

기술노트

# 조류를 이용한 수계모니터링 시스템에서 뉴럴 네트워크에 의한 실시간 독성물질 판단

정종혁 · 정하규 · 권원태<sup>†</sup>

서울시립대학교 기계정보공학과

## On-line Identification of The Toxicological Substance in The Water System using Neural Network Technique

Jonghyuk Jung · Hakyu Jung · Wontae Kwon<sup>†</sup>

Department of Mechanical and Information Engineering, University of Seoul

(Received 22 October 2007, Accepted 13 December 2007)

### Abstract

Biological and chemical sensors are the two most frequently used sensors to monitor the water resource. Chemical sensor is very accurate to pick up the types and to measure the concentration of the chemical substance. Drawback is that it works for just one type of chemical substance. Therefore a lot of expensive monitoring system needs to be installed to determine the safeness of the water, which costs too much expense. Biological sensor, on the contrary, can judge the degree of pollution of the water with just one monitoring system. However, it is not easy to figure out the type of contaminant with a biological sensor. In this study, an endeavor is made to identify the toxicant in the water using the shape of the chlorophyll fluorescence induction curve (FIC) from a biological monitoring system. Wem-tox values are calculated from the amount of fluorescence of contaminated and reference water. Curve fitting is executed to find the representative curve of the raw data of Wem-tox values. Then the curves are digitalized at the same interval to train the neural network model. Taguchi method is used to optimize the neural network model parameters. The optimized model shows a good capacity to figure out the toxicant from FIC.

**keywords** : Algae, Neural network, On-line identification, Taguchi's method

### 1. 서론

최근 환경을 오염시키는 물질을 감시하고 관리하는 제품들이 많이 출시되고 있다. 수질을 오염을 판별하는 방법에는 여러 가지가 있다. 이 방법을 크게 화학적인 방법과 생물학적인 방법으로 나눌 수 있다(환경부, 2004; 황인영, 1997). 화학적인 방법은 수계에 포함되어 있는 물질을 화학적인 방법을 이용하여 정확하게 알아내는 방법이고, 생물학적인 방법은 생물을 이용하여 전체적으로 그 물이 오염이 되었는지 아닌지를 판단하는 방법이다. 일반적으로 화학적인 방법은 수계가 어떤 물질을 얼마나 포함하는지를 정확하게 파악할 수 있지만, 각 화학물질 하나마다 모니터링 시스템이 필요하기 때문에 설치비가 많이 드는 단점이 있다.

반면 생물학적인 방법은 물이 오염되어 있는지를 판단하는 방법으로 어떤 독성물질에 의해 오염되었는지 정확히 알 수 없는 단점이 있다. 오염수에 있는 독성물질을 생물학적으로 판단할 때 식물체를 이용하는 방법은 식물성 플랑크톤의 성장저해, 상추씨의 발아율 및 뿌리 성장 저해

실험방법(Greene et al., 1989)과 물벼룩, 물고기, 조류 등의 동물체를 이용하는 방법(Keddy et al., 1995) 등이 있다. 식물체를 이용한 방법의 장점은 결과의 반복성 및 객관성을 유지하는데 동물성 시험체보다는 유지 및 보관이 쉽다. 하지만 식물체를 이용하는 방법은 동물체를 이용한 방법보다 시간이 훨씬 많이 소요되는 단점이 있다(이지나와 황인영, 1999). 식물체나 동물체를 사용하여 독성물질 여부를 판단하는 생물학적인 방법이 공통적으로 가지고 있는 단점은 오염원이 위험수위 이상 포함 여부를 판단할 수는 있지만, 그 오염원이 무엇인지를 알기가 매우 어렵다는 것이다.

본 연구에서는 생물학적인 방법을 이용하여 오염여부를 확인함과 동시에 그 오염원이 어떤 물질인지를 확인하는 방법을 제안한다. 사용한 생물로는 동물성 시험체인 반달말을 선택하였다. 반달말은 각각의 독성 물질에 대해 시간에 변함에 따라 광합성에 의한 형광량이 다르다. 시간이 지남에 따라 변하는 광합성 형광량의 변화하는 패턴을 측정하여 오염물질의 종류를 파악하였다. 오염물질의 종류를 판단하는 모델로 신경 회로망(Rumelhart et al., 1987)을 이용하였고 이 모델의 최적화를 위하여 다꾸지 방법(Phadke, 1989)을 사용하였다. 완성된 모델을 이용하여 오염수에 포함된 독성물질의 종류를 실시간으로 평가하였다.

<sup>†</sup> To whom correspondence should be addressed.  
kwon@uos.ac.kr

## 2. 연구방법

### 2.1. 실험장치

이 연구에서는 반달말을 이용하여 수질을 감시하는 Wem-M((주)환경바이오)을 이용하였다. 반달말이 빛 에너지를 받으면 광합성을 하면서 형광을 발생시키는데, 독성물질을 투입했을 때 나타나는 형광량의 변화를 감지하여 수계의 오염여부를 확인하는 원리가 사용되었다. Fig. 1에 나타나듯이 측정수에서의 반달말의 형광량을, 표준수에서의 반달말의 형광량과 비교하는 방법으로 수질의 오염도를 측정하였다. 표준수에서 얻어지는 정상 상태의 형광량(reference)과 독성물질이 함유된 측정수에서의 형광량(sample)이 각각 측정되었다. 실험장치는 채수부, 구동부, 제어부로 구성되어 있는데, 채수부는 표준수와 측정수의 채수를 조절하고, 실험에 사용된 반달말을 교체하는 부분이다. 그리고 구동부는 반달말에 빛을 가해서 광합성량을 측정하는 부분이다. 광합성량을 측정하는 방법은 일정한 빛의 세기가 10초 동안 주어졌을 때 발생하는 평균 형광량(F)과 그 10초 중 임의의 시간에 일정 빛의 세기에 2~3배의 빛을 순간적으로 주었을 때 형광량(Fm)을 측정한다. 마지막으로 제어부는 구동부와 채수부가 작동되고 있는 상황과 반달말의 형광량을 그래프로 보여주는 부분이다. 하나의 독성 물질로 실험할 때 소요되는 시간은 30분 정도이다.

실험에서 사용된 형광량의 값은 Wem-tox값의 계산에 사용된다. Wem-tox값은 식 (1), (2), (3)을 이용하여 구해진다 (Juneau et al., 2003).

$$Y_R = \frac{Fm_R - F_R}{Fm_R} \quad (1)$$

$$Y_S = \frac{Fm_S - F_S}{Fm_S} \quad (2)$$

$$WEM-tox = \frac{Y_S - Y_R}{Y_S} \quad (3)$$

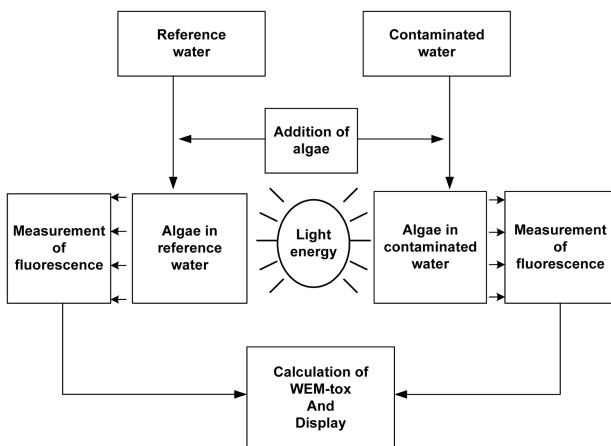


Fig. 1. The principle of the Wems-M device.

F<sub>MR</sub> : Maximum Fluorescence of Reference

F<sub>R</sub> : Fluorescence of Reference

F<sub>MS</sub> : Maximum Fluorescence of Sample

F<sub>S</sub> : Fluorescence of Sample

Y<sub>R</sub> : Quantum Yield of Reference

Y<sub>S</sub> : Quantum Yield of Sample

본 실험에서는 Table 1과 같이 7가지 오염물질이 다른 농도를 가진 16가지의 경우에 대하여 사용되었다.

Table 1. Toxicities and their quantity (ppm) used for the experiments

Cu (ppm)	Diazinon (ppm)	Cyanogen (ppm)	Atrazine (ppm)	Fe (ppm)	Waste water
1	1	1	0.01	10	Rubber
10	10	5	0.1	20	Pigment
100	100	10	1		

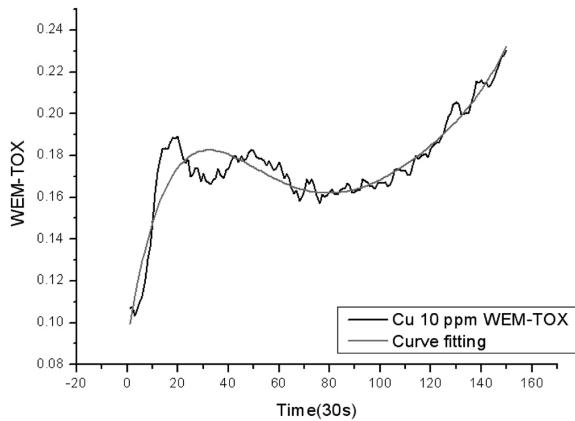
### 2.2. 독성 물질 확인을 위한 모델링

Table 1에 주어진 독성물질이 첨가된 측정수와 표준수의 형광량을 측정한 후 각각의 값을 식 (1), (2), (3)에 대입하면 Wem-tox 값을 구할 수 있다. 하지만 이 값들은 단위 시간에 대한 단속적인 값으로 나타나기 때문에 잡음(noise)에 너무 민감하여 원치 않는 변화를 나타낼 수 있다. 이 연구에서는 잡음에 대한 감도를 줄이기 위하여 구해진 단속적인 Wem-tox 값의 이동 평균값을 계산하였다. 또한 오차를 최소화하기 위하여 각각의 독성 물질에 대한 실험을 3회 이상 수행한 후, 각 독성 물질의 Wem-tox 평균값을 구하여 사용하였다. 최종적으로 각 독성 물질의 Wem-tox 평균값들이 단속적으로 나타나 있어 어려움이 있으므로 커브피팅을 적용하여 연속적인 그래프를 구하였다.

여러 개의 커브 중에서 실험값과 비교하여 결정계수(R<sup>2</sup>) 값은 raw data와 curve fitting한 결과값이 근사할수록 때 1에 근접하고, 카이스퀘어 검증값(Chi<sup>2</sup>/DoF)이 0에 가까워질 때 데이터간의 상관 관계가 없고 독립적인 것을 알 수 있다. 그러므로 결정계수(R<sup>2</sup>)가 1에 가깝고, 카이스퀘어 검증값(Chi<sup>2</sup>/DoF)이 0에 가까운 커브가 피팅에 사용될 커브로 선택되었다. Table 2에서 사용된 Polynomial, Biphasic, GaussAmp, Asym2sig, LogNormal 커브가 각각 식 (4), (5), (6), (7), (8)에 주어졌다. 구리 10 ppm을 포함하고 있는 수계의 Wem-tox의 평균 값의 변화와 polynomial로 커브피팅한 결과가 Fig. 2에 주어졌다. 커브피팅을 실시한 이유는 Wem-tox 값의 순간적인 변화가, 수계에 들어있는 독성물질을 판정하는데 미치는 영향을 최소화하기 위해서이다. 즉 순간적으로 급하게 변하는 Wem-tox 값을 따라가면서 판단하다가 전체 커브의 변화모양을 놓쳐 잘못 판단할 가능성을 최소화하기 위해서이다. 이렇게 피팅된 커브를 각 독성 물질에 대한 대표 커브로 하고 이를 바탕으로 독성물질의 종류를 판정할 수 있는 모델을 구성하였다. 모델구성을 위하여 신경회로망을 이용하였다. 이때 한 개의 모델이 16개 전부의 독성물질을 표현하도록 신경회로망을 구성할 수 있

**Table 2.** Curve fitting method of each experiment

Experiment	Curve fitting method	Experiment	Curve fitting method	Experiment	Curve fitting method	Experiment	Curve fitting method
Cu 1 ppm	Biphasic	Diazinon 1 ppm	Biphasic	Cyanogen 1 ppm	Asym2sig	Atrazine 0.01 ppm	LogNormal
Cu 10 ppm	Polynomial	Diazinon 10 ppm	GaussAmp	Cyanogen 5 ppm	Polynomial	Atrazine 0.1 ppm	Biphasic
Cu 100 ppm	Polynomial	Diazinon 100 ppm	Polynomial	Cyanogen 10 ppm	Polynomial	Atrazine 1 ppm	Polynomial
Fe 10 ppm	Polynomial	Fe 20 ppm	Polynomial	Rubber	Polynomial	Pigment	Polynomial



**Fig. 2.** Cu 10 ppm raw data and the curve fitted data graph.

으나 정확도를 높이기 위하여 출력값으로 주어지는 Wem-tox 최종값의 평균이 0.3보다 클 때와 작을 때의 2가지로 나누어서 모델을 구성하였다. 이때 Wem-tox 값이 0.3 부근에 있어서 어느 한 그룹에 속하면 잘못 판단할 가능성이 있으므로 이를 배제하기 위해 Wem-tox의  $m \pm 2\sigma$  (평균값  $\pm 2$ 표준편차)값이 0.3을 포함하는 독성물질은 그룹 1과 그룹 2에 동시에 포함시켰다. 이렇게 구분된 독성물질이 Table 3에 주어졌다.

$$y = A_0 + A_1 * x + A_2 * x^2 + A_3 * x^3 + A_4 * x^4 + A_5 * x^5 \quad (4)$$

$$y = A_{\min} + \frac{(A_{\max 1} - A_{\min})}{1 + 10^{(x-x_{0_1})/h^1}} + \frac{(A_{\max 2} - A_{\min})}{1 + 10^{(x-x_{0_2})/h^2}} \quad (5)$$

$$y = y_0 + A * \exp(-0.5 * ((x - x_c) / w)^2) \quad (6)$$

$$y = y_0 + A * (1 / (1 + \exp(-(x - x_c + w_1 / 2) / w_2))) * (1 - 1 / (1 + \exp(-(x - x_c - w_1 / 2) / w_3))) \quad (7)$$

**Table 3.** Two groups of toxicants sorted by the criterion of average value of Wem-tox = 0.3

Group 1	Group 2
Cu 1,10,100 ppm, Diazinon 1,10,100 ppm Cyanogen 1.5,10 ppm, Atrazine 0.01 ppm Fe 10 ppm, Fe 20 ppm	Cu 100 ppm, Diazinon 100 ppm Atrazine 0.1 ppm, Rubber Pigment

$$y = y_0 + A * (\text{sqrt}(2 * p_1) * w * x) * \exp(-\ln(x/x_c))^2 / (2 * w^2) \quad (8)$$

커브피팅된 Wem-tox 값을 기준 입력으로 하여, 각 그룹의 독성 물질을 대표하는 신경 회로망을 구성하였다. 신경 회로망을 구성하는 것은 신경회로망의 파라미터 중 은닉층 수, 은닉층의 뉴런 수, 입력값의 개수, 모멘텀, 학습률 등을 결정하는 것이다. 최적의 신경회로망을 구하기 위해서 많은 노력이 필요하다. 이러한 노력을 최소화하기 위하여 다꾸지 방법의  $L_{18}(2^1 * 3^7)$  직교 배열을 이용하였는데 이를 위한 입력표는 Table 4-1, 4-2와 같다. 이런 식으로 신경회로망의 파라미터 범위를 줄인 결과 결정된 모델의 파라미터는 Table 5와 같다.

**Table 4-1.** The optimal parameters for the model

	A	B	C
No. of hidden layer	1	2	3
No. of neuron in the hidden layer	3	6	10
No. of input	50	100	150
Momentum value	0.9	0.5	0.1
Learning rate	0.9	0.5	0.1

다꾸지 방법에 의한 최적 파라미터는 은닉층 수 1, 은닉층 뉴런수 22, 입력 개수 50, 모멘텀 0.1, 학습률 0.1로 나타났다. 하지만 해석 조건이 은닉층 1, 은닉층 뉴런수 10, 입력 개수 100, 모멘텀 0.9, 학습률 0.6일 때 실제값과 추정값의 오차가 더 작게 나왔다. 이는 다꾸지 방법에 의해 구해지는 최적 파라미터와 다른 결과이다. 이렇게 다른 결과가 나오는 이유는 해석하려는 시스템이 결정적 시스템(deterministic system)이 아니고 확률적인 시스템(stochastic system)이며, 또한 무작위로 결정되는 초기 값들에 의해 시스템의 성능이 좌우되기 때문인 것으로 추정된다.

**Table 4-2.**  $L_{18}(2^1 \times 3^7)$  orthogonal array

	No. of hidden layer	No. of neuron in hidden layer	No. of input	Momentum value	Learning rate
1	A	A	A	A	A
2	B	B	B	B	B
3	C	C	C	C	C
4	A	B	B	C	C
5	B	C	C	A	A
6	C	A	A	B	B
7	B	A	C	B	C
8	C	B	A	C	A
9	A	C	B	A	B
10	C	C	B	B	A
11	A	A	C	C	B
12	B	B	A	A	C
13	B	C	A	C	B
14	C	A	B	A	C
15	A	B	C	B	A
16	C	B	C	A	B
17	A	C	A	B	C
18	B	A	B	C	A

### 3. 결과 및 고찰

2장에서 얻어진 모델을 이용하여 오염물질을 추정하는 확인 실험을 수행하였고 결과가 Table 6에 주어졌다. 회로망 모델이 구리 1 ppm이 포함된 측정수를 다이아진은 1 ppm으로 판정할 가능성이 21%, 시안 10 ppm으로 판정할

가능성이 1%, 아트라진 0.01 ppm으로 판정할 가능성이 31%이며, 공백은 0%이고, 소수 셋째 자리에서 반올림한 값이다. Table 6을 살펴보면 구리 100 ppm, 다이아진은 100 ppm, 시안 10 ppm, 아트라진 0.1 ppm, 1 ppm, 고무 등과 같이 오염도가 높을 때는 그것이 어떤 물질인지 매우 정확하게 판정한다. 하지만, 구리 1 ppm, 철 10 ppm, 시안 1 ppm 등과 같이 오염물질이 매우 적을 때는 정확하게 판정하는 확률이 낮아진다. 사실 이때는 육안으로 그래프를 검사하여도 정확하게 판정하기란 쉽지 않다.

이를 보완하기 위하여 새로운 알고리즘을 고안하였다. Fig. 3이 그 알고리즘을 나타내는 흐름도이다. 이 알고리즘의 기본적인 아이디어는 가장 높은 확률이 연속 3번 나온 오염물질을 신경회로망이 판단한 오염물질로 결정하는 것이다. 이 방법은 첫 번째 확률을 가진 물질과 두 번째 확률을 가진 확률을 가진 물질의 비를 더 크게 한다. 즉, 첫 번째 확률을 가진 물질로 판단될 확률이 좀 더 커지는 것이 이 알고리즘이다. 단 이렇게 하면 판단하지 않고 넘어갈 확률도 커지므로 포함된 물질을 결정하는데 걸리는 시간이 길어진다. 다투지 방법이 아닌 실제값과 추정값의 오차가 가장 작게 나오는 파라미터 조건으로 구성된 신경 회로망을 이용하여 각각의 오염 물질을 입력(input) 값으로 실험한 결과가 Table 7에 주어졌다. Table 7에서 볼 수 있듯이 전체적으로 올바른 판정을 하는 확률이 향상되었다. 특히 구리 1 ppm과 철 20 ppm의 경우에는 각각 47%에서 74%로, 66%에서 98% 확률이 매우 향상되었음을 알 수 있고 다른 오염물질로 판정할 확률이 현저하게 줄었음을 알 수 있다. 즉 철 20 ppm의 경우 구리 1 ppm으로 판정할 확률 13%, 다이아진은 1 ppm으로 판정할 확률이 13%, 시안

**Table 5.** Neural Network model determined by Taguchi method

No. of hidden layer	No. of neuron in hidden layer	No. of input	Momentum	Learning rate
1	22	50	0.1	0.1

**Table 6.** Out put value by test input using the ninth row parameter at first experiment

In put	Out put	Cu	Cu	Cu	Dz	Dz	Dz	Cg	Cg	Cg	At	At	At	Rb	Pg	Fe	Fe
		1	10	100	1	10	100	1	5	10	0.01	0.1	1			10	20
Cu 1	0.47				0.21					0.01	0.31						
Cu 10		0.98								0.02							
Cu 100			1.00														
Dz 1					1.00												
Dz 10			0.13		0.00	0.87											
Dz 100							1.00										
Cg 1	0.47				0.12	0.19	0.02				0.14					0.07	
Cg 5								1.00									
Cg 10									1.00								
At 0.01										1.00							
At 0.1											1.00						
At 1												1.00					
Rb				0.05										0.90	0.05		
Pg															1.00		
Fe 10					0.11	0.21					0.11					0.27	0.32
Fe 20	0.13				0.13			0.09									0.66

Table 7. Out put value by test input using the complementing algorithm

	Cu 1	Cu 10	Cu 100	Dz 1	Dz 10	Dz 100	Cg 1	Cg 5	Cg 10	At 0.01	At 0.1	At 1	Rb	Pg	Fe 10	Fe 20
Cu 1	0.74			0.06					0.00	0.20						
Cu 10		1.00							0.00							
Cu 100			1.00													
Dz 1				1.00												
Dz 10		0.00			1.00											
Dz 100						1.00										
Cg 1	0.91			0.01	0.06	0.00				0.02					0.00	
Cg 5								1.00								
Cg 10									1.00							
At 0.01										1.00						
At 0.1											1.00					
At 1												1.00				
Rb			0.00										1.00	0.00		
Pg														1.00		
Fe 10				0.02	0.14					0.02					0.31	0.51
Fe 20	0.01			0.01				0.00								0.98

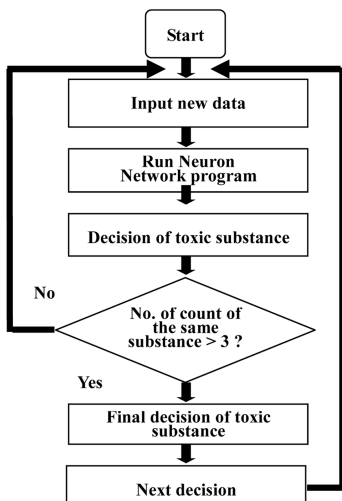


Fig. 3. Flow chart to determine the toxicological substance contaminating the water.

5 ppm으로 판정할 가능성이 9%였으나 새로운 알고리즘을 적용한 결과, 구리 1 ppm으로 판정할 확률 1%, 다이아지는 1 ppm으로 판정할 확률 1%, 시안 1 ppm으로 판정할 확률 0% 등으로 바뀌었다. 결과적으로 정상적으로 판정할 확률이 매우 높아졌다. 16개의 실험중에서 제대로 판정하지 못한 것이 시안 1 ppm의 경우 1개인데 이는 원래 데이터에서 얻어진 그래프 자체가 구리 1 ppm의 경우와 시안 1 ppm의 경우가 매우 구분하기 어렵기 때문이다. 그리고 철 10 ppm의 경우에는 처음부터 철 20 ppm으로 판정할 확률이 높기 때문에 확률이 32%에서 51%로 높아졌지만, 다이아지는 1 ppm, 10 ppm 및 아트라진 0.01 ppm으로 판정할 확률은 낮아진 것을 알 수 있다.

#### 4. 결론

이 연구에서는 오염물질을 정성적으로 판단하는 생물학

적인 방법을 이용하여 오염물질의 종류를 판정하는 새로운 방법을 제안하였다. 판정에 사용된 반달말이 각각의 독성 물질에 반응하는 광합성 형광량의 변화패턴을 이용하여 수계 시스템이 포함하고 있는 독성 물질을 판별하는 방법을 제안하였다. 독성물질을 판별하는 모델로 신경회로망을 이용하였고 정확성을 높이기 위하여 WEM-tox 값의 크기에 따라 2개의 모델 중 하나를 선정하도록 하였으며 경계선에서 잘못 판정할 가능성을 최소화하기 위해 경계값은 가진 물질은 양쪽에 다 포함시켰다. 신경 회로망 파라미터를 최적화시키기 위해 다투지 방법을 이용하였다. 신경회로망 모델의 판정확률을 높이기 위해서 또 다른 새로운 알고리즘을 도입하였다. 3번 연속 같은 물질로 판정될 때 그 물질이라 결정하는 이 알고리즘은 옳은 판정을 내릴 확률을 현격하게 증가시켰다.

#### 참고문헌

이지나, 황인영(1999). 생물학적 검정법을 이용한 소규모 수계내 수질 오염물질의 환경독성 평가. *환경독성학회지*, 14(4), pp. 135-144.

환경부(2004). 수질오염공정시험법.

황인영(1997). 생물학적 지시자에 의한 수질 오염 평가. *인제환경연구소 제 6회 심포지움 논문집*, pp. 37-50.

Greene, J. C., Bartels, C. J., Warren-Hicks, W. J., Park hurts, B. R., Linder, G. L., Peterson, S. A. and Miller, W. E. (1989). *Protocol for Short Term Toxicity Screening of Hazardous Waste Sites*, EPA, 600/3-88-029.

Juneau, P., Sumitomo, Hi., Matsui, S., Itoh, S., Kim, S. G. and Popovic, R. (2003). Use of Chlorophyll Fluorescence of *Closterium Ehrenbergii* and *Lemna gibba* for Toxic Effect Evaluation of Sewage Treatment Plant Effluent and Its Hydrophobic Components. *Ecotoxicology and Environmental Safety*, 55, pp. 1-8.

Keddy, C. J., Greene, J. C. and Bonnell, M. A. (1995). Re-

view of whole-Organism Bioassay: Soil, Freshwater Sediment, and Freshwater Assessment in Canada. *Ecotoxicology and Environmental Safety*, **30**, pp. 221-254.

Phadke, M. S. (1989). *Quality Engineering Using Robust Design*,

Prentice Hall, U.S.A.

Rumelhart, D. E., McClelland, J. L. and PDP Research group (1987). *Parallel Distributed Processing*. the MIT Press, U.S.A.