon examination of the features of the device, the fluorometer could potentially conduct a continuous assessment of water quality with minimal supervision since the device was equipped with an automated measuring function at defined intervals. The memory capacity of the device was 4 megabytes which could store up to 100,000 measurements or approximately 300 0JIP

curves. Further tests examining the performance of the device in a flow-through setting were recommended to assess whether the machine could effectively provide a real-time online analysis of water conditions.

4.3 Application of Multivariate Statistical Tools on the Endpoint Data Set

A number endpoint data were collected from a study by Marshall (2009), and one endpoint was obtained from the algal bioassay in the present study. The endpoint data for *Daphnia magna, Hyalella azteca* and *Lumbriculus variegatus* at t = 6 h are attached in Appendices B.1 – B.3, while the measurements for the effective photosynthetic yield of *Pseudokirchneriella subcapitata* are listed in Appendix A.3. These endpoints resulted in a 30-by-18 matrix of responses.

4.3.1 Results of Principal Component Analysis

After constructing the 30-by-18 matrix of responses, PCA was applied to this data set to reduce the number of dimensions to 2 principal components. All statistical calculations were performed using the MATLAB software and according to the steps outlined in Section 3.4. The MATLAB codes to perform the steps are listed in Appendix C, and all intermediate calculations for the PCA are included in Appendix D. As elaborated earlier, two different options were considered for the matrix preconditioning – non-normalization and normalization. The non-normalized data set is shown in Table 4.1 and Figure 4.5, and the normalized data set in Table 4.2 and Figure 4.6.

Table 4.1. Transformed non-normalized matrix of responses at t = 6 h.

Treatment	Principal Component 1	Principal Component 2
Dechlorinated tap water		
Replicate 1	-154.57	4.01
Replicate 2	-154.60	3.93
Replicate 3	-154.60	3.93
Replicate 4	-154.57	4.01
Replicate 5	-154.60	3.93
Replicate 6	-147.65	2.21
Dechlorinated tap water with 0.1% v/v DMSO		
Replicate 1	-152.63	0.39
Replicate 2	-152.63	0.39
Replicate 3	-154.53	4.09
Replicate 4	-154.13	1.58
Replicate 5	-154.50	4.17
Replicate 6	-147.58	2.36
Tributyltin (0.010 mg/L)		
Replicate 1	6.36	-55.92
Replicate 2	87.63	-17.44
Replicate 3	42.27	-14.24
Tributyltin (0.050 mg/L)		
Replicate 1	54.68	-68.68
Replicate 2	98.61	-24.74
Replicate 3	87.99	-46.86
Tributyltin (0.100 mg/L)		
Replicate 1	136.75	-95.46
Replicate 2	158.12	-26.60
Replicate 3	138.47	-42.89
Atrazine (0.005 mg/L)		
Replicate 1	40.00	49.33
Replicate 2	84.87	35.55
Replicate 3	89.83	34.61
Atrazine (0.050 mg/L)		
Replicate 1	121.12	67.07
Replicate 2	128.16	36.94
Replicate 3	109.59	15.09
Atrazine (0.100 mg/L)		
Replicate 1	157.51	79.78
Replicate 2	161.30	33.63
Replicate 3	133.35	5.83

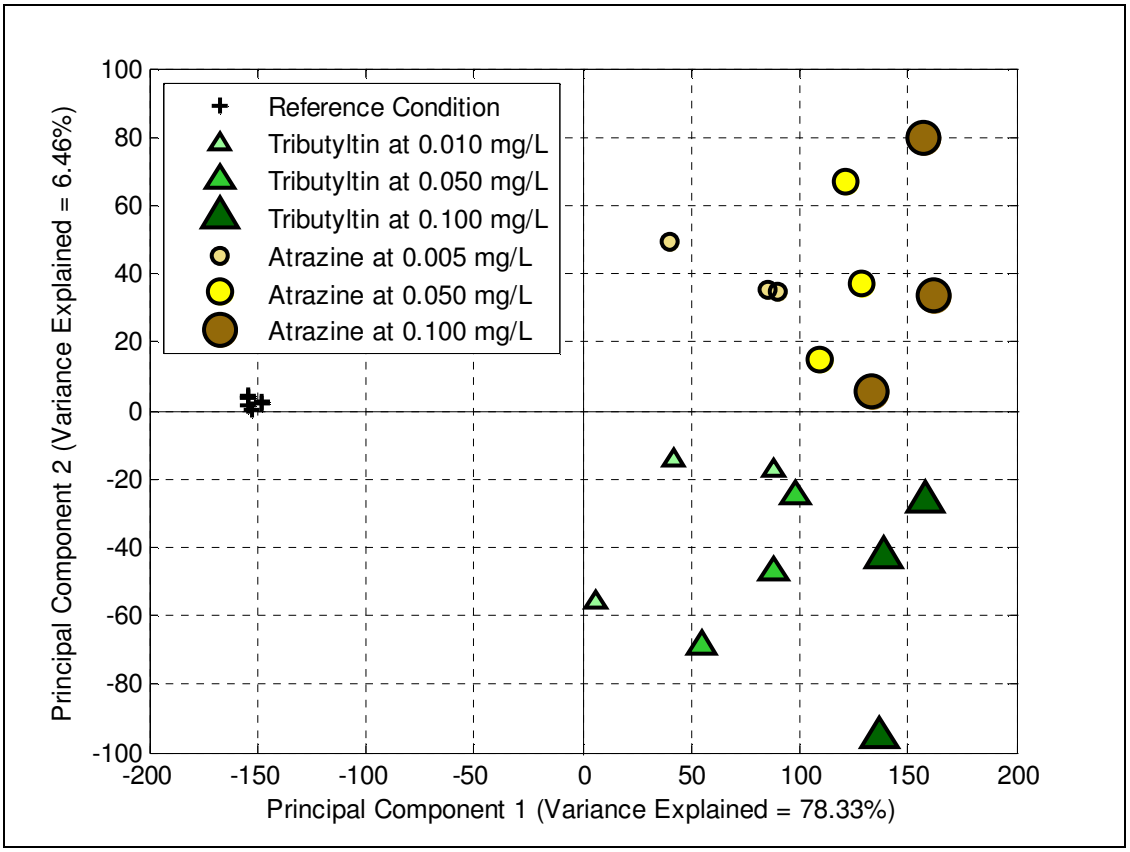


Figure 4.5. A plot of the transformed non-normalized matrix of responses containing endpoints at t = 6 h.

Table 4.2. Transformed normalized matrix of responses at t = 6 h.

Treatment	Principal Component 1	Principal Component 2
Dechlorinated tap water		
Replicate 1	-146.60	-5.68
Replicate 2	-146.82	-6.05
Replicate 3	-146.82	-6.05
Replicate 4	-146.60	-5.68
Replicate 5	-146.82	-6.05
Replicate 6	-141.17	-6.18
Dechlorinated tap water with 0.1% v/v DMSO		
Replicate 1	-143.13	-7.81
Replicate 2	-143.13	-7.81
Replicate 3	-146.38	-5.30
Replicate 4	-144.98	-9.85
Replicate 5	-146.16	-4.93
Replicate 6	-140.72	-5.43
Tributyltin (0.010 mg/L)		
Replicate 1	8.38	-46.50
Replicate 2	84.18	1.63
Replicate 3	30.63	5.97
Tributyltin (0.050 mg/L)		
Replicate 1	46.97	-48.52
Replicate 2	91.02	-14.96
Replicate 3	82.29	-32.06
Tributyltin (0.100 mg/L)		
Replicate 1	144.01	-74.02
Replicate 2	157.27	-33.64
Replicate 3	142.02	-43.17
Atrazine (0.005 mg/L)		
Replicate 1	27.03	28.30
Replicate 2	68.44	33.31
Replicate 3	86.73	35.39
Atrazine (0.050 mg/L)		
Replicate 1	115.39	53.98
Replicate 2	116.58	43.76
Replicate 3	104.78	30.95
Atrazine (0.100 mg/L)		
Replicate 1	147.08	80.99
Replicate 2	155.38	42.33
Replicate 3	131.18	13.06

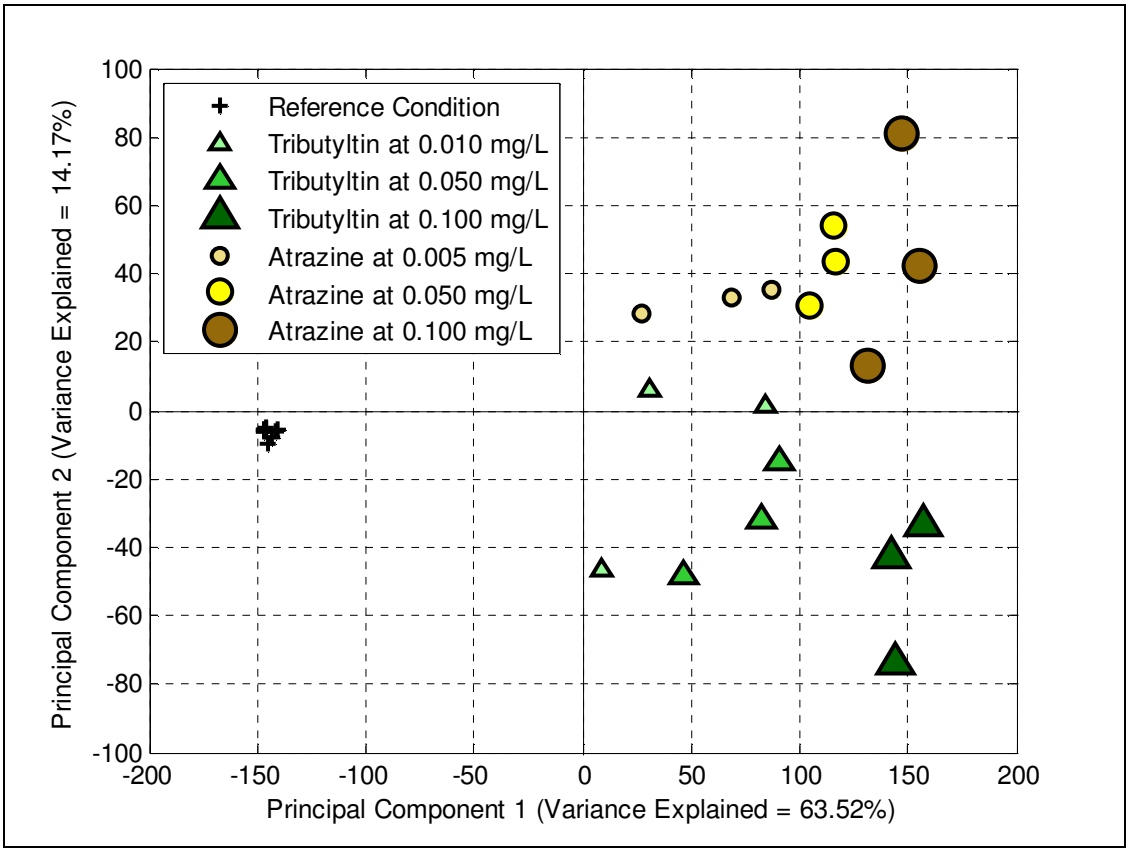


Figure 4.6. A plot of the transformed normalized matrix of responses containing endpoints at t = 6 h.

The transformed non-normalized matrix of responses contained approximately 85% of information from the original data set, where the first principal component (PC1) described 78.3% of the variance and the second principal component (PC2) described an additional 6.5% of variance. The allocated variance was determined by the eigenvalue of each eigenvector as described in Appendix D. For the transformed normalized matrix of responses, 63.5% of variance is contained in PC1 while 14.2% of variance was described by PC2, totalling approximately 78% of information from the original data set. In the non-normalized case, data points appeared to form three distinct groups with a relatively greater distance between the atrazine and tributyltin treatments. For the normalized case, significant separation was obtained between higher concentrations of atrazine and tributyltin, but not at lower concentrations of the toxicants.

A test to formally evaluate the number of non-trivial axes does not currently exist, and therefore the number of axes must be selected based on the purpose and objectives of each study (Shaw, 2003). Some methods have been discussed among statisticians to determine the appropriate number of principal components required to accurately describe the original data set without significant loss of information. An example of such evaluation method is the Kaiser-Guttman criterion which recommended discarding any principal components with an eigenvalue of less than 1% (Cliff, 1988). A major drawback to the Kaiser-Guttman criterion, however, is the tendency of the method to overestimate the number of non-trivial axes, resulting in a data set that would still contain a large number of variables (Jackson, 1993). Another method was suggested by Jackson (2003) where the total explained variability by different principal components should

exceed 80%. By following this method, selecting two principal components in the present study was sufficient for the non-normalized case as the total variance was 85%, but the normalized case did not meet this criterion as the total variance was only 78%. This finding suggested that reducing the original data set into two principal components was not appropriate for the normalized case, and only the non-normalized case was therefore considered for further analysis in the present work. Normalizing the original matrix of responses may also be inappropriate for the current study as the number of trials and replicates were relatively few, resulting in a more discrete distribution. The normalized case, however, could still be considered in future studies by including more principal components and rotating the axes which could increase the total variance.

In Figure 4.5, the two principal components displayed the data points as three different groups which corresponded to the three different treatments described in Table 3.2 of Section 3.3. The reference condition contained measurements conducted in dechlorinated tap water with and without 0.1% v/v DMSO, and data points from this condition formed a closely-packed cluster to the left of the origin. In the first quadrant of Figure 4.5, a loosely-formed group of data points was observed, and this group comprised measurements conducted in solutions of atrazine. To the opposite of the atrazine group, just across the PC1 axis and in the fourth quadrant of the graph, another loosely-formed group of data points was found. This second group was composed of measurements conducted in solutions containing tributyltin. Higher doses of either contaminant were also found to cause data points to shift further toward larger positive PC1 values.

The first principal component contained the largest variance, and this value was attributed to the responses exhibited by the selected aquatic organisms to the contaminants. Since data points were pre-conditioned by mean-centering, the overall mean of the two principal components was located at the origin of the graph. When the organisms were not stressed, they displayed very small deviations, close to 0%, in their parameters, and these below-average deviations resulted in the data points falling to the left of the average for PC1. When a stressor was introduced into the test solution, however, the organisms displayed rapid changes in their endpoints, which in turn caused a drift towards the positive PC1 direction. A greater intensity in response generally resulted in a drift further away from the origin. Based on this finding, PC1 could therefore be used as an indicator of a contamination or a stressed condition for the behavioural endpoints, regardless of the cause or the type of contamination.

The proposed early-warning biomonitoring system, however, aimed to not only accurately detect the presence of a contaminant, but also to identify a number of contaminants. In order to identify the contaminants in the present study, the second principal component was useful in differentiating the responses exhibited by the selected aquatic organisms. As observed in Figure 4.5, atrazine elicited certain responses in the aquatic organisms which caused a drift towards the positive PC2 direction. Conversely, the responses in solutions containing tributyltin resulted in data points leaning toward the negative PC2 direction. These findings reflected the different modes of action exhibited by the two chemicals on the endpoints of the aquatic organisms and the possibility to characterize contaminant-specific responses for atrazine and tributyltin.

Equation (4) in Section 2.4.1 describes the relationship between the original endpoint measurements and the new data points plotted in Figure 4.5. This equation also established PC1 and PC2 as aggregate variables, which signified that each principal component was composed of different weightings of the original 18 endpoints. These weightings were determined by the eigenvectors as the original matrix of responses was transformed through matrix multiplication with these eigenvectors. Each of the eigenvector was composed of coefficients as shown in Table 4.3.

By examining the eigenvector coefficients listed in Table 4.3, endpoints which acted as indicators and/or discriminators could then be identified. For example, endpoints with a large positive PC1 coefficient tended to strongly indicate the presence of a contaminant, and an example of such endpoint would be the percentage of *Daphnia magna* exhibiting a change in the swimming height. As discussed earlier, the exposure to atrazine generally caused a data point to fall on the positive PC2 axis, while exposure to tributyltin resulted in a drift towards the negative PC2 axis. Based on this observation, endpoints with a large positive PC2 coefficient, such as the percentage of *Hyalella azteca* burrowing, were more strongly affected by atrazine, while endpoints with a large negative PC2 coefficient, such as the percentage of *Lumbriculus variegatus* moving within groups, were better predictors of tributyltin contamination. By incorporating the contaminant indicators and discriminators into the design of the early-warning biomonitoring system, a more accurate and comprehensive analysis of water quality could be accomplished.

Table 4.3. Eigenvector coefficients which governed the contribution of each endpoint towards the construction of the transformed data set.

Measured Endpoint (%)	PC1 coefficients	PC2 coefficients
Daphnia magna		
Changing swimming height	0.35	-0.12
Spinning	0.09	-0.19
Changing in body orientation	0.18	0.21
Immobilized	0.23	0.03
Swimming using secondary antennae	0.24	-0.15
Changing swimming style	0.32	0.11
Hyalella azteca		
Changing swimming height	0.35	-0.09
Immobilized	0.35	0.02
Burrowing	0.21	0.43
Changing grouping behaviour	0.26	0.49
Shortening body length	0.33	0.06
Changing body orientation	0.04	-0.16
Lumbriculus variegatus		
Displaying abnormal behaviour	0.34	-0.18
Immobilized	0.07	-0.27
Shortening body length	0.16	-0.34
Changing body orientation	0.04	-0.26
Moving within groups	0.12	-0.36
Pseudokirchneriella subcapitata		
Effective photosynthetic yield	0.03	0.08

4.3.2 Further Analysis of Endpoints Using a Biplot

In order to visualize the significance and the extent of contribution by each endpoint to the transformed data set, a biplot can be constructed. A biplot is an exploratory graph that displays the eigenvector coefficients as unit vectors and then superimposes the vectors on a plot of the transformed data points (Kohler and Luniak, 2005; Reimann *et al.*, 2008). Figure 4.7 depicts a biplot of the eigenvector coefficients in Table 4.3 and the data points from Table 4.1 which had been scaled by a factor of 1/200 to more clearly illustrate the relationship between the unit vectors and the data points. By applying the analysis in a study of biplots by Kohler and Luniak (2005), two properties in Figure 4.7 were examined.

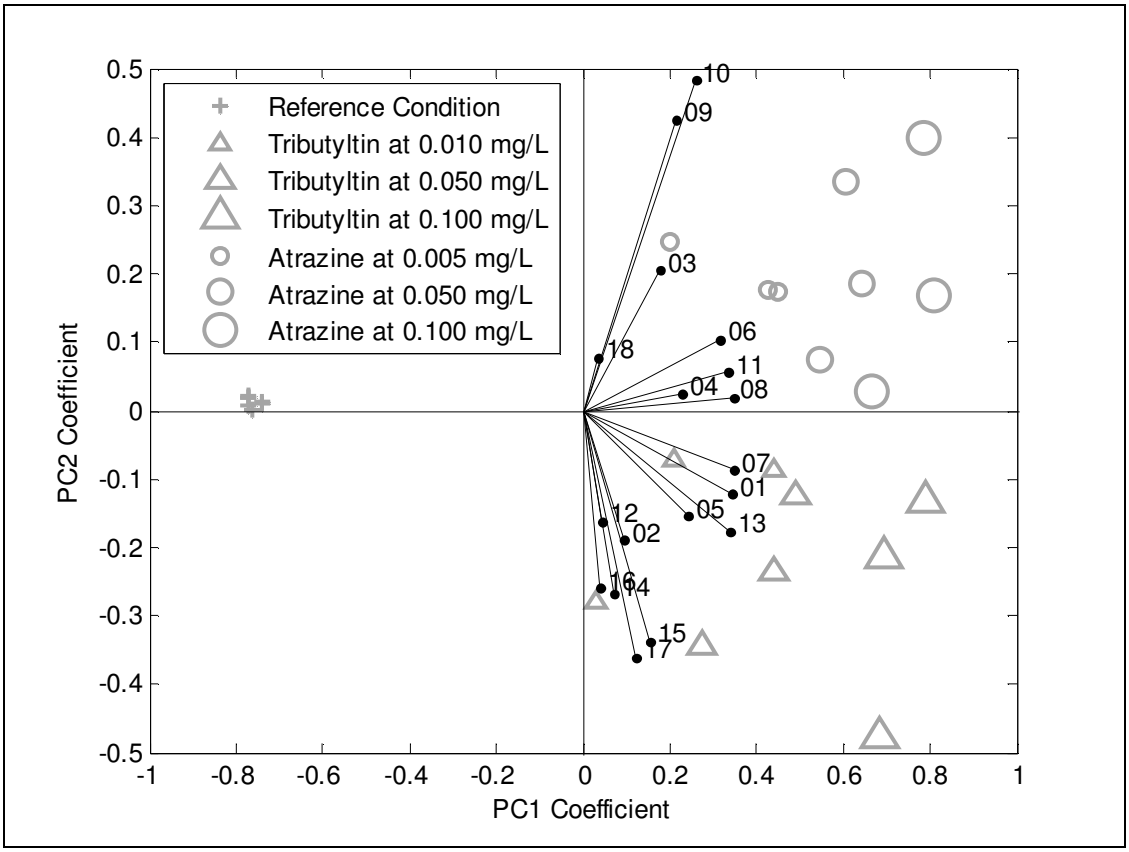
1) Length of the unit vectors

The length of each unit vector in Figure 4.7 describes the variance for each endpoint. In the present study, each endpoint was measured in reference condition as well as in solutions containing atrazine or tributyltin, and the results from all conditions were tabulated in a single column of observations. A large variance in an endpoint therefore indicated a more pronounced departure of that behavioural parameter from the reference condition when exposed to one of the two contaminants. Since sensitivity is often correlated to the intensity of responses, any endpoint exhibiting a long unit vector could be inferred as being sensitive to the test contaminants. From Figure 4.7, unit vectors 09 and 10, which correspond to the percentages of *Hyalella azteca* burrowing and grouping, respectively, were the longest unit vectors. These endpoints were therefore the most susceptible parameters to the contaminants.

Conversely, endpoints with short unit vectors, such as the change in the effective photosynthetic yield of *Pseudokirchneriella subcapitata* (vector 18) and the percentages of *Hyalella azteca* changing body orientation (vector 12), showed relatively small deviations from the reference condition in the presence of the two contaminants. For endpoints with low sensitivities, a highly precise automated monitoring system must be implemented in order to adequately capture and minimize incorrect readings, or these endpoints could be eliminated altogether from the library of responses since their contribution to the overall model was relatively low.

2) *Angle between two unit vectors*

The angle between two unit vectors in a biplot is a measure of their correlation. For example, the percentages of *Hyalella azteca* burrowing and changing grouping behaviour, or unit vectors 09 and 10, resulted in two unit vectors which were almost identical in magnitude and direction. This finding indicates that the two endpoints were very strongly related and behaved very similarly when either atrazine or tributyltin was added to the test solution. Some other endpoints also exhibited very strong correlations such as vectors 15 and 17, or the percentages of *Lumbriculus variegatus* shortening their body lengths and moving within groups, respectively. The pairs of strongly-correlated endpoints could either be maintained to provide system redundancy, or one of the endpoints could be eliminated to streamline the biomonitoring system. Having redundancy is advantageous as the system becomes more robust, but if the cost to measure more endpoints outweighs the benefits, then some endpoints can be eliminated.



LEGEND FOR UNIT VECTORS:

Daphnia magna

01 Changing swimming height
02 Spinning
03 Changing body orientation
04 Immobilized
05 Using secondary antennae
06 Changing swimming style

Hyalella azteca

07 Changing swimming height
08 Immobilized
09 Burrowing
10 Grouping
11 Shortening body length
12 Changing body orientation

Lumbriculus variegatus

13 Displaying abnormal swimming
14 Immobilized
15 Shortening body length
16 Changing body orientation
17 Moving within groups

Pseudokirchneriella subcapitata

18 Changing effective photosynthetic yield

Figure 4.7. Biplot of the unit vectors formed by the coefficients in Table 4.3 and data points scaled by a factor of 1/200 from Table 4.1.

In addition, vectors which are directly superimposed on top of one another indicate multicollinearity, and these vectors describe the same information. If multicollinearity occurs in a future experiment, only one of the endpoints should be recommended for the library of responses. As discussed in Section 2.3.2, Marshall (2009) recommended measuring both the percentages of *H. azteca* crawling and immobilized, but only the former endpoint was included in the present study because the two endpoints were exact complements of each other. Including both endpoints would have caused multicollinearity and resulted in two superimposed and identical unit vectors. Similar to having strongly-correlated endpoints, multicollinear endpoints can be advantageous as the endpoints create a more robust system. Most biomonitoring systems, however, are only capable of measuring a limited number of endpoints, and including extra endpoints may necessitate some instrumental upgrades. In this case, measuring the percentages of *H. azteca* crawling may require a more detailed camera and image analysis software, and thus a measurement of immobilized *H. azteca* was preferred.

On the contrary, unit vectors which are almost perpendicular to each other signify a weak relationship between the two endpoints, and the vectors could be assumed to be independent of each other. An example of such a situation is the relationship between the percentages of *Daphnia magna* changing swimming height and body orientation, or vectors 01 and 03, respectively. These two vectors were weakly correlated as the swimming height of the *D. magna* was more susceptible to a contamination by tributyltin, while the body orientation of the invertebrate changed more rapidly in the presence of

atrazine. Having several pairs of perpendicular unit vectors indicated some contaminant-specific sensitivity of different endpoints to the test contaminants.

4.3.3 Results of Cluster Analysis

By visual inspection, the data points in Figure 4.5 appeared to separate into three distinct clusters on the basis of treatment. As previously elaborated, however, an unbiased evaluation must be performed to assess whether the data points were actually separable into a number of well-defined clusters. To perform this evaluation, K-means clustering was applied to objectively classify the data points into three unspecified clusters, and the number of correctly assigned data points would be a measure of cluster quality. The algorithms for K-means clustering were conducted using MATLAB, and the resulting groups of data points are shown in Figure 4.8.

The application of K-means clustering on the data points in Figure 4.5 resulted in three groups of data points which were analogous to the conditions listed in Table 3.2 in Section 3.3. Cluster 1 matched the reference condition, while clusters 2 and 3 corresponded to the tributyltin and atrazine treatments, respectively. K-means clustering also assigned data points from similar treatments into the same group resulting in an accuracy of 100% as no data point was incorrectly classified into the wrong cluster of data points. This finding signified a strong evidence of contaminant-specific patterns exhibited by the organisms to the test contaminants in the present study, and these organisms could potentially be employed to distinguish the two contaminants by monitoring different changes in their endpoints.

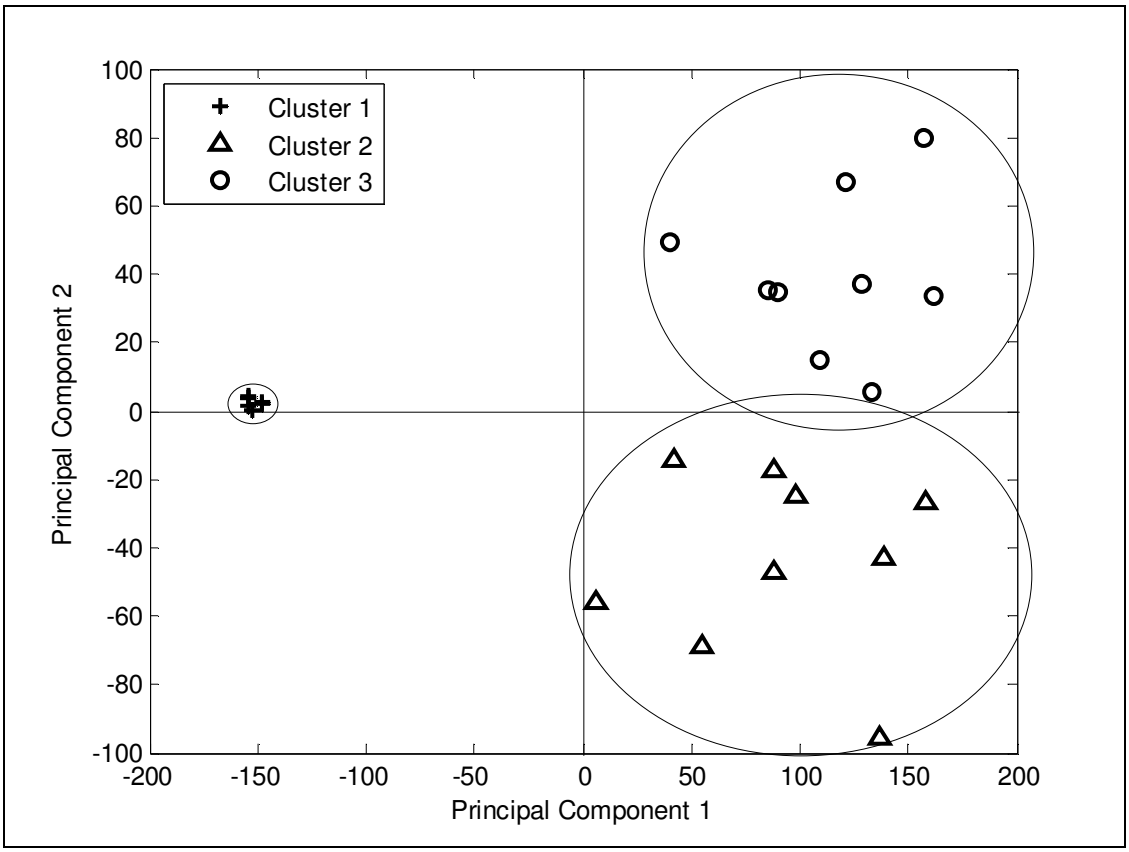
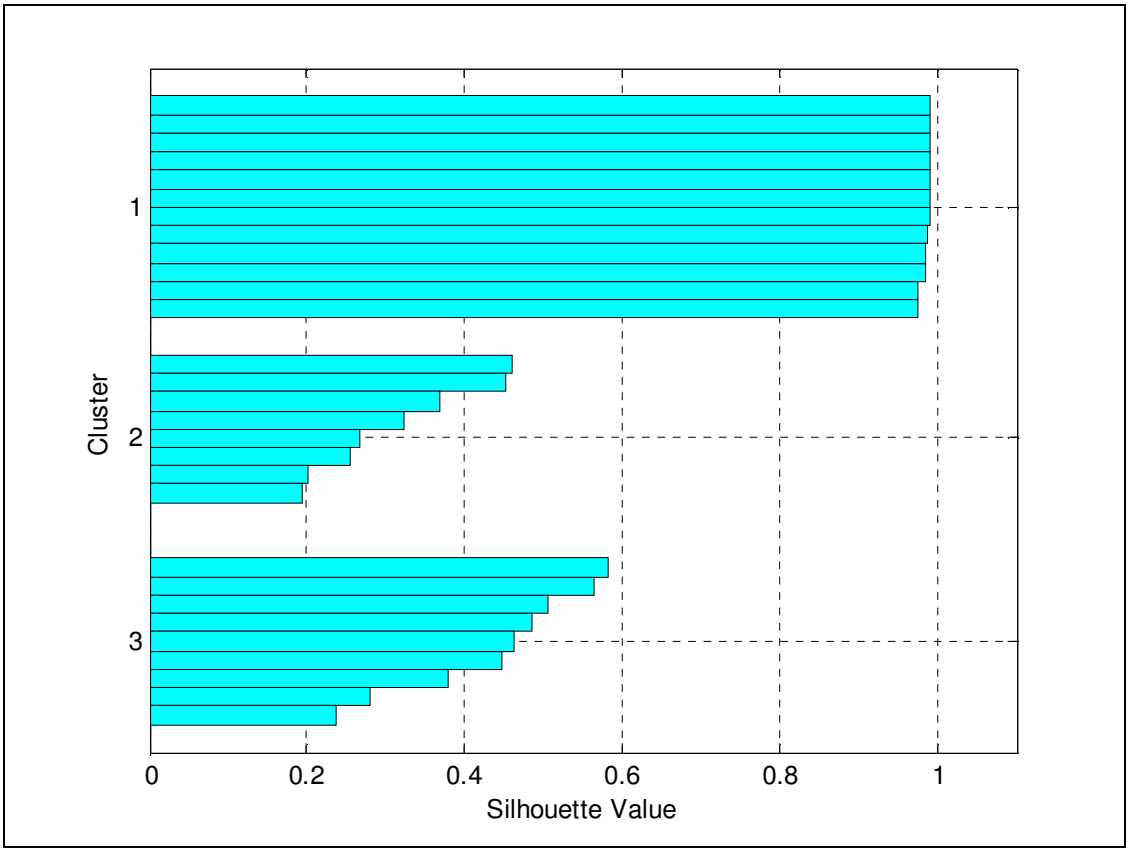


Figure 4.8. Classification of data points in Figure 4.5 using K-means clustering.

The clusters in Figure 4.8 were further evaluated using the Average Silhouette Plot method as outlined in Section 3.4. By comparing the average distance of a data point to other points in the same cluster and the average distance of the same data point to other point in the nearby cluster, an assessment of the cluster quality could be obtained as governed by Equation (5). Figure 4.9 displays the Average Silhouette Plot constructed from data points in the present study. Raw data for the plot are included in Appendix F.

Data points for the reference condition formed a ball cluster in Figure 4.9, resulting in a high average silhouette coefficient of 0.98. This number indicated strong evidence of a cluster for the data points measured in reference condition. The clusters for the atrazine and tributyltin treatments in Figure 4.5 were more loosely-formed, and as a result, the average silhouette coefficient for the two clusters was lower at 0.43 and 0.28, respectively. Despite the lower values, these coefficients were still higher than 0.25, and thus there was still some evidence of a cluster structure (Kaufman and Rousseeuw, 1990).

One of the tributyltin treatments, specifically the second replicate at 0.100 mg/L, resulted in a silhouette coefficient of almost 0. This value denoted an ambiguity for that particular data point where the point could be correctly classified as a contamination by tributyltin or misclassified as a contamination by atrazine. Despite this ambiguity, however, several runs of K-means clustering were still able to correctly classify this data point, and so this data point was considered acceptable for the present study.



Cluster	Condition	Average Silhouette Coefficient
1	Reference Condition	0.98
2	Contamination by Tributyltin	0.28
3	Contamination by Atrazine	0.43

Note: A silhouette coefficient of greater than 0.7 indicates strong cluster quality while a coefficient less than 0.25 signifies the absence of a cluster.

Figure 4.9. Average Silhouette Plot describing the cluster quality for data points in Figure 4.8.

As elaborated in Section 2.4.2, data points collected from environmental samples are more likely to form loose clusters with a few data points lying in between several main groups (Shaw, 2003). Based on this assumption, encountering data points with low silhouette coefficients was reasonable for the present study. A factor which could have contributed to these low coefficient values was the limited number of trials in the present study where only 9 trials were conducted for each stressor condition. These small sets were more strongly influenced by any unforeseen and unaccounted factors which could have affected the central mean of each cluster. In addition, as higher doses of atrazine and tributyltin were added, the responses exhibited by the aquatic organisms started to become similar as the organisms were distressed in general. These similarities in stressed responses could also present some ambiguities, and more experimental trials involving different concentrations are highly recommended for future experiments in order to improve the separation between different clusters.

4.3.4 Results of Discriminant Analysis

Cluster analysis revealed some contaminant-specific patterns in the data set from the present study as the data points formed three relatively-defined groups corresponding to the three test conditions listed in Table 3.2. Based on this finding, the data set could be employed as a preliminary library of responses to distinguish contaminations by atrazine and tributyltin. As discussed in Section 3.5, two matrices of responses were created to simulate real *in-situ* readings, and each of the matrices contained 30 observations (12 for reference condition and 9 for each contaminant). Figure 4.10 and Table 4.4 illustrate the classification of the endpoint data measured at t = 2 h by the library of responses.

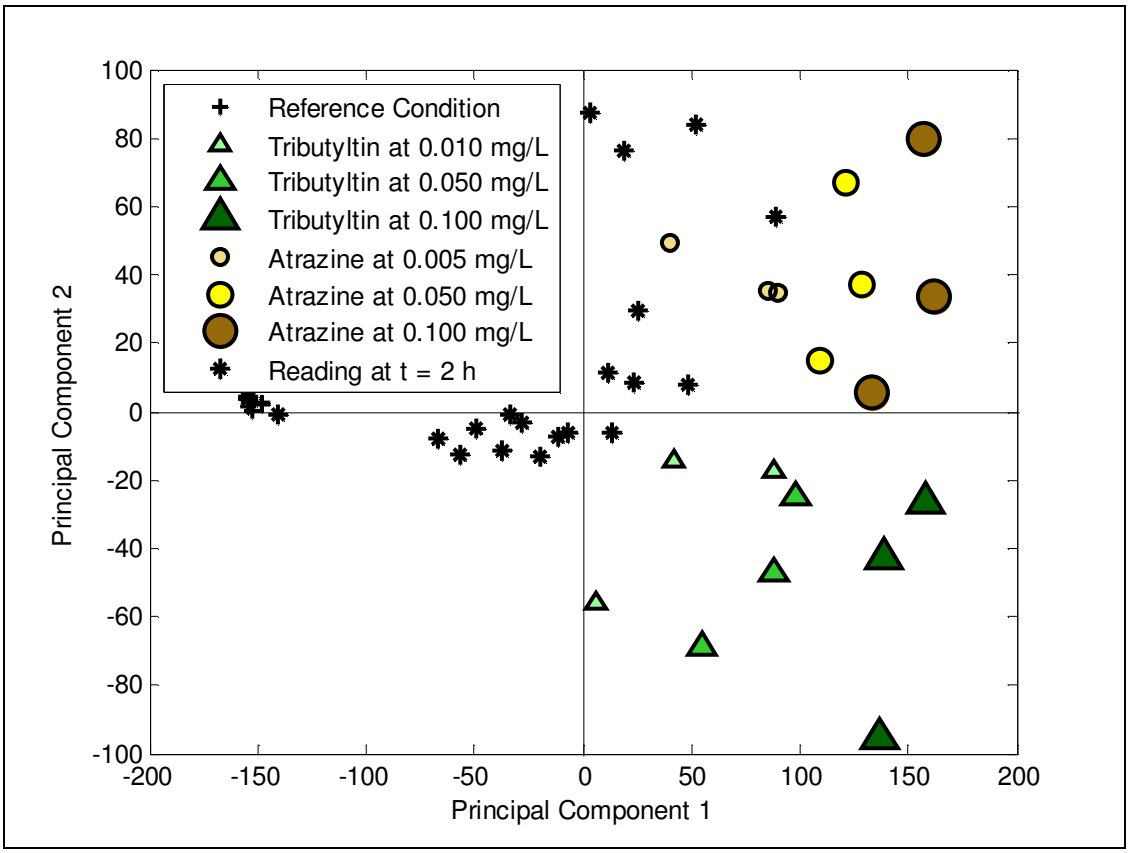


Figure 4.10. Plot of the endpoint measurements at t = 2 h against the library of responses.

Table 4.4. Classification of the endpoint measurements at t = 2 h using the current library of responses.

Observation	Actual Condition	Classified Condition
1	Reference Condition	Reference Condition
2	Reference Condition	Reference Condition
3	Reference Condition	Reference Condition
4	Reference Condition	Reference Condition
5	Reference Condition	Reference Condition
6	Reference Condition	Reference Condition
7	Reference Condition	Reference Condition
8	Reference Condition	Reference Condition
9	Reference Condition	Reference Condition
10	Reference Condition	Reference Condition
11	Reference Condition	Reference Condition
12	Reference Condition	Reference Condition
13	*Contamination by Tributyltin*	*Reference Condition*
14	*Contamination by Tributyltin*	*Reference Condition*
15	*Contamination by Tributyltin*	*Reference Condition*
16	*Contamination by Tributyltin*	*Reference Condition*
17	*Contamination by Tributyltin*	*Reference Condition*
18	Contamination by Tributyltin	Contamination by Tributyltin
19	Contamination by Tributyltin	Contamination by Tributyltin
20	*Contamination by Tributyltin*	*Contamination by Atrazine*
21	Contamination by Tributyltin	Contamination by Tributyltin
22	Contamination by Atrazine	Contamination by Atrazine
23	*Contamination by Atrazine*	*Reference Condition*
24	Contamination by Atrazine	Contamination by Atrazine
25	Contamination by Atrazine	Contamination by Atrazine
26	Contamination by Atrazine	Contamination by Atrazine
27	*Contamination by Atrazine*	*Contamination by Tributyltin*
28	Contamination by Atrazine	Contamination by Atrazine
29	Contamination by Atrazine	Contamination by Atrazine
30	Contamination by Atrazine	Contamination by Atrazine

Note:

- Italicized entries denote incorrectly classified data points.
- Total accuracy = 22/30 = 73.3%
- The following treatments were applied on the respective observations:

Observation	Treatment
1-12	Dechlorinated tap water with and without 0.1% v/v DMSO
13-15	Tributyltin at 0.010 mg/L
16-18	Tributyltin at 0.050 mg/L
19-21	Tributyltin at 0.100 mg/L
22-24	Atrazine at 0.005 mg/L
25-27	Atrazine at 0.050 mg/L
28-30	Atrazine at 0.100 mg/L

From Figure 4.10, many of the data points measured at t = 2 h were located near the origin which signified that these endpoints were similar in magnitudes to the overall mean of the library of responses. As discussed in previous sections, the library of responses contained one extreme, where the endpoints are close to 0% during the reference condition, and another extreme, where the endpoints may approach 100% in the presence of contaminants. In addition, the library of responses was created using measurements collected after 6 h of exposure to the contaminants. As a result, any readings collected between t = 0 and t = 6 h would result in data points falling in between the two extremities as the organisms were still reacting to the contaminants and the full extent of the stressor had not been reached.

Despite having been exposed for only 2 h to the contaminants, a number of observations could still be correctly classified by the library of responses as summarized in Table 4.4. All the reference conditions were correctly identified, while 33% of the tributyltin contaminations and 78% of the atrazine contaminations were properly classified, resulting in an overall accuracy of 73%. The majority of misclassified observations occurred at low concentrations of tributyltin (observations 13 – 17) and atrazine (observation 23) where these observations resulted in false negative readings. From the raw data in Appendix E, many of the endpoints at t = 2 h did not exhibit large departures yet from the reference condition, and this delayed response could be attributed to the low doses of the contaminants which required a longer incubation time to elicit a significant response from the organisms. After a longer exposure at 4 h, more of the observations could be correctly classified as shown in Figure 4.11 and Table 4.5.

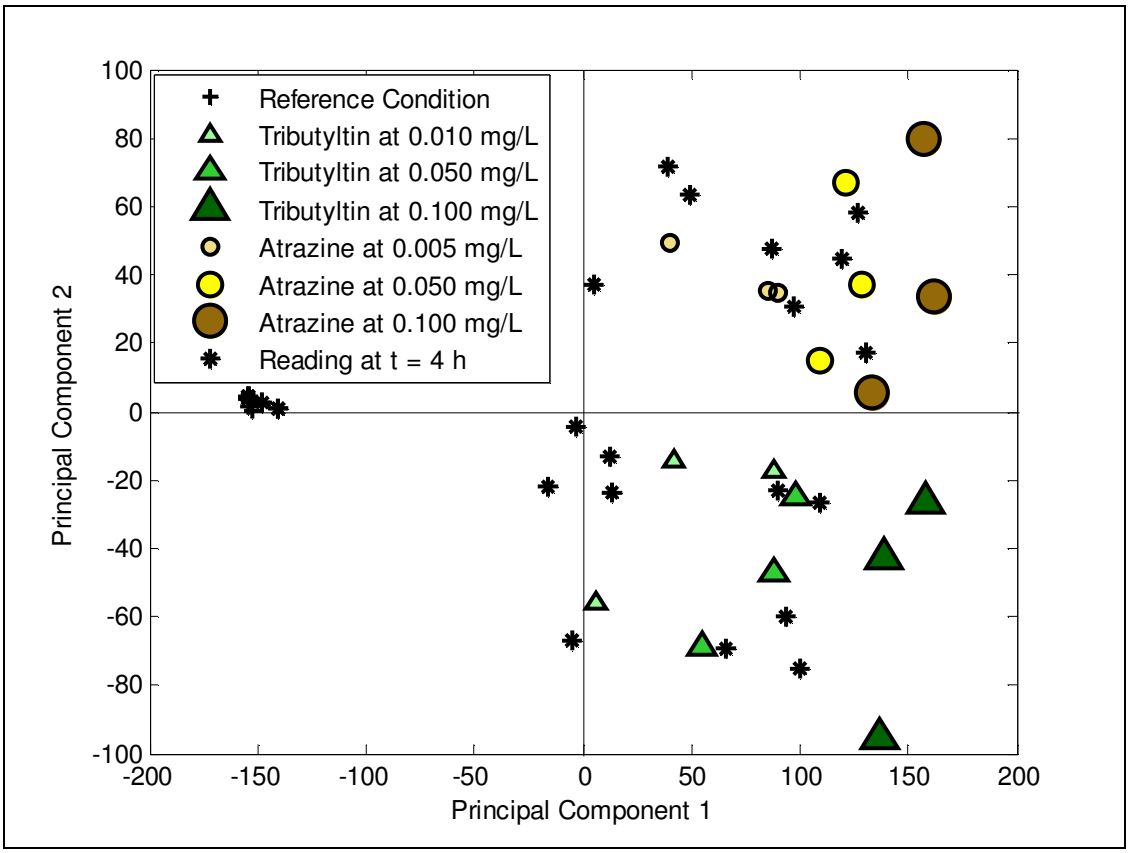


Figure 4.11. Plot of the endpoint measurements at t = 4 h against the library of responses.

Table 4.5. Classification of endpoint measurements at t = 4 h using the current library of responses.

Observation	Actual Condition	Classified Condition
1	Reference Condition	Reference Condition
2	Reference Condition	Reference Condition
3	Reference Condition	Reference Condition
4	Reference Condition	Reference Condition
5	Reference Condition	Reference Condition
6	Reference Condition	Reference Condition
7	Reference Condition	Reference Condition
8	Reference Condition	Reference Condition
9	Reference Condition	Reference Condition
10	Reference Condition	Reference Condition
11	Reference Condition	Reference Condition
12	Reference Condition	Reference Condition
13	Contamination by Tributyltin	Contamination by Tributyltin
14	Contamination by Tributyltin	Contamination by Tributyltin
15	Contamination by Tributyltin	Contamination by Tributyltin
16	Contamination by Tributyltin	Contamination by Tributyltin
17	Contamination by Tributyltin	Contamination by Tributyltin
18	Contamination by Tributyltin	Contamination by Tributyltin
19	Contamination by Tributyltin	Contamination by Tributyltin
20	Contamination by Tributyltin	Contamination by Tributyltin
21	Contamination by Tributyltin	Contamination by Tributyltin
22	Contamination by Atrazine	Contamination by Atrazine
23	Contamination by Atrazine	Contamination by Atrazine
24	Contamination by Atrazine	Contamination by Atrazine
25	Contamination by Atrazine	Contamination by Atrazine
26	Contamination by Atrazine	Contamination by Atrazine
27	Contamination by Atrazine	Contamination by Atrazine
28	Contamination by Atrazine	Contamination by Atrazine
29	Contamination by Atrazine	Contamination by Atrazine
30	*Contamination by Atrazine*	*Contamination by Tributyltin*

Note:

- Italicized entries denote incorrectly classified data points.
- Total accuracy = 29/30 = 96.7%
- The following treatments were applied on the respective observations:

Observation	Treatment
1-12	Dechlorinated tap water with and without 0.1% v/v DMSO
13-15	Tributyltin at 0.010 mg/L
16-18	Tributyltin at 0.050 mg/L
19-21	Tributyltin at 0.100 mg/L
22-24	Atrazine at 0.005 mg/L
25-27	Atrazine at 0.050 mg/L
28-30	Atrazine at 0.100 mg/L

From Figure 4.11, many of the data points measured at t = 4 h had shifted from their previous positions at t = 2 h and were situated closer to the measurements at t = 6 h (the actual library of responses). This shift in position could be attributed to the fact that the organisms had been incubated for a longer period in the contaminated solutions, and the harmful effects of the contaminants were starting to affect many of the behavioural and physiological endpoints of the organisms. A visual inspection of the raw data in Appendix E confirmed that after a longer exposure of 4 h, many of the endpoints displayed more significant departures from the reference condition even at lower concentrations of atrazine and tributyltin.

Since the endpoints were more susceptible to the two contaminants after 4 h of exposure, more observations could be correctly classified by the library of responses as summarized in Table 4.5. All the reference conditions and the tributyltin contaminations were correctly identified, while 88% of the atrazine contaminations were properly classified, resulting in an overall accuracy of 97%. Observation 30, which was conducted in a solution containing 0.100 mg/L atrazine, was incorrectly classified as a tributyltin contamination by the library of responses. While this classification was inaccurate, the library of responses still categorized observation 30 as a non-reference condition. During the *in-situ* deployment of the early-warning biomonitoring system, the consequences of such misclassification would not be critical because an alarm would still be raised to warn the water safety operator of a contamination in the water. Furthermore, the proposed EWBS is a flow-through, online monitoring system, so a continuous measurement of water parameters would still be performed. When additional

measurements are collected at t = 6 h, observation 30 would be correctly identified as the organisms would have been exposed even longer to the contaminants. The preliminary library of responses in the current study was therefore capable of accurately detecting either atrazine or tributyltin within 4 h of exposure, but a longer exposure of at least 6 h was recommended to obtain a correct identification of contaminants.

4.4 Potential Implementation of the Library of Responses

The present study provided a model for a library of responses which could be incorporated into the design of an automated early-warning biomonitoring system. A potential configuration for the EWBS is shown in Figure 4.12(a), where a central computer is connected to *n* biomonitors which continuously measure the endpoints of *m* aquatic organisms. To process the incoming endpoint data, a computer algorithm was developed as shown in Figure 4.12(b) where automated measurements can be conducted at specified time intervals. As previously elaborated, endpoint data are obtained from the biomonitors and compared against the library of responses to determine the current water condition. To minimize false-positives, the system can be configured to issue a first warning (*e.g.* a yellow screen) when a contamination is detected and a second warning (*e.g.* a red screen) when the same contaminant is registered again in the next set of reading. When a second warning is issued, an audible alarm can also be deployed to warn the water safety operators to take some corrective actions. By implementing these configurations, the proposed EWBS can provide a real-time, comprehensive, and automated monitoring of water quality.

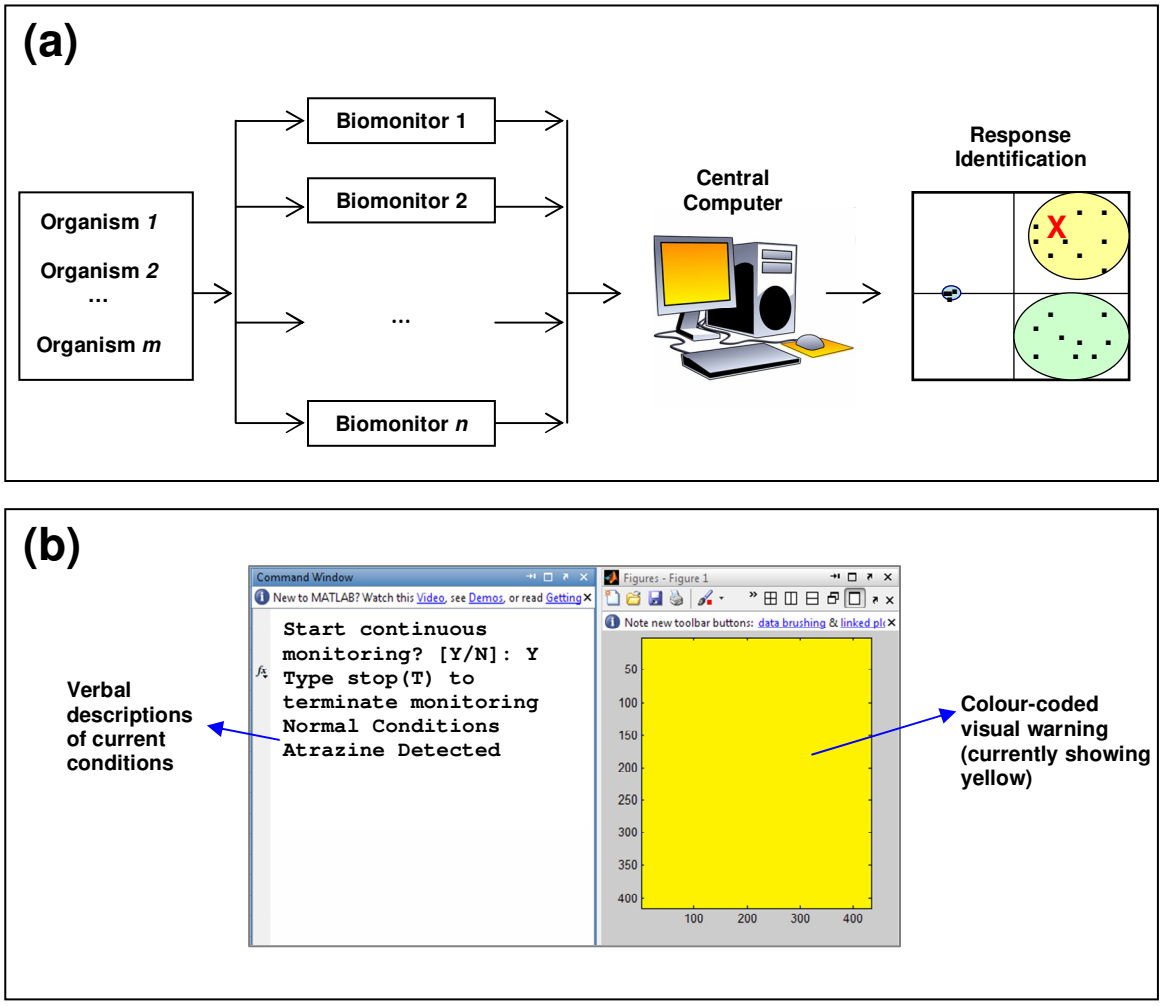


Figure 4.12. Illustration of the early-warning biomonitoring system in action where (a) depicts the overall system configuration, and (b) displays the MATLAB program to process the endpoint data.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The following conclusions can be drawn from this thesis:

1) In order to develop the library of responses, a number of endpoints were extracted from a number of previous studies. In a previous study of *Daphnia magna, Hyalella azteca,* and *Lumbriculus variegatus*, only 17 endpoints were considered appropriate for inclusion in the present study. Another published study involving a number of aquatic plants could not be incorporated at all into the present study due to the inconsistencies in the doses of atrazine and tributyltin used in that study.

2) An algal bioassay was conducted in the present study where the effective photosynthetic yield of *Pseudokirchneriella subcapitata* was measured over a period of 6 h in various test solutions. Effective photosynthetic yield was found to be a sensitive endpoint as the parameter was strongly affected by various doses of atrazine and less-strongly affected by tributyltin. A performance analysis of the Algae Online Monitor™ (AOM) 2800 also found the instrument to be satisfactory for quantifying various photosynthetic parameters and could potentially be deployed for continuous online measurements.

3) Data-simplification by PCA was successful in reducing the original matrix of responses, which contained 18 intercorrelated variables, into a transformed matrix

of 2 independent variables. The reduction process resulted in the creation of two principal components that contained ~85% of the original information. This amount of original information was considered adequate for the present study as three separate clusters of data sets could be visually distinguished. The first principal component separated data sets according to the presence of contaminants, while the second principal component differentiated data sets according to the type of the contaminant.

4) An objective assessment of the data points using some cluster analysis techniques revealed that the groups of points were quite well separated. An application of the K-means clustering algorithm was able to correctly classify all data points into 3 distinct clusters according to contaminant type. Results of the silhouette plot revealed strong evidence of a cluster for the data points measured at reference condition, while the clusters for the atrazine and tributyltin conditions could be further improved.

5) Analysis of the endpoints using a biplot revealed specific properties associated with each endpoint. A number of endpoints were more sensitive than others to the contaminants used in the present study, and these highly-sensitive endpoints could be identified by examining the length of each unit vector. Some endpoints were also useful for identifying any contamination by atrazine or tributyltin, and these discriminators could be identified by noting the direction of each unit vector. Unit vectors which pointed toward the atrazine cluster were better predictors of

atrazine, while unit vectors facing the tributyltin cluster were more susceptible to tributyltin.

6) Preliminary evaluation of the discriminating ability of the library of responses revealed that endpoint measurements at t = 2 h could be classified with an accuracy of 73%, while endpoint measurements at t = 4 h were classified with an accuracy of 97%. This finding demonstrated the potential capability of the library of responses to accurately detect contaminants and correctly identify a number of chemicals in a water sample.

5.2 Recommendations for Future Research

The following recommendations are suggested to guide future research on the development of an early-warning biomonitoring system:

1) Results from the algal bioassay were promising as the effects of tributyltin and atrazine on the photosynthetic parameters of the algae were detected after only 2 h of exposure. More algal bioassays should be conducted to measure the effects of other contaminants. The present study only measured the effective photosynthetic yield, but a number of other parameters could also be examined. Last, the algae online monitor should be tested in a continuous, flow-through setting to assess whether the instrument was indeed capable of performing real-time assessments of water quality.

2) The library of responses created in this study was still preliminary because it only comprised a relatively small number of endpoints measured for two contaminants at three different concentrations. This library of responses to atrazine and tributyltin could be expanded by including more measurements at different concentrations. Smaller increments and more trials should be conducted to improve the separation between clusters of data points. Other chemicals which are commonly found in aquatic systems could also be tested on the suite of aquatic organisms to increase the number of contaminants that could be identified by the early-warning biomonitoring system.

3) For future experiments, all bioassays must be performed at the same doses so that the endpoints measured from different studies could be compared and incorporated into the library of responses.

4) Some endpoint measurements in previous studies exhibited very large standard deviations which may have resulted from an experimental error or indicated that the behaviour of such endpoint was poorly understood. These endpoints should be carefully re-examined, and some bioassays could be repeated to verify whether the endpoints were suitable for monitoring contaminants.

5) Data normalization was not performed to pre-condition the matrix of responses in the current study because such procedure resulted in less defined boundaries between the two treatments and could potentially require more than two principal

components. Normalization could be considered in future experiments where a large number of replicates are conducted, more than two principal components are employed, and more than two contaminants are tested.

6) A number of other tests could be employed to further assess cluster quality, and one example is the leave-one-out cross validation technique. This technique was not applied in the current study as several cluster validation techniques were already applied, but this method could be considered in future studies if required.

7) After creating the library of responses, blind experiments should be conducted to assess whether the library of responses was capable of resolving different situations. To conduct these blind experiments, one researcher should prepare several test solutions pertaining to either one of the established conditions, and a second researcher would then conduct a series of bioassays to try to identify the unknown test solution. By performing such experiment, the discriminating ability of the suite of biomonitoring organisms could thus be truly and objectively evaluated.

REFERENCES

Anderberg, M.R. (1973). *Cluster Analysis for Applications*. Academic Press, NY.

Anton, Howard. (2000). *Elementary Linear Algebra*. Ed. 8. John Wiley and Sons, Inc. NJ.

Ahmed, H. and Häder, D.-P. (2010). Rapid ecotoxicological bioassay of nickel and cadmium using motility and photosynthetic parameters of *Euglena gracilis*. Environmental and Experimental Botany 69: 68–75.

Akbal, F., Gürel, L., Bahadır, T., Güler, İ., Bakan, G. and Büyükgüngör, H. (2011). Multivariate Statistical Techniques for the Assessment of Surface Water Quality at the Mid-Black Sea Coast of Turkey. Water, Air, and Soil Pollution 216: 21–37.

Alzieu, C. (1998). Tributyltin: case study of a chronic contaminant in the coastal environment. Ocean and Coastal Management 40: 23–36.

Balan, S.A., Kausch, M.F., Stüeken, E.E., and von Appen, W.-J. (2006). The Effects of Microgravity on the Photosynthetic Yield of *Nannochloropsis Salina* and *Spirulina Platensis*. Accessed on June 29, 2011 from http://mit.edu/~wilken/www/Papers/Neptunas_Paper.pdf.

Barrera, M. (2011). Photochemical treatment of organic constituents and bacterial pathogens from synthetic slaughterhouse wastewater by combining vacuum-UV and UV-C . M.A.Sc. Thesis. Ryerson University. Toronto, ON.

bbe Moldaenke. (n.d.). Manual for the Daphnia Toximeter II. Accessed on June 27, 2011 from http://www.bbe-moldaenke.de/fileadmin/Documents/Brochures/DaphTox_II.pdf.

bbe Moldaenke. (n.d.). Manual for the Fish Toximeter. Accessed on June 27, 2011 from http://www.bbe-moldaenke.de/fileadmin/Documents/Brochures/FishToximeter.pdf.

Berden-Zrimec, M., Drinovec, L., Zrimec, A., and Tisler, T. (2007). Delayed fluorescence in algal growth inhibition tests. Central European Journal of Biology 2 (2): 169–181.

Bednarska, A., Gerhardt, A. and Laskowski, R. (2010). Locomotor activity and respiration rate of the ground beetle, *Pterostichus oblongopunctatus* (Coleoptera: Carabidae), exposed to elevated nickel concentration at different temperatures: novel application of Multispecies Freshwater Biomonitor. Ecotoxicology 19: 864–871.

Behera, S.K., Lalitendu, N., and Biswal, B. (2003). Senescing leaves possess potential for stress adaptation: the developing leaves acclimated to high light exhibit increased tolerance to osmotic stress during senescence. Journal of Plant Physiology 160: 125–131.

Borcherding, J. (2006). Ten years of practical experience with the Dreissena-Monitor, a biological early warning system for continuous water quality monitoring. Hydrobiologia 556: 417–426.

Borcherding, J. and Volpers, M. (1994). The Dreissena-monitor – 1st results on the application of this biological early warning system in the continuous monitoring of water quality. Water Science and Technology 29: 199–201.

Bunn, S.E. (1995). Biological monitoring of water quality in Australia: Workshop summary and future directions. Australian Journal of Ecology 20: 220–227.

Bund/Länder-Projektgruppe 'Wirkungstests Rhein'. (1994). Continuous Biotests for Water Monitoring of the River Rhine – Summary, Recommendations, Description of Tests Methods. Umweltbundesamt, Berlin 27 pp.

Cairns J. Jr, Dickson K.L., Sparks R.E., and Waller W.T. (1970). A preliminary report on rapid biological information systems for water pollution control. Journal of Water Pollution Control Federation 42: 685–703.

Calder, R.S.D. and Schmitt, K.A. (2010). Role of Detection Limits in Drinking Water Regulation. Environmental Science & Technology 44 (21): 8008–8014.

Clark, S. T. (2010). Design and validation of oligonucleotide primers suitable for waterborne bacterial pathogen detection via real-time qPCR. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Clark, S.T., Gilbride, K.A., Mehrvar, M., Laursen, A.E., Bostan, V., Pushchak, R., and McCarthy, L.H. (2011). Evaluation of low-copy genetic targets for waterborne bacterial pathogen detection via qPCR. Water Research 45 (11): 3378–3388

Cliff, N. (1988). The eigenvalues-greather-than-one rule and the reliability of components. Psychology Bulletin 103: 276–279.

Diehl, P., Gerke, T., Jeuken, A., Lowis, J., Steen, R., van Steenwijk, J., Stoks, P., and Willemsen, H.-G. (2006). Early Warning Strategies and Practices along the River Rhine. Handbook of Environmental Chemistry 5, Part L: 99–124.

Dillon, W. R., and Goldstein, M. (1984). *Multivariate Analysis Methods and Applications*. John Wiley and Sons, Inc, NY.

Donna, A., Betta, P., Robutti, F., Crosignani, P., Berrino, F. and Bellingeri, D. (1984). Ovarian mesothelial tumors and herbicides: a case-control study. Carcinogenesis 5: 941–942.

Dort, A. (2010). The application of an early-warning biomonitoring system (EWBS) in a Canadian Context. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Durrieu, C., C. Tran-Minh, J. M. Chovelon, L. Barthet, C. Chouteau, and C. Védrine (2006). Algal biosensors for aquatic ecosystems monitoring. European Physical Journal of Applied Physics 36: 205–209.

El Jay, A. (1996). Effects of organic solvents and solvent-atrazine interactions on two algae, *Chlorella vulgaris* and *Selenastrum capricornutum*. Archives of Environmental Contamination and Toxicology 31: 84–90

Environment Canada. (2007). Biological Test Method: Growth Inhibition Test Using a Freshwater Algae. Accessed on June 25, 2011 from http://www.ec.gc.ca/Publications/default.asp?lang=Enandxml=D0BE08B2-DB41-45D6-A9E7-21783CCF3FBE

Evans G.P., Briers M.G., and Rawson, D.M. (1986). Can Biosensors Help to Protect Drinking-Water. Biosensors 2 (5): 287–300.

Fai, P.B., Grant, A., and Reid, B. (2007). Chlorophyll a fluorescence as a biomarker for rapid toxicity assessment. Environmental Toxicology and Chemistry 26 (7): 1520–1531

Fargasová, A. (1998). Comparison of tributyltin compound effects on the alga *Scenedesmus quadricauda* and the benthic organisms *Tubifex tubifex* and *Chironomus plumosus*. Ecotoxicology and Environmental Safety 41(3): 222–230.

Fent, K. (1996). Ecotoxicology of Organotin Compounds. Critical Reviews in Toxicology 26: 3–117.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics 7 (2): 179–188

Fleet, V. (2010). Evaluation of the multispecies freshwater biomonitor to determine behavioural effects of tributyltin and atrazine on *Daphnia magna* and *Hyalella azteca*. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Gaylord Chemical Company. (2007). Dimethyl sulfoxide (DMSO) Health and Safety Information, Bulletin # 106. Accessed on June 28, 2011 from http://www.gaylordchemical.com/bulletins/bulletin106b/Bulletin106B.pdf.

Gerhardt, A., Ingram, M.K., Kang, I.J. and Ulitzur, S. (2006). In-situ on-line toxicity biomonitoring in water: recent developments. Environmental Toxicology and Chemistry 25: 2263–2271.

Gerhardt, A., Jassens de Bisthoven, L. and Penders, E. (2003). Quality control of drinking water from the River Rhine with the Multispecies Freshwater Biomonitor. Journal of Aquatic Ecosystem Health and Management Society 6(2): 159–166.

Gerhardt, A., Kienle, C., Allan, I.J., Greenwood, R., Guigues, N., Fouillac, A., Mills, G. and Gonzalez, C. (2007). Biomonitoring with *Gammarus pulex* at the Meuse (NL), Aller, (GER) and Rhine (F) rivers with the online Multispecies Freshwater Biomonitor. Journal of Environmental Monitoring 9: 979–985.

Gerhardt, A., Schmidt, S., and Hoss, S. (2002). Measurement of movement patterns of *Caenorhabditis elegans* (Nematoda) with the Mulitspecies Freshwater Biomonitor® (MFB) – a potential new method to study a behavioural toxicity parameter of nematodes in sediments. Environmental Pollution 120: 513–516.

Gingerich, P. D. (1987). Evolution and the fossil record: patterns, rates, and processes. Canadian Journal of Zoology 65 (5): 1053–1060.

Grayman, W, Deininger, R, Males, and Gullick, R. (2004). Source water early warning systems. In: Mays, Larry. (eds). *Water Supply Systems Security*, McGraw-Hill and Companies, NY.

Graymore, M., Stagnitti, F. and Allinson, G. (2001) Impacts of atrazine in aquatic ecosystems. Environment International 26: 483–495.

Green, U, Kremer, J. H., Zillmer, M. and Moldaenke, C. (2003). Detection of Chemical Threat Agents in Drinking Water by an Early Warning Real-Time Biomonitor. Environmental Toxicology 18: 368–374.

Gullick R.W., Grayman W.M., Deininger R.A., and Males R.M. (2003). Design of early-warning monitoring systems for source waters. Journal AWWA 95 (11): 58–72.

Gunkel, G. and Streit, B. (1980). Mechanisms of bioaccumulation of a herbicide (atrazine, s-triazine) in a freshwater mollusc (*Ancylus fluviatilis* Mull.) and a fish (*Coregonus fera* Jurine). Water Research 14 (11): 1573–1584.

Haap, T., Triebskorn, R., and Köhler, H.-R. (2007). Acute effects of diclofenac and DMSO to *Daphnia magna*: Immobilization and hsp70-induction. Chemosphere 73(3): 353–359.

Hasan, J, States, S, and Deininger, R. (2004). Safeguarding the security of public water supplies using early warning systems: A Brief Review. Journal of Contemporary Water Research and Education 129: 27–33.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernández, J.M., and Fernández, L.. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research 34: 807–816.

Health Canada. (1993). Atrazine. Accessed on June 28, 2011 from http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/atrazine/index-eng.php.

Hoar, Z., Blair, A., Holmes, F.F., Boysen, C., Robel, R.J., Hoover, R. and Fraumeni, J.F. (1988). A case-control study of non-Hodgkin's lymphoma and agricultural factors in eastern Nebraska. American Journal of Epidemiology 128: 901–910.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24: 417–441.

Huberty, C. J. (1994). *Applied Discriminant Analysis*. John Wiley and Sons, Inc. NY.

ILSI. (1999). Early Warning Monitoring to Detect Hazardous Events in Water Supplies. International Life Sciences Institute-Risk Science Institute. Accessed on June 28, 2011 from http://www.mosselmonitor.nl/Links/ilsi_org/EWM.pdf.

Jackson, D.A. (1993). Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches. Ecology 74: 2204–2214.

Jensen, H.F., Holmer, M., and Dahllöf, I. (2004). Effects of tributyltin (TBT) on the seagrass *Ruppia maritima*. Marine Pollution Bulletin 49 (7–8): 564–573.

Jeon, J., Kim, J.H., Lee, B.C., and Kim, S.D. (2008). Development of a new biomonitoring method to detect the abnormal activity of *Daphnia magna* using automated Grid Counter device. Science of the Total Environment, 389: 545–556.

Johnson, R.A. and Wichern, D. W. (2002). Applied Multivariate Statistical Analysis. 5th ed. Prentice Hall, Upper Saddle River, NJ.

Juneau, P. and Popovic, R. (1999). Evidence for the rapid phytotoxicity and environmental stress evaluation using the PAM fluorometric method: Importance and future application. Ecotoxicology 8: 449–455.

Kaufman, L. and Rousseeuw, P.J. (1990). *Finding Groups in Data: an introduction to Cluster Analysis*. John Wiley and Sons, Inc, NY.

Kieu, N.D., Michels, E. and De Meester, L. (2001). Phototactic Behavior of *Daphnia* and the continuous monitoring of water quality: interference of fish kairomones and food quality. Environmental Toxicology and Chemistry 20: 1098–1103.

Kirkpatrick, A.J., Gerhardt, A., Dick, J.T.A., McKenna M. and Berges, J.A. (2006). Use of the multispecies freshwater biomonitor to assess behavioural changes of *Corophium volutator* (Pallas, 17766) (Crustacea, Amphipoda) in response to toxicant exposure in sediment. Ecotoxicology and Environmental Safety 64: 298–303.

Kiss, I., Kovats, N. and Szalay, T. (2003). Evaluation of some alternative guidelines for risk assessment of various habitats. Toxicology Letters 140: 411–417.

Knie J. (1982). Der Dynamische Daphnientest. Decheniana, 26, 82–86.

Kohler, U., and Luniak, M. (2005). Data inspection using biplots. The Stata Journal 5 (2): 208–223.

Konstantinou, I.K. and Albanis, T.A. (2004). Worldwide occurrence and effects of antifouling paint booster biocides in the aquatic environment: a review. Environment International 30: 235–248.

Kramer K.J.M, and Foekema E.M. (2001). The ‘‘Mussel Monitor’’ as a biological early warning system. In: Butterworth F.M., Gunatilaka A., Gonsebatt M.E. (eds). *Biomonitors and Biomarkers as Indicators of Environmental Change*, Vol 2. Kluwer Academic/Plenum, New York, NY, USA, 59–89.

Labbaf, Z. (2010). Identifying the potential sources of contaminants to the Welland Canal, the major source of drinking water in the Niagara region. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Lebert, M. and Häder, D.P. (1996). How Euglena tells up from down. Nature 379: 590.

LeBlanc, G.A. and Surprenant, D.C. (1983). The acute and chronic toxicity of acetone, dimethyl formamide and triethylene glycol to *Daphnia magna* (Straus). Archives of Environmental Contamination and Toxicology 12: 305–310.

Ma, J. (2005). Differential sensitivity of three cyanobacterial and five green algal species to organotins and pyrethroids pesticides. Science of the Total Environment 341: 109–117.

Marshall, G. C. (2009). Assessing behavioural and physiological responses of three aquatic invertebrates to atrazine and tributyltin in a multi-species, early-warning biomonitoring technology. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Martins, J.C., Saker, M.L., Teles, L.F.O. and Vasconcelos, V.M. (2007). Oxygen consumption by *Daphnia magna* Straus as a marker of chemical stress in the aquatic environment. Environmental Toxicology and Chemistry 26: 1987–1991.

Miller, W.E., Peterson, S.A., Greene, J.C., and Callahan, C.A. (1985). Comparative toxicology of laboratory organisms for assessing hazardous waste sites. Journal of Environmental Quality, 14(4), 569–574.

Netto, I. (2010). Assessing the usefulness of the automated monitoring system ECOTOX and DaphniaTox in an integrated early-warning system for drinking water. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Oberdöster, E., Rittschof, D. and LeBlanc, G.A. (1998). Alteration of [^{14}C]-testosterone metabolism after chronic exposure of *Daphnia magna* to tributyltin. Archives of Environmental Contamination and Technology 34: 21–35.

Ontario Ministry of the Environment. (2010). Spill Action Centre: 2009 Summary Report. Accessed on June 28, 2011 from http://www.ene.gov.on.ca.

Osbild, D., Babut, M. and Vasseur, P. (1995). Etat de l'Art. Les biocapteurs appliqués au controle des eaux. Revue des Services de l'Eau 8: 505–538.

Pandard, P., Vasseur, P., and Rawson, D.M. (1993). Comparison of two types of sensors using eukaryotic algae to monitor pollution of aquatic systems. Water Research 27 (3): 427–431.

Pearce, C. (2009). Investigating the Sub-acute Responses of *Lemna minor*, *Pseudokirchneriella subcapitata*, *Euglena gracilis* and *Anodonta grandis* to Tributyltin-hydride and Atrazine in Freshwater. M.A.Sc. Thesis. Ryerson University. Toronto, ON.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2: 559–572.

Rao, V.R. (1952). *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, Inc, London.

Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R. (2008). *Statistical Data Analysis Explained*. John Wiley and Sons, Inc. Chichester, West Sussex, England.

Ren, Z.M., Li, Z.L., Zha, J.M., Rao, K.F., Ma, M., Wang, Z. and Fu, R.S. (2008). The Avoidance Responses of *Daphnia magna* to the Exposure of Organophosphorus Pesticides in an On-Line Biomonitoring System. Environmental Modelling and Assessment 14: 405–410.

Ren, Z., Li, Z., Ma, M., Wang, Z. and Fu, R. (2009). Behavioural responses of *Daphnia magna* to Stresses of Chemicals with Different Toxic Characteristics. Bulletin of Environmental Contamination and Toxicology 82: 310–316.

Roig, B., Allan, I.J., and Greenwood, R. (2003). A toolbox of existing and emerging methods for water monitoring under the WFD. Accessed on August 17, 2011 from http://www.swift-wfd.com/Local/swift/dir/doc/WP1_OM%20mai%202005%20version%20public.pdf.

Roig, B., Valat,C., Berho, C., Allan, I., Guigues, N., Mills, G.A., Ulitzer, N., and Greenwood, R. (2007). The use of field studies to establish the performance of a range of tools for monitoring water quality. Trends in Analytical Chemistry 26 (4): 274–282.

Schmidt, K., Steinberg, C.E.W., Staaks, G.B.O. and Pflugmacher, S. (2005). Influence of Xenobiotic Mixture (PCB and TBT) Compared to Single Substances on Swimming Behaviour or Reproduction of *Daphnia magna*. Acta Hydrochimistria and Hydrobiologia 33: 287–300.

Shaw, P. (2003). *Multivariate Statistics for the Environmental Sciences*. Hodder Arnold. London.

Shrestha, S and Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modelling and Software 22: 464–475.

Sneath, P.H.A. and Sokal, R.R. (1973). Numerical Taxonomy. Freeman, San Francisco, CA.

Solomon, K.R., Baker, D.B., Richards, R.P., Dixon, D.R., Klaine, S.J., LaPoint, T.W., Kendall, R.J., Weisskopf, C.P., Giddings, W., Giesy, J.P., Hall, L.W., and Williams, W.M. (1996). Ecological risk assessment of atrazine in North American surface waters. Environmental Toxicology and Chemistry 15: 31–74.

Strasser R. J., Srivastava A., and Tsimilli-Michael M. (2000) The fluorescence transient as a tool to characterize and screen photosynthetic samples. In: Mohanty, P., Yunus, M., and Pathre, U. (Eds). *Probing Photosynthesis: Mechanism, Regulation and Adaptation*, Taylor and Francis, London, pp. 443–480.

Stratton, G.W. (1985). The influence of solvent type on solvent-pesticide interactions in bioassays. Archives of Environmental Contamination and Toxicology 14: 651–658.

Streb, C., Richter, P., Ntefidou, M., Lebert, M. and Häder, D.P. (2002). ECOTOX – Biomonitoring Based on Real Time Movement Analysis of Unicellular Organisms. Proceedings of Life on Earth, 8th European Symposium of Life Sciences Research in Space, 23rd Annual International Gravitational Physiology Meeting, Karolinska Institute, Stockholm, Sweden. pp. 241–242.

Tahedl, H. and Häder, D.P. (1998). Fast Examination of Water Quality Using the Automatic Biotest ECOTOX Based on the Movement Behavior of a Freshwater Flagellate. Water Research, 33(2), 426–432.

Tahedl, H., Häder, D.-P. (1999). Fast examination of water quality using the automatic biotest ECOTOX based on the movement behavior of a freshwater flagellate. Water Research 33: 426–432.

Tahedl, H., Häder, D.-P., (2001). Automated biomonitoring using real time movement analysis of *Euglena gracilis*. Ecotoxicology and Environmental Safety 48: 161–169.

United States Army Center for Environmental Health Research (USACEHR). (2004). Aquatic Biomonitoring for Rapid, Continuous Toxicity Assessment. Environmental Sentinel Biomonitor (ESB) Environmental Program. Accessed on June 28, 2011 from http://www.potomacriver.org/potomacdwspp/Meetings/Aug25-2004/van_der_Shalie_%20biomonitoring.pdf.

United States Environmental Protection Agency. (1997). Toxicological Review. Tributyltin Oxide. In support of Summary Information on the Integrated Risk Information System (IRIS). Accessed on June 28, 2011 from http://www.epa.gov/iris/toxreviews/0349tr.pdf.

United States Environmental Protection Agency (US EPA). (2005). *Technologies and Techniques for Early Warning Systems to Monitor and Evaluate Drinking Water Quality: State-of-the-Art Review*. Office of Research and Development, National Homeland Security Research Centre.

United States Environmental Protection Agency (US EPA). (2011). Re-Evaluation of Human Health Effects of Atrazine: Review of Non-Cancer Effects, Drinking Water Monitoring Frequency, and Cancer Epidemiology. Accessed on June 28, 2011 from http://www.regulations.gov/#!docketDetail;dct=FR+PR+N+O+SR;rpp=10;po=0;D=EPA-HQ-OPP-2011-0399.

Van Der Schalie, W. H., Shedd, T. R., Knechtges, P. L. and Widder, M. W. (2001). Using Higher Organisms in Biological Early Warning Systems for Real-Time Toxicity Detection. Biosensors and Bioelectronics 16: 457–465.

Van Der Schalie, W.H., Shedd, T.R., Widder, M.W., and Brennan, L.M. (2004) Response characteristics of an aquatic biomonitor used for rapid toxicity detection. Journal of Applied Toxicology, 24(5), 387–394.

Van der Schalie, W.H., James,R.R., and Gargan II, T. P. (2006). Selection of a battery of rapid toxicity sensors for drinking water evaluation. Biosensors and Bioelectronics 22: 18–27.

White, B., Gibert, M., and Deininger, R. (2002). Biological Monitors. In: Awrf & Crs Proaqua and Hargesheimer, Erika E. (eds). Online Monitoring for Drinking Water Utilities. American Water Works Association, Denver, CO.

World Health Organization. (1996). WHO/FAO Data Sheets on Pesticides No. 82. Atrazine. Accessed on June 28, 2011 from http://www.inchem.org/documents/pds/pds/pest82_e.htm.

Zhang, J., Jiang, W., Dong, Z., Zhao, S., and Wei, F. (2006). Simultaneous determination of triazine herbicide residues in maize by gas chromatography. China Journal of Chromatography/Se Pu 24(6): 648–651.

APPENDICES

APPENDIX A: Raw Data for Algal Bioassay

A.1 Sample Calculations for Preparing Test Solutions

A number of test solutions were prepared for the algal bioassay, and these solutions contained an algal suspension and other chemical constituents depending on the specifications. These test solutions were prepared in 250-mL Erlenmeyer flasks, and the total volume of the test solutions was 50 mL. Stocks of tributyltin and atrazine were available, where the contaminants were dissolved in dimethyl sulfoxide at a concentration of 100 mg tributyltin or atrazine per litre dimethyl sulfoxide. A stock of pure dimethyl sulfoxide was also prepared for mixing. Test solutions were prepared in triplicates according to the following specifications:

1. *Reference condition*

   These test solutions contained only 50 mL of algal suspension. The concentration of the algae was approximately 1.25×10^6 cells/mL

2. *Reference condition with 0.1% v/v dimethyl sulfoxide*

   These test solutions were similar to the reference condition, except that dimethyl sulfoxide was added at a ratio of 1:1000. To achieve this specification, the test solutions were filled with 50 mL of algal suspension and 50 μL, or 0.05 mL, dimethyl sulfoxide.

$$\frac{0.05mL}{50mL} = 0.001 = 0.1\%$$

3. *Atrazine solutions at 0.005, 0.050, and 0.100 mg/L with the addition of 0.1% v/v dimethyl sulfoxide*

Atrazine test solutions were prepared at concentrations 0.005, 0.050 and 0.100 mg/L by adding the stock solutions into the algal suspension. In addition, the total dimethyl sulfoxide concentration must also be 0.1% v/v. To achieve these specifications for the 0.005 mg/L solutions, the following steps were taken:

i. Determine the total mass of atrazine.

$$\frac{0.005\ mg}{L} \times 50\ mL \times \frac{1\ L}{1000\ mL} = 2.5 \times 10^{-4}\ mg \text{ atrazine}$$

ii. Determine how much atrazine stock was required to obtain the amount of atrazine in step (i).

Atrazine stock concentration = 100 *mg/L*

$$2.5 \times 10^{-4}\ mg \times \frac{1\ L}{100\ mg} \times \frac{1000000\ \mu L}{1\ L} = 2.5\ \mu L \text{ atrazine stock}$$

iii. Determine how much dimethyl sulfoxide should be added to obtain a concentration of 0.1% v/v.

To obtain 0.1% v/v, a total of 50 μL DMSO must be added

Remainder = 50 μL – 2.5 μL = 47.5 μL

iv. Summarize

To prepare test solutions at 0.005 mg/L atrazine and 0.1% v/v dimethyl sulfoxide, the following amounts were required: 50 mL algal suspension + 47.5 μL dimethyl sulfoxide + 2.5 μL atrazine stock at 100 mg/L.

4. Tributyltin solutions at 0.010, 0.050, and 0.100 mg/L with the addition of 0.1% v/v dimethyl sulfoxide

   Tributyltin test solutions were prepared at concentrations 0.010, 0.050 and 0.100 mg/L by adding the stock solutions into the algal suspension. In addition, the total dimethyl sulfoxide concentration must also be 0.1% v/v. To achieve these specifications for the 0.010 mg/L solutions, the following steps were taken:

   i. Determine the total mass of tributyltin.

$$\frac{0.010\ mg}{L} \times 50\ mL \times \frac{1\ L}{1000\ mL} = 5 \times 10^{-4}\ mg \text{ tributyltin}$$

   ii. Determine how much tributyltin stock was required to obtain the amount of tributyltin in step (i).

Tributyltin stock concentration = 100 *mg/L*

$$5 \times 10^{-4}\ mg \times \frac{1\ L}{100\ mg} \times \frac{1000000\ \mu L}{1\ L} = 5\ \mu L \text{ tributyltin stock}$$

   iii. Determine how much dimethyl sulfoxide should be added to obtain a concentration of 0.1% v/v.

To obtain 0.1% v/v, a total of 50 μL DMSO must be added

Remainder = 50 μL – 5 μL = 45 μL

iv. Summarize

To prepare test solutions at 0.005 mg/L tributyltin and 0.1% v/v dimethyl sulfoxide, the following amounts were required: 50 mL algal suspension + 45 μL dimethyl sulfoxide + 5 μL tributyltin stock at 100 mg/L.

A.2 Sample Calculations and Raw Data for Algal Growth Curve

A haemocytometer was used to estimate the number of algal cells in given batch, and thus the overall concentration of the entire flask could be inferred by assuming the sample was well-mixed. To count the number of algal cells, the flask was swirled several times to ensure the algal cells were uniformly distributed, and a small aliquot of algal suspension was aseptically extracted using a sterile pipette. This aliquot was often diluted according to different ratio since the algal suspension could contain a very high concentration. A drop from this aliquot was then placed on the haemocytometer, and the cells were counted by using a microscope. A haemocytometer typically comprises the following grids:

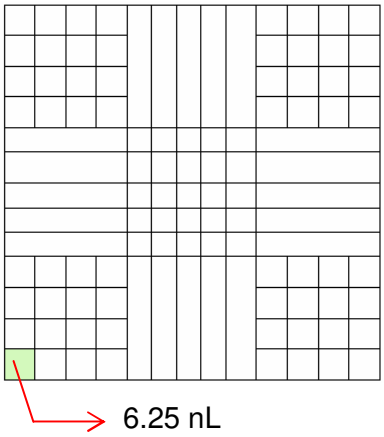


Figure A1: Grids commonly found in a haemocytometer.

The four corner squares are typically divided into 16 smaller squares, each containing approximately 6.25 nL of liquid (See Figure A1), and therefore an entire corner square contains 0.0001 mL of liquid. To determine the concentration of the algal suspension, three of the four corner squares were examined, and the number of algal cells was counted.

Day 0:

1	1	0	0
1	0	1	2
0	0	1	1
1	2	0	1

0	0	0	1
0	1	1	0
2	1	0	1
0	1	0	0

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

Total for corner square #1: 12 cells

Total for corner square #2: 8 cells

Total for corner square #3: 10 cells

Average count: $\frac{12+8+10}{3} = 10\,\text{cells}$

Dilution ratio: 4 times

Total number of cells = $\frac{10\,cells \times 4}{0.0001\,mL} = 400{,}000\,\text{cells/mL}$ or $4\times10^5\,\text{cells/mL}$

Day 1:

0	5	2	4
3	1	2	1
2	2	0	2
3	1	2	1

1	3	0	0
0	4	0	1
0	2	1	3
1	0	4	3

4	2	1	2
0	1	0	3
0	1	1	2
1	1	0	3

Total for corner square #1: 31 cells

Total for corner square #2: 23cells

Total for corner square #3: 22 cells

Average count: 25.33 cells

Dilution ratio: 4 times

Total number of cells = 1.01×10^6 cells/mL

Day 2:

2	4	4	4
3	3	3	10
7	6	5	4
5	2	7	3

4	2	8	10
6	4	8	2
6	2	5	6
6	9	2	4

5	6	7	6
3	5	9	9
4	9	7	2
11	7	7	1

Total for corner square #1: 72 cells

Total for corner square #2: 84cells

Total for corner square #3: 98 cells

Average count: 84.67 cells

Dilution ratio: 4 times

Total number of cells = 3.39×10^6 cells/mL

Day 4:

4	6	5	12
10	5	11	10
5	10	13	7
10	12	5	12

9	12	8	10
12	12	11	8
13	6	12	7
9	11	10	9

12	10	9	5
10	10	7	12
12	9	9	5
7	8	8	9

Total for corner square #1: 137 cells

Total for corner square #2: 159 cells

Total for corner square #3: 142 cells

Average count: 146 cells

Dilution ratio: 5 times

Total number of cells = 7.30×10^{6} cells/mL

Day 5:

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

8	10	4	4
8	3	4	7
1	10	3	3
5	5	6	3

4	2	13	3
5	3	6	7
4	4	5	3
7	5	3	3

Total for corner square #1: 77 cells

Total for corner square #2: 85 cells

Total for corner square #3: 77 cells

Average count: 79.33 cells

Dilution ratio: 10 times

Total number of cells = 7.93×10^{6} cells/mL

For Algal Bioassay:

On Day 4, some of the algal cultures were taken for use in the bioassay. The culture was diluted approximately 5 times using the enriched growth medium, and the following concentration was obtained:

1	1	0	5
4	2	4	2
4	0	3	1
3	3	0	5

1	2	6	3
2	2	5	0
0	2	1	1
0	1	0	2

4	0	2	0
1	2	3	2
7	0	3	0
1	2	0	1

Total for corner square #1: 38 cells

Total for corner square #2: 28 cells

Total for corner square #3: 28 cells

Average count: 31.33 cells

Dilution ratio: 4 times

Total number of cells = 1.25×10^6 cells/mL

A.3 Endpoint Data for Effective Photosynthetic Yield Experiments

All measurements were reported as percentages of yield

Treatment	0 h	2 h	4 h	6 h
Dechlorinated tap water				
Replicate 1	13	18	20	23
Replicate 2	14	20	22	24
Replicate 3	15	21	23	24
Dechlorinated tap water with 0.1% v/v DMSO				
Replicate 1	13	18	19	21
Replicate 2	14	18	19	21
Replicate 3	16	18	21	22
Tributyltin (0.010 mg/L)				
Replicate 1	15	17	19	19
Replicate 2	16	18	18	18
Replicate 3	15	18	19	19
Tributyltin (0.050 mg/L)				
Replicate 1	14	17	17	19
Replicate 2	14	18	19	20
Replicate 3	15	19	20	19
Tributyltin (0.100 mg/L)				
Replicate 1	14	19	18	19
Replicate 2	15	18	19	20
Replicate 3	16	16	18	18

Atrazine (0.005 mg/L)				
Replicate 1	15	15	13	15
Replicate 2	16	16	14	13
Replicate 3	16	17	16	11
Atrazine (0.050 mg/L)				
Replicate 1	13	11	11	10
Replicate 2	14	12	12	9
Replicate 3	13	13	13	8
Atrazine (0.100 mg/L)				
Replicate 1	15	9	9	5
Replicate 2	14	11	10	8
Replicate 3	15	12	11	9

Note: For implementation in the library of responses, the effective photosynthetic yield measurements at t = 6 h (last row) were used. In order to convert this endpoint into a stress indicator, an average value for the reference condition was calculated (22.5), and this value was subtracted from each replicate in the last column and multiplied by -1.

APPENDIX B: Raw Data for Invertebrates Bioassay

B.1 Daphnia magna at t = 6 h

All columns are expressed as the percentages of organisms exhibiting the stressed endpoints. Data were obtained from Marshall (2009)

Treatment	Changing swimming height	Spinning	Changing body orientation	Immobilized	Using secondary antennae	Changing swimming style
Dechlorinated tap water						
Replicate 1	0	0	0	0	0	0
Replicate 2	0	0	0	0	0	0
Replicate 3	0	0	0	0	0	0
Replicate 4	0	0	0	0	0	0
Replicate 5	0	0	0	0	0	0
Replicate 6	0	0	0	0	0	0
Dechlorinated tap water with 0.1% v/v DMSO						
Replicate 1	0	20	0	0	0	0
Replicate 2	0	20	0	0	0	0
Replicate 3	0	0	0	0	0	0
Replicate 4	0	0	0	0	0	0
Replicate 5	0	0	0	0	0	0
Replicate 6	0	0	0	0	0	0
Tributyltin (0.010 mg/L)						
Replicate 1	80	60	0	40	40	40
Replicate 2	80	40	60	60	60	80
Replicate 3	100	20	20	40	40	60
Tributyltin (0.050 mg/L)						
Replicate 1	100	0	0	40	80	60
Replicate 2	100	40	40	60	60	100
Replicate 3	100	0	20	60	80	80
Tributyltin (0.100 mg/L)						
Replicate 1	100	100	40	60	80	60
Replicate 2	100	40	20	60	80	100
Replicate 3	100	0	60	60	60	80
Atrazine (0.005 mg/L)						
Replicate 1	100	0	0	0	0	60
Replicate 2	100	40	40	0	0	60
Replicate 3	60	40	60	60	60	60

Atrazine (0.050 mg/L)						
Replicate 1	80	60	40	80	60	100
Replicate 2	100	20	60	60	60	100
Replicate 3	80	20	60	80	80	100
Atrazine (0.100 mg/L)						
Replicate 1	100	0	80	100	100	100
Replicate 2	100	0	80	80	80	100
Replicate 3	100	40	60	60	60	100

B.2 Hyalella azteca at t = 6 h

All columns are expressed as the percentages of organisms exhibiting the stressed endpoints. Data were obtained from Marshall (2009)

Treatment	Changing swimming height	Immobilized	Burrowing	Grouping	Shortening body length	Changing body orientation
Dechlorinated tap water						
Replicate 1	0	0	0	0	0	0
Replicate 2	0	0	0	0	0	0
Replicate 3	0	0	0	0	0	0
Replicate 4	0	0	0	0	0	0
Replicate 5	0	0	0	0	0	0
Replicate 6	20	0	0	0	0	0
Dechlorinated tap water with 0.1% v/v DMSO						
Replicate 1	0	0	0	0	0	0
Replicate 2	0	0	0	0	0	0
Replicate 3	0	0	0	0	0	0
Replicate 4	0	0	0	0	0	0
Replicate 5	0	0	0	0	0	0
Replicate 6	20	0	0	0	0	0
Tributyltin (0.010 mg/L)						
Replicate 1	80	60	0	0	80	20
Replicate 2	100	100	40	40	40	20
Replicate 3	100	60	40	0	60	20
Tributyltin (0.050 mg/L)						
Replicate 1	100	80	0	0	60	0
Replicate 2	100	80	0	60	80	0
Replicate 3	100	80	0	40	80	40

Tributyltin (0.100 mg/L)						
Replicate 1	100	100	40	40	100	20
Replicate 2	100	100	100	80	100	0
Replicate 3	100	100	60	60	100	40
Atrazine (0.005 mg/L)						
Replicate 1	80	80	80	80	80	0
Replicate 2	100	100	40	100	100	0
Replicate 3	100	100	60	80	100	0
Atrazine (0.050 mg/L)						
Replicate 1	100	100	100	100	100	0
Replicate 2	100	100	60	80	100	0
Replicate 3	80	80	40	60	100	20
Atrazine (0.100 mg/L)						
Replicate 1	100	100	100	100	100	0
Replicate 2	100	100	80	100	100	0
Replicate 3	100	100	60	80	80	40

B.3 Lumbriculus variegatus at t = 6 h

All columns are expressed as the percentages of organisms exhibiting the stressed endpoints. Data were obtained from Marshall (2009)

Treatment	Displaying abnormal behaviour	Immobilized	Shortening body length	Changing body orientation	Moving within groups
Dechlorinated tap water					
Replicate 1	0	0	0	0	0
Replicate 2	0	0	0	0	0
Replicate 3	0	0	0	0	0
Replicate 4	0	0	0	0	0
Replicate 5	0	0	0	0	0
Replicate 6	0	0	0	0	0
Dechlorinated tap water with 0.1% v/v DMSO					
Replicate 1	0	0	0	0	0
Replicate 2	0	0	0	0	0
Replicate 3	0	0	0	0	0
Replicate 4	0	0	0	10	0
Replicate 5	0	0	0	0	0

Replicate 6	0	0	0	0	0
Tributyltin (0.010 mg/L)					
Replicate 1	33	33	17	17	25
Replicate 2	100	0	0	0	80
Replicate 3	100	0	0	0	0
Tributyltin (0.050 mg/L)					
Replicate 1	100	40	40	40	0
Replicate 2	100	20	40	40	0
Replicate 3	100	25	25	25	33
Tributyltin (0.100 mg/L)					
Replicate 1	100	33	100	33	100
Replicate 2	100	60	75	25	83
Replicate 3	100	60	75	25	83
Atrazine (0.005 mg/L)					
Replicate 1	50	20	0	0	20
Replicate 2	100	0	0	0	30
Replicate 3	50	0	30	0	30
Atrazine (0.050 mg/L)					
Replicate 1	50	10	10	0	10
Replicate 2	100	10	20	0	10
Replicate 3	100	0	33	0	10
Atrazine (0.100 mg/L)					
Replicate 1	100	0	0	0	0
Replicate 2	100	10	100	0	10
Replicate 3	100	20	60	0	20

APPENDIX C: MATLAB Codes

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%% PRINCIPAL COMPONENT ANALYSIS %%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Step 1: Reading from Excel file
DATA_BASE_RAW = xlsread('input_newset2.xls');
MEAN_DATA_BASE = mean(DATA_BASE_RAW);

%Step 2: Mean-centering of raw data to create a modified matrix
for n = 1:size(DATA_BASE_RAW,1)
    DATA_BASE(n,:) = (DATA_BASE_RAW(n,:) - MEAN_DATA_BASE);
end

%Step 3: Calculating the covariance matrix
COV_INPUT = cov((DATA_BASE));

%Step 4: Calculating the eigenvectors and eigenvalues
[EIG_VECTORS, EIG_VALUES] = eig(COV_INPUT);
LATENT = diag(EIG_VALUES);

%Step 5: Two largest eigenvectors were selected, and the original data
set was transformed
FEAT_VECTORS = [-EIG_VECTORS(:,18) EIG_VECTORS(:,17)];
DATA_TRANSFORMED = (DATA_BASE*FEAT_VECTORS);



%Graphing
figure(1);
plot(DATA_TRANSFORMED(1:12,1),DATA_TRANSFORMED(1:12,2),'+','MarkerEdgeCo
lor','k','LineWidth',2)
hold on
plot(DATA_TRANSFORMED(13:15,1),DATA_TRANSFORMED(13:15,2),'^','MarkerEdge
Color','k','LineWidth',2,'MarkerSize',6)
plot(DATA_TRANSFORMED(16:18,1),DATA_TRANSFORMED(16:18,2),'^','MarkerEdge
Color','k','LineWidth',2,'MarkerSize',9)
plot(DATA_TRANSFORMED(19:21,1),DATA_TRANSFORMED(19:21,2),'^','MarkerEdge
Color','k','LineWidth',2,'MarkerSize',12)
plot(DATA_TRANSFORMED(22:24,1),DATA_TRANSFORMED(22:24,2),'o','MarkerEdge
Color','k','LineWidth',2,'MarkerSize',6)
plot(DATA_TRANSFORMED(25:27,1),DATA_TRANSFORMED(25:27,2),'o','MarkerEdge
Color','k','LineWidth',2,'MarkerSize',9)
plot(DATA_TRANSFORMED(28:30,1),DATA_TRANSFORMED(28:30,2),'o','MarkerEdge
Color','k','LineWidth',2,'MarkerSize',12)
legend('Reference Condition','Tributyltin at 0.010 mg/L','Tributyltin at
0.050 mg/L', 'Tributyltin at 0.100 mg/L', 'Atrazine at 0.005
mg/L','Atrazine at 0.050 mg/L','Atrazine at 0.100
mg/L','Location','NW');
grid on
AXIS1 = [1000;-1000];
AXIS2 = [0;0];
plot(AXIS1,AXIS2,'-k','LineWidth',1)
plot(AXIS2,AXIS1,'-k','LineWidth',1)
xlabel('Principal Component 1 (Variance explained = 78.33%)')
```

```
ylabel('Principal Component 2 (Variance explained = 6.46%)')
set(gca,'box','on')
ylim([-100 100]);
xlim([-200 200]);
hold off

%Constructing the biplot
figure(2);
LABEL =
['01';'02';'03';'04';'05';'06';'07';'08';'09';'10';'11';'12';'13';'14';'
15';'16';'17';'18'];
DATA_TRANSFORMED2 = DATA_TRANSFORMED./200;
plot(DATA_TRANSFORMED2(1:12,1),DATA_TRANSFORMED2(1:12,2),'+','MarkerEdge
Color',[0.65 0.65 0.65],'LineWidth',2)
hold on
plot(DATA_TRANSFORMED2(13:15,1),DATA_TRANSFORMED2(13:15,2),'^','MarkerEd
geColor',[0.65 0.65 0.65],'LineWidth',2,'MarkerSize',6)
plot(DATA_TRANSFORMED2(16:18,1),DATA_TRANSFORMED2(16:18,2),'^','MarkerEd
geColor',[0.65 0.65 0.65],'LineWidth',2,'MarkerSize',9)
plot(DATA_TRANSFORMED2(19:21,1),DATA_TRANSFORMED2(19:21,2),'^','MarkerEd
geColor',[0.65 0.65 0.65],'LineWidth',2,'MarkerSize',12)
plot(DATA_TRANSFORMED2(22:24,1),DATA_TRANSFORMED2(22:24,2),'o','MarkerEd
geColor',[0.65 0.65 0.65],'LineWidth',2,'MarkerSize',6)
plot(DATA_TRANSFORMED2(25:27,1),DATA_TRANSFORMED2(25:27,2),'o','MarkerEd
geColor',[0.65 0.65 0.65],'LineWidth',2,'MarkerSize',9)
plot(DATA_TRANSFORMED2(28:30,1),DATA_TRANSFORMED2(28:30,2),'o','MarkerEd
geColor',[0.65 0.65 0.65],'LineWidth',2,'MarkerSize',12)
h =
biplot(FEAT_VECTORS(:,1:2),'LineWidth',1,'Color','k','VarLabels',LABEL);
legend('Reference Condition','Tributyltin at 0.010 mg/L','Tributyltin at
0.050 mg/L', 'Tributyltin at 0.100 mg/L', 'Atrazine at 0.005
mg/L','Atrazine at 0.050 mg/L','Atrazine at 0.100
mg/L','Location','NW');
xlabel('PC1 Coefficient')
ylabel('PC2 Coefficient')
set(gca,'box','on')
ylim([-0.5 0.5]);
xlim([-1 1]);
plot(AXIS1,AXIS2,'-k','LineWidth',1)
plot(AXIS2,AXIS1,'-k','LineWidth',1)
grid off
hold off


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%% CLUSTER ANALYSIS %%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

[CA_SCORE C] = kmeans(DATA_TRANSFORMED,3);
SIZE_SCORE = size(CA_SCORE,1);

A = CA_SCORE(1,1);
B = CA_SCORE(13,1);

for Z=1:SIZE_SCORE
    if CA_SCORE(Z,1)==A
```

```
        CA_SCORE(Z,1)=1;
    elseif CA_SCORE(Z,1)==B
        CA_SCORE(Z,1)=2;
    else
        CA_SCORE(Z,1)=3;
    end
end

%Silhouette plot
figure(3);
[silh3,h] = silhouette(DATA_TRANSFORMED,CA_SCORE,'Euclidean');
hold on
set(gca,'box','on')
grid on
hold off


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%% DISCRIMINANT ANALYSIS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Step 1: Reading from Excel file
DATA_BASE_RAW = xlsread('input_newset2.xls');
MEAN_DATA_BASE = mean(DATA_BASE_RAW);
%for reading at t = 2 h:
READING_RAW = xlsread('input_newset3.xls');
%for reading at t = 4 h:
READING_RAW = xlsread('input_newset4.xls');


%Step 2: Mean-centering of raw data to create a modified matrix
for n = 1:size(DATA_BASE_RAW,1)
    DATA_BASE(n,:) = (DATA_BASE_RAW(n,:) - MEAN_DATA_BASE);
    READING(n,:) = (READING_RAW(n,:) - MEAN_DATA_BASE);
end

%Step 3: Calculating the covariance matrix
COV_INPUT = cov((DATA_BASE));

%Step 4: Calculating the eigenvectors and eigenvalues
[EIG_VECTORS, EIG_VALUES] = eig(COV_INPUT);
LATENT = diag(EIG_VALUES);

%Step 5: Two largest eigenvectors were selected, and the original data
set was transformed
FEAT_VECTORS = [EIG_VECTORS(:,18) -EIG_VECTORS(:,17)];
DATA_TRANSFORMED = (DATA_BASE*FEAT_VECTORS);
READING_TRANSFORMED = (READING*FEAT_VECTORS);

CLUSTER = [1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 2; 2; 2; 2; 2; 2; 2; 2;
2; 3; 3; 3; 3; 3; 3; 3; 3; 3];

DA_DATA = classify(READING_TRANSFORMED, DATA_TRANSFORMED, CLUSTER);
```

APPENDIX D: Intermediate Calculations in PCA

The following steps describe the intermediate calculations in PCA

1) *Data points were tabulated into a matrix of responses*

		Endpoint																	
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Ref	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.5
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.5
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.5
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.5
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.5
	6	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	-1.5
DMSO	**7**	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5
	8	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5
	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	1.5
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5
	12	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0.5
TBT 0.010 mg/L	**13**	80	60	0	40	40	40	80	60	0	0	80	20	33	33	17	17	25	3.5
	14	80	40	60	60	60	80	100	100	40	40	40	20	100	0	0	0	80	4.5
	15	100	20	20	40	40	60	100	60	40	0	60	20	100	0	0	0	0	3.5
TBT 0.050 mg/L	**16**	100	0	0	40	80	60	100	80	0	0	60	0	100	40	40	40	0	3.5
	17	100	40	40	60	60	100	100	80	0	60	80	0	100	20	40	40	0	2.5
	18	100	0	20	60	80	80	100	80	0	40	80	40	100	25	25	25	33	3.5
TBT 0.100 mg/L	**19**	100	100	40	60	80	60	100	100	40	40	100	20	100	33	100	33	100	3.5
	20	100	40	20	60	80	100	100	100	100	80	100	0	100	60	75	25	83	2.5
	21	100	0	60	60	60	80	100	100	60	60	100	40	100	60	75	25	83	4.5

ATZ	**22**	100	0	0	0	0	60	80	80	80	80	80	0	50	20	0	0	20	7.5
0.005	**23**	100	40	40	0	0	60	100	100	40	100	100	0	100	0	0	0	30	9.5
mg/L	**24**	60	40	60	60	60	60	100	100	60	80	100	0	50	0	30	0	30	11.5
ATZ	**25**	80	60	40	80	60	100	100	100	100	100	100	0	50	10	10	0	10	12.5
0.050	**26**	100	20	60	60	60	100	100	100	60	80	100	0	100	10	20	0	10	13.5
mg/L	**27**	80	20	60	80	80	100	80	80	40	60	100	20	100	0	33	0	10	14.5
ATZ	**28**	100	0	80	100	100	100	100	100	100	100	100	0	100	0	0	0	0	17.5
0.100	**29**	100	0	80	80	80	100	100	100	80	100	100	0	100	10	100	0	10	14.5
mg/L	**30**	100	40	60	60	60	100	100	100	60	80	80	40	100	20	60	0	20	13.5

LEGEND FOR ENDPOINTS 1 -18

Daphnia magna		*Hyalella azteca*		*Lumbriculus variegatus*		*Pseudokirchneriella subcapitata*	
1	Changing swimming height	**7**	Changing swimming height	**13**	Displaying abnormal swimming	**18**	Changing effective photosynthetic yield
2	Spinning	**8**	Immobilized	**14**	Immobilized		
3	Changing body orientation	**9**	Burrowing	**15**	Shortening body length		
4	Immobilized	**10**	Grouping	**16**	Changing body orientation		
5	Using secondary antennae	**11**	Shortening body length	**17**	Moving within groups		
6	Changing swimming style	**12**	Changing body orientation				

2) *Variables were pre-conditioned*

The mean across each variable was calculated, and each data point was subtracted by their respective mean:

	Endpoint																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Avg	56	18.7	24.7	33.3	36	48	59.3	54	30	36.7	52	7.33	52.8	11.4	20.8	7.17	18.1	4.9

The original matrix of responses then became a modified, pre-conditioned matrix:

	Endpoint																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-5.4
2	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-6.4
3	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-6.4
4	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-5.4
5	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-6.4
6	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-39.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-6.4
7	-56.0	1.3	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-3.4
8	-56.0	1.3	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-3.4
9	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-4.4
10	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	2.8	-18.1	-3.4
11	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-59.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-3.4
12	-56.0	-18.7	-24.7	-33.3	-36.0	-48.0	-39.3	-54.0	-30.0	-36.7	-52.0	-7.3	-52.8	-11.4	-20.8	-7.2	-18.1	-4.4
13	24.0	41.3	-24.7	6.7	4.0	-8.0	20.7	6.0	-30.0	-36.7	28.0	12.7	-19.8	21.6	-3.8	9.8	6.9	-1.4
14	24.0	21.3	35.3	26.7	24.0	32.0	40.7	46.0	10.0	3.3	-12.0	12.7	47.2	-11.4	-20.8	-7.2	61.9	-0.4
15	44.0	1.3	-4.7	6.7	4.0	12.0	40.7	6.0	10.0	-36.7	8.0	12.7	47.2	-11.4	-20.8	-7.2	-18.1	-1.4
16	44.0	-18.7	-24.7	6.7	44.0	12.0	40.7	26.0	-30.0	-36.7	8.0	-7.3	47.2	28.6	19.2	32.8	-18.1	-1.4
17	44.0	21.3	15.3	26.7	24.0	52.0	40.7	26.0	-30.0	23.3	28.0	-7.3	47.2	8.6	19.2	32.8	-18.1	-2.4
18	44.0	-18.7	-4.7	26.7	44.0	32.0	40.7	26.0	-30.0	3.3	28.0	32.7	47.2	13.6	4.2	17.8	14.9	-1.4
19	44.0	81.3	15.3	26.7	44.0	12.0	40.7	46.0	10.0	3.3	48.0	12.7	47.2	21.6	79.2	25.8	81.9	-1.4
20	44.0	21.3	-4.7	26.7	44.0	52.0	40.7	46.0	70.0	43.3	48.0	-7.3	47.2	48.6	54.2	17.8	64.9	-2.4
21	44.0	-18.7	35.3	26.7	24.0	32.0	40.7	46.0	30.0	23.3	48.0	32.7	47.2	48.6	54.2	17.8	64.9	-0.4
22	44.0	-18.7	-24.7	-33.3	-36.0	12.0	20.7	26.0	50.0	43.3	28.0	-7.3	-2.8	8.6	-20.8	-7.2	1.9	2.6
23	44.0	21.3	15.3	-33.3	-36.0	12.0	40.7	46.0	10.0	63.3	48.0	-7.3	47.2	-11.4	-20.8	-7.2	11.9	4.6
24	4.0	21.3	35.3	26.7	24.0	12.0	40.7	46.0	30.0	43.3	48.0	-7.3	-2.8	-11.4	9.2	-7.2	11.9	6.6
25	24.0	41.3	15.3	46.7	24.0	52.0	40.7	46.0	70.0	63.3	48.0	-7.3	-2.8	-1.4	-10.8	-7.2	-8.1	7.6
26	44.0	1.3	35.3	26.7	24.0	52.0	40.7	46.0	30.0	43.3	48.0	-7.3	47.2	-1.4	-0.8	-7.2	-8.1	8.6
27	24.0	1.3	35.3	46.7	44.0	52.0	20.7	26.0	10.0	23.3	48.0	12.7	47.2	-11.4	12.2	-7.2	-8.1	9.6
28	44.0	-18.7	55.3	66.7	64.0	52.0	40.7	46.0	70.0	63.3	48.0	-7.3	47.2	-11.4	-20.8	-7.2	-18.1	12.6
29	44.0	-18.7	55.3	46.7	44.0	52.0	40.7	46.0	50.0	63.3	48.0	-7.3	47.2	-1.4	79.2	-7.2	-8.1	9.6
30	44.0	21.3	35.3	26.7	24.0	52.0	40.7	46.0	30.0	43.3	28.0	32.7	47.2	8.6	39.2	-7.2	1.9	8.6

3) *A covariance matrix was calculated from the pre-conditioned matrix in Step 2*

The covariance matrix was determined using the MATLAB software

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	2245.5	532.4	929.7	1255.2	1390.3	1881.4	2176.6	2085.5	1158.6	1420.7	2008.3	292.4	2137.7	487.7	865.5	280.0	697.8	181.5
2	532.4	632.6	227.1	349.4	353.1	438.6	592.2	598.6	275.9	340.2	582.1	79.1	463.1	128.1	317.7	100.2	416.0	35.7
3	929.7	227.1	818.9	818.4	777.9	1009.7	996.3	1063.4	737.9	940.2	990.3	130.1	1032.9	69.3	471.1	4.0	320.0	137.5
4	1255.2	349.4	818.4	1140.2	1172.4	1310.3	1319.5	1310.3	841.4	956.3	1269.0	188.5	1294.9	235.6	634.0	137.7	400.2	144.6
5	1390.3	353.1	777.9	1172.4	1307.6	1357.2	1431.7	1395.9	786.2	896.6	1332.4	209.7	1459.7	333.9	762.8	230.3	492.3	133.9
6	1881.4	438.6	1009.7	1310.3	1357.2	1837.2	1881.4	1870.3	1186.2	1462.1	1790.3	229.0	1873.7	356.3	782.1	180.7	540.3	192.1
7	2176.6	592.2	996.3	1319.5	1431.7	1881.4	2220.2	2120.0	1179.3	1446.0	2008.3	280.9	2093.6	441.6	842.0	255.3	724.9	186.5
8	2085.5	598.6	1063.4	1310.3	1395.9	1870.3	2120.0	2128.3	1289.7	1579.3	2005.5	252.4	2017.2	423.0	872.8	210.7	785.0	200.2
9	1158.6	275.9	737.9	841.4	786.2	1186.2	1179.3	1289.7	1317.2	1275.9	1255.2	48.3	1052.1	203.1	488.6	-39.0	440.0	159.7
10	1420.7	340.2	940.2	956.3	896.6	1462.1	1446.0	1579.3	1275.9	1629.9	1558.6	66.7	1343.3	186.8	578.0	11.6	420.5	195.7
11	2008.3	582.1	990.3	1269.0	1332.4	1790.3	2008.3	2005.5	1255.2	1558.6	2057.9	226.2	1873.2	427.9	881.4	209.7	667.3	201.7
12	292.4	79.1	130.1	188.5	209.7	229.0	280.9	252.4	48.3	66.7	226.2	178.9	312.1	104.1	166.1	49.1	198.3	13.1
13	2137.7	463.1	1032.9	1294.9	1459.7	1873.7	2093.6	2017.2	1052.1	1343.3	1873.2	312.1	2243.4	427.4	909.7	276.4	724.8	175.4
14	487.7	128.1	69.3	235.6	333.9	356.3	441.6	423.0	203.1	186.8	427.9	104.1	427.4	318.2	389.0	180.4	338.8	3.2
15	865.5	317.7	471.1	634.0	762.8	782.1	842.0	872.8	488.6	578.0	881.4	166.1	909.7	389.0	980.4	230.5	555.1	58.6
16	280.0	100.2	4.0	137.7	230.3	180.7	255.3	210.7	-39.0	11.6	209.7	49.1	276.4	180.4	230.5	172.8	165.6	-12.2
17	697.8	416.0	320.0	400.2	492.3	540.3	724.9	785.0	440.0	420.5	667.3	198.3	724.8	338.8	555.1	165.6	863.0	15.4
18	181.5	35.7	137.5	144.6	133.9	192.1	186.5	200.2	159.7	195.7	201.7	13.1	175.4	3.2	58.6	-12.2	15.4	32.7

4) *The eigenvectors and eigenvalues for the covariance matrix in Step 3 were calculated*

Eigenvectors

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	-0.06	0.07	-0.22	0.65	-0.08	0.35	0.05	0.06	0.08	0.18	0.14	-0.05	-0.07	-0.16	0.34	-0.20	0.12	0.35
2	-0.04	0.02	-0.09	-0.13	0.02	0.16	0.03	0.05	0.20	-0.04	0.32	0.12	-0.30	0.71	0.18	0.34	0.19	0.09
3	-0.22	0.23	-0.10	-0.01	-0.04	0.22	0.55	-0.07	-0.15	0.16	-0.10	0.38	0.35	0.19	-0.32	0.01	-0.21	0.18
4	0.41	-0.51	-0.01	0.25	0.17	-0.03	0.16	-0.01	0.09	0.03	-0.13	-0.25	-0.06	0.28	-0.46	-0.13	-0.03	0.23
5	-0.30	0.29	-0.20	-0.09	-0.08	0.18	-0.43	0.13	-0.23	-0.18	-0.05	-0.31	-0.08	0.13	-0.46	-0.19	0.15	0.24
6	-0.14	0.26	0.15	0.07	-0.23	-0.48	0.12	-0.13	0.59	-0.27	0.07	-0.04	-0.03	-0.01	-0.09	-0.16	-0.11	0.32
7	0.05	0.14	-0.16	-0.18	0.44	-0.30	-0.01	-0.55	-0.28	0.15	-0.02	-0.08	-0.02	0.09	0.26	-0.15	0.09	0.35
8	0.01	-0.30	0.39	-0.12	-0.59	0.19	-0.13	-0.33	-0.23	-0.06	-0.12	0.02	0.07	0.07	0.17	0.05	-0.02	0.35
9	-0.06	0.06	0.20	-0.12	0.13	0.11	0.08	0.07	-0.02	0.23	0.43	-0.51	0.05	-0.18	-0.05	0.38	-0.43	0.21
10	0.06	-0.09	-0.23	0.02	0.27	0.20	-0.31	-0.05	0.14	-0.39	-0.23	0.31	-0.06	-0.14	0.05	0.26	-0.48	0.26
11	-0.07	0.02	0.12	-0.10	-0.02	-0.28	0.09	0.56	-0.16	0.29	-0.43	0.07	-0.36	-0.01	0.15	0.07	-0.06	0.33
12	-0.08	0.03	0.20	-0.19	0.19	0.26	-0.30	-0.07	0.54	0.49	-0.30	0.01	0.25	0.03	-0.02	-0.04	0.16	0.04
13	0.15	-0.21	0.00	-0.35	0.12	0.00	-0.02	0.40	0.00	-0.15	0.40	0.23	0.37	-0.12	0.08	-0.30	0.18	0.34
14	0.05	-0.09	-0.44	-0.41	-0.16	0.21	0.40	-0.09	0.21	-0.10	-0.19	-0.25	-0.16	-0.34	0.02	0.14	0.27	0.07
15	0.03	-0.04	0.06	0.09	-0.01	-0.13	-0.13	-0.18	-0.04	0.22	0.26	0.41	-0.27	-0.36	-0.43	0.32	0.34	0.16
16	0.01	0.20	0.59	0.05	0.39	0.32	0.26	-0.01	-0.05	-0.39	-0.11	-0.02	-0.19	-0.10	0.00	-0.04	0.26	0.04
17	0.07	0.09	-0.05	0.25	0.02	-0.22	-0.04	0.12	-0.07	-0.18	-0.20	-0.16	0.54	0.05	0.07	0.55	0.36	0.12
18	0.79	0.54	0.00	-0.07	-0.20	0.11	-0.07	0.04	-0.01	0.08	-0.02	0.03	-0.02	0.03	-0.03	-0.02	-0.08	0.03

The corresponding eigenvalue and variance for each eigenvector were also calculated. Based on the table below, eigenvectors 17 and 18 had the two largest eigenvalues, each describing 6.46% and 78.33% of variance, respectively.

Eigenvector	Eigenvalue	Variance
1	0.75	0.003%
2	1.66	0.01%
3	6.32	0.03%
4	18.53	0.08%
5	19.72	0.09%
6	25.50	0.12%
7	34.40	0.16%
8	67.54	0.31%
9	108.10	0.49%
10	131.47	0.59%
11	150.27	0.68%
12	303.70	1.37%
13	365.28	1.65%
14	480.71	2.17%
15	740.50	3.35%
16	910.13	4.11%
17	1429.68	6.46%
18	17330.81	78.33%

5) *The matrix of responses from Step 1 was transformed into a new data set of 2 variables and 30 observations*

Eigenvectors 18 and 17 were then combined to form an 18-by-2 matrix, and matrix multiplication was performed between the original matrix of responses

from Step 1 and this combined matrix of eigenvectors. A 30-by-2 transformed matrix was then created.

Obs.	PC1	PC2
1	-154.57	4.01
2	-154.60	3.93
3	-154.60	3.93
4	-154.57	4.01
5	-154.60	3.93
6	-147.65	2.21
7	-152.63	0.39
8	-152.63	0.39
9	-154.53	4.09
10	-154.13	1.58
11	-154.50	4.17
12	-147.58	2.36
13	6.36	-55.92
14	87.63	-17.44
15	42.27	-14.24
16	54.68	-68.68
17	98.61	-24.74
18	87.99	-46.86
19	136.75	-95.46
20	158.12	-26.60
21	138.47	-42.89
22	40.00	49.33
23	84.87	35.55
24	89.83	34.61
25	121.12	67.07
26	128.16	36.94
27	109.59	15.09
28	157.51	79.78
29	161.30	33.63
30	133.35	5.83

APPENDIX E: Raw Data for Discriminant Analysis

The following matrices were constructed to test the capability of the library of responses to resolve different contaminants.

- Matrix #1 was constructed using endpoint data at t = 2 h from the current algal bioassay and the previous study by Marshall (2009)

	Endpoint																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.33
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.33
4	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.33
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.33
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
13	40	40	0	20	0	0	100	80	0	0	40	0	0	0	0	0	0	2.67
14	60	0	40	40	40	60	100	60	0	0	0	0	0	0	0	0	0	1.67
15	80	0	40	0	40	60	100	0	0	0	20	0	0	0	0	0	0	1.67
16	80	40	0	20	0	80	80	40	0	0	40	20	0	0	0	0	0	2.67
17	100	0	0	0	20	60	80	60	20	0	40	0	0	0	10	0	0	1.67
18	100	60	20	40	40	60	80	20	20	0	40	20	0	0	0	0	0	0.67
19	100	40	0	40	40	80	100	60	40	0	60	40	0	0	0	0	0	0.67
20	80	80	20	40	60	80	100	100	60	0	100	0	0	0	0	0	0	1.67
21	100	20	0	40	40	60	100	40	20	0	60	0	0	0	0	0	0	3.67
22	80	0	0	0	0	0	100	100	80	100	100	0	0	0	0	0	0	4.67

23	80	20	0	0	20	80	60	0	0	0	20	0	0	0	0	0	0	3.67
24	0	40	0	20	0	0	100	100	100	100	100	0	0	0	0	0	0	2.67
25	60	40	20	0	40	60	100	100	100	100	100	0	0	0	0	0	0	8.67
26	80	20	20	40	60	80	80	60	40	0	80	0	0	0	0	0	0	7.67
27	40	60	20	60	20	60	80	60	0	0	80	0	20	0	0	0	0	6.67
28	100	60	40	60	0	80	100	100	60	100	100	0	40	0	10	20	0	10.67
29	80	0	40	0	40	80	80	60	20	60	80	0	40	10	0	10	0	8.67
30	100	40	0	40	60	80	80	60	0	60	60	0	20	0	0	0	0	7.67

- Matrix #2 was constructed using endpoint data at t = 4 h from the current algal bioassay and the previous study by Marshall (2009)

	Endpoint																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.33
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.33
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.67
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.33
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.33
7	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	2.67
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.67
9	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0.67
10	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	2.67
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.67
12	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0.67
13	100	40	0	20	20	20	60	60	20	0	60	0	40	40	0	20	100	2.67
14	80	40	20	0	40	40	100	60	0	60	40	20	20	20	0	20	0	3.67
15	80	20	0	40	40	60	100	60	0	0	40	20	0	20	0	0	0	2.67

16	100	20	0	40	80	100	80	0	0	0	80	0	50	0	0	0	0	4.67
17	80	40	0	20	40	60	100	100	0	0	100	0	0	0	0	0	0	2.67
18	100	40	0	40	80	100	100	60	0	0	80	40	100	0	0	100	0	1.67
19	100	40	0	0	0	60	100	100	40	80	100	60	100	100	100	100	20	3.67
20	80	40	60	0	40	80	100	100	40	60	100	0	66	67	67	67	0	2.67
21	100	0	0	40	60	60	100	100	40	60	60	40	100	80	80	80	0	3.67
22	100	40	20	0	0	100	80	80	80	80	80	0	0	0	0	0	0	8.67
23	80	40	0	0	0	100	100	60	40	60	40	0	10	0	0	0	0	7.67
24	60	60	0	40	40	40	100	100	80	100	100	0	0	10	0	0	0	5.67
25	100	40	40	80	80	100	80	100	80	100	80	0	100	0	0	0	0	10.67
26	80	40	20	20	20	100	80	80	60	100	100	0	100	0	25	0	0	9.67
27	80	60	20	40	40	100	100	80	80	60	100	0	100	10	0	0	10	8.67
28	100	60	40	40	40	100	100	100	80	100	80	0	100	20	0	30	0	12.67
29	100	0	60	40	40	100	100	100	40	100	100	40	100	0	100	0	0	11.67
30	80	40	0	40	40	100	100	80	40	80	100	40	100	20	100	0	30	10.67

Observation	**Treatment**	**Observation**	**Treatment**	**Observation**	**Treatment**
1-12	Dechlorinated tap water with and without 0.1% v/v DMSO	13-15	Tributyltin at 0.010 mg/L	22-24	Atrazine at 0.005 mg/L
		16-18	Tributyltin at 0.050 mg/L	25-27	Atrazine at 0.050 mg/L
		19-21	Tributyltin at 0.100 mg/L	28-30	Atrazine at 0.100 mg/L

LEGEND FOR ENDPOINTS 1 -18

Daphnia magna	*Hyalella azteca*	*Lumbriculus variegatus*	*Pseudokirchneriella subcapitata*
1 Changing swimming height	**7** Changing swimming height	**13** Displaying abnormal swimming	**18** Changing effective photosynthetic yield
2 Spinning	**8** Immobilized	**14** Immobilized	
3 Changing body orientation	**9** Burrowing	**15** Shortening body length	
4 Immobilized	**10** Grouping	**16** Changing body orientation	
5 Using secondary antennae	**11** Shortening body length	**17** Moving within groups	
6 Changing swimming style	**12** Changing body orientation		

APPENDIX F: Raw Data for Average Silhouette Plot

Treatment	Silhouette Coefficient
Dechlorinated tap water	
Replicate 1	0.99
Replicate 2	0.99
Replicate 3	0.99
Replicate 4	0.99
Replicate 5	0.99
Replicate 6	0.97
Dechlorinated tap water with 0.1% v/v DMSO	
Replicate 1	0.98
Replicate 2	0.98
Replicate 3	0.99
Replicate 4	0.99
Replicate 5	0.99
Replicate 6	0.97
Tributyltin (0.010 mg/L)	
Replicate 1	0.32
Replicate 2	0.20
Replicate 3	0.20
Tributyltin (0.050 mg/L)	
Replicate 1	0.46
Replicate 2	0.27
Replicate 3	0.45
Tributyltin (0.100 mg/L)	
Replicate 1	0.37
Replicate 2	0.00
Replicate 3	0.26
Atrazine (0.005 mg/L)	
Replicate 1	0.24
Replicate 2	0.46
Replicate 3	0.49
Atrazine (0.050 mg/L)	
Replicate 1	0.58
Replicate 2	0.56
Replicate 3	0.38
Atrazine (0.100 mg/L)	
Replicate 1	0.50
Replicate 2	0.45
Replicate 3	0.28