

**NEURAL NETWORK SOLUTION FOR  
ASSESSMENT OF EUTROPHICATION IN  
LAKE TENKILLER**

*Thesis Approved*

**By**

**ANUPAM KUMAR**

**Bachelor of Technology**

**I.I.T-KANPUR**

**KANPUR, INDIA**

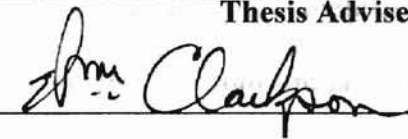
**1993**

**Submitted to the Faculty of the  
Graduate College of the  
Oklahoma State University  
in partial fulfillment of  
the requirements for  
the Degree of  
MASTER OF SCIENCE  
May, 2000**

**NEURAL NETWORK SOLUTION FOR  
ASSESSMENT OF EUTROPHICATION IN  
LAKE TENKILLER**

**Thesis Approved:**

  
\_\_\_\_\_  
**Thesis Adviser**

  
\_\_\_\_\_

  
\_\_\_\_\_

  
\_\_\_\_\_  
**Dean of the Graduate College**

## **PREFACE**

This study was conducted to ascertain the possibility of applying neural networks to environmental engineering problems. Eutrophication of Lake Tenkiller (Oklahoma) due to inflow of phosphorus from various point and non-point sources was selected for this case study. The phosphorus sources are distributed across Oklahoma and Arkansas and are transported to the Lake by the Illinois River. This study attempts to determine the contribution of point and non-point sources to the eutrophication of the Lake and reductions in nutrient inflow required for remediation of the Lake. Further, results obtained are compared with results from similar studies carried out using different methodologies.

I sincerely thank my advisor Dr. William F. McTernan and my committee members, Drs. William W. Clarkson and Greg Wilber – for guidance and support in completion of this research. I also thank the Environmental Institute, Oklahoma State University for providing me with the necessary data to research. Finally, I thank my parents for making possible higher education at Oklahoma State University.

Chapter	Page
1. Introduction	1
2. Eutrophication	5
3. Previous Studies and Background Data	18
4. Neural Network	29
5. Materials and Methods	62

## Table of Contents

Chapter	Page
1. Introduction	
1.0 Background	1
1.1 Model Selection	3
1.2 Specific Objectives	4
2. Eutrophication	
2.0 Introduction	5
2.1 Important Factors in Eutrophication	6
2.2 Nutrients of Concern	7
2.2.1 Phosphorus	7
2.2.2 Nitrogen	8
2.3 Nutrient Availability at Tenkiller	8
2.4 Eutrophication in Reservoirs	14
3. Previous Studies and Background Data	18
4. Neural Network	
PART - I	
4.0 Introduction	29
4.1 Artificial Neural Network	29
4.2 Artificial Neural Network Structure	32
4.2.1 Neural Network Model	35
4.3 Backpropagation Network	37
4.4 Neural Network Operation	37
4.5 Learning	40
4.6 Genetic Algorithm	46
4.7 Neural Network and GA Nomenclature	47
PART - II	
4.8 Neural Network Applications in Environmental Engineering	54
4.8.1 Case Studies	54
5. Materials and Methods	
5.0 Introduction	62
5.1 Model Development	78

<b>Chapter</b>	<b>Page</b>
6. Results and Discussions	87
7. Summary and Conclusions	136
Bibliography	138
Appendix - I	142

	Page
	88

### List of Tables

<b>Table</b>	<b>Page</b>
2-1. Land Uses for the Illinois River Basin Based on 1985 Data	10
2-2. Nutrient Export Coefficients	12
2-3. List of Permitted Industrial and Municipal Sewage Treatment Plants along Illinois River Relevant to this Study	13
2-4. Lake Tenkiller Morphometry	17
3-1. Summary of Point Source Data at Lake Tenkiller Ferry	20
3-2. Mean, Maximum and Minimum of Sampled Parameters	26
3-3. Estimates of Phosphorus Discharges from Point Sources for the Period 1991 to 1993	28
4-1. Computed Aquifer Parameter Values from Different Methods	57
5-1. Available Data	63
5-2. Training Data Set for Network Model - I	67
5-3. Training Data Set for Network Model - II	68
5-4. Training Data Set for Secchi Disk Simulations	71
5-5. Point Source Contribution at Tahlequah	76

<b>Table</b>	<b>Page</b>
6-1. Alternatives for Building Network Model - II ( <i>without</i> with <i>genetic supervisor</i> )	88
6-2. Performance of a 3-layer Network with Varying Learning Rate and Momentum Inputs	96
6-3. Alternatives for Building Network Model - II ( <i>with genetic supervisor</i> )	98
6-4. Performance of Various Simulations for Developing Network Model - I	102
6-5. Network Parameters for Model - I and Model - II	108
6-6. Chlorophyll <i>a</i> Predictions in the Lake with Network Model - II	111
6-7. Chlorophyll <i>a</i> Predictions at Tahlequah Using Network Model - II without Point Source Phosphorus Contribution	113
6-8. Chlorophyll <i>a</i> Predictions at Tahlequah Using Network Model - II with 50% Reduced Non-Point Source Contribution	115
6-9. Chlorophyll <i>a</i> Predictions at Tahlequah Using Network Model - II with 75% Reduced Non-Point Source Contribution	118
6-10. Chlorophyll <i>a</i> Predictions at Tahlequah Using Network Model - II with 80% Reduced Non-Point Source Contribution	119
6-11. Chlorophyll <i>a</i> Predictions at Tahlequah Using Network Model - II with 85% Reduced Non-Point Source Contribution	120
6-12. Secchi disk in the Lake Using Network Model - III	125
6-13. Secchi disk at Tahlequah Using Network Model - III after Total Point Source P-reduction from Inputs	126
6-14. Secchi disk at Tahlequah Using Network Model - III with 50% Non-Point Source Reduction in Phosphorus	127
6-15. Secchi disk at Tahlequah Using Network Model - III with 75% Non-Point Source Reduction in Phosphorus	128
6-16. Secchi disk at Tahlequah Using Network Model - III with 80% Non-Point Source Reduction in Phosphorus	129

<b>Table</b>	<b>Page</b>
6-17. Secchi disk at Tahlequah Using Network Model - III with 85% Non-Point Source Reduction in Phosphorus	131

Figure

6

Page

Figure 6-17  
Secchi disk at Tahlequah

131



Figure	Page
2-1. Lake Tenkiller	15
3-1. Concentration Probability for Total Nitrogen and Total Phosphorus at Tahlequah, Oklahoma Water Quality Gauge	19
3-2. Illinois River Schematic	21
3-3. Phosphorus Channel Losses Probability Plot	23
3-4. Management Alternatives	24
4-1. Schematic Drawing of Biological Neuron	31
4-2. Multiple-Input Neuron	33
4-3. Multi-Layer Feed Forward Model	36
4-4. Backpropagation Network	38
4-5. Schematic of Network Operation	39
4-6. Feed Forward Neural Net	42
4-7. Differentiable Transfer Function: <i>Sigmoid Function</i>	48
4-8. Non-Differentiable Transfer Function: <i>Step Function</i>	49

## List of Figures

Figure	Page
2-1. Lake Tenkiller	15
3-1. Concentration Probability for Total Nitrogen and Total Phosphorus at Tahlequah, Oklahoma Water Quality Gauge	19
3-2. Illinois River Schematic	21
3-3. Phosphorus Channel Losses Probability Plot	23
3-4. Management Alternatives	24
4-1. Schematic Drawing of Biological Neuron	31
4-2. Multiple-Input Neuron	33
4-3. Multi-Layer Feed Forward Model	36
4-4. Backpropagation Network	38
4-5. Schematic of Network Operation	39
4-6. Feed Forward Neural Net	42
4-7. Differentiable Transfer Function: <i>Sigmoid Function</i>	48
4-8. Non-Differentiable Transfer Function: <i>Step Function</i>	49

<b>Figure</b>	<b>Chapter 1. Introduction</b>	<b>Page</b>
5-1. Flow Chart of Activities in Present Study		64
5-2. Probability Plot for Phosphorus Concentrations for the Period 1991 to 1993 at Tahlequah		66
5-3. Probability Plot for Monthly Peak Flow Data for the Period 1977 to 1985 at Tahlequah		75
5-4. A General Methodology for the Development of Neural Network Systems		79
6-1. RMS Error Plot of Various Alternatives		89
6-2. RMS Error Plot of Various Alternatives Simulated with Genetic Supervisor for Network Model - II		99
6-3. RMS Error Plots of Simulation for Network Model - I		104
6-4. RMS Error Plot for Network Model - III		110
6-5. Chl. <i>a</i> Values with Varying Phosphorus Reductions (Plot 1)		114
6-6. Management Alternatives with Chl. <i>a</i> Predictions (Pie 1)		117
6-7. Chl. <i>a</i> Values with Varying Phosphorus Reductions (Plot 2)		122
6-8. Management Alternatives with Chl. <i>a</i> Predictions (Pie 2)		123
6-9. Secchi Values with Varying Phosphorus Reductions		132
6-10. Management Alternatives with Secchi Disk Predictions		133

## Chapter 1. Introduction

### 1.0 Background

“The escalating costs of environmental cleanup together with conflicting concerns of various stake holders motivate the search for improved management methodologies to reduce costs” (Rogers et al., 1995). Realization that the pristine state at contaminated sites cannot be restored within feasible economic limits prompted the need for an optimization of resources utilized in environmental cleanups (Hellman and Hawkins, 1988, Wang, 1995). Management approaches are required to clean a contaminated site to a level acceptable depending on its end use. This level should preclude the possibility of any harmful effects to human populations or to impacted ecosystems. Formulating such decision making approaches is not simple and must take into account existing environmental policies, resource constraints and effective allocation of limited resources, defining ultimate goals and convincing the public about the efficacy and safety of the management strategies (Wang, 1995). Stochastic and deterministic models are used to aid in the decision-making process (Gorelick, 1993). A main challenge facing any such modeling technique is its ability to handle uncertainty relating to inputs and uncertainty inherent to the particular modeling technique (Freeze et al., 1990).

This study evaluated the effect of phosphorous (biostimulant) loading in the Illinois River basin of Arkansas and Oklahoma. This river system leads to Lake Tenkiller, which has experienced eutrophication. Eutrophication is a natural process of decay of a water body and is enhanced in the presence of certain macro and trace nutrients (Schindler, 1977). Excessive algal growth because of high nutrient content in the water

body leads to overall deterioration in the water quality. Several nutrients have been generally identified as contributors to algal growth. Phosphorus is considered to be one of the primary nutrients in the process of eutrophication, and Lake Tenkiller is considered to be phosphorus limited, as was determined by studies performed earlier (Harton, 1989, Haraughty, 1995 and OWRB et al., 1996). Phosphorus is contributed from point as well as non-point sources of pollution. Point sources of phosphorus loading include wastewater treatment plant discharges while non-point sources include surface water runoff high in nutrients from poultry and cattle litter (OWRB et al., 1996). "A 1989 Soil Conservation Service inventory estimated there were approximately 230 million poultry and 300,000 cattle and swine residing in the basin during an average year. These animals produced an estimated 8.8 billion pounds of manure per year", (OWRB et al., 1996). This manure is used as fertilizer on forage lands, and a significant fraction is transported to surface water bodies during runoff events (OWRB et al., 1996).

The OWRB study also evaluated some general management alternatives to stem the accelerated eutrophication of Lake Tenkiller. Several options of phosphorus load reductions to the Illinois River system were evaluated. The approach for control of phosphorus loading to the river was divided into two parts: 1) control from point source contribution, and 2) control from non-point source contributions. The effectiveness of each management alternative was judged by simulated chlorophyll *a* and secchi disk values obtained after reduction in phosphorus loading to the lake. Chlorophyll *a* and secchi disk readings are considered to be indicators of the planktonic growth responsible for accelerated eutrophication (Haraughty, 1995).

**1.1 Model Selection** evaluated with data from the period of 1977 through 1985. These

To analyze the effects of phosphorus reduction on eutrophication parameters (measured by chlorophyll *a* and secchi disk) it was necessary to establish relationships between phosphorus and chlorophyll *a* concentrations, and phosphorus concentrations and secchi disk values. As eutrophication is a gradual process with a significant time step, deterministic steady state artificial neural network models were applied for predictions of phosphorus, chlorophyll *a* and secchi disk values.

The choice of artificial neural network (ANN) application to the problem was governed by two factors: 1) the ability of neural networks to “learn” or adapt to new conditions or situations and 2) incomplete data sets. ANNs do not assign a particular technique for modeling a given set of data. The equations used for making final predictions are totally data dependent. The artificial neural network builds a model with a combination of linear and non-linear equations which it formulates as it attempts to model the output data with data inputs. This improvisation technique enables the network to learn the underlying intricacies between input and output relationships unique to the environment. Since artificial neural networks have a data dependent adaptive learning approach, which simultaneously preserves locational and temporal biases in a data set, it was chosen as the analytical tool for this environmental problem.

Due to the availability of incomplete data sets, there was a need to generate phosphorus data for lower reaches of the lake for the period of 1977 through 1985 to evaluate the resultant trophic state. The network was conditioned on data between 1992 to 1993 collected by OWRB (1996) to establish relationships between phosphorus and chlorophyll *a* data, and additionally, phosphorus and secchi disk data. The trophic state of

the system was then evaluated with data from the period of 1977 through 1985. These different data sets were employed as no single, available data set described all of the phosphorus discharges and subsequent trophic alterations simultaneously. The availability of point source data for the period 1977 through 1985 dictated the period of trophic state evaluation. Since there were no sampling data available for the Lake for the period 1977-1985, an artificial neural net model was conditioned to establish the phosphorus relationships between the upper and lower reaches of the lake with data available for the period of 1992 - 1993. This network was then used for phosphorus predictions in the lower reaches of the reservoir by employing available data for the upper reaches as input to the model. In this manner a complete data set was generated by the artificial neural network modeling approach taken for the period 1977 through 1985.

## **1.2 Specific Objectives**

The specific objectives of this study was to develop models for:

- 1) Characterization of eutrophic state of Lake Tenkiller,
- 2) Evaluating reductions in phosphorus loading for controlling eutrophication, and
- 3) Comparing the performance of neural network approach with other methodologies applied to research the eutrophication of Lake Tenkiller.

## Chapter 2. Eutrophication

### 2.0 Introduction

A water body goes through three stages (trophic levels) in the eutrophication process. Starting with the cleanest water quality levels, these three states of eutrophication are *oligotrophic*, *mesotrophic* and *eutrophic* (Reynolds, 1984). There have been several studies that define the trophic level of a water body based on different parameters. Rosas et al., 1993, based the trophic level of a lake on the response of biota to changing conditions. Other studies have correlated the trophic level to nutrient availability in a water body (Shannon and Brezonik, 1972). The Oklahoma Water Resources Board (1996) measured the trophic level by the parameters chlorophyll *a* and secchi disk observations. These data were employed in this study.

Chlorophyll *a* has been widely used as a correlative parameter for estimations of phytoplankton biomass and productivity (Reynolds, 1984). It accounts for 0.5 to 2% of algal dry weight (Reynolds, 1984). From an EPA National Eutrophic Survey, a chlorophyll *a* value of 10µg/L has been accepted as the breakpoint between mesotrophic and eutrophic state of a lake (Gakstatter et al., 1974).

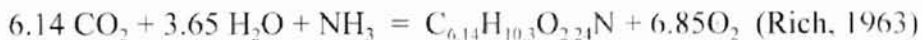
Secchi disk is used for measuring the depth of visibility of a disk lowered in a water body. Secchi disk is a weighted circular plate 8" in diameter painted white or alternatively in black and white quadrants (Reynolds, 1984). A classification adopted by the Michigan Department of Environmental Quality (DEQ) of the trophic state of lakes based on secchi disk values is:

- Oligotrophic - Greater than 4.88 meters. The rate of oxygen and carbon are always
- Mesotrophic - Between 1.98 meters and 4.88 meters. The rate of oxygen and carbon are
- Eutrophic - Less than 1.98 meters (<http://www.iserv.net/~mlsa/secchi.html>).

## 2.1 Important Factors in Eutrophication

There has been research relating the chemical composition of water to the growth and presence of phytoplankton in a body of water. Pearsall related diatoms to levels of silica, and various kinds of algae to levels of silica, phosphorus and nitrogen (Reynolds, 1984). Other factors such as optimum temperature, light availability as related to photoperiod and intensity, and their effects on algal productivity have been well documented (Haraugthy, 1995). Routinely these are not taken into account in engineering studies because of a lack of control over these factors. That is, for engineering studies only those factors that can be manipulated with available technology are considered.

Phytoplankton growth has been demonstrated to be a result of bioavailability of nutrients, nutrient content of algal cells, and bioavailable macronutrient ratios in the media (Reynolds, 1984). The equation for cell synthesis of algae with ammonia as nitrogen source is:



From the cell synthesis equation, nitrogen comprises 19.3 % of the cell dry weight.

“Algae are chlorophyll-bearing organisms and their growth is influenced greatly by fertilizing elements in the water” (Sawyer and McCarty, 1978). “Research has shown that nitrogen and phosphorus are both essential for growth of algae and that limitation in amounts of these elements is usually the factor that controls their rate of growth” (Sawyer



and McCarty, 1978). The other elements (hydrogen, oxygen and carbon) are always found in abundance in typical environments and therefore are not considered as “limiting” nutrients for cell growth. Since the availability of nitrogen and phosphorus form the major limitation for phytoplankton yields (Reynolds, 1984), the accepted control approach has been to remove either nitrogen or phosphorus from waters draining into lakes and reservoirs. The choice between the control of nitrogen and phosphorus is normally determined by the ratio of nitrogen to phosphorus. If the nitrogen to phosphorus ratio (N:P) in the lake is less than 5, algal growth is considered to be nitrogen limited. If the ratio is between 5 and 10, either or both may control eutrophication. If N:P is greater than 10, phosphorus should control eutrophication (U.S. EPA, 1977).

## **2.2 Nutrients of Concern**

The following sections elaborate on the role of important biostimulants in the process of eutrophication. They deal with phosphorus and nitrogen in terms of their necessity for algal growth and forms of these nutrients required for growth.

### **2.2.1 Phosphorus**

Phosphorus is an essential element in the production of energy molecules, such as adenosine triphosphate (ATP), used for growth and maintenance purposes by algae (Harris, 1986). Among the available forms of phosphorus, orthophosphates are the most readily assimilated form and are metabolized without further breakdown (Metcalf and Eddy, 1991). Dissolved orthophosphates are the primary sources of phosphorus available to phytoplankton (Reynolds, 1984). Other forms of available phosphorus are polyphosphate and organic phosphate. The forms of phosphorus usually measured in water are total phosphorus (TP) and soluble reactive phosphorus (SRP) (Reynolds, 1984).

A substantial part of the total phosphorus (TP) in soil water runoff is in the form of dissolved organic phosphorus (DOP). The DOP may be more mobile than inorganic orthophosphate and thus can be a more important phosphorus source for surface water eutrophication (Chardon et al., 1997). Since phosphorus concentrations as low as 10 µg/l can stimulate algal growth, phosphorus inputs need to be controlled (Haygarth and Jarvis, 1997). Therefore, understanding the nature and mechanism of phosphorus release into waters is integral to any management control strategy (Haygarth and Jarvis, 1997).

### **2.2.2 Nitrogen**

Nitrogen is mainly required as a component for synthesis of amino acids and proteins in aquatic organisms (Reynolds, 1984). Commonly occurring forms of nitrogen available to algae are: nitrate, nitrite, ammonium ions, and certain dissolved organic nitrogenous compounds (Reynolds, 1984). Several species of blue-green algae use ammonium in cells for fulfilling nitrogen requirements to manufacture amino acids (Lee, 1989) and ammonium is considered as the most energetically favorable nitrogen form for algal growth (Reynolds, 1984). "Ammonium primarily results from bacterial degradation of organic matter and animal excrement" (Haraughty, 1995). In general, because of the complex processes in the nitrogen cycle, the abundance of nitrogen is much greater than phosphorus, causing phosphorus to often be considered to be the "limiting" nutrient.

### **2.3 Nutrient Availability at Tenkiller**

Lake Tenkiller is characterized as having generally high water quality which promotes various recreational, power generation, flood control and municipal uses (OWRB et al., 1996). The upstream Illinois River is Oklahoma's only state-designated

“scenic” river, and concerns exist that accelerated eutrophication of the lake from various anthropogenic activities in the drainage basin is occurring (OWRB et al., 1996). Nutrient availability to the lake is quantified based on origin as point sources and non-point sources.

#### Non-point sources

The drainage area of the lake consists of 4170 sq. kilometers, with approximately 55% in eastern Oklahoma and the remainder in western Arkansas (OWRB et al., 1996). Land use for the Illinois River Basin, derived from map work sponsored by U.S. EPA (OWRB et al., 1996), is listed in Table 2-1, which presents percentage of land cover based on land surface development or use. From the table it is seen that pasture/range and forest lands account for 92.88% of the Illinois River watershed, with forest lands accounting for 43.48% and pasture/range lands comprising 49.40%. Croplands comprise 1.33% of the basin and are of limited concern (OWRB et al., 1996).

It has been observed that nutrient loading to the lake from forest lands is the least of the various land covers even though maximum land coverage falls under this category (OWRB et al., 1996). This is because the vegetative cover of the forest lands stems the flow of surface water runoff and allows most of it to percolate or evapotranspire (OWRB et al., 1996). Besides the dense vegetation, forest lands have a greater water retention capacity in the layer of mulch and organic debris covering the top soil layer. Water captured by the organic layers on top soil percolates into the underlying soil layers. Overall, vegetation cover and organic layers on the soil surface prevent the flow of water enriched with dissolved and suspended nutrient materials to make its way to nearby surface water bodies. Pasture and range lands are a pollution hazard as they provide

**Table 2-1. Land Uses for the Illinois River Basin Based on 1985 Data  
(OWRB et al., 1996)**

Land Use	Illinois River Basin Area, Hectares(%)	Oklahoma Portion Area, Hectares (%)	Arkansas Portion Area, Hectares (%)
Crop	5713.20(1.33)	1675.44(0.72)	4037.76(2.06)
Confined Animal	1647.99(0.38)	232.65(0.10)	1415.34(0.72)
Forest	186199.20(43.48)	128955.69(55.49)	57243.51(29.24)
Pasture/Range	211521.87(49.40)	91679.76(39.45)	119842.11(61.21)
Roads & ROW	1227.15(0.29)	572.40(0.25)	654.75(0.33)
Urban	14980.77(3.50)	3028.23(1.30)	11952.54(6.10)
Water	6910.09(1.61)	6258.15(2.69)	652.14(0.33)
Totals	428200.47(100.00)	232402.32(100.00)	195798.15(100.00)
		(54.27% of Total)	(45.73% of Total)

nutrients for forage growth (Haraughty and Burks, 1996). This litter then becomes available for transport to surface water bodies during runoff events. Confined animal operations, which include poultry rearing, dairy, and hog rearing, comprise 0.38% of the total basin area, but they are the largest generators of nutrients in the form of litter and waste produced by animals (OWRB et al., 1996).

The amount of nutrients reaching the waterways from each land use operation is defined by nutrient export values. The value is an estimate in kilograms per hectare per year (kg/ha/yr) of a particular nutrient that is transported from a particular land use of the watershed. The corresponding range of values estimated by OWRB et al., 1996, for nitrogen and phosphorus from land uses of concern are given in Table 2-2. From the ranges presented, it can be seen that the estimated export value from pasture land is roughly 2 to 3 times that from forest lands and export values from confined animal operations are orders of magnitude higher. Based on nutrient export values it was determined that confined animal operations are the largest non-point contributors of nutrients to Lake Tenkiller, while pasture and range lands are the second largest. Based on data reported for the period 1991 to 1993, the average annual total phosphorus load from upstream land use was 270,557 kg-P/yr and total nitrogen was 2,578,978 kg-N/yr to Lake Tenkiller at the Horseshoe Bend monitoring station (OWRB et al., 1996).

#### Point sources

Table 2-3 lists the permitted industrial and municipal wastewater treatment plants along the reaches of Illinois River relevant to the present study. There are a total of 14 permitted treatment plants in Arkansas and Oklahoma discharging through various creeks into the Illinois River (OWRB et al., 1996). Based on data reported for the period of

Table 2-2: Nutrient Export Coefficients (OWRB et al., 1996) for Major Land Use Types of the Illinois River Relevant to this Study (OWRB et al., 1996)

Land Use	Total Phosphorus (kg/ha/yr)	Total Nitrogen (kg/ha/yr)
Row Crops	0.9 - 5.3	4.0 - 21.8
Non Row Crops	0.6 - 1.5	4.1 - 6.5
Pasture	0.2 - 2.6	2.4 - 10.9
Mixed Agriculture	0.5 - 1.4	9.4 - 25.5
Urban	0.6 - 2.7	4.0 - 11.2
Forest	0.1 - 0.3	2.2 - 3.3
Confined Animal Operations	120	750

**Table 2-3. List of Permitted Industrial and Municipal Sewage Treatment Plants along Illinois River Relevant to this Study (OWRB et al., 1996)**

NPDES Permit ID	Name
OK0001198	Cavenham Forest Industries
OK0027456	Cherokee Nation of Oklahoma
OK0034070	Cherokee Nation Of Oklahoma (Sequoyah High School)
OK0030341	Stilwell Area Development Authority
OK0026964	City of Tahlequah
OK0028126	City of Westville
AR0020010	City of Fayetteville
AR0033910	USDAFS – Lake Wedington Recreation Area
AR0035246	City of Lincoln
AR0022098	City of Prairie Grove
AR0043397	City of Rogers
AR0022063	City of Springdale
AR0020273	City of Siloam Springs
AR0020184	City of Gentry

1991 – 1993, point source contribution was 12547 kg-P/yr and 61605 kg-N/yr entering Lake Tenkiller at Horseshoe Bend.

## 2.4 Eutrophication in Reservoirs

Reservoirs can be classified into three zones (Thornton et al., 1990). Because of the changing depths and changing flow velocities in the lake, the dynamics of the eutrophication process differ in the three zones known as: *riverine*, *transition* and *lacustrine* zones. The *riverine* zone is a well-mixed system with plenty of nutrients and high turbidity (Haraughty, 1995). High turbidity hinders light penetration and therefore impedes the photosynthetic process needed for plankton growth. The transition zone begins with the flow velocity reductions and vertical mixing because of increased depths (Haraughty, 1995). The nutrients and particles start settling in this zone, resulting in lower turbidity. This zone has high nutrient content and good light penetration and therefore supports maximum algal growth. Anoxic conditions can be created because of improper mixing which also aids in the eutrophication process (Haraughty, 1995). The lacustrine zone is characterized by clear water and low algal growth (Haraughty, 1995). Though light penetration is good, nutrient availability is at its minimum, which stems excessive plankton growth.

Construction of Lake Tenkiller began in 1947 for the purpose of flood control and hydroelectric power generation in the region (OWRB et al., 1996). The Lake was built by the United States Army Corps of Engineers and is located in Cherokee and Sequoyah counties of eastern Oklahoma. It is along the Illinois River with the river flowing into and out of the Lake as shown in Figure 2-1. The Lake is located 20.6 km upstream of the confluence of the Illinois and Arkansas rivers. There are several lateral tributaries



the Lake as shown in the figure. These tributaries are about 200 m long  
 (see OWRB et al., 1996). The study the upper reaches of

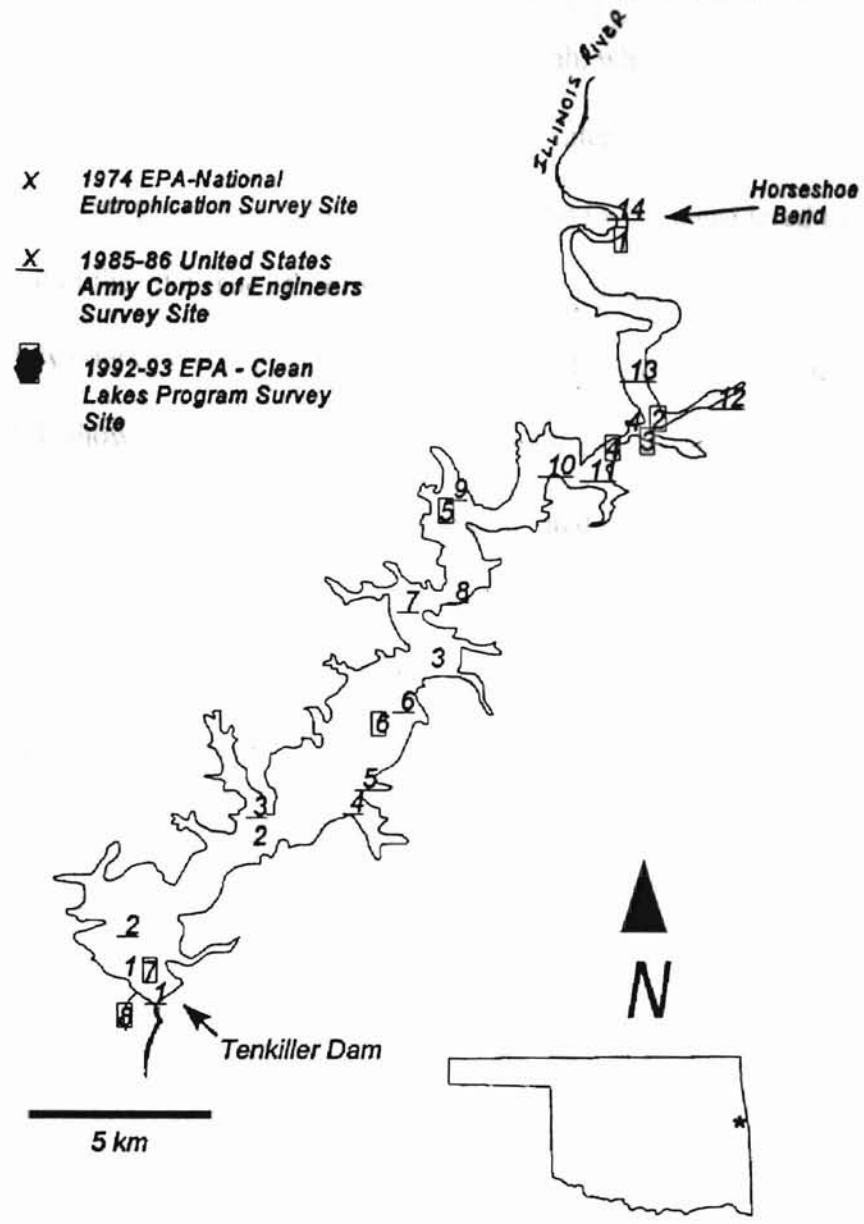


Figure 2-1. Lake Tenkiller (OWRB et al., 1996)

flowing directly into the lake as shown in the figure. These tributaries account for 25% of the watershed of the lake (OWRB et al., 1996). For this study the upper reaches or riverine zone of the lake were considered to be above Tahlequah, which is 6 miles upstream of Horseshoe Bend and is the beginning of the headwaters of the lake. The conditions at Tahlequah resemble a well mixed turbid system and were considered to model the riverine zone of the lake. The other two zones were modeled at the locations marked by OWRB study (stations 4, 5, 6 and 7 in Figure 2-1). Morphometric data for the lake, given in Table 2-4, shows the depth of the lake varies over a range of 40 feet, and an average hydraulic residence time of 1.76 years has been estimated. Additionally, from Figure 2-1 it is observed that the width of the water body increases as the river flows into the lake and moves downstream. Therefore, it is expected that flow velocity reductions and particle settling will occur as water moves from riverine to lacustrine zones of the lake. The above observations provide a basis for dividing Lake Tenkiller into three zones for studying the effects of eutrophication.

**Table 2-4. Lake Tenkiller Morphometry (OWRB et al., 1996)**

Parameter	Value
Elevation (NGVD <sup>**</sup> )	
@Conservation pool	632.00
@Flood pool	667.00
Capacity (km <sup>3</sup> )	
@Conservation pool	0.81
@Flood pool	1.52
Area (km <sup>2</sup> )	
@Conservation pool	52.2
@Flood pool	84.2
Depth (m)	
Mean	15.5
Maximum	46.3
Relative	0.57
Shoreline length (km)	209
Shoreline development	8.17
Volume development	1.00
Average hydraulic residence time (yr)	1.76

\*\* - National Geodetic Vertical Datum

### **Chapter 3. Previous Studies and Background Data**

Eutrophication in Lake Tenkiller has been a cause for concern since 1974 (OWI et al., 1996). Several studies were performed to assess the trophic state of the lake and provide management solutions for the existing problem. This chapter highlights two study performed on the lake that provided data to this effort.

#### **Study I**

A stochastic analysis was performed using QUAL2E - UNCAS and existing data to generate probabilistic phosphorus concentrations and Vollenweider eutrophication plots characterizing the trophic state of Lake Tenkiller (Harton, 1989). The hydraulic steady state model was applied in a Monte Carlo simulation to generate stochastic output of existing and planned point source, low flow phosphorus distributions. Historical water quality data plotted as probability density functions (pdf's), are presented in Figure 3-1. These data were randomly accessed using Monte Carlo sampling techniques and combined with modeled data to generate a non steady state phosphorus distribution at the Tahlequah station. Pdf's generated for point source phosphorus concentrations from available statistical data (Table 3-1) along the reaches of Illinois River were used as inputs to the model. Table 3-1 lists the treatment plants considered in this effort and give details on their flow, phosphorus loads and distance from Tenkiller along the river. Figure 3-2 is a schematic of all the point sources evaluated in this study and their river mile distances from Tenkiller. Non-point source distributions were determined from in-stream phosphorus levels and from a stochastic phosphorus loss function developed in this

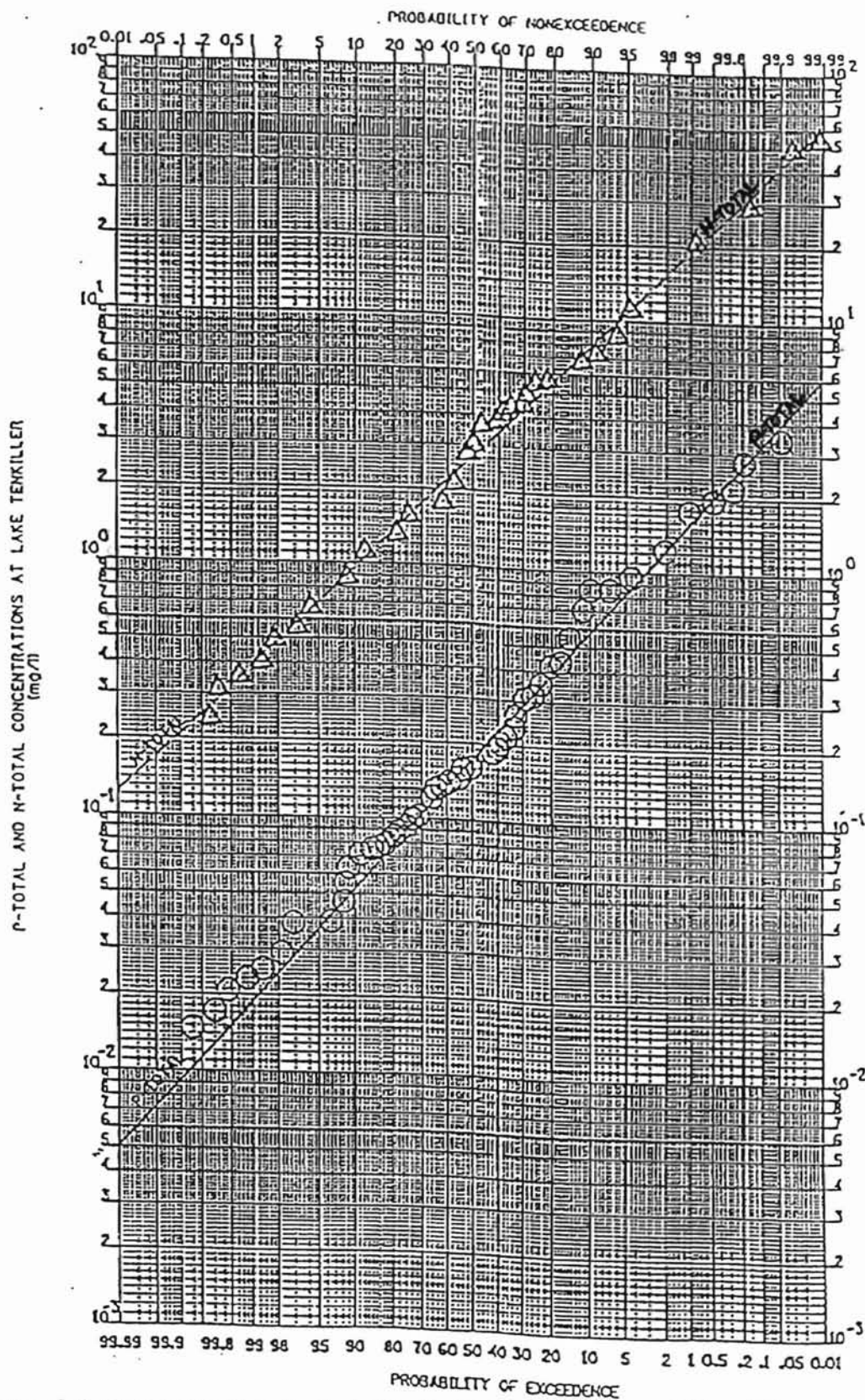


Figure 3-1. Concentration Probability for Total Nitrogen and Total Phosphorus at Tahlequah, Oklahoma Water Quality Gauge (Harton, 1989)

Table 3-1. Summary of Point Source Data at Lake Tenkiller Ferry (Harton, 1989)

Location City	Tributary Receiving Effluent	Plant Flow (MGD)	Phosphorous Loading Parameters						Distance to Tenkiller (miles)	Distance Watts (miles)
			Min.	Max.	Mean (mg/l)	Std. Dev.	95% UCL	95% LCL		
Fayetteville Arkansas	Clear Creek	6	0.00	0.23	0.158	0.054	0.181	0.134	106	47
Rogers Arkansas	Osage Creek	3.5	0.00	11.05	3.829	2.810	5.031	2.627	95	37
Springdale Arkansas	Osage Creek	7	0.00	6.05	3.345	1.467	3.972	2.718	95	37
Siloam Springs Arkansas	Flint Creek	2.4	0.00	4.86	1.609	1.427	2.219	0.999	68	-
Tahlequah Oklahoma	Tahlequah Creek	1.0	0.00	6.58	4.632	1.128	5.110	4.150	20	-
Stillwell Oklahoma	Caney Creek	0.7	0.00	16.54	8.80	5.129	10.990	6.610	35	-

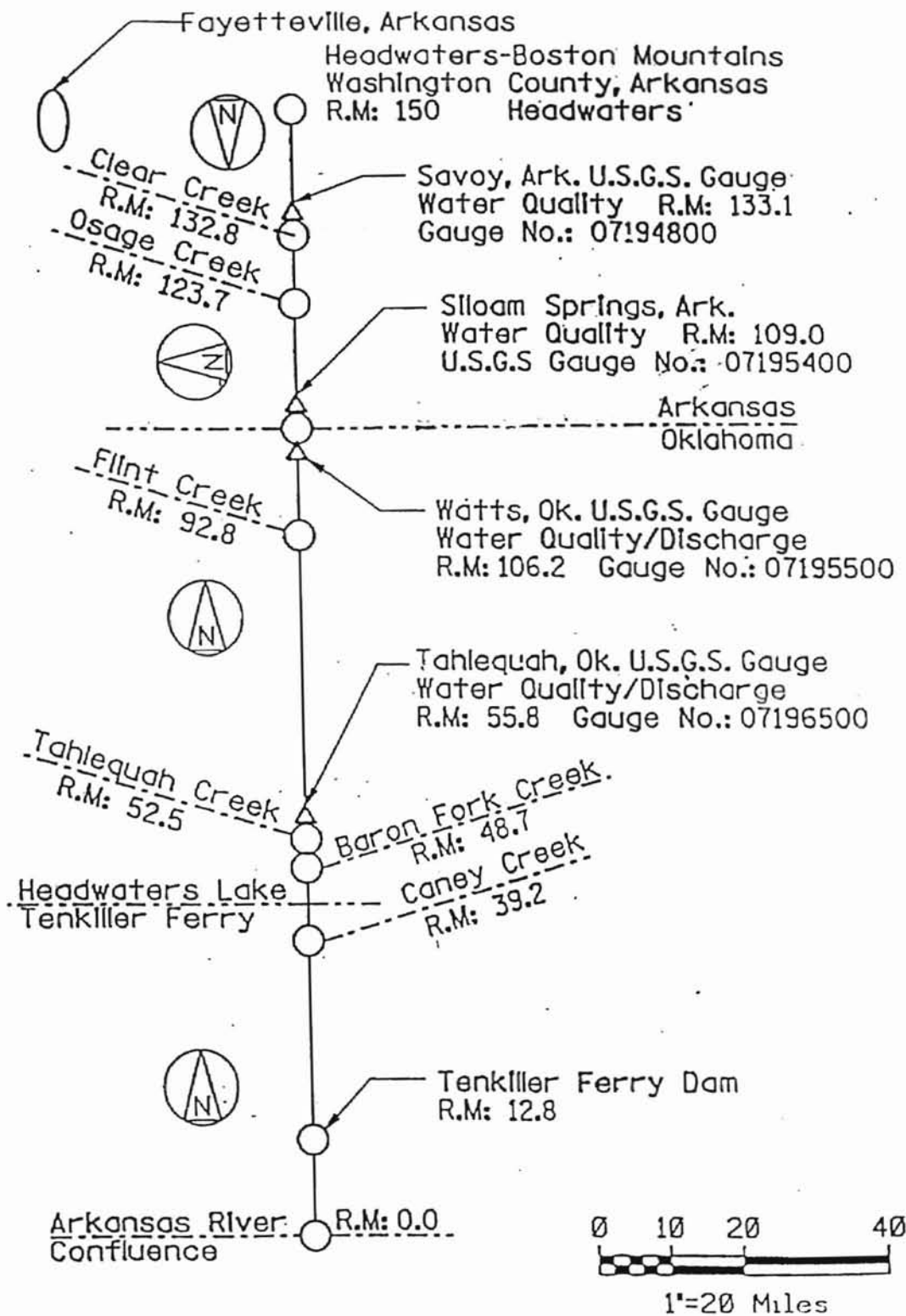


Figure 3-2. Illinois River Schematic (Harton, 1989)

earlier effort. This function is presented in Figure 3-3. The figure shows a variation of 0.0001 mg/l per mile to 0.04 mg/l per mile of phosphorus loss in the river. In-stream phosphorus levels were generated by randomly accessing the channel loss function.

Stochastic outputs generated in the previous effort (Harton, 1989) included:

1) Oklahoma point source contribution, 2) Arkansas point source contribution, 3) Oklahoma non-point source contribution, 4) Arkansas non-point source contribution, 5) total point source contribution and, 6) total non-point source loads of phosphorus at Lake Tenkiller.

Vollenweider methods with randomly selected lake morphometric data were used for generating a Vollenweider distribution for each set of stochastic phosphorus inputs employed. This established a relationship between probable phosphorus inputs and the probable trophic state of the water body. Various management alternatives were subsequently evaluated to suggest methods of stemming the eutrophication of Lake Tenkiller. Some of the management alternatives considered were 50%, 70%, and 90% removal of total non-point source phosphorus loading (Harton, 1989).

The study concluded that point source loading from the individual states did not significantly contribute to the eutrophication levels observed in the reservoir. Among the management alternatives suggested, a 70% to 90% total non-point source phosphorus reduction seemed warranted for effective control of eutrophication at Lake Tenkiller. Figure 3-4 includes pie charts of the effect of various alternatives of reduction in non-point source phosphorus loading on eutrophication of the lake. The pie charts highlight that 70% - 90% reductions in non-point source phosphorus loading are required for any significant impact in controlling eutrophication



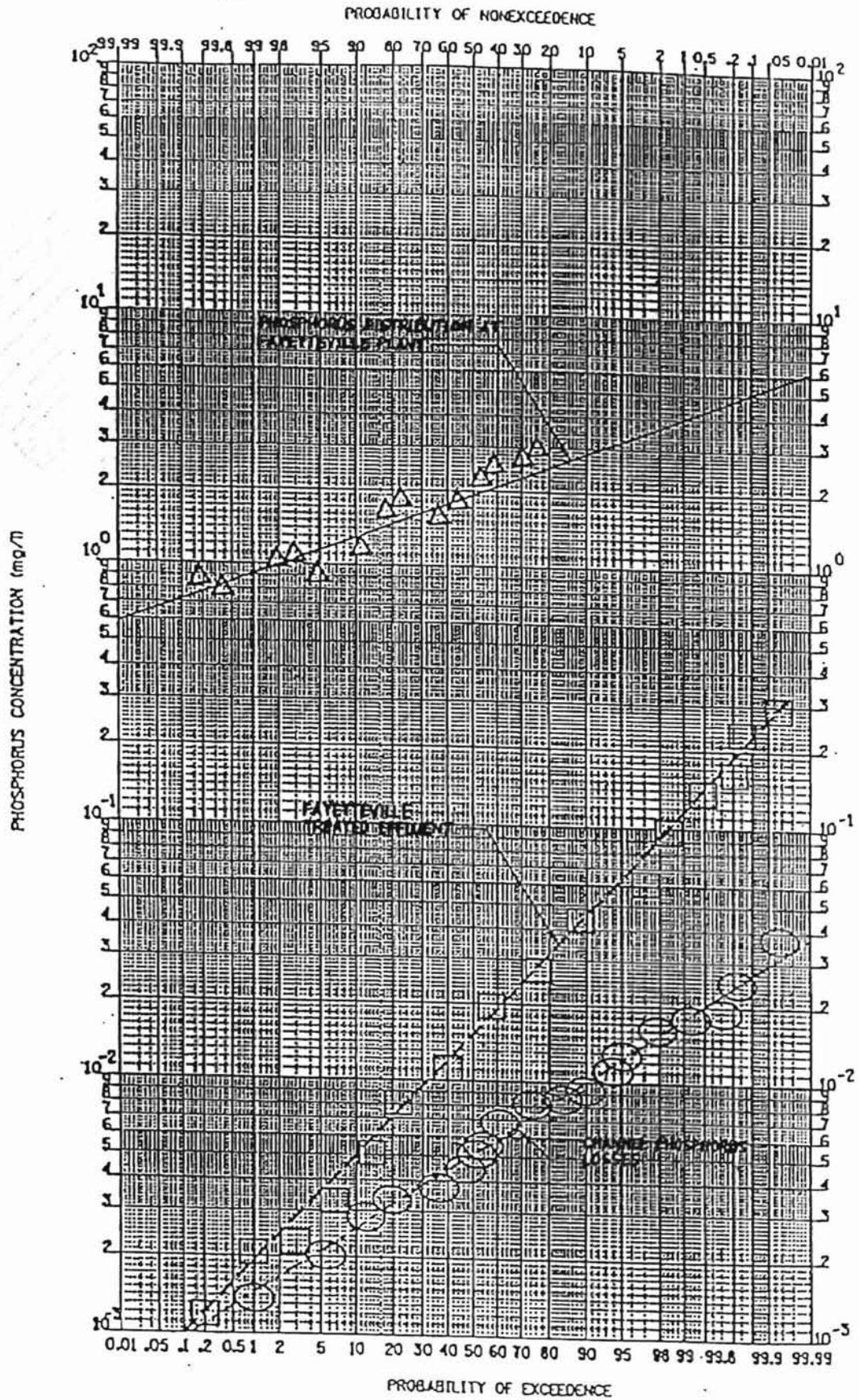


Figure 3-3. Phosphorus Channel Losses Probability Plot (Harton, 1989)

70% REMOVAL ARKANSAS  
NON-POINT SOURCE LOAD



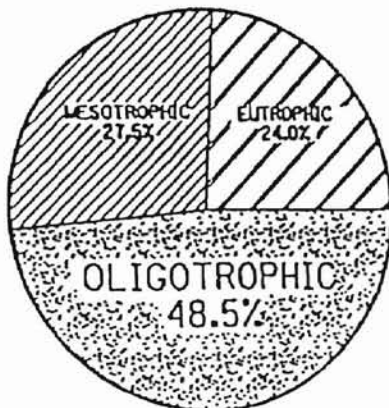
70% REMOVAL ARKANSAS  
TOTAL LOAD



HISTORICAL LOAD  
WITH FAYETTEVILLE LOAD



70% REMOVAL  
TOTAL LOAD



90% REMOVAL  
TOTAL LOAD

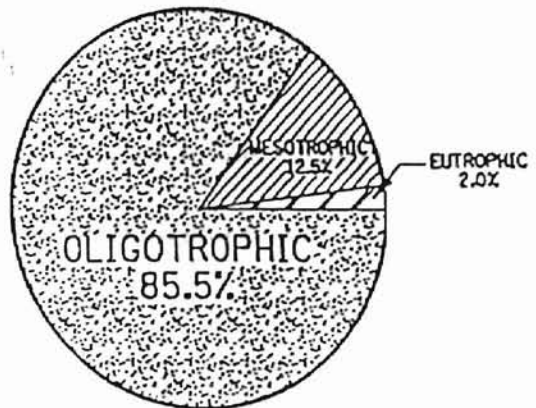


Figure 3-4. Management Alternatives (Harton, 1989)

## Study II

### 3.2. Mean, Maximum and Minimum of Sampled Parameters

Field sampling was performed in riverine, transition and lacustrine zones in the lake to measure phosphorus, nitrogen, chlorophyll *a*, secchi disk and turbidity to determine the trophic state of Lake Tenkiller by OWRB in 1996. The mean, maximum and minimum of each sampled parameter are given in Table 3-2. Some of the sampled data points used later for training neural nets are given in Chapter 5 on materials and methods. The results obtained showed that for chlorophyll *a*, 65.6% of all data points were in the eutrophic range; further, 70.8% of data points in the lacustrine zone and part of the transition zone (stations 4, 5, 6, and 7, Figure 2-1) were in the eutrophic range. Similarly, the percentages of values with secchi disk observations in the eutrophic range were 79.2% and 65.2% respectively in these same zones. These observations confirmed the highly eutrophic state of the lake (OWRB et al., 1996).

In this effort to determine the limiting nutrients, total nitrogen to total phosphorus ratios were correlated to the eutrophication state measured by chlorophyll *a* and secchi disk values. Vollenweider's eutrophication index was also developed in this study for phosphorus loads, to correlate phosphorus concentrations with the trophic state of the lake.

In the riverine zone (stations 1 and 2, Figure 2-1) the study concluded a possible nitrogen limitation to eutrophication (OWRB et al., 1996). High phosphorus concentrations but low chlorophyll *a* concentrations supported the above hypothesis. The planktonic growth in the upper part of the transition zone (station 3, Figure 2-1) was considered to be dual nutrient limited, where both phosphorus and nitrogen controlled the rate of planktonic growth. The lower part of the transition zone (station 4, Figure 2-1) and

**Table 3-2. Mean, Maximum and Minimum of Sampled Parameters**

	<b>Phosphorus (mg/l)</b>	<b>Chlorophyll a (µg/l)</b>	<b>Secchi (m)</b>
Mean	0.073	19.5	1.44
Maximum	0.343	47.7	5.50
Minimum	0.007	0.8	0.20

*data set conforms to the period Oct. 1992 to Sep. 1993  
(OWRB et al., 1996)*

the lacustrine zone (stations 5, 6 and 7, Figure 2-1) were concluded by the OWRB investigations to be phosphorus limited in terms of algal growth. As with the Harton study this effort concluded that there was no significant impact from point sources and approximately 70% - 80 % reduction in total phosphorus loads was required for control of eutrophication of Lake Tenkiller. Table 3-3 lists phosphorus loading values from all the point sources considered in the OWRB study. A total of 12 point source discharges was considered in this effort.

Besides providing a barometer to check the performance of the current study, the data gathered in the two studies were used for developing neural network models used in the present study. Data used in the present study included:

- 1) log-normal distributions developed as probability density functions for historical water quality data for the period 1977 - 1985 (Figure 3-1),
- 2) statistical data on point source flows and phosphorus concentrations for wastewater treatment plants along the reaches of the Illinois River (Table 3-1),
- 3) log-normal distributions developed as probability density function for channel mile loss functions of phosphorus (Figure 3-3),
- 4) field sampling data for the period of 1992 - 1993 to establish correlations between phosphorus and chlorophyll *a* and phosphorus and secchi disk values, and
- 5) point source loading data of phosphorus from wastewater treatment plants along reaches of the Illinois River (Table 3-3).

Table 3-3. Estimates of Phosphorus Discharges from Point Sources for the Period 1991 to 1993 (OWRB et al., 1996)

Discharger	Estimated Load at Source (kg P/yr)	Distance to Horseshoe Bend (mi)	Estimated Corrected Load at Horseshoe Bend			Estimated Annual Total Load (kg P/yr)
			Low Flow (kg P/yr)	Medium Flow (kg P/yr)	High Flow (kg P/yr)	
Prairie Grove	1200	100	19	28	23	70
Rogers	21600	99	355	519	417	1292
Fayetteville	4500	97	80	114	90	283
Springdale	43150	95	820	1150	893	2862
Lincoln	1200	81	38	46	31	115
Gentry	1700	68	85	91	56	232
Siloam Springs	10000	62	623	628	362	1614
Watts	500	62	31	31	18	81
Westville	2900	28	615	441	187	1243
Midwestern Nursery	600	14	211	131	49	391
Tahlequah	4700	6	2200	1267	441	3908
Cherokee Nation	530	5	257	147	51	454
Total	92580		5335	4593	2619	12547

#### 4.0 Introduction

Two primary strengths of the human brain are “massive interconnection and parallel processing architecture. Neural networks are an alternative computational approach based on such theories of the human brain and intelligence”, (Rogers et al., 1995). Similar to the human brain, artificial neural networks rely on interconnections in their structure. Parallel processing allows the network to attain high computational speeds because it is able to process the information through various interconnections at the same time like the human brain. Besides the structure, the similarity between the human brain and neural networks is further extended to their method of functioning. Both attempt to learn from a phenomenon and adapt their behavior based on their learning (Rogers et al., 1995). Learning is viewed as the establishment of new connections between neurons or the modifications of existing connections (Hagan et al., 1996). The neurons that are considered for neural network models are developed by computer algorithms and are not biological. They are iterative algorithms that are used to develop a combination of linear and non-linear equations for modeling real world situations. These computer-based models are called artificial neural networks or neural networks, as both terms are used interchangeably.

#### 4.1 Artificial Neural Network

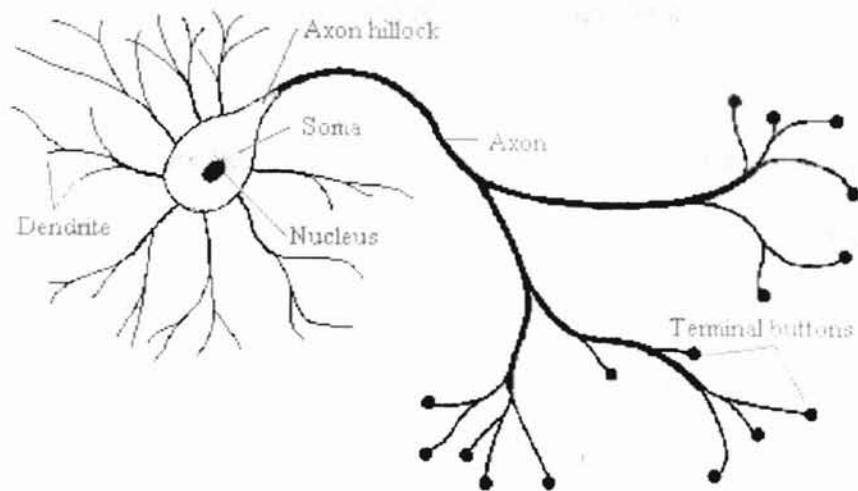
Artificial neural networks (ANNs) are morphologically very different from networks formed by biological neurons, but have some functional similarities to them. This section briefly describes those characteristics of a brain's functionality that have

inspired the development of artificial neural networks and presents a mathematical viewpoint of ANNs.

The brain is composed of a large number (approximately  $10^{11}$ ) of highly interconnected elements (approximately  $10^4$  connections per element) called neurons (Hagan et al., 1996). These neurons have three principal components: the dendrites, the cell body (soma and nucleus) and the axon as shown in Figure 4-1. Each of the components has a separate function. The dendrites (nerve fibers) behave as receptors carrying information in the form of electrical signals to the cell body which processes the information received. The cell body massages the data through incoming signals and passes the output on to the axon. The axon provides a communication route between cells. It can be thought of as a one-way multilane information highway where response from one cell is communicated to various others. The axon carries the output from a cell and passes it on to receivers or dendrons of other cells over a bridge called a synaptic junction. The communication over synaptic junctions is a complex chemical process and depends on the strength of incoming and outgoing signals. The arrangement of neurons and the complex chemical processes of communication over synaptic junctions establish the functioning of biological neural networks (Hagan et al., 1996).

Artificial neural networks are less intricate than the human brain. ANNs come close to emulating their biological counterparts because of two key characteristics. Both biological and artificial neural networks use computational devices to process input signals and develop outputs. The computational devices are highly interconnected and thereby able to model and understand complex situations. Secondly, the behavior of synaptic junctions (interconnection between neurons) in processing the information





**Figure 4-1. Schematic Drawing of Biological Neuron**  
(<http://vv.carleton.ca/~neil/neural/neuron-a.html>)

determines the functionality of both structures (Hagan et al., 1996). The strength of synaptic junctions are called weights in ANN terminology and “the computational power of the neural network lies in the interconnection weights that designate the strength of a node to produce the output at the node to which it is connected” (Basheer et al., 1995).

From a mathematical viewpoint it may be helpful to think of artificial neural networks (ANNs) as “nonparametric, nonlinear regression techniques” (Rogers et al., 1995). As opposed to traditional data analysis techniques, where a model is initially selected and then appropriate data are applied, an ANN builds the best combination of linear and nonlinear functions to which the data are fitted. This is possible with neural nets because of their ability to learn and then apply this learning in a generalized sense to similar situations. In this adaptive learning approach the net undergoes the training process and learns the significance of all data values, which includes peaks and plateaus. A neural network not only assigns a significance (or weight) to the magnitude of each point in the data, but also identifies connections and defines weights to establish useful relationships among all data points. As more operations are executed with a neural network model the base for making decisions increases. Consequently, the precision with which the network can make predictions increases.

#### **4.2 Artificial Neural Network Structure**

There are several ways of aligning the neurons and connections between the layers of an artificial neural network that can result in several network models. The starting point for most network models is a model neuron as presented in Figure 4-2. This neuron consists of multiple inputs  $p_i$ , a hidden layer  $f$  and a single output  $a$ . As is explained in the text by Hagan et al. (1996), each input  $p_i$  is modified by a *weight*  $w$  and a

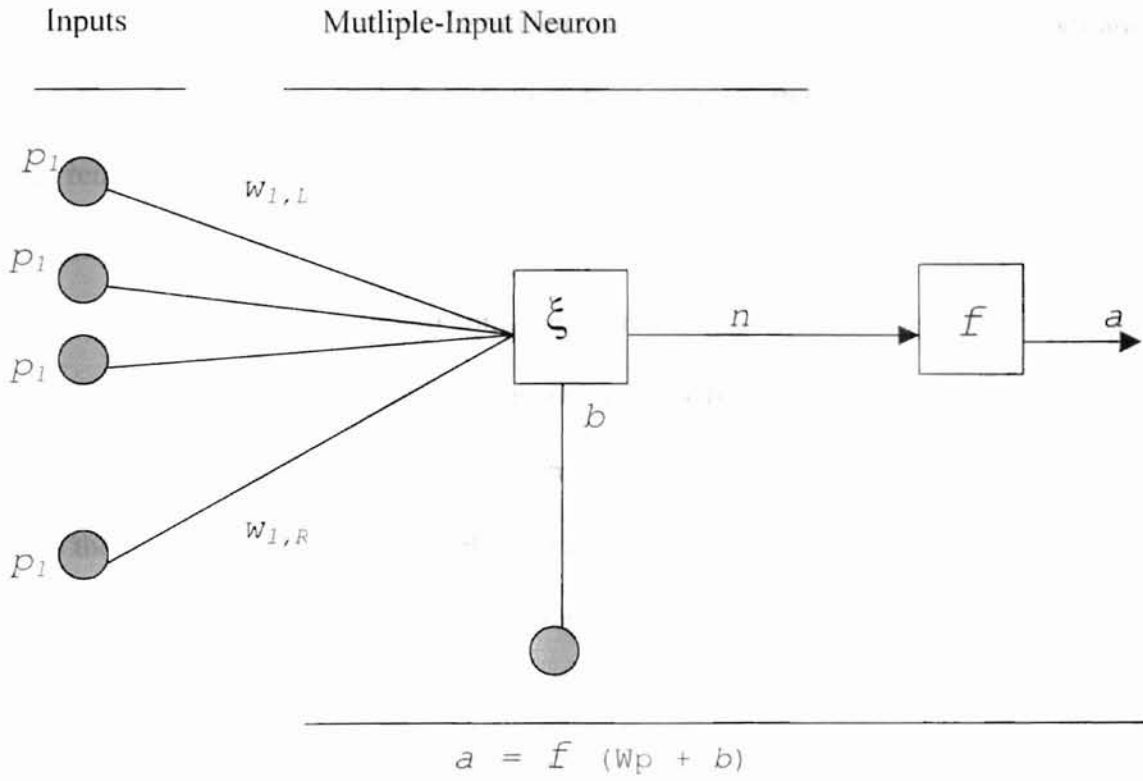


Figure 4-2. Multiple-Input Neuron (Hagan et. al., 1996)

*bias*  $b$  before being passed to the summer ( $\xi$ ). The summer adds all the modified inputs to generate a single input value for the hidden layer. The weight  $w$  signifies the importance of every input  $p_i$  with respect to the output. The bias gives the network an extra variable that removes the possibility of null inputs that provide no training. The output from the summer becomes the *net input*  $n$  for the *transfer function*  $f$  or the “activation function” in the hidden layer. The neuron will combine all the weighted inputs and, with reference to a *threshold value* (specified by the user) and *transfer function*, use them to produce the scalar neuron output  $a$ . In environmental modeling terminology this scalar output neuron,  $a$ , is the output answer and the combined structure of the inputs, weights, biases, summer and transfer functions are the model elements. These take the place of the systems of solved partial differential equations more typically used in simulation modeling.

There are a large number of *transfer functions* that have been employed in neural network architectures (Hagan et al., 1996). Transfer functions play a very important role in generating accurate neural predictions. They control the outputs and functioning of the network (Hagan et al., 1996). Changing the neuron *transfer function* will change the nature and characteristic behavior of the network. Some commonly used *transfer functions* are: the step, the gaussian and the sigmoid, among others (Cheshire Engineering Corporation, 1996). The nature and type of transfer function to be selected for a particular data set come from experience, although there are some broad guidelines available.

### 4.2.1 Neural Network Model

There is a large number of existing neural network models with different learning algorithms (Rogers et al., 1995). Several distinct neural models can be built based on the connectivity they share between input layers, hidden layers and output layers (Rogers et al., 1995). The following discussion deals with multilayer feed forward networks with a back propagation learning algorithm, which is routinely employed in neural network modeling and has been used in this study.

#### Multi-layer, feed forward neural networks

These models consist of a multiple layer of neurons where information is passed through the network in one direction from the inputs to the outputs, in a "forward feed". The model starts with the input layers and traverses the hidden layers towards the output layers. Neurons of each layer are connected only to neurons of a subsequent layer as shown in Figure 4-3. Between an input layer receiving external inputs  $x_i$  and an output layer generating outputs  $y_n$ , there are several hidden layers. The hidden layers of neurons in this type of network model allow the neural network to develop its own internal representation of input to output. The non-linear behavior of neurons allows the network model to learn many different types of input-output relationships, thereby addressing the complexity of the underlying data sets. It is observed that multilayer networks are more powerful than single layer networks (Hagan et al., 1996). A multilayer network with a sigmoidal transfer function for the first layer and a linear transfer function for the second layer can simulate most functions reasonably well, as opposed to single layer networks (Hagan et al., 1996). The back propagation learning algorithm allows the network to upgrade and optimize its network connections before the next input forward feed. This

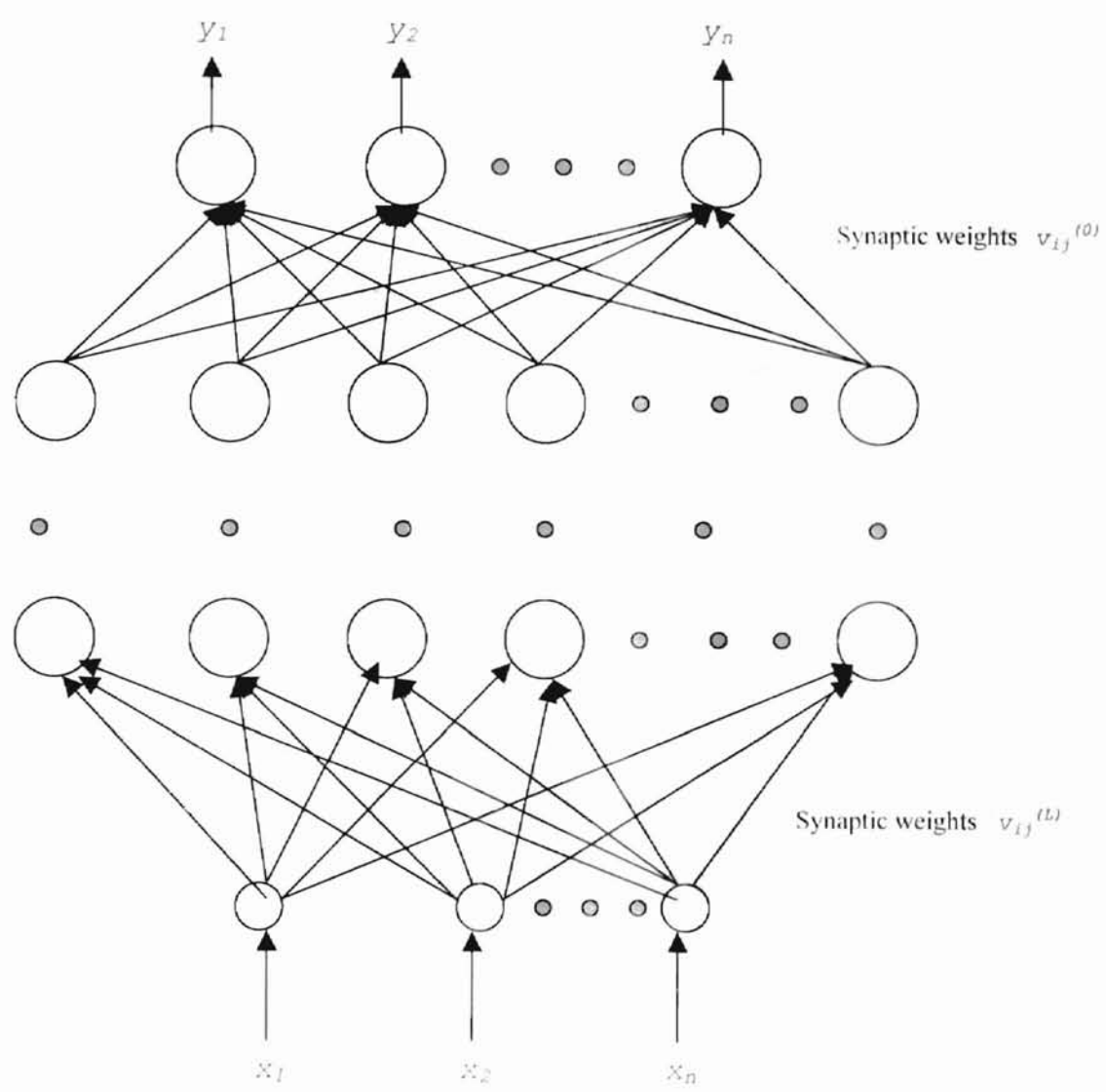


Figure 4-3. Multi-Layer Feed Forward Model (Karayiannis et. al., 1993)

process has become a standard in training the multilayer, feed forward neural network (Rogers et al., 1995). The back propagation network and the arithmetic behind back propagation is elaborated in the following sections.

### 4.3 Back Propagation Network

A back propagation network has a number of neurons arrayed to form a layer as presented in Figure 4-4. As shown in the figure, each layer has all its inputs connected to either a preceding layer or inputs from the external world, but not within the same layer. Next, multiple layers are then arrayed one succeeding the other so that there is an input layer, multiple intermediate layers and finally an output layer. Layers that have no connection to the external world are called *hidden layers*. Input layers are connected to the external world and are therefore distributors of incoming signals from the outside world. The hidden layers process the incoming signals as they categorize them by their individual signatures. Output layers collect all the signals and make an appropriate response to the input signals. Input, hidden and output layers are also called “distributors, categorizers and collectors”, respectively (Cheshire Engineering Corporation, 1996).

### 4.4 Neural Network Operation

The output of each neuron is a function of its inputs as shown in Figure 4-5. This figure is a schematic presentation of the manner in which outputs are generated, the errors initially calculated, and are then applied to modify the network weights. In particular, the output of the  $j$ th neuron in any layer is described by two sets of equations:

$$U_j = \sum (X_i * w_{ij}), \text{ and}$$

$$Y_j = F_{th}(U_j + t_j)$$

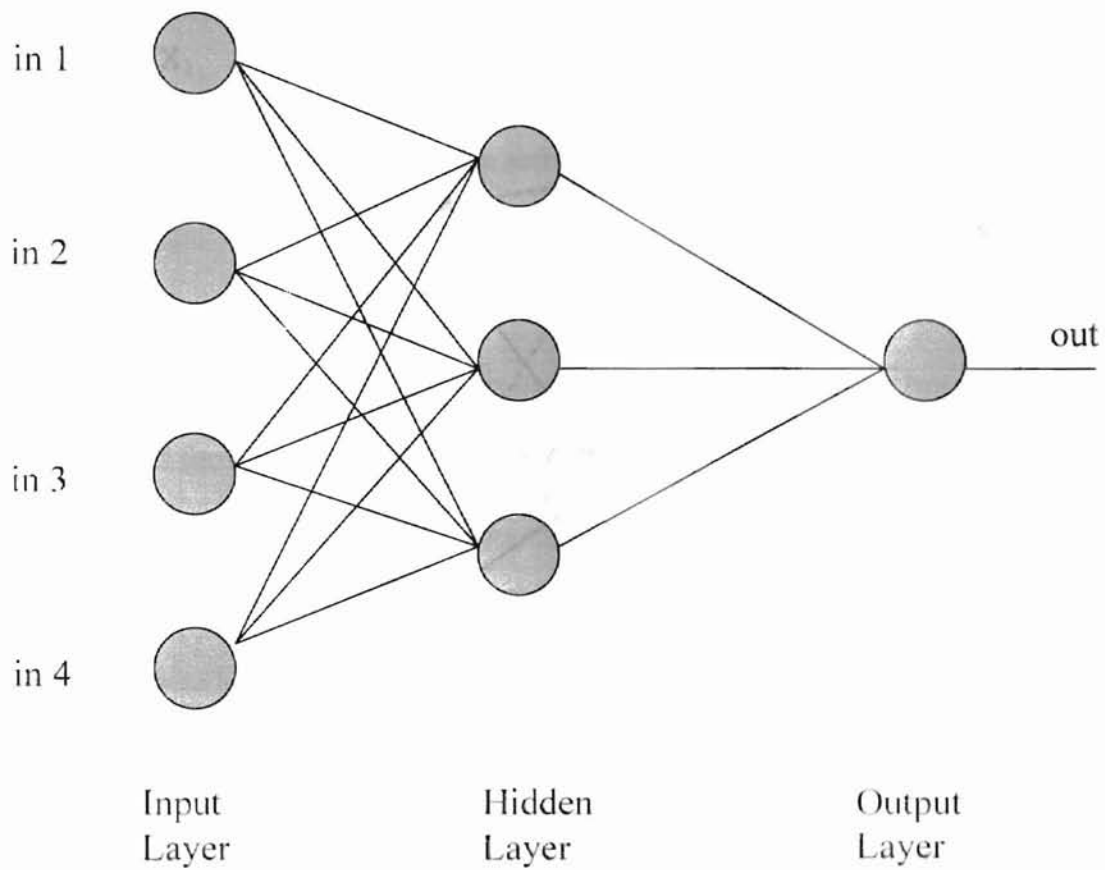


Figure 4-4. Backpropagation Network (Cheshire Engineering Manual, 1996)



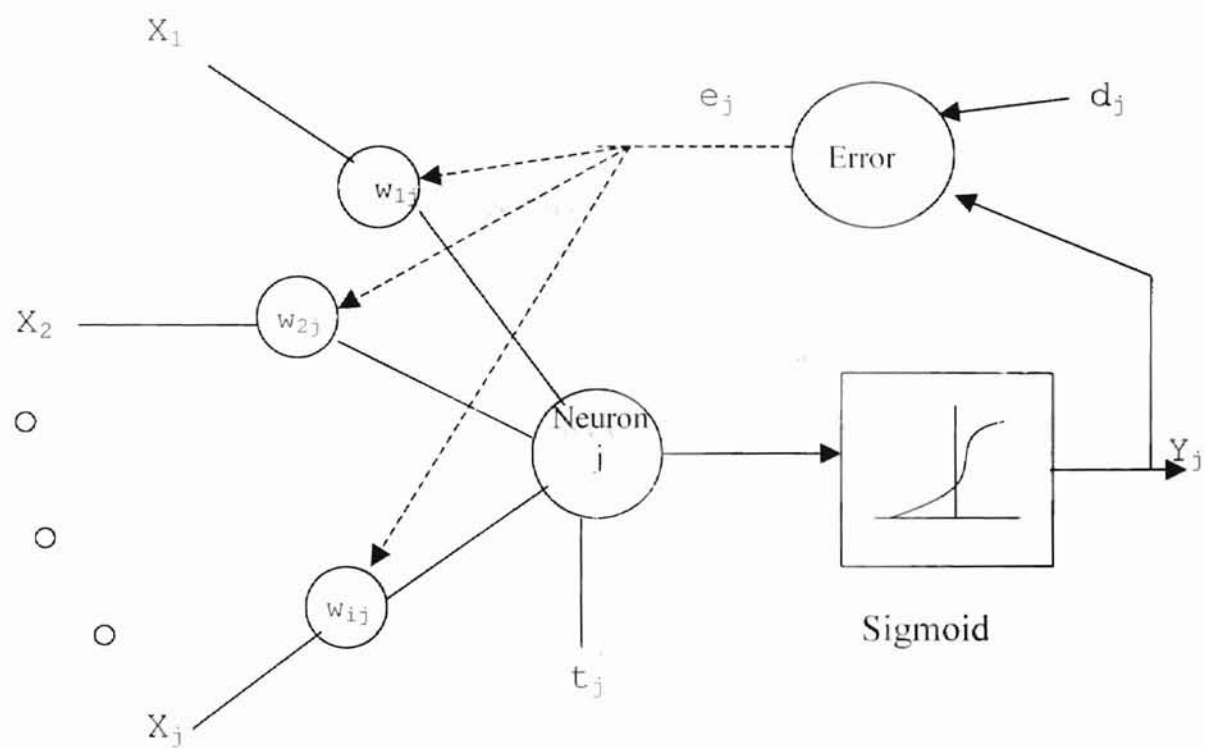


Figure 4-5. Schematic of Network Operation (Cheshire Engineering Manual, 1996)

Every neuron  $j$ , takes inputs from all the  $i$  neurons connected to it. Each input  $X_i$ , from the input layer is multiplied to a weight,  $w_{ij}$ . The weights are constantly updated after being randomly assigned initially. All weighted inputs are summed together, resulting in the internal value,  $U_j$ , or net input  $n$ . Each resulting internal value from all  $j$  neurons are biased by a value,  $t_j$ , and passed to an activation function,  $F_m$ . The output from the activation function,  $Y_j$ , is the neural net response for the given input.

This operation is repeated several times until the network can produce outputs within a user-specified tolerance. This occurs when the network reaches a plateau in its learning and further training runs do not improve its performance. The process of repeating the entire sequence of events explained earlier to update the value of weights is known as training the network. The accuracy of predictions made by a network depends on its ability to learn during the training process and can be considered the critical stage in the development of the neural network model.

#### 4.5 Learning

Learning or training is a process by which the network becomes familiar with a data set in detail, before responding to other, similar situations. The driving factor in neural network *training* is the error computed as the difference between the desired and the actual response. The error is the sum of the squares of the differences between actual and desired outputs in each of the nodes in the output layer and can be expressed as a function of the connection weights. A portion of the root-mean-square of the error (RMS error) is passed back through the hidden layers of the network to the input so that the connection weights on all previous neurons can be altered in such a manner as to minimize the quadratic error between desired and actual outputs (Hagan et al., 1996). For

a multilayer feed forward network with back propagation training algorithm the learning is achieved in three phases:

- Forward propagation or forward feed - input training set, proceed through network from layer to layer applying weights, calculate output.
- Calculate total error.
- Back propagation - error is passed back through net causing each connection weight to be modified.

The sequence is repeated with next training set and the network is said to have converged when the error is less than a specified tolerance value as defined:

$$e(W) = \sum_{i=1}^m (y_i - a_i) \quad \text{where } m = \text{number of inputs,}$$

where  $e(W)$  is the current error defined as a function of the connection weights,  $y$  is the desired output and  $a$  is the actual output.  $W$  is a matrix of connection weights (Dowd and Sarac, 1994). The input to each node consists of the weighted sums of outputs from the previous layers as shown in Figure 4-6.

Notation describing the sequence of events to develop the network shown in Figure 4-6 and complete the above chronology is:

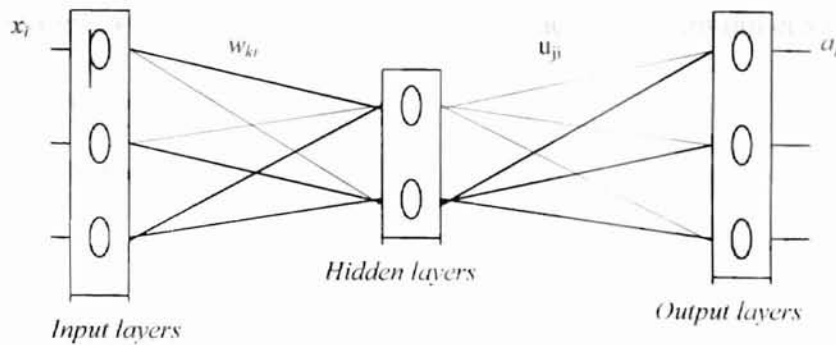
(a) Forward Propagation

When  $x$  is presented for the  $t^{\text{th}}$  time:

In the hidden layer:

$$\text{inputs are } U_k = \sum_i (w_{ki}x_i + t_k) \quad \forall k \in S_H \text{ (over all hidden neurons)}$$

$$\text{outputs are } Y_k = F_{th}(U_k) \quad \forall k \in S_H$$



**Figure 4-6. Feed Forward Neural Net (Dowla et al., 1996).**

$x_i$  is the external input to the first or input layer. The external input is modified by a weight  $w_{ki}$  and bias  $t_k$  to generate the input  $U_k$  for the hidden layer. The output  $Y_k$  from the hidden layer is generated by an activation function  $F_{th}(U_k)$ .

In the output layer:

$$\text{inputs are } U_j = \sum_k (u_{jk} Y_k + t_j) \quad \forall j \in S_o \text{ (over all the output neurons)}$$

$$\text{outputs are } a_j = F_{th}(U_j) \quad \forall j \in S_o$$

$U_j$  is the input generated for the output layer by modifying the output from hidden layer  $Y_k$  by a weight factor  $u_{jk}$  and a bias  $t_j$ . The output from the network is generated by the activation function  $F_{th}(U_j)$  (Dowd and Sarac, 1994).

*(b) Back Propagation*

The network works backward from the output layer to the input layer and modifies the weights according to the gradient descent rule (Dowd and Sarac, 1994) which updates weights as a function of the subsequent outputs. In the gradient descent method, the weights are updated proportionally to the first derivative of the error  $e(W)$  with respect to the weights evaluated at the current value. The first derivative of the error gives the direction of increase on the error surface. A direction opposite to the derivative

is chosen to decrease the error in order to reach the minimum error in repeated runs with the network.

In the output layer:

$$u_{jk}(t) = u_{jk}(t-1) - g(t) \cdot d_j \cdot a_k \quad \forall j \in S_o, \forall k \in S_H$$

where

$$d_j = 2(a_j - Y_j) F_{th}'(U_j) \quad \forall j \in S_o$$

the weight in the output layer  $u_{jk}$  is modified by an error term  $d_j$  (derivative of the error  $e(W)$ ), a gain term  $g(t)$  and the output from the hidden layer  $Y_k$ .

In the hidden layer:

$$w_{ki}(t) = w_{ki}(t-1) - g(t) \cdot d_k \cdot x_i \quad \forall k \in S_H, \forall i \in S_I$$

where

$$d_k = F_{th}'(U_k) \sum d_j W_{kj} \quad \forall k \in S_H$$

$$W_{kj} = \text{matrix of weights } w_{kj}$$

$$S_I = \text{number of inputs}$$

These weights are updated using steepest descent techniques (Dowd and Sarac, 1994) which allow for the quickest convergence of the error surface to reach the minima by specifying a direction of descent opposite to that generated by the derivative. The connection weights between nodes  $i$  and  $j$  in the current iteration ( $t$ ) is found by subtracting a gain term ( $g(t)$ ) which controls the rate of descent from the previous iteration ( $t-1$ ). The gain term defines a multiple of the error correction term that the user applies to update the current weight in successive iteration through the network. It changes over time to allow for changes in the rate of descent and depends on the error calculation for each simulation or local error. "The goal is to decrease (descend) the error

function, avoiding the local minima and reaching the actual or global minimum” (Kartalopoulos, 1996).

There are many types of neural network *learning rules*. Four categories of network learning that are discussed below are: *supervised learning*, *unsupervised learning*, *reinforcement* (or *graded*) learning, and *competitive learning*. The type used depends upon the type and quality of the data as well as overall objectives of the modeling exercise.

### Supervised Learning

In *supervised learning*, the network is trained to provide outputs with an example or collected data set (the *training set*) (Hagan et al., 1996). There is a *target* output specified to which the network is trained. The output of the network is compared to the *target* outputs and the *weights* and *biases* are modified to move the network outputs closer to the targets (Hagan et al., 1996). Once the network is able to make predictions close to *target* outputs within specified margins of prediction error the network is considered trained.

### Unsupervised Learning

In *unsupervised learning*, there is no example data set or target output. The network works with the *input* set of data and *weights* and *biases* are modified based on the *inputs* only. In unsupervised learning a clustering operation takes place where the net learns about outputs by clustering the inputs into categories for recognition. “They learn to categorize the input patterns into a finite number of classes” (Hagan et al., 1996).

### Reinforcement Learning

This is similar to *supervised learning*, except that, instead of being provided a *target* output, the network is given a grade of pass/fail. This type of learning requires at least one neuron in the output layer and a teacher (Kartalopoulos, 1996). Every output is graded by the teacher which generates a binary error signal of pass/fail. The grade determines the performance of the network for a given set of inputs. Every iteration modifies the network algorithm to improve the grade or performance score.

### Competitive Learning

In this scheme, several neurons are at the output layer. When an input stimulus is applied, each output neuron is compared with the others to produce the closest output signal to the target (Kartalopoulos, 1996). This output then becomes the dominant one, and the other output neurons become dormant for that particular stimulus. A different stimulus leads to some other output neuron in the dominant role. Each output neuron in this way is trained to respond to a particular stimulus.

The type of learning employed affects the rate at which the net learns. Learning rate determines the speed at which a neural network is trained (Rogers et al., 1995). It is one of the parameters that indicate the success or failure of a neural network. There are a number of reasons why slow learning may occur. The gradient descent technique used for updating weights in back propagation networks consumes significant time especially when the gradient is small (Rogers et al., 1995). Learning is slow with large data sets because the number of function evaluations, the most time consuming operation, is very large. A large data set, however, is still desirable as it may contain more complexity allowing for a more accurate model (Rogers et al., 1995).

↘ The number of interconnections in the network model influences the learning rate. Complicated networks with large number of layers or a large number of neurons per layer increases the training time of the network. Finally, “contradictory” data sets might stall the learning process with a stagnant root-mean-square (RMS) error. Deciding on the optimal setting of parameters is important in getting the best and quickest results from a neural network. An optimal set of parameters can be achieved through trial and error, which may be time consuming. “Genetic algorithms.” (GA) have been developed to eliminate this trial and error process (Yip and Pao, 1995).

#### **4.6 Genetic Algorithm**

The genetic algorithm attempts to build the optimum network by selecting the best alternative from a pool of alternatives. By optimizing the input and training parameters, subsequent usage of the model can be performed with a minimum amount of training. This reduces the computational load and model development time.

Genetic algorithms evaluate network performance over the entire set of data and record the parameter values that generated the best results. These values are continuously updated as better results are obtained by varying the parameters. The process of parameter value upgrade also involves addition and removal of interconnections with changes in the previously recorded parameter values. In this way the algorithm builds better networks by weaning out the weak links in a model as it evolves. The analogy can be drawn from the evolution of ape to man, which presumably, was a process of selective evolution to create a better model of an ape by a process of continuous upgrade (Yip and Pao, 1995).



#### 4.7 Neural Network and GA Nomenclature

This section deals with the terms commonly used in the field of neural networks. The section provides a description of the significance of terms and their role in network development.

Transfer Function: simulates the process of biological activation or firing of synaptic junctions in a biological neuron, as signals are processed and responses generated in the biological context. The shape and differentiability of transfer functions is important in deciding their usability. Transfer functions in artificial neural networks are basically categorized in two ways: differentiable (Figure 4-7) and non-differentiable (Figure 4-8). A differentiable function is essential for use in back propagation networks. The reason is that for modification of the interconnection weights as a network undergoes training, the error passed back for upgrading the weights is a derivative of the difference between target output and network output. Since network output is the output from the transfer function, it has to be differentiable for its derivative to exist.

The other differences are attributed to the shape of the functional curve, and though there are no set rules for selecting a particular differentiable function; one can decide between a differentiable and a non-differentiable function. Non-differentiable functions are used in boolean situations where the outputs are in terms of zero and one. Though differentiable functions can also be used in such situations, their accuracy is not as high as non-differentiable functions because of the range of outputs which are not close to maximums for inputs close to zero as is illustrated in Figure 4-7. Therefore, non-differentiable functions are able to perform much better as switch functions, while differentiable functions perform much better as continuously discriminating functions.

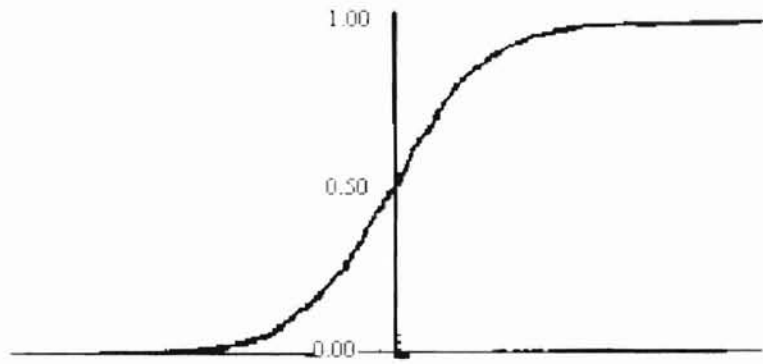
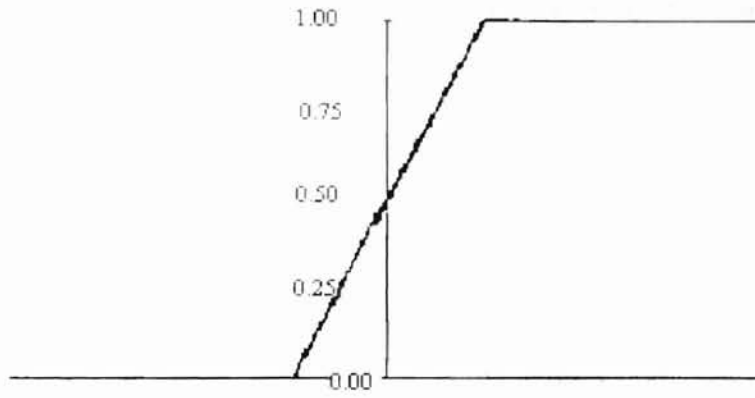


Figure 4-7. Differentiable Transfer Function: *Sigmoid Function*  
(Cheshire Engineering Manual)



**Figure 4-8. Non-Differentiable Transfer Function: *Step Function***  
(Cheshire Engineering Manual, 1996)

Between differentiable functions there is a lack of set rules identifying suitability of a particular differentiable function. Sometimes one function fits a particular set of data better than the other, therefore, one has to be careful in deciding on a transfer function.

Epoch: is one complete simulation over all data points. During network training it is one complete processing run through all defined training cases. Similarly, during network testing it is one complete processing run over all test data points.

Learning Rate (LR): determines the amount of weight adjustment to be made based on the error passed back. If the factor is set to a large value, then the neural network may learn faster, but a large variability in the input data may lead to restricted learning or no learning at all for the neural net.

Momentum: allows a change to the weights to persist for a number of adjustment cycles. It determines the existence time of a correction term, that is, determines the number of epochs for which a correction term will have a continued influence. Momentum must be greater than or equal to 0 but less than 1. Values of momentum closer to one result in greater influence of previously applied correction terms. Values of momentum closer to 0 result in modification of weights by current correction term only. Momentum helps in taking care of the peaks in the data set, which could be aberrant or faulty data. By applying a momentum term the model ensures that a peak data value does not unduly influence the training of the network.

Input Noise: imparts a slight variation to each input value for every training epoch. By providing a different set of inputs to each training epoch we ensure that the learning process remains dynamic and the network does not memorize any situation. This prevents the network from overtraining (increase in error after the network has already

achieved the lowest possible error). Input noise should be set between 0 and 1, meaning 0% to 100% of the input range will be set as a noise range. For example, if the input ranges from 10 to 20 and the input noise is set to 0.1, then random values ranging from 1 to 2, will be added to or subtracted from each input value.

Training Tolerance: provides a basis for stopping the training process. When the training output falls within the target output, plus or minus the training tolerance, the network stops training on that set of input and scores the output as "Right". Training tolerance should be between 0 and 1. For example, if values in the target columns ranged from 100 to 300, a tolerance of 0.2 corresponding to 20%, would allow errors of 20% of 200 or 40 (i.e.,  $0.2(300-100) = 40$ ). The training tolerance value has no effect on the learning algorithm. However, when the network finds 100% "Right", as defined by the training tolerance, it will automatically stop training.

Testing Tolerance: it is similar to Training Tolerance, but it is applied only to the test data. Testing Tolerance is often set to a less stringent value compared to training tolerance since it is assumed that prediction is more difficult than training to target values (Cheshire Engineering Corporation, 1996).

Epoch per Update: controls the number of epochs between weight modifications. To save time the network can be set to update its interconnection weights only after training for a certain number of epochs. This saves computational time while training the network.

Epoch Limit: sets the maximum number of training epochs a neural network will undergo while training.

Time Limit: sets a time limit to the neural network training process.

Error Limit: sets a maximum increase allowed in the error value over the lowest achieved value earlier. Under normal circumstances a neural network steadily reduces the training error (RMS error) as it undergoes training. This general trend can be offset and the RMS error can increase under any of the following conditions:

1. if the neural network begins overtraining, or,
2. if the neural network has insufficient capacity, or,
3. if the network is inappropriately configured.

Genetic algorithms support their own terminology. The following section deals with the terms used in the application of genetic supervisor in this study.

The number of desired iterations and the method of new *structure* (network architecture) generations were controlled by specifying two parameters: *population pool size* and the *population pool mode* control. The *population pool size* controlled the number of *structures* created in each generation count (parameter used to control the number of times networks with specified pool size were to be evaluated). Setting a large value for this parameter evaluates a larger number of possible *structures* for optimum network configuration but takes greater computational time. The *population pool mode* specified one of the three modes or methods by which new structures were generated during genetic simulation. The three modes considered were 1) *closed pool*, 2) *immigration*, and 3) *emigration*. These modes specify the replacement strategy used for generating new population (or *structures*) of networks.

Closed Pool: existing population pool evolved with no addition of new *structures* except through crossbreeding and mutation. Crossbreeding and mutation are mechanisms for developing genetic networks and are discussed later.

Immigration: in each new generation entirely new *structures* will replace the weakest *structures* of the current pool with remaining *structures* being crossbred and mutated.

Emigration: the best members of the current pool are emigrated to an entirely new population and crossbred with the new population.

During simulation time each *population pool size* is evaluated and each new *structure* generated is ranked by its fitness. These *structures* then evolve to produce new generations with the most fit *structures* having a greater probability of passing on their characteristics (or *features*) to new generations. There are two mechanisms by which *features* are passed to successive generations: *crossbreeding* and *mutation*.

Cross-breeding: is determined by a crossover setting which determines the frequency of intermingling of *features* on the same string to generate new *structures*.

Mutation: *structures* and *features* are chosen at random to be changed to new values. The mutation rate parameter controls the percentage of *structures* that will undergo a mutation.

To determine the best fit *structure* there are four options: *Train error*, *Train epoch*, *Test error*, and *Test epoch*. *Train error* finds the *structure* with least RMS error when run on training data. *Train epoch* finds the *structure* that requires the least number of epochs to achieve the error specified in fitness criteria when run on training data. *Test error* and *Test epoch* perform the same evaluation as *Train error* and *Train epoch*, except that the evaluations are performed on test as well as train data. The genetic supervisor stops at the generation count that is specified by the user. It then reports the best *structure* identified in the current generations.

## 4.8 Neural Network Applications in Environmental Engineering

Neural networks have been gaining acceptance in a majority of areas with requirements for forecasting or predictions and as decision support systems. Applications in the areas of finance and management, engineering, pure sciences, medicine, and even literature have been reported (Hagan et al., 1996). However, there have not been many neural network applications reported in the open literature in the field of environmental engineering. This section examines some of the cases in which neural networks have been applied for ecological or environmental management issues.

### 4.8.1 - Case Studies

Farid U. Dowla and Leah L. Rogers, in a book titled *Solving Problems in Environmental Engineering and Geosciences with Artificial Neural Networks* (1995) have explored some possible applications of neural networks to environmental problems. Specifically, ANN was suggested to be used as an alternative to groundwater flow and transport codes to evaluate pumping alternatives in a remediation scheme. Pumping pattern generation was carried out using a genetic algorithm, which was linked to the ANN. Based on outcomes describing remedial success it attempted to generate better pumping patterns.

Back propagation was used to train the ANNs with the error calculations made by the conjugate gradient variation method. The conjugate gradient method has the same objective of error minimization as steepest descent techniques except that it is much faster (Hagan et al., 1996). The ANN-GA approach was able to generate an effective combination of 28 pumping wells for the groundwater remediation problem.



In related work, Ranjithan et al., (1993) have used a neural network to design and evaluate groundwater remediation alternatives. In the design of groundwater remediation, the uncertainty arising from the spatial variability of hydraulic conductivity and other aquifer parameters is important. Stochastic models based upon conditional simulation evaluate equally probable multiple realizations of hydraulic conductivity patterns to decide on the critical realization which would influence the outcome (Freeze et al., 1990). As an alternative, an ANN model can be based on the number of possible hydraulic conductivity patterns that can satisfy an output that is a management objective (Ranjithan et al., 1993). Neural networks with their adaptive learning approach, were applied as a pattern classification technique to identify critical hydraulic conductivity patterns and yield reliable remediation alternatives for ground water contaminated sites.

Rizzo and Dougherty (1994) developed a model called Neural Kriging by applying an artificial neural network to a traditional spatial estimation problem. This was a new method of pattern completion for geohydrological applications. The methodology was applied to construct maps of "discrete spatially distributed field" (e.g., log hydraulic conductivity). The network underwent supervised training with a counter-propagation feed-forward learning algorithm on field sampled data. The network employed an interpolating algorithm for generating values at unsampled locations. Neural kriging demonstrated the possibility of applying a simple technique to obtain fast and reasonably accurate statistical results on aquifer parameters.

Johnson and Rogers (1995) used the Dowla - Rogers approach for location analysis in groundwater remediation using neural networks. In groundwater remediation when a new pumping strategy was considered, a flow and transport model was called to

evaluate its effectiveness. It is common for optimization techniques to examine hundreds of strategies, at the very least, often with the computational time becoming prohibitive. Typical flow and transport models of contaminated aquifers can require hours to predict the effectiveness of a single prospective pumping strategy in pump-and-treat remediation. A method was developed for speeding the simulation step so that millions of alternative strategies could be evaluated within practical time limits.

The method used artificial neural networks trained to predict selected time and effectiveness information that normally were generated by the stochastic flow and transport models. The networks were then linked to a search technique based on the genetic algorithm to select pumping patterns which balanced the competing goals of timely and effective cleanup at minimum cost. This method was applied to a 28-location pump-and-treat Superfund site. The network was trained to predict mass-extraction and containment information generated by a 2-D model. Subsequently it was used in an optimization procedure to identify 250 (out of over 4 million) pumping patterns which met restoration goals at minimum costs. This study demonstrated the way in which ANNs enable planners to adopt a hypothesis testing approach for analysis and design that would otherwise be computationally impractical.

Additionally, the pattern-matching capabilities of neural networks have been tried for determining aquifer parameters by Aziz et al. (1992). The network was trained on drawdown data as input and corresponding aquifer parameters as output. Theis and Hantush-Jacob solutions were used to derive aquifer parameters for comparison with neural network outputs. The advantage with neural network application is the simple approach of obtaining aquifer parameters without computational complexities. Table 4-1

**Table 4-1. Computed Aquifer Parameter Values from Different Methods (Rashid et al., 1992)**

Method	Transmissivity (m <sup>2</sup> /min)	Storage coefficient	Leakage factor (m)
Hantush	1.156	0.0017	600
Walton	1.201	0.0019	900
Discrete	1.167	0.0015	850
Kalman	1.151	0.0017	668
Slope	1.094	0.0024	505
Neural network	1.179	0.0018	698

- Data were analyzed by Kruseman and de Ridder (1979) using Hantush and Walton (1970) methods.
- \*\* - Data were analyzed by Rushton and Chan (1976) using discrete numerical model.
- Data were analyzed by Chander et al. (1981) using an iterated Kalman filter.
- Data were analyzed by Sen (1986) using the slope-matching method.

is a comparison of aquifer parameter values obtained by different methods. The values obtained by neural network compared well with values obtained by other statistical techniques.

El-Hawary (1995) used a multi-layer network for predicting pollution concentration levels from electric power generation using fossil fuel. The ANN developed could be used as an alarm processor for sounding warnings or it could be used as an analyzer to process data and predict pollution levels. The input to the model consisted of site coordinates, plume dispersion model parameters relevant to the site, atmospheric conditions, and active power conditions. The neural network model was shown to be a feasible option by comparing the net output with computational approaches on the test data.

Another paper examined the effectiveness of using artificial neural networks (ANNs) for real-time data analysis of a sensor array (Hashem et al., 1995). The motive behind using a neural network coupled with a sensor array was to enhance the analyzing capabilities of a single sensor to identify contaminants in the field without requiring highly selective sensors. A prototype sensor array that consisted of nine tin-oxide sensors, a temperature sensor, and a humidity sensor was used in the study (Hashem et al., 1995). The study showed that using a neural network data analysis technique enhanced the selectivity of the sensor array in identifying contaminants.

Researchers at Sandia National Laboratory have completed investigations into the application of artificial neural networks for reliability and risk analysis (Robinson, 1995). The model inputs consisted of probability density functions of parameters describing

system characteristics and output was a statistical description of the level of system performance. The ANN model was used to describe the boundary between safe and unsafe system configurations. This methodology was involved in the comparison of various low-residue, lead-free soldering processes to minimize wastes without compromising on product reliability. The inputs to the model consisted of statistical descriptions of material properties like coefficients of thermal expansion of substrate and solder. Model output gave the fatigue life of the surface mounted components. The model was used to make decisions to maximize component life.

In a study by Basheer et al. (1995), neural networks were used for landfill site identification. Neural networks were used for mapping the soil permeability and delineating boundaries for construction of landfills. Most site exploration studies involve sampling through boreholes at various locations and then drawing contours for a particular material property. Most mapping models, for contouring, enable visualization of continuity in a property from information derived from limited borehole data. They depend on the reliability (or probability) of finding a certain material or property at an unknown point given that material or property has been observed at a known point. Such probabilistic models share some common shortcomings; “they need to estimate many model parameters (through model calibration) and impose several simplifying assumptions before use” (Basheer et al., 1995). This leads to a distortion of the pattern embedded in the data set of several variables.

Neural network techniques capture this pattern through learning. Distortions induced by simplifying assumptions are therefore reduced. In this study a back propagation network was applied to data from a site located in Orlando, Florida. An

earthen landfill was planned to be constructed in the existing natural soil. A three layer network (4 neurons in hidden layer), was successfully used to determine the spatial distribution of permeability in an attempt to delineate the boundaries of the earthen landfill. In this study the neural network was used as a decision-making tool that aided in demarcating additional borehole locations for preparing a thorough sampling strategy.

The U.S. Army Corps of Engineers' procedures to clean up buried ordnance have been designed to proceed with minimum digging (Millhouse and Gifford, 1997). The Corps and its contractors are developing a neural network model that would read the input from sensors to identify the presence of buried ordnance and estimate the depth of burial (Millhouse and Gifford, 1997). The methodology incurs minimum risk as the sensors do not have to be placed below ground, eliminating the possibility of accidental detonation. This model can be linked with a geographic information systems (GIS) package to provide map locations and other site related information to help in making remediation decisions. In the testing phase, the network was used to estimate the depth of 39 detected, but unknown, pieces of ordnance. The net predicted 36 out of those 39 within a weight error threshold of 5 pounds and depth error threshold of one half foot.

Mukhopadhyay (1999) has used ANN technology for estimation of transmissivity values of the Damman Formation in Kuwait. The pattern matching capability of neural networks was used to output transmissivity values of the Damman formation based on varying combinations of UTM (Universal Transverse Mercator) well coordinates, height of formation above sea level, logarithm of total dissolved solids and parameters indicating absence or occurrence of loss of circulation as input parameters. The outputs were tested against outputs obtained by kriging techniques. It was observed that

correlation between estimated and actual values at known data points were much better with neural network application (0.83 to 0.97) than kriging (0.75).

Neural network applications in environmental engineering are in their incipient stages. As compared to some modeling techniques that have a theoretical basis to which data are adapted, neural networks work the other way. As the net evolves on a set of data it adapts to these data. This adaptability of the network is compared to plasticity of materials. As plastic materials undergo deformation without losing their plasticity properties the network should undergo additional learning without losing its previous learning. This creates the *plasticity-stability dilemma* (Karayiannis and Venetsanopoulos, 1993) for the network, i.e., to maintain stability in its learning and still be able to undergo additional learning. With highly disaggregated data presented by environmental engineering problems, neural networks face this dilemma. At the same time, adaptive learning capabilities of neural networks hold a lot of promise for environmental engineering problems, which are difficult to describe by set rules and classical theories on transport phenomena.

## Chapter 5. Materials and Methods

### 5.0 Introduction

The trophic state in Lake Tenkiller was checked at stations in the upper reaches (at Tahlequah, Oklahoma) and in the lower reaches or lacustrine zone of the lake. Stations numbered 4, 5, 6 and 7 in Figure 2-1 identify the lower reaches. The aim of this exercise was to develop linked neural network models to evaluate several watershed phosphorus reduction alternatives for controlling eutrophication in the lake, using available data sets for the period 1977 - 1985.

Three linked artificial neural network models were needed for this effort to link data collected during separate sampling events as shown in Table 5-1. The Harton data set described instream and point source phosphorus levels at locations upstream from Tahlequah, OK. These data consisted of pdf's developed from USGS water quality monitoring data which were collected from 1977 through 1985 (Harton, 1989). OWRB personnel between 1992 and 1993 collected the other two data sets in the lake (OWRB, 1996). An initial neural network model was required to establish a relationship between phosphorus concentrations upstream and in the lake to generate phosphorus concentrations in the lake for the period 1977 - 1985. This was needed to maintain consistency with the point source and non-point source data studied in this effort, which was available for the period 1977 - 1985. Further, to characterize the trophic state of the lake based on phosphorus concentrations, relationships had to be established through neural network models between phosphorus and chlorophyll *a*, and phosphorus and secchi disk data, since the two were used as indicators of the trophic state.



**Table 5-1. Available Data**

Source	Parameter	Location	Form	Time Period
USGS	Phosphorus	Tahlequah	Values	1992 - 1993
OWRB	Phosphorus	Lake	Values	1992 - 1993
Harton	Phosphorus	Tahlequah	Log-normal distribution	1977 - 1985
OWRB	Chlorophyll <i>a</i>	Lake	Values	1992 - 1993
OWRB	Secchi disk	Lake	Values	1992 - 1993
Harton	Phosphorus	Point Sources	Values	1977 - 1985
Harton	Phosphorus Channel Loss	Instream	Log-normal distribution	1977 - 1985

The flow chart of activities employed to study the phosphorus reduction alternatives is given in Figure 5-1. The flow chart highlights the steps involved and data sets used in developing the linked artificial neural network models.

Steps 1 - 4 in Figure 5-1 were used to prepare a larger data set of phosphorus concentrations at Tahlequah for the period 1992 - 1993 using Latin-Hypercube sampling on log-normal plots prepared from limited data. The Harton data set (step 7, Figure 5-1) included statistical density functions of total phosphorus at the USGS gage at Tahlequah, Oklahoma. These distributions included point and non-point source contributions from the contributing watersheds and were developed using historic monitoring and discharge data as well as modeled data that generated transport losses to the gauging station. To be

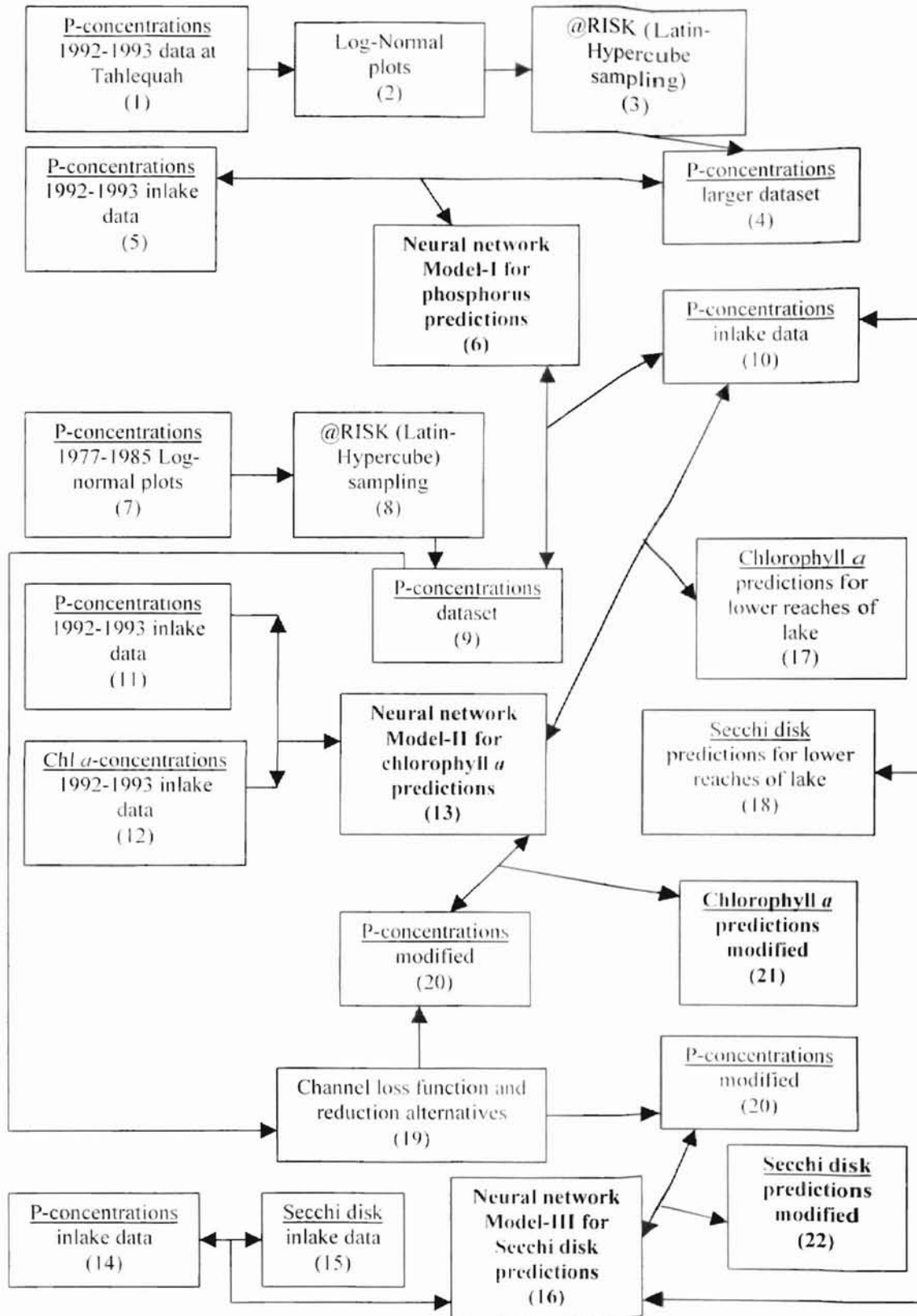


Figure 5-1. Flow Chart of Activities in Present Study

consistent with these distributions the OWRB data collected during the eutrophication survey were cast into a similar statistical format (Figure 5-2) at step 2 in Figure 5-1. Figure 5-2 is a probability plot developed for the sampled OWRB data and is similar to the Harton data set format. The two data sets could then be randomly sampled by Latin-Hypercube methods as shown in steps 4 and 8 in Figure 5-1. The data generated at step 4 was then used as input to train a neural network (Neural Network Model-I) for predicting in-lake phosphorus concentrations using phosphorus concentrations at Tahlequah as inputs as shown in steps 4, 5 and 6 in Figure 5-1. Thirty-two data points were used for conditioning the ANN to establish a relationship between phosphorus concentration in the upper and lower reaches of the lake. The data set used for conditioning the neural network is presented in Table 5-2. The table contains input phosphorus concentration data at Tahlequah and target phosphorus concentration in the lake for the period 1992 – 1993 used for training Neural Network Model – I. Steps 7, 8 and 9 in Figure 5-1 prepared a data set of phosphorus concentrations at Tahlequah for the period 1977-1985 from log-normal phosphorus concentration plots (Harton, 1989) available for the same period. These data were used as inputs for Neural Network Model-I for predicting in-lake phosphorus concentrations as shown in steps 9 and 10 in Figure 5-1 for the period 1977-1985. This data set was used to characterize the trophic state of the lake as shown in step 17 of Figure 5-1.

Steps 11, 12 and 13 in Figure 5-1 developed a neural network (Neural Network Model – II) for chlorophyll *a* predictions by using in-lake phosphorus and chlorophyll *a* data sets available for the period 1992 - 1993 from the OWRB study. The training data set for this network model is given in Table 5-3. The table contains input phosphorus

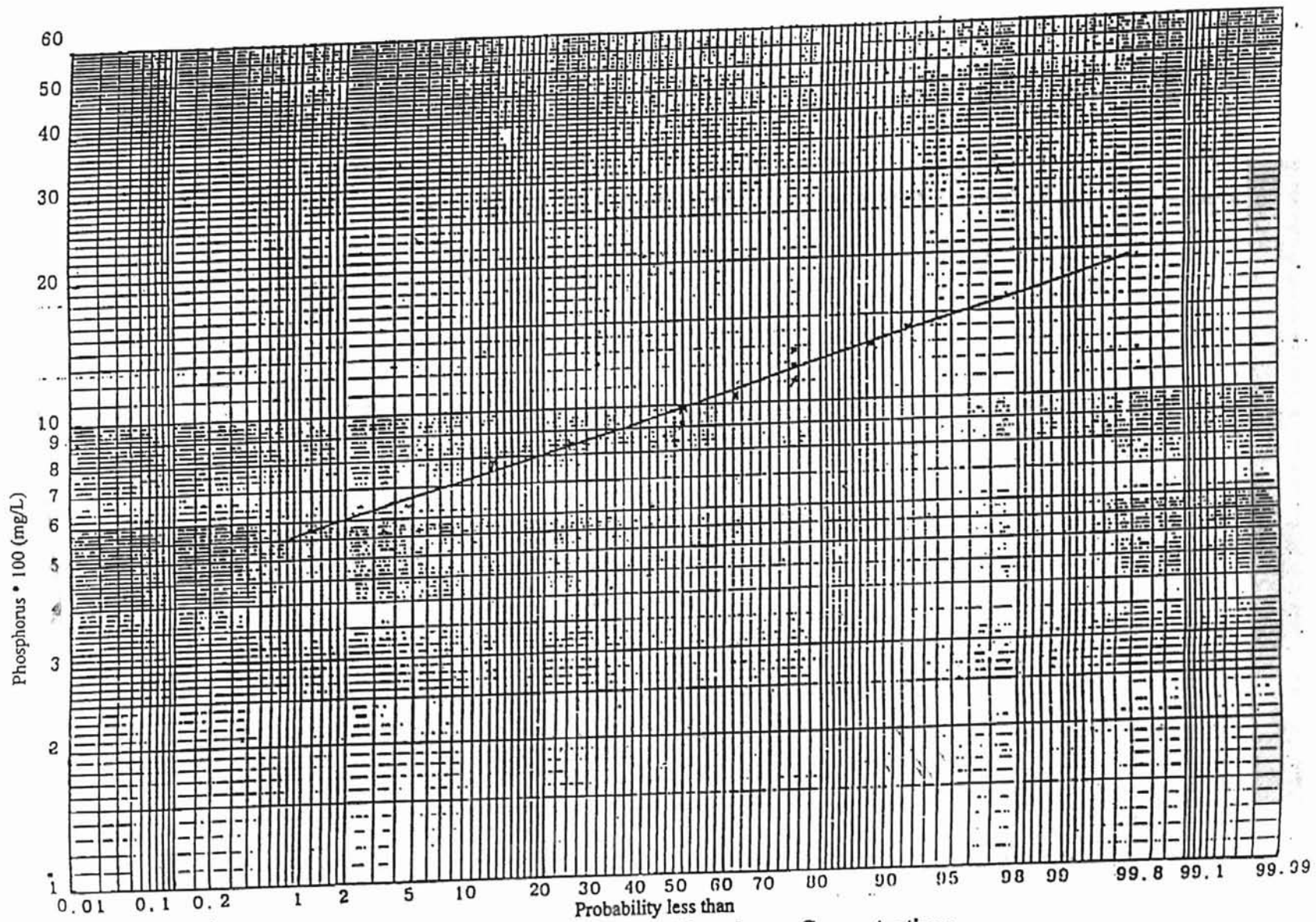


Figure 5-2. Probability Plot for Phosphorus Concentrations  
for the Period 1991 to 1993 at Tablequah (USGS, 1993)

**Table 5-2. Training Data Set for Network Model - I**

<b>Serial #</b>	<b>Input P-Tahlequah (mg/l)</b>	<b>Target P-Lake (mg/l)</b>
#1	0.0685	0.039
#2	0.0700	0.025
#3	0.0713	0.023
#4	0.0749	0.023
#5	0.1040	0.084
#6	0.0960	0.081
#7	0.0962	0.057
#8	0.1000	0.097
#9	0.1400	0.081
#10	0.1572	0.118
#11	0.1139	0.085
#12	0.1370	0.087
#13	0.0800	0.176
#14	0.0823	0.124
#15	0.0763	0.078
#16	0.0779	0.067
#17	0.1300	0.081
#18	0.1197	0.064
#19	0.1202	0.040
#20	0.1320	0.033
#21	0.0900	0.086
#22	0.0876	0.041
#23	0.0911	0.015
#24	0.0927	0.015
#25	0.1100	0.165
#26	0.1168	0.067
#27	0.1120	0.013
#28	0.1068	0.009
#29	0.0898	0.075
#30	0.0962	0.059
#31	0.0911	0.164
#32	0.0927	0.051

*data set conforms to the period Oct. 1992 to Sep. 1993  
(OWRB et al., 1996)*

**Table 5-3. Training Data Set for Network Model - II**

Serial #	Input P-Lake (mg/l)	Target Chlorophyll a ( $\mu\text{g/l}$ )
#1	0.061	34.3
#2	0.027	15.7
#3	0.023	16.5
#4	0.016	9.4
#5	0.058	47.2
#6	0.056	45.6
#7	0.051	39.6
#8	0.031	28.0
#9	0.069	31.0
#10	0.034	28.0
#11	0.020	13.3
#12	0.014	12.1
#13	0.069	33.4
#14	0.057	18.8
#15	0.042	13.4
#16	0.038	10.1
#17	0.057	29.2
#18	0.041	15.8
#19	0.017	11.0
#20	0.025	6.3
#21	0.039	23.4
#22	0.025	11.8
#23	0.023	14.2
#24	0.023	12.6
#25	0.085	1.0
#26	0.073	2.7
#27	0.079	3.9
#28	0.081	4.3
#29	0.118	4.1
#30	0.085	9.4
#31	0.087	2.5
#32	0.287	2.6
#33	0.159	1.6
#34	0.146	8.9
#35	0.176	2.7
#36	0.124	5.8

**Table 5-3. Training Data Set for Network Model – II**  
*(continued)*

#37		0.078		11.9
#38		0.067		30.6
#39		0.056		36.4
#40		0.056		15.5
#41		0.046		5.8
#42		0.037		8.5
#43		0.081		38.0
#44		0.064		40.3
#45		0.040		32.4
#46		0.033		39.1
#47		0.059		21.7
#48		0.038		12.6
#49		0.022		3.7
#50		0.020		6.4
#51		0.076		22.9
#52		0.042		26.5
#53		0.023		13.8
#54		0.016		8.2
#55		0.050		31.1
#56		0.034		16.9
#57		0.028		10.5
#58		0.021		7.0
#59		0.086		28.1
#60		0.041		18.9
#61		0.015		22.3
#62		0.015		24.9
#63		0.165		26.0
#64		0.067		23.5
#65		0.013		11.6
#66		0.009		10.7
#67		0.075		33.9
#68		0.059		13.7
#69		0.164		8.9
#70		0.051		5.7
#71		0.056		9.5
#72		0.051		11.4
#73		0.023		5.3
#74		0.017		1.3

*data set conforms to the period 1992-1993 (OWRB et al., 1996)*

and target chlorophyll *a* level data in the lake for the period 1992 - 1993 used to train Neural Network Model - II. This model was used for chlorophyll *a* predictions in the upper and lower reaches of the lake using phosphorus concentrations generated by Neural Network Model - I for the period 1977 - 1985 as shown in steps 9, 17, 20 and 21 (Figure 5-1). Predictions at step 17 were used to characterize the trophic state of the lake using chlorophyll *a* as indicator before applying phosphorus reduction alternatives. The model was further used for making chlorophyll *a* predictions at Tahlequah to evaluate the trophic state of the lake against different input phosphorus reduction alternatives (step 19) as shown in steps 20 and 21 (Figure 5-1). Tahlequah was modeled as upper reaches of the lake and reduction alternatives were evaluated at Tahlequah because of availability of p-concentration data and the ability to calculate transport losses and phosphorus contributions from point and non-point sources at Tahlequah for the period 1977 - 1985.

Steps 14, 15 and 16 (Figure 5-1) developed a neural network (Neural Network Model - III) for secchi disk value predictions using in-lake phosphorus and secchi disk value data sets available for the same period (OWRB et al., 1996). The training data set for conditioning this network is given in Table 5-4. The table contains input phosphorus data and target secchi disk values in the lake for the period 1992 - 1993 used to train Neural Network Model - III. The model was then used for predicting secchi disk values for characterizing the trophic state of the lake before applying phosphorus reductions and evaluating the trophic state of the lake against different phosphorus reduction alternatives as shown in steps 18, 20 and 22 (Figure 5-1). The input phosphorus values corresponded with values used for the phosphorus vs. chlorophyll *a* model, enabling comparison of the



**Table 5-4. Training Data Set for Secchi Disk Simulations**

<b>Serial #</b>	<b>Input - P Lake (mg/l)</b>	<b>Target Secchi Lake (m)</b>
#1	0.061	1.00
#2	0.027	1.60
#3	0.023	1.70
#4	0.016	2.40
#5	0.058	0.70
#6	0.056	1.30
#7	0.051	1.60
#8	0.031	1.60
#9	0.069	0.70
#10	0.034	1.23
#11	0.020	1.65
#12	0.014	1.70
#13	0.069	0.90
#14	0.057	1.50
#15	0.042	1.70
#16	0.038	2.20
#17	0.057	1.00
#18	0.041	1.60
#19	0.017	1.80
#20	0.025	2.30
#21	0.039	1.90
#22	0.025	2.30
#23	0.023	2.60
#24	0.023	4.30
#25	0.085	1.70
#26	0.073	0.70
#27	0.079	0.75
#28	0.081	0.80
#29	0.118	0.30
#30	0.085	1.70
#31	0.087	1.45
#32	0.287	0.00
#33	0.159	0.30
#34	0.146	0.27
#35	0.176	0.30

**Table 5-4. Training Data Set for Secchi Disk Simulations  
(continued)**

#36		0.124		0.48
#37		0.078		1.20
#38		0.067		1.40
#39		0.056		1.20
#40		0.056		1.40
#41		0.046		2.00
#42		0.037		2.20
#43		0.081		0.95
#44		0.064		1.15
#45		0.040		1.55
#46		0.033		1.40
#47		0.059		1.20
#48		0.038		1.50
#49		0.022		2.10
#50		0.020		2.00
#51		0.076		0.85
#52		0.042		1.20
#53		0.023		2.10
#54		0.016		3.00
#55		0.050		1.30
#56		0.034		1.95
#57		0.028		2.50
#58		0.021		2.80
#59		0.086		0.80
#60		0.041		1.40
#61		0.015		2.30
#62		0.015		2.00
#63		0.165		0.45
#64		0.067		1.00
#65		0.013		2.40
#66		0.009		3.10
#67		0.075		0.90
#68		0.059		1.40
#69		0.164		2.80
#70		0.051		3.30
#71		0.056		1.10
#72		0.051		2.20
#73		0.023		2.10
#74		0.017		2.80

*data set conforms to the period 1992 – 1993 (OWRB et al., 1996)*

rophic state as predicted by the two indicative parameters.

Seventy-four data points were used to condition the ANN, which defined relations between phosphorus and chlorophyll *a*, and phosphorus and secchi disk data. These numbers were governed by data availability.

Finally, the three neural network models described three relationships needed for characterizing the eutrophication state of the lake and evaluating phosphorus reduction alternatives. All three neural network models were generated using commercially available software (*Neuralyst* – Cheshire Engineering Corporation). The relationships established using the ANN models were:

1. Phosphorus concentration at Tahlequah versus phosphorus concentration in the lake (P vs. P model, Neural Net Model I – Figure 5-1).
2. Phosphorus concentration versus chlorophyll *a* concentration (P vs. chlorophyll *a* model, Neural Net Model II – Figure 5-1).
3. Phosphorus concentration versus secchi disk values (P vs. secchi disk model, Neural Net Model III – Figure 5-1).

Once these models were developed, they were applied to five alternative phosphorus reduction levels, as illustrated in steps 13, 16, 20, 21 and 22 in Figure 5-1, respectively. A total of ten predictions were obtained, five each for chlorophyll *a* and secchi disk values. These predictions were assigned a serial number of C1 through C5 and S1 through S5, respectively. C1 through C5 were related to evaluation with chlorophyll *a* predictions while alternatives S1 through S5 were related to evaluation with secchi disk predictions. The phosphorus reduction alternatives evaluated in this study were as follows:

*C1 and S1.* Complete phosphorus removal from point sources. Peak flow data at Tahlequah and channel phosphorus losses per mile were required to perform this analysis. A log-normal distribution was developed for peak flow data at Tahlequah from available records for monthly peak flow values in cubic feet per second from 1977 through 1985 (USGS) and is presented in Figure 5-3. The 95<sup>th</sup> percentile value was used for modeling the reasonable worst-case situation of phosphorus loading. Further, the contribution from point sources at Tahlequah was calculated with the help of 50<sup>th</sup> percentile channel mile loss value adopted from log-normal distributions of channel loss functions developed in an earlier study and presented in Figure 3-2 (Harton, 1989). The channel mile loss allows calculation of the amount of phosphorus assimilated and degraded in the river per mile. Since eutrophication is a gradual process, the use of the 50<sup>th</sup> percentile value for channel loss function was considered justified.

The contribution to phosphorus loading from each point source was estimated by calculating the total mass of phosphorus in the discharge of individual point sources reported by Harton in 1989. The channel mile loss function was then used to establish the total amount of point source phosphorus lost in transmission from the respective point source to Tahlequah. The phosphorus mass lost was then subtracted from the total phosphorus load discharged by the individual point source to calculate the phosphorus load without point source contribution at Tahlequah. Sample calculations for estimation of phosphorus contribution at Tahlequah from a point source are provided in Appendix - I. Point source contributions calculated from available data for all treatment plants upriver of Tahlequah are given in Table 5-5. The table presents the phosphorus amounts

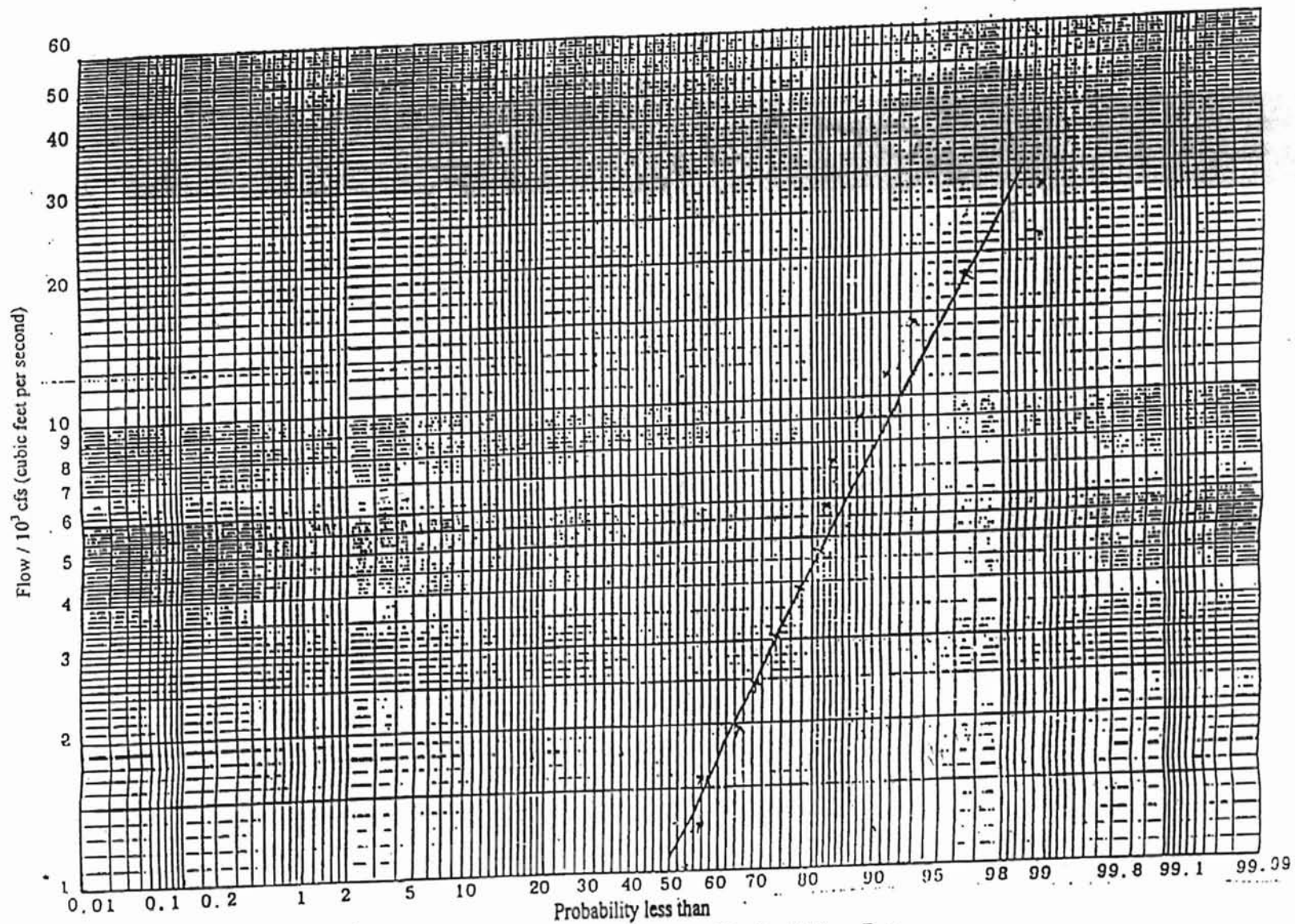


Figure 5-3. Probability Plot for Monthly Peak Flow Data  
for the Period 1977 to 1985 at Tahlequah (USGS)

Table 5-5. Point Source Contribution at Tahlequah (OWRB et al., 1996)

Serial #	Name of Creek	Location	Amount Input (Kg)	Amount Lost in Transport (Kg)	Amount Reaching Tahlequah (Kg)	Concentration at Tahlequah (mg/l)
1	Clear	Fayetteville (Arkansas)	3.5818	3.5818	0	0
2	Osage	Rogers (Arkansas)	50.7245	4.9678	45.7567	0.0014
3	Osage	Springdale (Arkansas)	75.9649	9.9251	66.0398	0.0021
4*		Lincoln (Arkansas)	3.2876	2.9726	0.3151	0.0000099
5*		Gentry (Arkansas)	4.6575	4.0219	0.6356	0.000019
6	Flint	Siloam Springs (Arkansas)	14.6007	2.1779	12.4228	0.0004
7*		Watts (Oklahoma)	1.3699	1.1479	0.2219	0.0000069
8*		Westville (Oklahoma)	7.9452	4.5397	3.4054	0.0001
9*		Midwestern Nursery (Oklahoma)	1.6438	0.5726	1.0712	0.0000033
10	Osage (Total)	Rogers + Springdale	126.6894	14.8929	111.7965	0.0035
11	Total		144.8719	20.6526	124.2193	0.0041

\* - from available data

flowing into the river, amounts lost in transport to Tahlequah, residual amounts and concentration left at Tahlequah from the treatment plants considered in this study.

In this alternative, the contribution from all point sources was summed and subtracted from the total phosphorus concentrations at Tahlequah. P vs. chlorophyll *a* and P vs. secchi disk ANN models developed earlier were applied on the modified phosphorus concentration data to predict chlorophyll *a* concentrations and secchi disk values for evaluating the effect of point source loading on the eutrophication of the water body at Tahlequah.

*C2 and S2* - Fifty percent reduction of phosphorus contribution from non-point sources above Tahlequah. Point source contributions were subtracted from the total phosphorus concentrations and the value obtained was the non-point source contribution at Tahlequah. This value was then reduced by 50% to obtain fifty percent reduced non-point source phosphorus mass at Tahlequah. Total point source contribution at Tahlequah was then added to the fifty percent reduced non-point source value to obtain the total phosphorus mass at Tahlequah, corresponding to a 50% reduction in non-point source loading. The reductions were subsequently applied to input phosphorus values for predicting chlorophyll *a* (alternative C2) and secchi disk values (alternative S2). This methodology of calculating phosphorus concentrations with non-point source phosphorus reductions at Tahlequah was applied for all the remaining alternatives.

*C3 and S3* - Seventy-five percent reduction of phosphorus contribution from non-point sources above Tahlequah. The reductions were applied to input phosphorus values for predicting chlorophyll *a* (alternative C3) and secchi disk values (alternative S3).

*C4 and S4* - Eighty percent reduction of phosphorus contribution from non-point sources above Tahlequah. The reductions were applied to input phosphorus values for predicting chlorophyll *a* (alternative C4) and secchi disk values (alternative S4).

*C5 and S5* - Eighty-five percent reduction of phosphorus contribution from non-point sources above Tahlequah. The reductions were applied to input phosphorus values for predicting chlorophyll *a* (alternative C5) and secchi disk values (alternative S5).

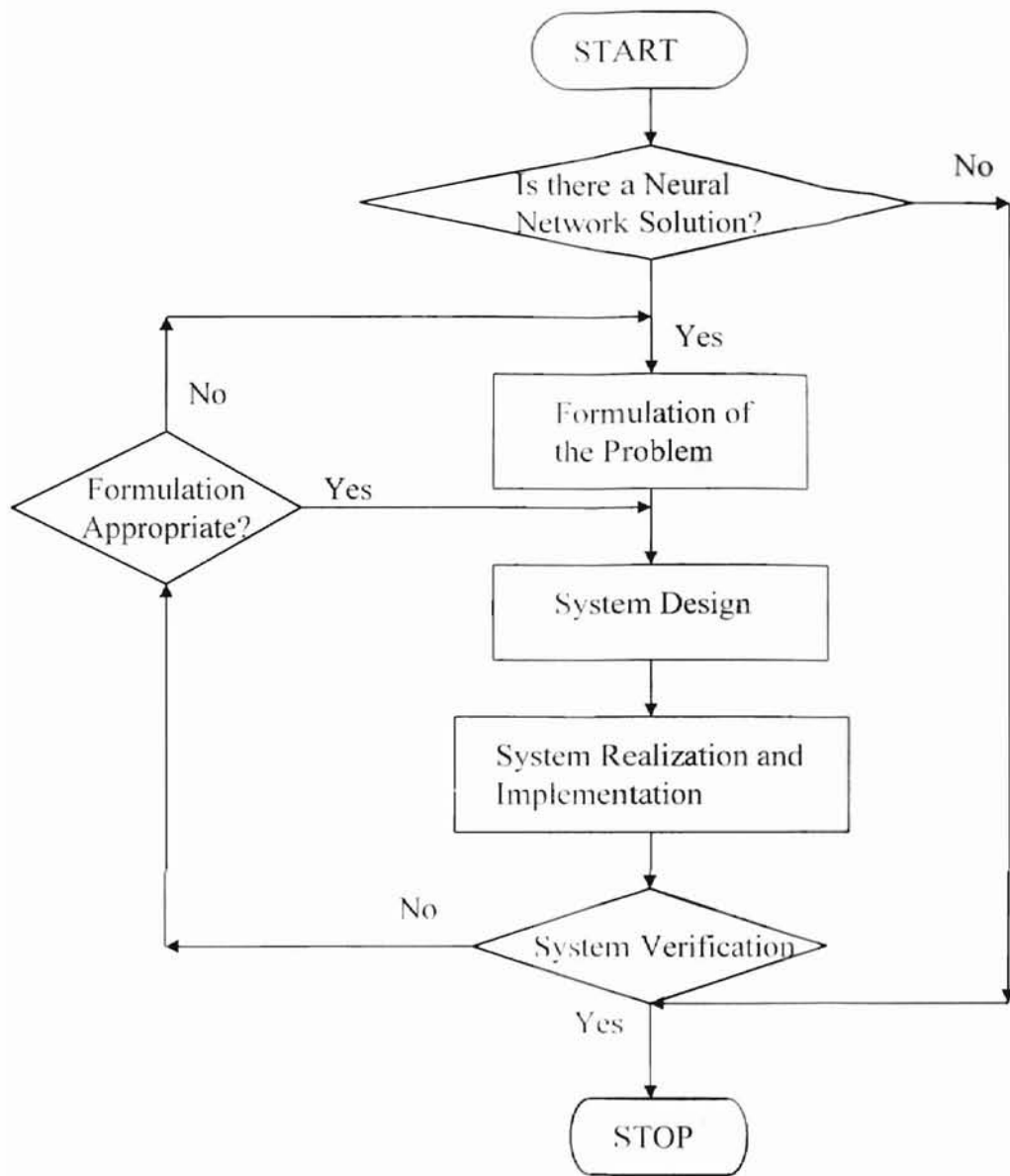
Initially, total point source reduction was considered to determine if these sources had a significant impact on eutrophication of the lake. This step was performed as a confirmation of results obtained from previous studies which showed the insignificance of point source contribution to overall eutrophication of the lake (Harton, 1989 and OWRB et al., 1996). No further work with point source contributions was deemed necessary.

For non-point sources, the reductions were chosen to match the reductions evaluated in earlier studies for comparison purposes (75% and 80%). Fifty percent reduction of non-point source contribution was chosen arbitrarily to provide a benchmark for assessing further reductions that would be required for controlling eutrophication.

## **5.1 Model Development**

The critical factor in this study was the development of stable and reliable neural networks for making precise predictions. The flow chart in Figure 5-4 presents a generalized methodology for development of neural network models. The networks for this study were prepared in four steps as highlighted by the flow chart and discussed subsequently.





**Figure 5-4. A General Methodology for the Development of Neural Network Systems (Karayiannis and Venetsanopoulos, 1993)**

- 1) *Step 1: Problem Formulation:* This step checks if the problem could be solved using a neural network. In the present study, available data to establish relationships between phosphorus vs. chlorophyll *a* and phosphorus vs. secchi disk values (OWRB et al., 1996) made the problem amenable to neural network analysis.
- 2) *Step 2: System Design:* This step involved the data conditioning for neural network model development. The model established a relationship between phosphorus in the upper and lower reaches of the lake. The sampled data used for conditioning the network model for establishing this relationship between phosphorus and chlorophyll *a* was edited to include some riverine zone data, entire lacustrine zone or lower reaches of the lake and part of the transition zone data (Table 5-3). This was done because some data exhibited a negative relationship between chlorophyll *a* and phosphorus, where high relative p-concentrations produced very low chlorophyll *a* readings. This seeming anomaly was explained on reach-specific and chemical bases by OWRB et al. (1996), which suggested that some reaches may be nitrogen or sunlight limited. The majority of the other sampling locations, however, was shown to be strongly P-limited (OWRB et al., 1996). The data from these reaches were employed in this effort. This meant that the subsequent eutrophication models more closely simulated the lacustrine sections of the reservoir (stations 5, 6 and 7 in Figure 2-1) where OWRB efforts indicated phosphorus limitations (OWRB et al., 1996).

3) *Step 3: System Realization and Implementation:* This step included various simulations carried out to develop an optimum network. Several permutations and combinations were tested for developing the optimum networks.

Specifically, in the P vs. chlorophyll *a* model (Neural Net Model II) the approach for building a network model was divided into two parts: with and without the genetic supervisor. Model development without the genetic supervisor was further divided into two phases. In the first, several network architectures were tested to arrive at the most appropriate configuration of layers and number of layers per neuron. Different alternatives were tried by varying the network architecture until some consistency in results in terms of prediction accuracy and numbers of iterations were achieved. Changing the number of hidden layers and the associated neurons per layer varied the network architecture. This was done to find an optimum number of interconnections in the network model for the most accurate training results. This step was not only important from the point of view of accurate predictions but also in terms of computational and training time utilized. Highly interconnected networks tend to have a very slow learning rate and carry a computational time burden. Extremely simple networks have less computational time but may end up without any learning after training on a data set. The various parameters specified in the first phase of model development were as follows:

*Transfer Function:* Since the concentrations of chlorophyll *a*, secchi disk data and phosphorus were not discretized, this study required functions with continuously discriminating outputs. Therefore, only differentiable functions were

evaluated for performance. Since the sigmoid function is a commonly used function in neural network applications, it was used for all simulations in this study.

Learning Rate (LR): A value of 0.7 was used for testing several network architectures in the first phase of model development. This value is the default value provided by the software. Since there was no basis for selecting this value except by trial and error, it was left unchanged as a first attempt for model development.

Momentum: A value of 0.9 was used in the first phase of model development. This value was chosen based on the available training data set, which had a lot of inconsistency (different chlorophyll *a* values for similar phosphorus values). The value of 0.9 was chosen so that no particular data point unduly affected the learning of the network.

Input Noise: Input Noise is only in effect when training a network, and a value of 0.01 was specified for the first phase of model development. This value ensured that the inputs for training to the network model constantly varied as  $\text{input} \pm \text{input} * 0.01$ .

Training Tolerance: The training tolerance for each simulation was set at 0.2, based on standard experimental errors obtained in measuring chlorophyll *a*. This value was adopted from *Standard Methods for Examination of Water and Wastewater*, 18<sup>th</sup> edition (Standard Methods, 1992).

Testing Tolerance: A value of 0.3 was chosen for this study. This value had to be kept greater than the training tolerance (0.2), based on guidelines provided in the software manual.

Epoch per Update: The value for this parameter was set to one as the number of training data points was not very large and updating after every run over the entire data set was not time consuming.

Error Limit: A value of 0.01 was used in the first phase of model development to avoid overtraining of the network model. The error limit behaves as a boundary and terminates training when the limit is exceeded. Time Limit and Epoch Limit were set to a value of zero (no limit), as each simulation duration was manually controlled. The training was stopped when it was observed from error plots that the network had stabilized (plateau reached in the error vs. epoch graph).

The second phase of model development without the genetic supervisor was a parameter uncertainty analysis to establish optimum values for the Learning Rate and Momentum to be used for network development. This phase was contingent on the first phase, as the permutations and combinations with the two parameters were performed on the optimum network architecture developed in the first phase of study. A range of values were tested for the two parameters to discern any underlying trends in network performance with increasing or decreasing values of these parameters.

The next step was to generate optimum networks using the genetic supervisor. In the present study three basic controlling parameters or three *strings* were available for developing an optimum network. The three *strings* controlled:

1) input column, 2) network configuration, and 3) network parameters. The combination of all three *strings* was used to define a neural network configuration. Each of the three *strings* mentioned above was treated as a *feature*. By manipulating or varying these *features* the software was able to create new generations (different networks) with the resulting *structure* of each generation being evaluated for fitness. The input column *string* is manipulated by treating each column of data as an individual candidate for inclusion or exclusion. Since there is a single column of input data (phosphorus as inputs only), used for this study this control parameter was rendered redundant. The network configuration *string* is varied by keeping the first (input) and last (output) layers constant and changing the number of hidden layers with the number of neurons per hidden layer. The network parameters *string* is varied by changing momentum, learning rate, and input noise. The software did the variation in string parameters during simulation time with no interference from the user. The user inputs required for the genetic supervisor were as follows:

*Population Pool Size*: A value of 3 was chosen for this study. The value was chosen after several permutations with different values. It was observed that there was no increase in percentage of “Right” scores by increasing the value above 3. Exceeding 3 resulted in additional computational times.

*Population Pool Mode*: *Immigration* mode was used in this study as it corresponds to the most versatile population management mode. In *Immigration* mode, weak *structures* are continuously culled, improving the *structure* with every run.

Crossbreeding: A value of 1 was used for crossover rate for reasons mentioned below in the Mutation Rate value selection.

Mutation Rate: A low value of 0.1 was used for this study because of the randomness of the method of selecting *structure features* for mutation. Chances of losing good structures dictated a low value of mutation rate with more consideration assigned to crossbreeding (by selecting a larger value) for generating a robust network.

To determine the fitness of a generated structure, Train Error was specified for generation of the best structure with the same basis for selecting an error tolerance as explained in the section for developing networks without genetic supervisor.

Generation Count: A value of 10 was specified for the current study and this value was chosen after evaluating the performance of networks generated with higher and lower generation counts.

P vs. P model (Neural Net Model I): With the experience gained from developing P vs. chlorophyll *a* and to avoid guessing the optimum combination of parameters for the best network by trial and error, the genetic supervisor was employed for developing the optimum network relationship between phosphorus concentrations in the upper and lower reaches of the lake. The training tolerance used in the development of this model was 0.2 again, based on standard experimental errors obtained in measuring phosphorus. This value was adopted from *Standard Methods for Examination of Water and Wastewater*, 18<sup>th</sup> edition (Standard Methods, 1992).

P vs. Secchi disk model (Neural Net Model III): The network was developed to establish a relationship between phosphorus concentrations and secchi disk values. The model was developed using the genetic supervisor. Parameter specifications for developing this network model corresponded with the parameter specification for P vs. P model or P vs. chlorophyll *a* model developed using genetic supervisor. The data for P vs. secchi disk data obtained from the lake were not as disaggregated as the P vs. chlorophyll *a* data obtained from the lake. This made the model development task easier, having developed the P vs. chlorophyll *a* model.

- 4) *Step 4: System Verification*. The system verification step checks the efficiency of the network model developed. The percentage of “Right” scores was used as an indicator of the model efficiency in every phase of model development.



Results obtained are divided into two categories: 1) those associated with model development, and 2) those obtained by evaluation of management alternatives. Each of these is further divided to reflect specific activity.

#### **P vs. Chlorophyll *a* model development without genetic supervisor**

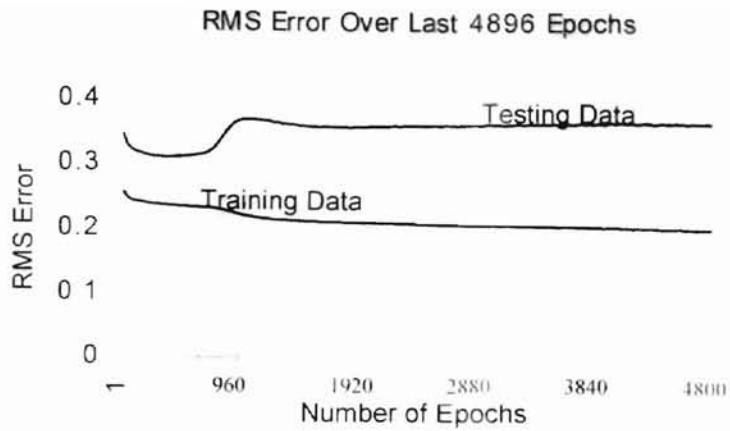
Table 6-1 lists the model architecture and results obtained in terms of number of layers, neurons per layer, percentage of “Right” scores, and number of training epochs in this first phase of model development. The percentage of “Right” scores was measured by the number of predictions within specified tolerance limits of the training data. Among the 2, 3, and 4-layer back propagation network models, the best results measured by percentage of “Right” scores were obtained with three layer network models in terms of consistency of prediction. A consistent prediction of 66% percent “Right” was obtained with three layer networks. Four layer networks lacked consistency in making right predictions (Table 6-1, Serial Number’s 6, 7, 8 and 10), and a 2 layer network exhibited little or no learning with the input data set. Predictions from four layer networks varied from 61% to 69% “Right”.

Figure 6-1 includes the root-mean-square (RMS) error plots of the various alternatives. The RMS error plots are used to monitor the network training and testing process. It is a representation of the difference between the actual output and the target output in case of training data after every simulation. The testing error curve is arrived by testing the training data after training is done. The error is the root-mean-square of the difference between tested predictions and target outputs. These plots serve two purposes: 1) they tell us the maximum learning achieved and 2) they tell us the consistency in the

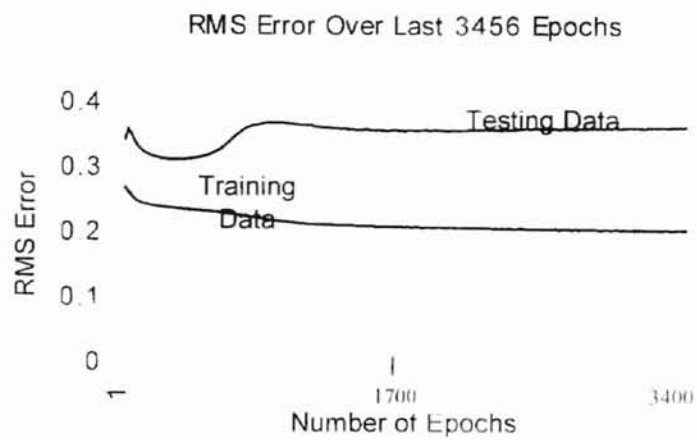
**Table 6-1. Alternatives for Building Network Model -II (without genetic supervisor)**

Serial #	# of Layers	Neurons per layer*	% Right	# of Epochs
1	3	1 7 1	66	4896
2	3	1 15 1	66	3456
3	2	1 1	61	4000
4	3	1 5 1	66	4064
5	3	1 3 1	66	4016
6	4	1 10 1 1	62	8448
7	4	1 1 10 1	69	8000
8	4	1 5 5 1	66	4480
9	3	1 29 1	66	4064
10	4	1 29 5 1	61	5504

\* - each value corresponds to number of neurons in each layer sequentially  
 Learning Rate = 0.7; Momentum = 0.9; Input Noise = 0.01; Training Tolerance = 0.2;  
 Testing Tolerance = 0.3; Error Limit = 0.01; other parameters had default values of 0.

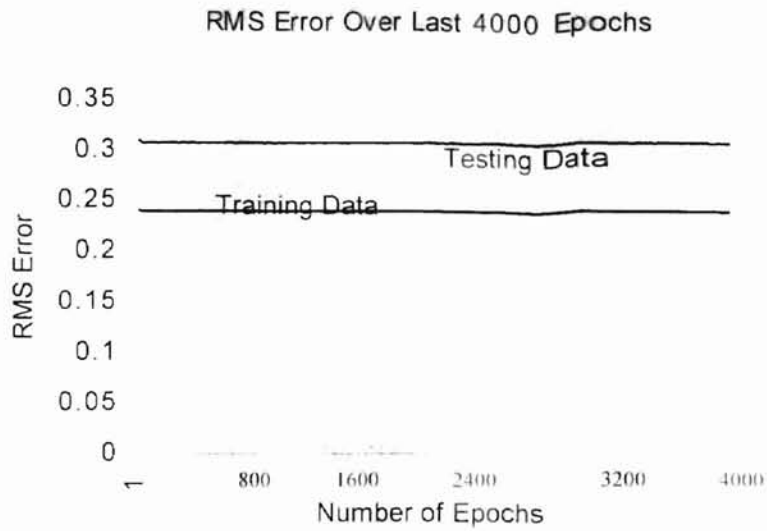


**Serial #1**

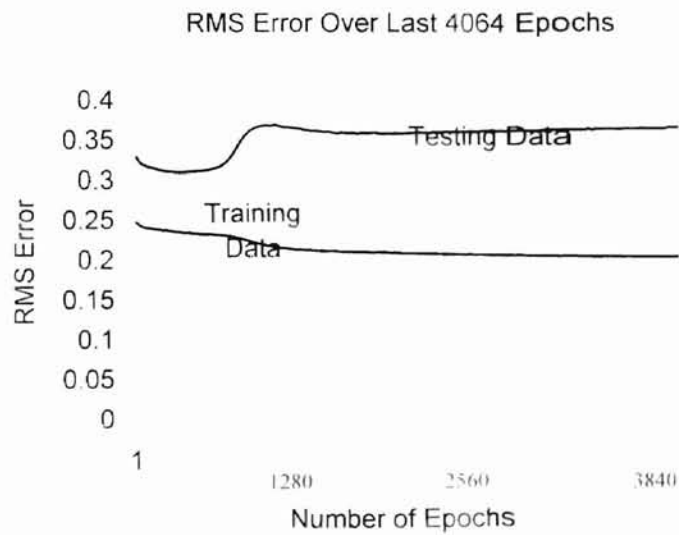


**Serial #2**

**Figure 6-1. RMS Error Plots of Various Alternatives (cont'd on next page)**

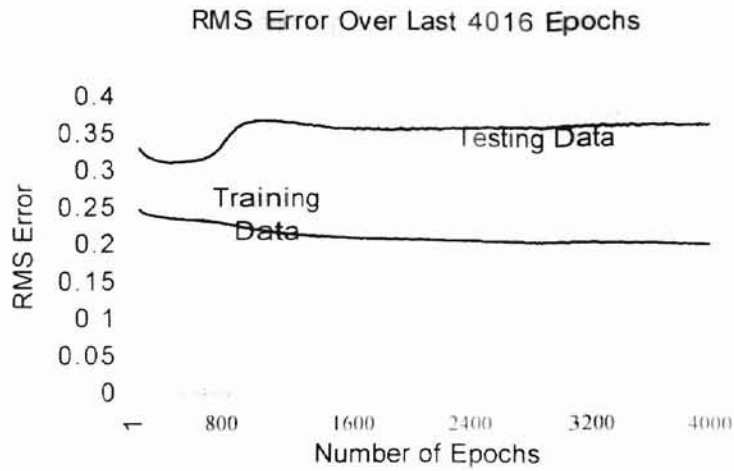


**Serial #3**

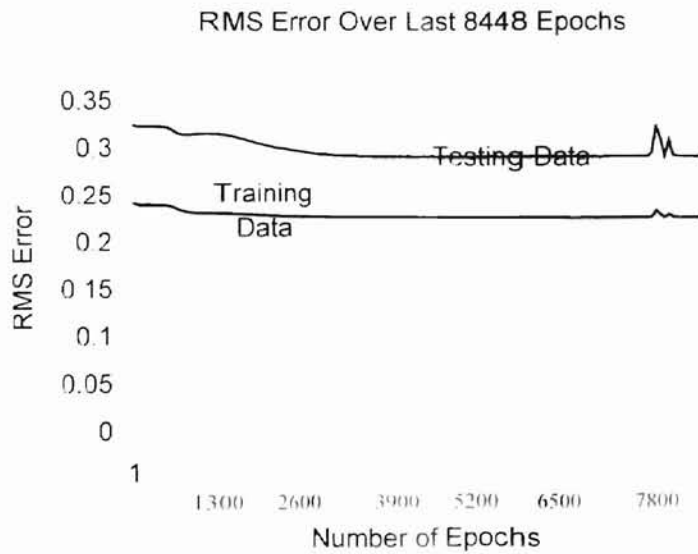


**Serial #4**

**Figure 6-1. RMS Error Plots of Various Alternatives (cont'd on next page)**

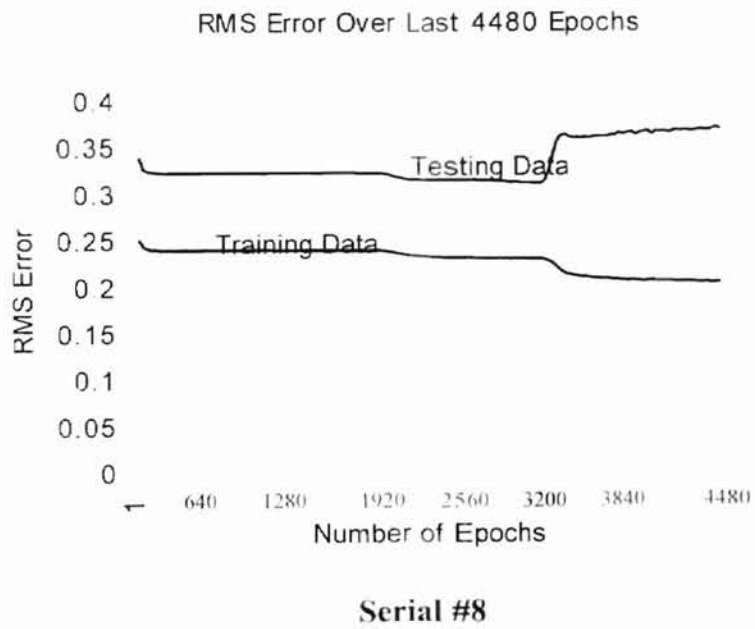
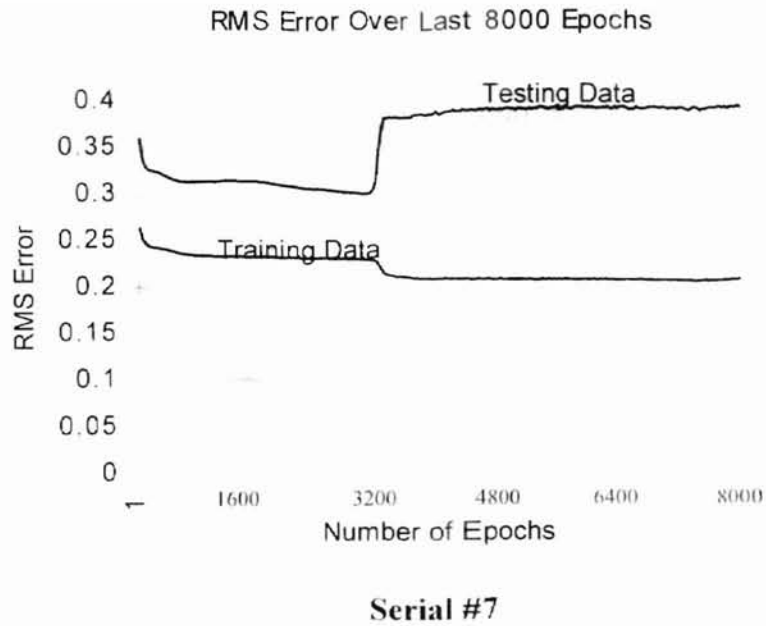


**Serial #5**

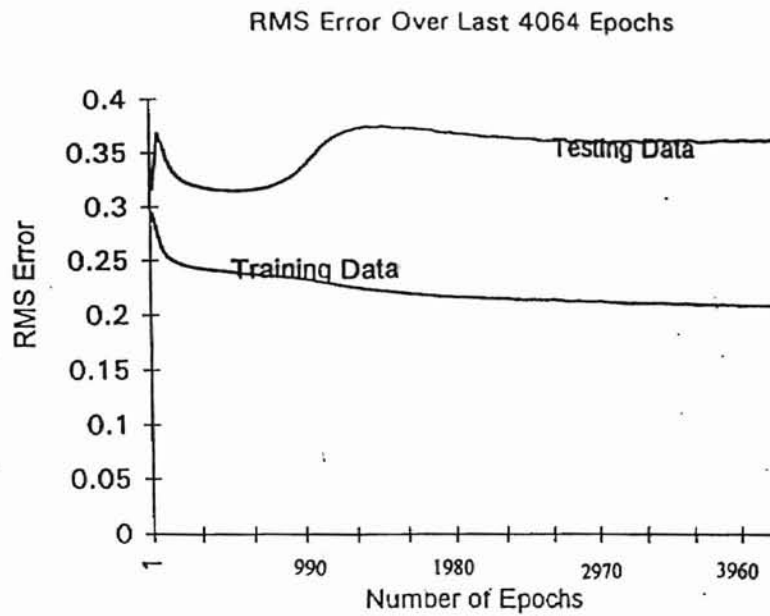


**Serial #6**

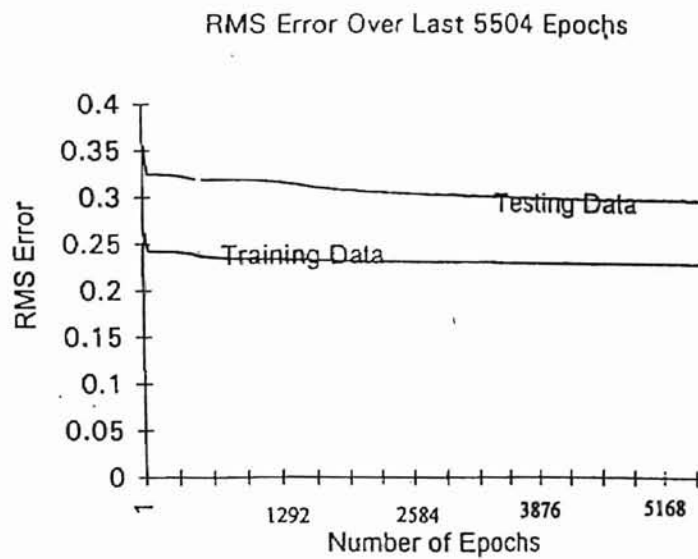
**Figure 6-1. RMS Error Plots of Various Alternatives (cont'd on next page)**



**Figure 6-1. RMS Error Plots of Various Alternatives (cont'd on next page)**



Serial #9



Serial #10

Figure 6-1. (cont,d from previous page)  
 (Learning Rate = 0.7; Momentum = 0.9; Input Noise = 0.01;  
 Training Tolerance = 0.2; Testing Tolerance = 0.3; Error Limit = 0.01)

learning approach. Theoretically complete learning can be said to have occurred when the training curve dips to a zero on the y-axis. In that case it can be said that the net is predicting the exact target values specified and is behaving as the phenomenon that it is supposed to learn. If the error curve indicates a value of 0.2 on the y-axis, it implies that the predicted value from the net is 20% off the mark from target values in case of network training. Secondly, if there are sharp peaks in the curve and not a gradual descent in the curve, it implies that the network predictions are not reliable. In such a case, the network learning is inappropriate and accurate predictions are more by chance than actual learning. “Testing” and “Training” error was plotted against the number of epochs in RMS error plots to evaluate network learning. RMS error plots for 3 layer networks (Serial #1, #2, #4, #5, and #9) exhibited similar characteristics, with gradual stabilization of the “Testing” and “Training” error plots. That is, the error curves became straight lines after an initial variation. The similarity of error plots obtained with three layer networks also highlighted the consistency in predictions obtained with these networks. The indifferent behavior of a two layer network was manifested in the straight lines of RMS “Testing” and “Training” data error plots (Figure 6-1, Serial #3) where no variations were observed in the percentage of correct predictions during simulation runtime. Though the four layer network with 1, 1, 10, and 1 neurons per layer produced the most accurate results, the network lacked stability, as shown in the RMS “Testing Data” error plot for the network (Figure 6-1, Serial #6). The spikes and undulations in the curve established the instability of the network with increased chances of erratic predictions on subsequent use of the network model. From this first phase of study, it was concluded that a 3-layer network was the most appropriate for this application.



The number of neurons per layer for optimized performance, however, could not be decided in this initial phase of study because of the similarity in results obtained from three layer networks, as presented in serials #'s 1, 2, 4, 5 and 9 in Table 6-1. A second phase of model development without the genetic supervisor for the P vs. chlorophyll *a* model employed a three layer network model to determine the most appropriate configuration. Any of the three layer models could have been chosen for this parameter uncertainty analysis, as all performed similarly. For this effort, the three layer model chosen had 1, 7, and 1 neurons per layer. Table 6-2 highlights the effect of modifying the learning rate and momentum on performance measured by percentage of "Right" scores and number of training epochs required to achieve the corresponding percentage. The values for the learning rate and momentum were chosen randomly to cover a wide range of typical user-specified input possibilities. The network did not perform well at values of 0.1 and 0.05 for learning rate. Similarly, beyond 0.15 the network performance was similar, with approximately the same number of training epochs. This made the selection of a definite value for learning rate extremely difficult. Variations in momentum had very little effect in improving the performance of the three layer network, as presented in the data corresponding to serial # 1 and serial #'s 13 through 20 in Table 6-2. It was therefore concluded that learning rate and momentum do not have a significant effect on network performance as long as values selected did not specifically fall within the ranges mentioned above where the network performance was noticeably poor.

Even with this lack of variation in model performance given alternative operating configurations, the two phases of this part of the study did provide an idea of general effects of network architecture and network parameters for developing an optimum

**Table 6-2. Performance of a 3-layer Network with Varying Learning Rate and Momentum Inputs**

Serial #	Learning Rate	Momentum	% Right	Epochs
1	0.70	0.90	66	4896
2	0.50	0.90	65	4189
3	0.80	0.90	62	4050
4	0.10	0.90	55	4063
5	0.15	0.90	68	5989
6	0.20	0.90	66	5084
7	0.27	0.90	66	5578
8	1.00	0.90	66	5500
9	0.05	0.90	57	4945
10	0.90	0.90	65	4641
11	0.60	0.90	66	5105
12	0.60	0.50	66	5225
13	0.70	0.30	64	4827
14	0.70	0.70	64	4116
15	0.70	0.10	64	5011
16	0.70	0.55	64	4472
17	0.70	0.85	64	4200
18	0.70	0.05	66	4175
19	0.70	1.00	65	4467
20	0.70	0.00	64	4503

network. It was concluded that the best network that could be developed by trial and error with the given training data was a three layer network. The best type of three layer network, however, remained unclear. These two phases of model development without the help of the genetic supervisor did serve as a basis for evaluating the optimum network when the genetic supervisor was employed.

### **P vs. Chlorophyll $a$ Model development with genetic supervisor**

Table 6-3 lists the performance of six simulations in terms of number of layers, neurons per layer, percentage of "Right" scores and number of epochs developed with genetic supervisor. The genetic parameters used for the simulations are detailed below the table and include initial value specifications for momentum, learning rate, input noise, mutation rate and crossover setting. The variations in the network size and in performance were probably due to the method of learning employed by the neural nets. The genetic supervisor specified the best model as a 3-layer network with 11 neurons in the hidden layer. Although there was no significant difference in the performance of networks suggested by the genetic supervisor, serial #3 (Figure 6-2) was selected as the network for chlorophyll  $a$  predictions because of the highest "Right" prediction scores. Smooth curves without spikes characterized the stability of the network model. The network architecture and network parameters suggested by the genetic supervisor was used for training on the given phosphorus vs. chlorophyll  $a$  training data. This network was adopted for all chlorophyll  $a$  prediction subsequently made in this study.

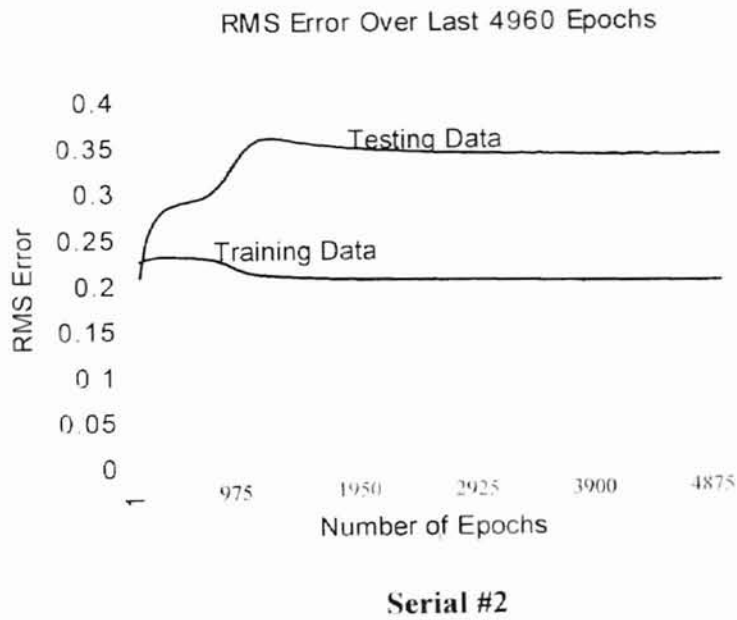
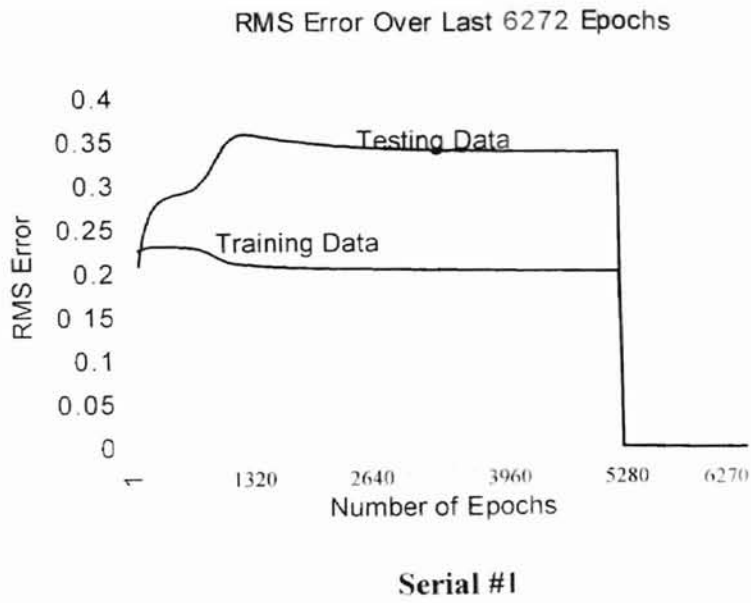
### **P vs. P network model**

Table 6-4 lists the outputs from several simulations developed with the genetic supervisor. Initial parameter specifications are detailed below the table and include

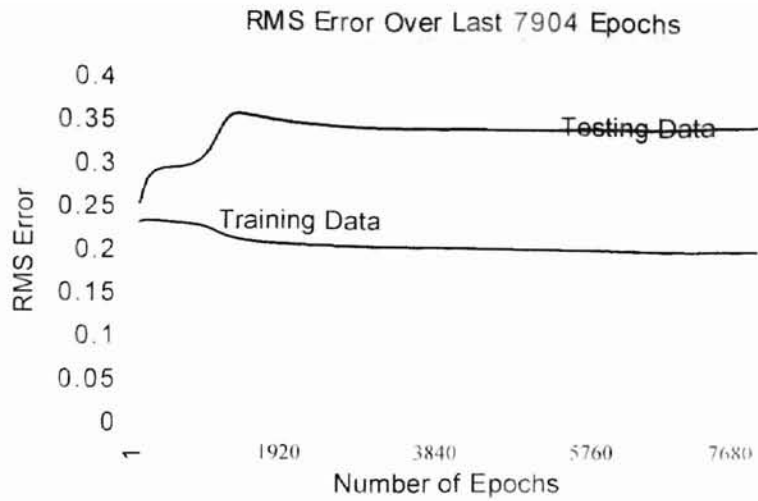
**Table 6-3. Alternatives for building Network Model -II  
(with genetic supervisor)**

Serial #	# of Layers	Neurons per layer	% Right	# of Epochs
1	3	1 18 1	66	6272
2	3	1 18 1	66	4960
3	3	1 11 1	68	7904
4	3	1 14 1	64	6912
5	3	1 26 1	64	4352
6	2	1 1	61	7424

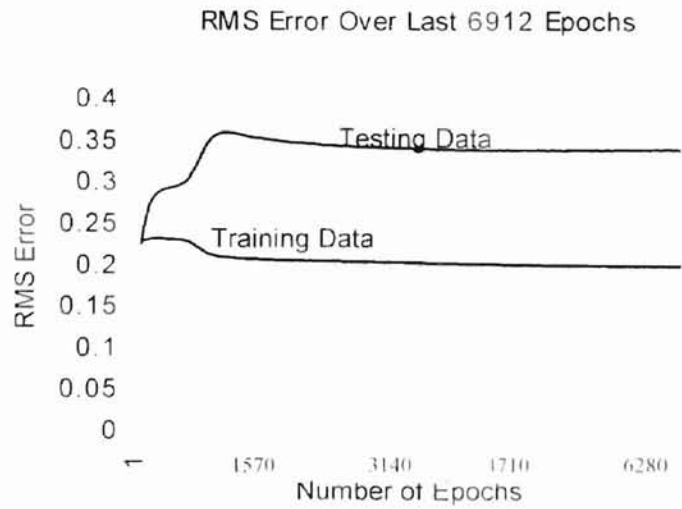
(Inclusion Rate = 0.75; Learning Rate = 0.5; Momentum = 1; Input Noise = 0.01;  
Crossovers = 1; Mutation Rate = 0.1 (except Serial #6 (0.2)); Fitness Limit = 200  
(except Serial #3 (100)); Pool Size = 3)



**Figure 6-2. RMS Error Plots of Various Alternatives Simulated with Genetic Supervisor for Network Model - II (cont'd on next page)**

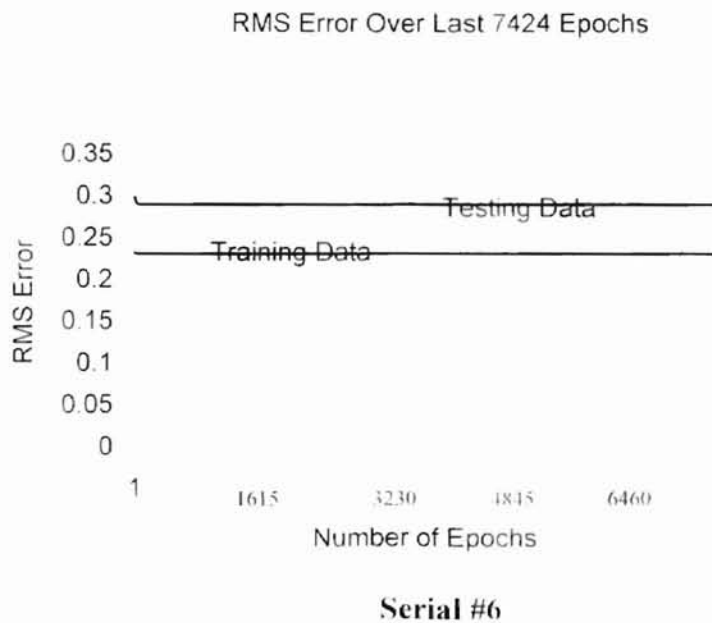
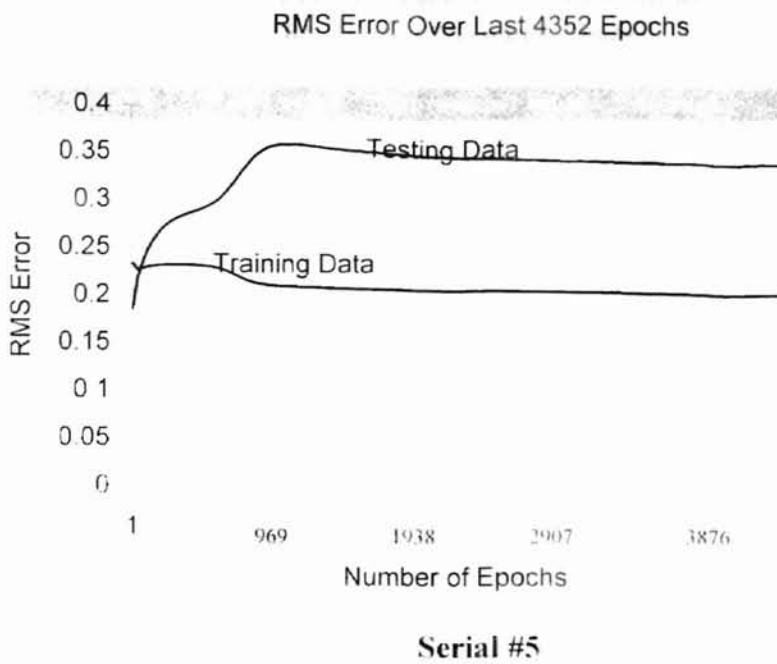


**Serial #3**



**Serial #4**

**Figure 6-2. RMS Error Plots of Various Alternatives Simulated with Genetic Supervisor for Network Model - II (cont'd on next page)**



**Figure 6-2. (cont'd from previous page) RMS Error Plots of Various Alternatives Simulated with Genetic Supervisor for Network Model - II. Fitness Limit = 200 (except Serial #3 (100)); Mutation Rate = 0.1 (except Serial #6 ( 0.2)); Crossovers = 1; Learning Rate = 0.5; Momentum = 1; Input Noise = 0.01**

**Table 6-4. Performance of Various Simulations for  
Developing Network Model – I**

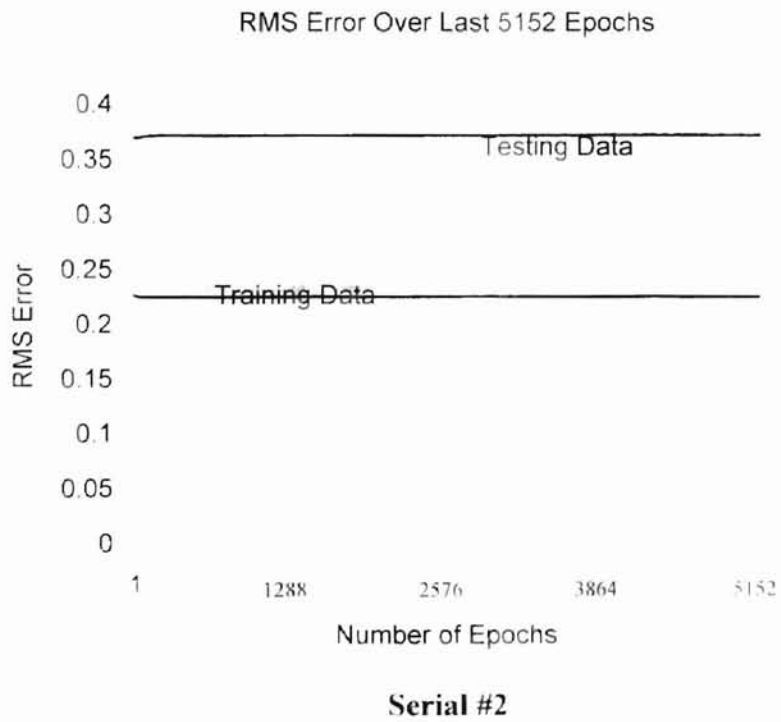
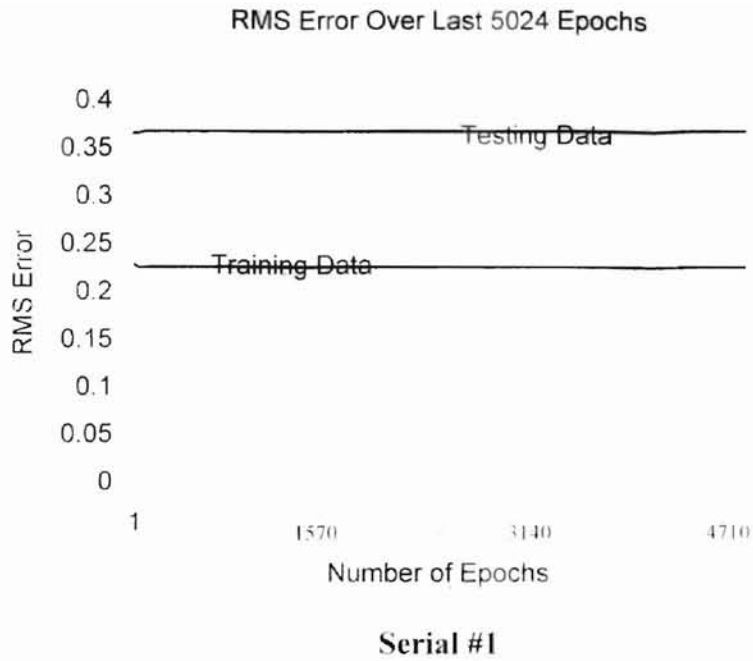
Serial #	# of Layers	Neurons per layer	% Right	# of Epochs
1	2	1 1	69	5024
2	2	1 1	72	5152
3	2	1 1	69	10624
4	2	1 1	69	6656
5	2	1 1	69	8000
6	2	1 1	72	10048
7	2	1 1	69	9536
8	2	1 1	69	5504

(Genetic Parameters : Inclusion Rate = 0.75; Learning Rate = 0.5; Momentum = 1; Input Noise = 0.01; Pool Size = 3; Crossovers = 1; Mutation Rate = 0.1; Fitness Limit = 100)

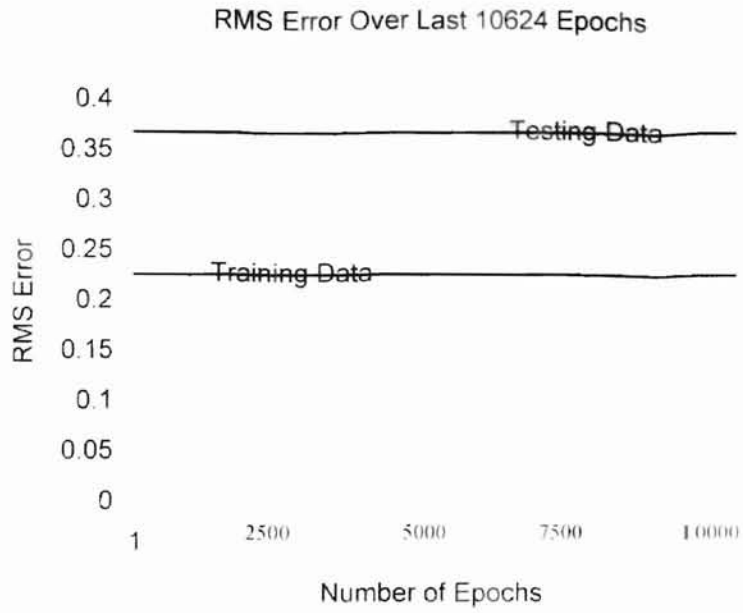


genetic parameters specified for building the model. These include mutation rate, crossover setting, initial specifications of momentum, learning rate and input noise. From the percentage of “Right” scores recorded in Table 6-4, which varied from 69% to 72%, and the similarities between RMS error plots (Figure 6-3), it was concluded that these simulations could be used interchangeably for making phosphorus predictions in the lake based on inputs from available data for the Illinois River at Tahlequah, Oklahoma. Although the percentage of “Right” predictions were approximately the same for each of these runs, the best scores were recorded with the 2<sup>nd</sup> and 6<sup>th</sup> serial runs. For this study, the network suggested by serial #2 was chosen for all phosphorus predictions and was based on the lesser number of epochs needed to attain the percentage of “Right” scores. This implied a much lower computational load on the network model and a faster learning rate.

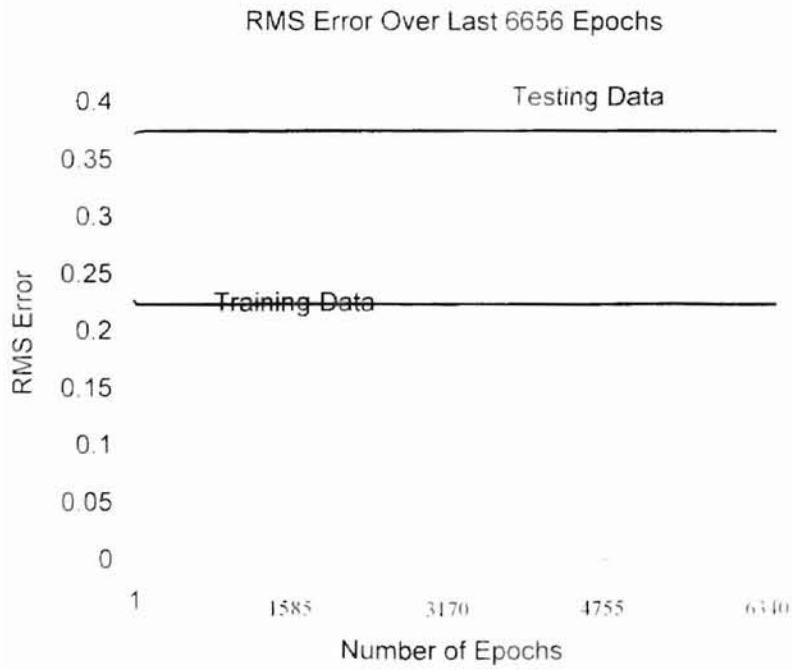
The network parameters generated for both networks with the genetic supervisor are given in Table 6-5. The parameter values differed from the initial values assigned before training the network. That is, the number of significant digits to which a genetic supervisor generated values was much higher than the values that were specified in the first and second phase of model development without the genetic supervisor. This also implied that the genetic supervisor provided a better network with a higher precision level than the trial and error process. This further established the appropriateness of selecting the 2<sup>nd</sup> simulation with the genetic supervisor for all phosphorus predictions in this study.



**Figure 6-3. RMS Error Plots of Simulations for Network Model - I**  
*(cont'd on next page)*

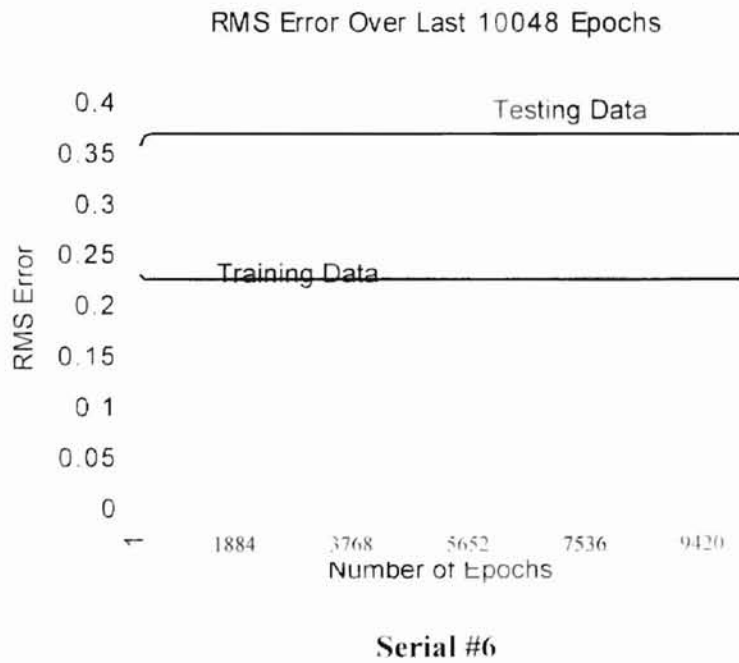
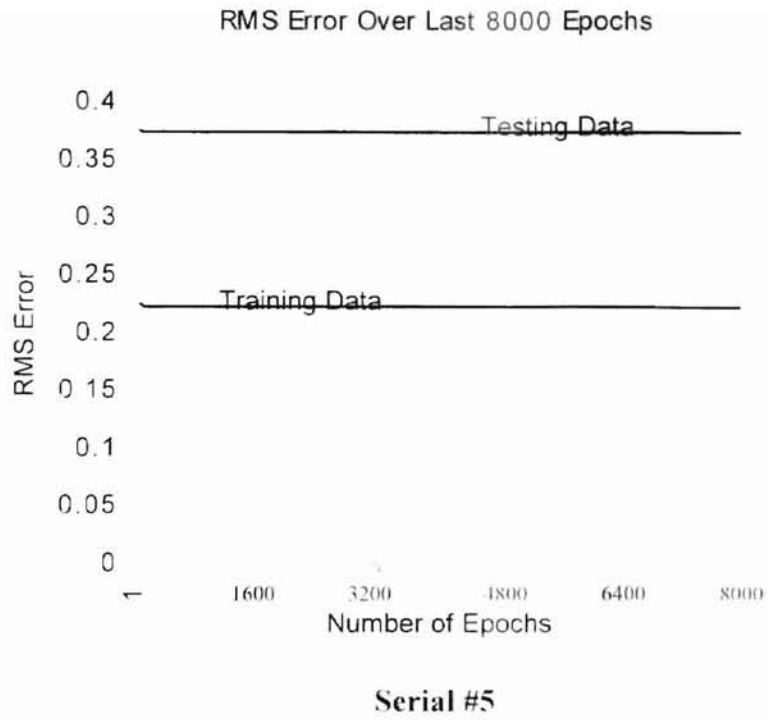


**Serial #3**



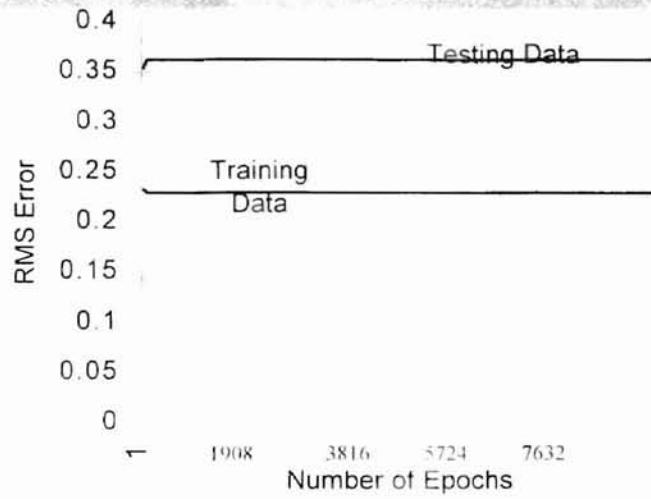
**Serial #4**

**Figure 6-3. RMS Error Plots of Simulations for Network Model - I**  
*(cont'd on next page)*



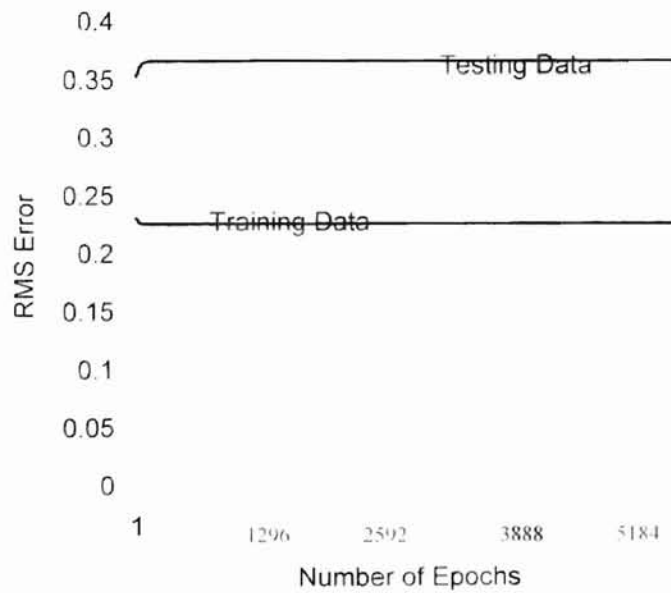
**Figure 6-3. RMS Error Plots of Simulations for Network Model - I**  
*(cont'd on next page)*

RMS Error Over Last 9536 Epochs



Serial #7

RMS Error Over Last 5504 Epochs



Serial #8

Figure 6-3. (cont'd from previous page) RMS Error Plots of Simulations for Network Model - 1

**Table 6-5. Network Parameters for Model - I and Model – II**

	<b>Network Model - I</b>	<b>Network Model - II</b>
Learning Rate	0.524476	0.589297
Momentum	0.882992	0.379467
Input Noise	0.00992	0.001749
Training Tolerance	0.2	0.2
Testing Tolerance	0.3	0.3
Epochs per Update	1	1
Epoch Limit	0	0
Time Limit	0	0
Error Limit	0.01	0.01

## **P vs. Secchi disk model**

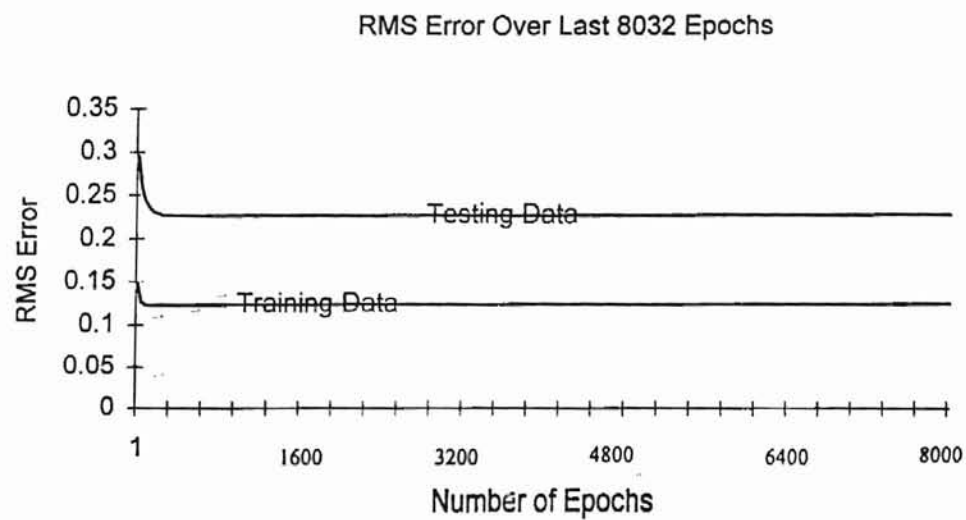
The genetic parameters specified for this model had identical values as those used for the P vs. chlorophyll *a* model to maintain consistency in approach. The network obtained from a single simulation with the genetic supervisor had a 95% “Right” score. Moreover, the horizontal “Training” error curve obtained in the RMS error plot displayed no spikes nor undulations, as shown in Figure 6-4, implying a highly stable network. The performance results in terms of percentage “Right” and stability established the selection of this network model for all secchi disk predictions in this study.

## **Network Models for Eutrophication Potentials under Phosphorus Reduction**

### **Alternatives**

These P vs. chlorophyll *a* and P vs. secchi disk value network models developed were used to evaluate the various phosphorus reduction alternatives. As mentioned, the aim was to evaluate point and non-point source phosphorus reductions needed to lower chlorophyll *a* concentrations and raise secchi disk values to levels where eutrophication in Lake Tenkiller would be reduced. Table 6-6 gives the results of chlorophyll *a* projections from the P vs. chlorophyll *a* network model before any phosphorus reductions were initiated in the lake. The input phosphorus concentrations were obtained by running the P vs. P network model on the phosphorus data set generated from pdf's prepared by Harton (1989) for phosphorus concentrations at Tahlequah.

It was observed that for higher values of phosphorus concentrations ( $> 0.09$  mg/L) the corresponding chlorophyll *a* values were less than  $10\mu\text{g/L}$ . This anomaly was due to the inclusion of a few data points in the training data set, as explained in Chapter 5, which had negative correlation between phosphorus and chlorophyll *a*. As these were



**Figure 6-4. RMS Error Plot for Network Model - III**



**Table 6-6. Chlorophyll *a* Predictions in the Lake Using Network Model - II**

Serial #	Input P (mg/l)	Chlorophyll <i>a</i> (µg/l)
#X1	0.0667000	23.21814209
#X2	0.0726000	21.00621826
#X3	0.0906000	12.18229004
#X4	0.1315756	7.13416016
#X5	0.0745000	20.11732910
#X6	0.0876000	13.41025635
#X7	0.0747000	20.00007813
#X8	0.1297175	7.16268066
#X9	0.0831000	15.61425781
#X10	0.1394723	7.04859863
#X11	0.0950000	10.70239258
#X12	0.0813000	16.51106934
#X13	0.0792000	17.67407227
#X14	0.1060977	8.47462402
#X15	0.0842000	15.05177002
#X16	0.1629305	6.96937500
#X17	0.1709889	6.95828369
#X18	0.0712000	21.63683838
#X19	0.1124243	7.84558838
#X20	0.1145910	7.69664795
#X21	0.1091919	8.12604004
#X22	0.1233258	7.30845215
#X23	0.0990000	9.67882324
#X24	0.1195519	7.44154785
#X25	0.1699780	6.95986816
#X26	0.1041196	8.74873779
#X27	0.0686000	22.63822510
#X28	0.0700000	22.12327148
#X29	0.1169611	7.56038330
#X30	0.0925000	11.51364258
#X31	0.1502395	6.99472656
#X32	0.1456125	7.01215576
#X33	0.1558299	6.98046631
#X34	0.1552087	6.98205078
#X35	0.1778595	6.95194580
#X36	0.1247218	7.26884033
#X37	0.1018011	9.12901123
#X38	0.0624000	24.03731445
#X39	0.0773000	18.67070557

region-specific within the lake, the analysis focused upon those areas where the OWRB (1996) study indicated phosphorus limitation. Overall, chlorophyll *a* values in the lake indicated a highly eutrophic state of the water body.

Chlorophyll *a* predictions obtained with the P vs. chlorophyll *a* model with phosphorus concentrations modified to remove the various point source contributions (alternative C1) at Tahlequah are given in Table 6-7. Figure 6-5 is a plot of the chlorophyll *a* output obtained with point source reductions in phosphorus loading with plots of chlorophyll *a* without any phosphorus reductions and breakpoint level of chlorophyll *a* between mesotrophic and eutrophic range. The similarity between chlorophyll *a* predictions without phosphorus reductions and with point source phosphorus reductions can be gauged by the similarity of the two plots. Since total point source contribution did not have any effect on eutrophication of the lake, it was removed from additional analysis. The remainder of the phosphorus control alternatives evaluated focused on non-point source reductions in phosphorus loading.

Chlorophyll *a* predictions obtained with 50% reduction in non-point source phosphorus loading (alternative C2) are given in Table 6-8. The value 50% was chosen to provide a benchmark for assessing reductions required in phosphorus loading from non-point sources for controlling eutrophication in the lake. From the neural network model predictions, it was concluded that a 50% reduction in non-point sources did not improve the eutrophic state of the water body as more than 95% of the randomly simulated chlorophyll *a* values remained above the threshold level of 10µg/L. It is to be noted that in terms of number of data points there seems to be an increase in the number of simulated output data points greater than 10µg/L when we compare chlorophyll *a*

**Table 6-7. Chlorophyll *a* Predictions at Tahlequah Using Network Model - II without Point Source Phosphorus Contribution**

Serial #	Input-P Tahlequah (mg/l)	Chlorophyll <i>a</i> Tahlequah (µg/l)
#X1	6.28E-02	21.15357422
#X2	6.87E-02	20.28369873
#X3	8.67E-02	15.29894775
#X4	1.28E-01	7.72675293
#X5	7.06E-02	19.89391846
#X6	8.37E-02	16.23854004
#X7	7.08E-02	19.85272217
#X8	1.26E-01	7.85509522
#X9	7.92E-02	17.62495361
#X10	1.36E-01	7.34331055
#X11	9.11E-02	13.94897705
#X12	7.74E-02	18.15575195
#X13	7.53E-02	18.73883789
#X14	1.02E-01	11.08741943
#X15	8.03E-02	17.29538330
#X16	1.59E-01	6.90916504
#X17	1.67E-01	6.86955322
#X18	6.73E-02	20.53879883
#X19	1.09E-01	9.74061768
#X20	1.11E-01	9.42689209
#X21	1.05E-01	10.46155273
#X22	1.19E-01	8.43342773
#X23	9.51E-02	12.80023438
#X24	1.16E-01	8.75982910
#X25	1.66E-01	6.87272217
#X26	1.00E-01	11.54691650
#X27	6.47E-02	20.93333252
#X28	6.61E-02	20.73368896
#X29	1.13E-01	9.14010254
#X30	8.86E-02	14.70952393
#X31	1.46E-01	7.07078125
#X32	1.42E-01	7.15951172
#X33	1.52E-01	6.97888184
#X34	1.51E-01	6.98997315
#X35	1.74E-01	6.85370850
#X36	1.21E-01	8.24645996
#X37	9.79E-02	12.06187012
#X38	5.85E-02	21.42451904
#X39	7.34E-02	19.23636230

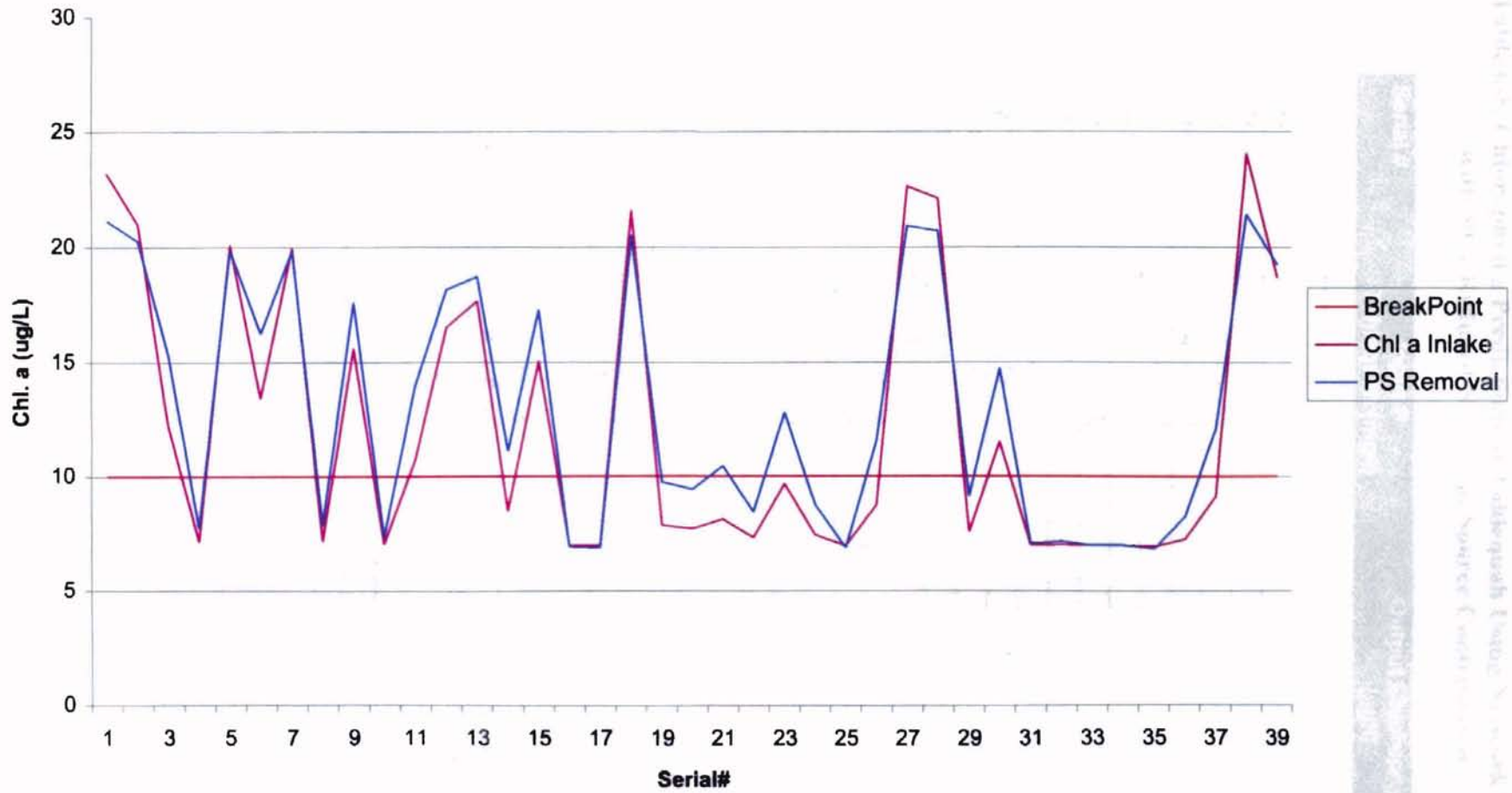


Fig 6-5. Chl. a Values with Varying Phosphorus Reductions (Plot 1)

City of New York  
Department of Environmental Protection  
Office of Water  
Water Quality Control Division  
110 Broadway, 11th Floor  
New York, NY 10038  
Tel: (212) 312-2000  
Fax: (212) 312-2001  
www.doe.state.ny.us

**Table 6-8. Chlorophyll *a* Predictions at Tahlequah Using Network Model - II  
with 50% Reduced Non-Point Source Contribution**

Serial #	Input-P Tahlequah (mg/l)	Output Chlorophyll <i>a</i> ( $\mu\text{g/l}$ )
#X1	3.53E-02	15.71249512
#X2	3.83E-02	17.12742920
#X3	4.73E-02	21.29617676
#X4	6.77E-02	23.78062988
#X5	3.92E-02	17.55682129
#X6	4.58E-02	20.65446533
#X7	3.93E-02	17.60593994
#X8	6.68E-02	23.99770264
#X9	4.35E-02	19.61029785
#X10	7.17E-02	22.36094238
#X11	4.95E-02	22.16446777
#X12	4.26E-02	19.18724365
#X13	4.15E-02	18.66436768
#X14	5.50E-02	23.82657959
#X15	4.40E-02	19.83846191
#X16	8.34E-02	13.92204102
#X17	8.74E-02	10.07811035
#X18	3.76E-02	16.78993652
#X19	5.82E-02	24.38114502
#X20	5.92E-02	24.48572021
#X21	5.65E-02	24.12921387
#X22	6.36E-02	24.48572021
#X23	5.15E-02	22.86163574
#X24	6.17E-02	24.57761963
#X25	8.69E-02	10.56612793
#X26	5.40E-02	23.58732422
#X27	3.62E-02	16.13238037
#X28	3.70E-02	16.50314697
#X29	6.04E-02	24.55860596
#X30	4.82E-02	21.66535889
#X31	7.71E-02	19.21417969
#X32	7.48E-02	20.72893555
#X33	7.99E-02	17.04503662
#X34	7.96E-02	17.29379883
#X35	9.09E-02	6.83311035
#X36	6.43E-02	24.41441895
#X37	5.29E-02	23.29261230
#X38	3.32E-02	14.76656494
#X39	4.06E-02	18.23180664

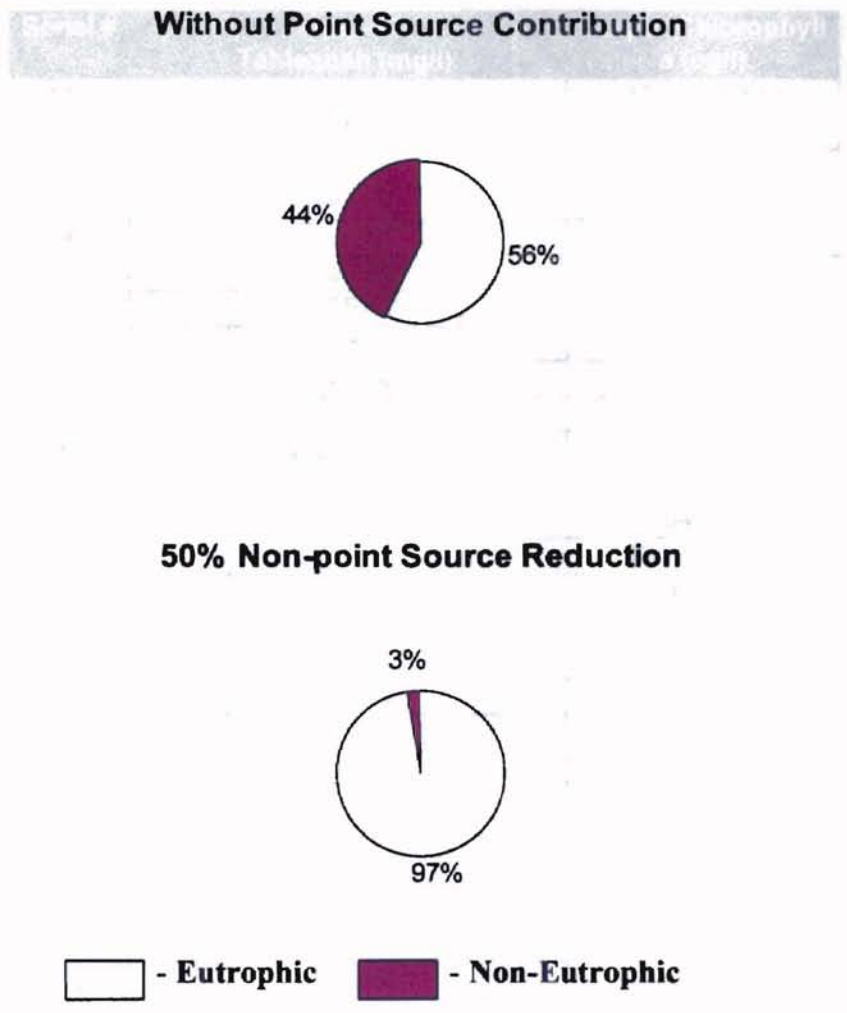
predictions after point source phosphorus reductions as against 50% non-point source reductions. This is demonstrated by comparison of the two pie diagrams in Figure 6-6, which has pie chart representations of all the chlorophyll *a* predictions as eutrophic and non-eutrophic categories for point source reduction, and 50% non-point source phosphorus reduction alternatives. This can be ascribed to the anomalous points in the training data set, where higher input phosphorus values had lower chlorophyll *a* readings against them. Consequently with reductions in phosphorus inputs the chlorophyll *a* values were higher. This anomaly stabilizes with further reductions in phosphorus values because with further reductions the ranges of input phosphorus values correspond to training data areas where the anomaly between phosphorus and chlorophyll *a* relation was absent.

Table 6-9 gives the simulated chlorophyll *a* values generated with a 75% reduction in non-point source phosphorus loading (alternative C3). Even with 75% reduction in phosphorus loading from the non-point sources, more than 95% of chlorophyll *a* values obtained were still above the breakpoint level of 10µg/L, between a eutrophic and mesotrophic state.

Table 6-10 gives the chlorophyll *a* values obtained with an 80% reduction in phosphorus loading from non-point sources (alternative C4). With this reduction in non-point source phosphorus contribution, approximately 72% of the values obtained for chlorophyll *a* were above the breakpoint level of 10µg/L, indicating a significant water quality problem.

An 85% reduction in non-point source loading of phosphorus inputs was simulated (alternative C5). The values generated by this are given in Table 6-11, where

Table 6-6. Chlorophyll *a* Predictions at Tablequah Using Scenario Model - II  
in 25% Reduced Non-Point Source Contribution



**Figure 6-6. Management Alternatives with Chl. *a* Predictions (Pie 1)**

**Table 6-9. Chlorophyll *a* Predictions at Tahlequah Using Network Model – II  
with 75% Reduced Non-Point Source Contribution**

Serial #	Input-P Tahlequah (mg/l)	Output Chlorophyll <i>a</i> (µg/l)
#X1	1.96E-02	10.14465820
#X2	2.11E-02	10.51542480
#X3	2.56E-02	11.83370605
#X4	3.58E-02	15.94382813
#X5	2.16E-02	10.63267578
#X6	2.49E-02	11.59286621
#X7	2.16E-02	10.64852051
#X8	3.54E-02	15.73467773
#X9	2.37E-02	11.24111328
#X10	3.78E-02	16.88342041
#X11	2.67E-02	12.19971924
#X12	2.33E-02	11.10801758
#X13	2.27E-02	10.94798584
#X14	2.95E-02	13.21378174
#X15	2.40E-02	11.31399902
#X16	4.37E-02	19.68001465
#X17	4.57E-02	20.58950195
#X18	2.08E-02	10.42669434
#X19	3.11E-02	13.85232422
#X20	3.16E-02	14.05830566
#X21	3.02E-02	13.50849365
#X22	3.38E-02	15.00740479
#X23	2.77E-02	12.55305664
#X24	3.28E-02	14.59385742
#X25	4.54E-02	20.47700439
#X26	2.90E-02	13.01889160
#X27	2.01E-02	10.25240234
#X28	2.05E-02	10.35063965
#X29	3.22E-02	14.31023682
#X30	2.61E-02	11.98264648
#X31	4.05E-02	18.18427246
#X32	3.94E-02	17.62970703
#X33	4.19E-02	18.85291992
#X34	4.18E-02	18.78161865
#X35	4.74E-02	21.33895752
#X36	3.41E-02	15.16268311
#X37	2.84E-02	12.81291016
#X38	1.86E-02	9.90381836
#X39	2.23E-02	10.82122803



**Table 6-10. Chlorophyll *a* Predictions at Tahlequah Using Network Model - II with 80% Reduced Non-Point Source Contribution**

Serial #	Input-P Tahlequah (mg/l)	Output Chlorophyll <i>a</i> (µg/l)
#X1	1.65E-02	9.46967285
#X2	1.77E-02	9.71209716
#X3	2.13E-02	10.55662109
#X4	2.94E-02	13.20110596
#X5	1.80E-02	9.78498291
#X6	2.07E-02	10.40292725
#X7	1.81E-02	9.79607422
#X8	2.91E-02	13.06008789
#X9	1.97E-02	10.17793213
#X10	3.10E-02	13.83647949
#X11	2.21E-02	10.79112305
#X12	1.94E-02	10.09395508
#X13	1.89E-02	9.99413330
#X14	2.43E-02	11.43125000
#X15	1.99E-02	10.22546631
#X16	3.57E-02	15.89470947
#X17	3.73E-02	16.64891846
#X18	1.74E-02	9.65347168
#X19	2.56E-02	11.83845947
#X20	2.60E-02	11.97472412
#X21	2.49E-02	11.61821777
#X22	2.78E-02	12.58791504
#X23	2.29E-02	11.01928711
#X24	2.70E-02	12.31380127
#X25	3.71E-02	16.55226563
#X26	2.39E-02	11.31083008
#X27	1.68E-02	9.53938965
#X28	1.71E-02	9.60276856
#X29	2.65E-02	12.13158691
#X30	2.16E-02	10.65485840
#X31	3.32E-02	14.75547363
#X32	3.23E-02	14.35935547
#X33	3.43E-02	15.25458252
#X34	3.42E-02	15.20229492
#X35	3.87E-02	17.31598145
#X36	2.81E-02	12.68615234
#X37	2.35E-02	11.17931885
#X38	1.56E-02	9.30805664
#X39	1.86E-02	9.90857178

**Table 6-11. Chlorophyll *a* Predictions at Tahlequah Using Network Model - II  
with 85% Reduced Non-Point Source Contribution**

Serial #	Input-P Tahlequah (mg/l)	Output Chlorophyll <i>a</i> ( $\mu$ g/l)
#X1	1.33E-02	8.90876953
#X2	1.42E-02	9.05770996
#X3	1.69E-02	9.55681885
#X4	2.30E-02	11.03513184
#X5	1.45E-02	9.10841309
#X6	1.65E-02	9.47601074
#X7	1.45E-02	9.10841309
#X8	2.28E-02	10.97650635
#X9	1.58E-02	9.34133057
#X10	2.42E-02	11.38846924
#X11	1.76E-02	9.70100586
#X12	1.55E-02	9.28745850
#X13	1.52E-02	9.23200195
#X14	1.92E-02	10.04958984
#X15	1.59E-02	9.36034424
#X16	2.78E-02	12.59108398
#X17	2.90E-02	13.03948975
#X18	1.40E-02	9.02443604
#X19	2.02E-02	10.28884521
#X20	2.05E-02	10.36331543
#X21	1.97E-02	10.16842529
#X22	2.18E-02	10.70239258
#X23	1.82E-02	9.82934815
#X24	2.12E-02	10.54077637
#X25	2.88E-02	12.96026611
#X26	1.89E-02	9.98145752
#X27	1.36E-02	8.95788818
#X28	1.38E-02	8.99116211
#X29	2.09E-02	10.46630615
#X30	1.72E-02	9.61702881
#X31	2.59E-02	11.93352783
#X32	2.52E-02	11.70377930
#X33	2.67E-02	12.19971924
#X34	2.66E-02	12.16802979
#X35	3.00E-02	13.42927002
#X36	2.20E-02	10.75309570
#X37	1.86E-02	9.91332520
#X38	1.27E-02	8.81687012
#X39	1.49E-02	9.17971436

approximately 50% of the values generated were below the threshold level of  $10\mu\text{g/L}$ , increase the mean trophic response in approximately 50% to 75% of projected samples while having a greater relative impact on eutrophication control in the lacustrine zone of the lake.

The results of the remaining chlorophyll  $a$  simulations obtained from the 75%, 80%, and 85% non-point source phosphorus reductions considered are plotted in Figure 6-7. Figure 6-5 demonstrated that the plot obtained by point source removal follows the chlorophyll  $a$  at Tahlequah plot. This tells us that point source removal alone can have no bearing on the overall improvement of the eutrophication state of the lake. Figure 6-7 shows the improvement in the eutrophication state of the lake with greater reductions in non-point source phosphorus contributions at Tahlequah. This is concluded from the fact that with greater reductions the plots move closer to the threshold line of  $10\mu\text{g/L}$ . The maximum effect is achieved by 85% reduction in non-point source phosphorus reductions. This deduction is made by observing that the 85% reduction of non-point source plot has the maximum consistency in staying below the breakpoint plot of  $10\mu\text{g/L}$ . Similar deductions can be made from Figure 6-8 which is a pie chart representation of chlorophyll  $a$  predictions distributed into eutrophic and non-eutrophic categories using the breakpoint values of  $10\mu\text{g/L}$  for the 75%, 80%, and 85% non-point source phosphorus reduction alternatives. It is observed that the non-eutrophic slice increases with greater reductions in non-point source phosphorus loads.

Each of the phosphorus reduction levels was evaluated with the P vs. secchi disk model previously developed. Secchi disk values greater than 1.98 meters would establish a non-eutrophic state in the water body. The initial secchi disk values without phosphorus

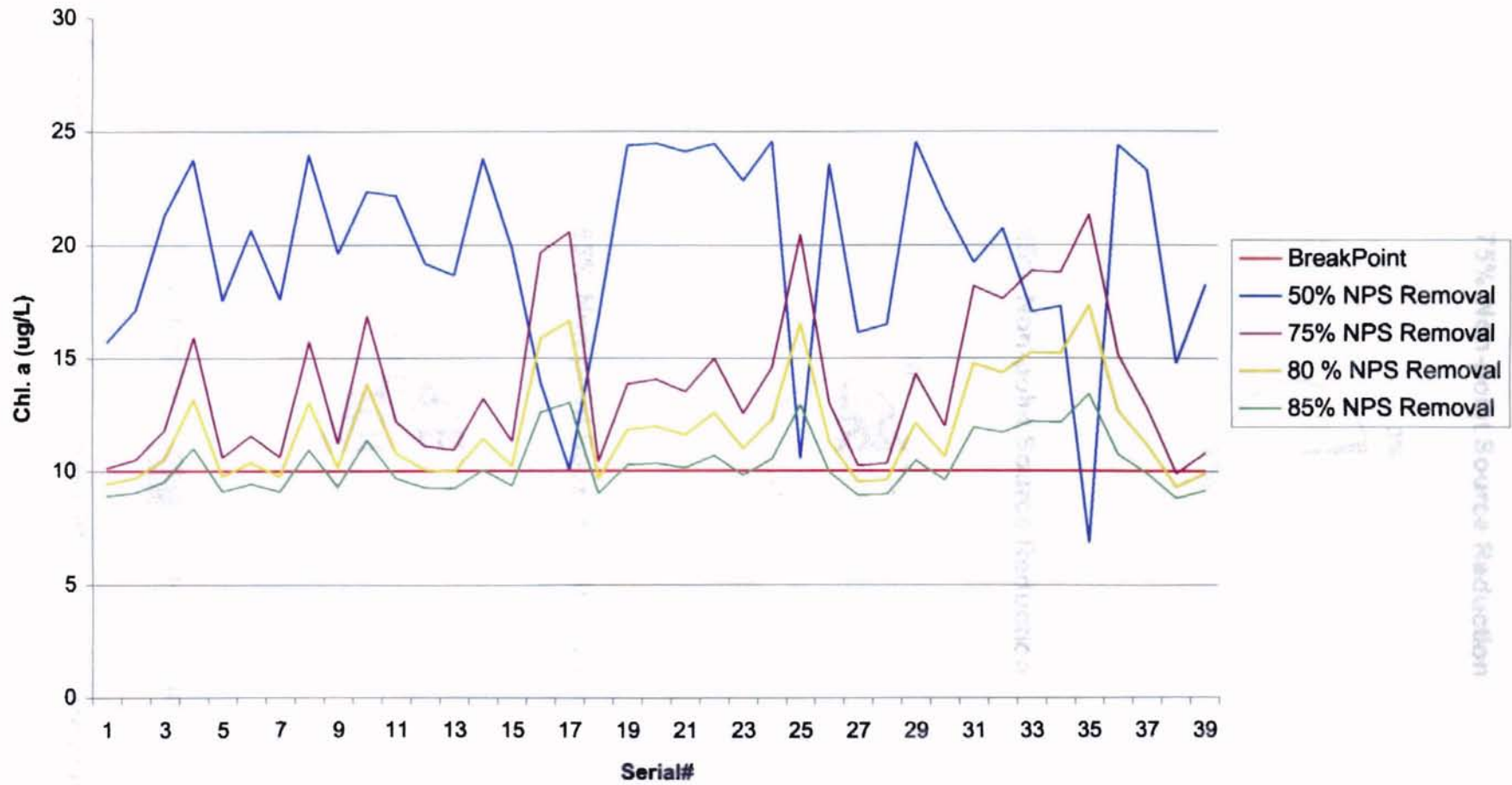
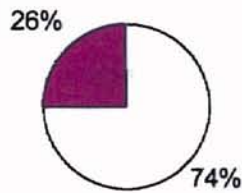


Fig 6-7. Chl. a Values with Varying Phosphorus Reductions (Plot 2)

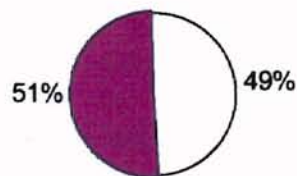
75% Non-point Source Reduction



85% Non-point Source Reduction



85% Non-point Source Reduction



□ - Eutrophic    ■ - Non- Eutrophic

Figure 6-8. Management Alternatives with Chl. a Predictions (Pie 2)

reductions in the lake are presented in Table 6-12. From Table 6-12 it was evident that the lake was highly eutrophic as 100% of the secchi values predicted was less than the threshold value of 1.98 meters.

Results of secchi disk predictions with the removal of point source contribution at Tahlequah are given in Table 6-13 (alternative S1). All the predicted values with the implementation of this alternative were still less than the threshold value, again establishing that point source contributions had minimal impact on the eutrophication of the lake.

The effect of 50% non-point source phosphorus reduction is given in Table 6-14 (alternative S2). Though 99% of the secchi values were still below the threshold level of 1.98 meters, there was a significant improvement in the water quality compared to alternative S1, as observed by increases in the actual secchi disk values.

Table 6-15 gives the secchi values predicted with 75% reduction in non-point source phosphorus loading (alternative S3). Approximately 72% of the secchi disk values were greater than the threshold value of 1.98 meters. Contrary to observations made by chlorophyll *a* predictions, a 75% reduction in non-point source phosphorus loading suggested a significant improvement in eutrophication control in the lake. This goes to the eutrophication indicator used, as the phosphorus levels used were the same. Obviously chlorophyll *a* is a more restrictive metric.

Table 6-16 gives the secchi disk values predicted with an 80% reduction in non-point source phosphorus loading (alternative S4). Approximately 89% of the secchi disk values were greater than the threshold value. Secchi disk values, if used as an indicator of the eutrophication state, strongly favored the implementation of 80% non-point source

Table 6-12. Secchi disk in the Lake Using Network Model – III

Serial #	Input - P (mg/l)	Secchi (m)
#X1	0.0667000	1.31754425
#X2	0.0726000	1.21794403
#X3	0.0906000	0.94050690
#X4	0.1315756	0.45405365
#X5	0.0745000	1.18763092
#X6	0.0876000	0.98395569
#X7	0.0747000	1.18387787
#X8	0.1297175	0.47180847
#X9	0.0831000	1.05266541
#X10	0.1394723	0.38274567
#X11	0.0950000	0.87858154
#X12	0.0813000	1.07908112
#X13	0.0792000	1.11300293
#X14	0.1060977	0.73278992
#X15	0.0842000	1.03592102
#X16	0.1629305	0.20837311
#X17	0.1709889	0.15987213
#X18	0.0712000	1.24161713
#X19	0.1124243	0.65628540
#X20	0.1145910	0.63131317
#X21	0.1091919	0.69468201
#X22	0.1233258	0.53589905
#X23	0.0990000	0.82401794
#X24	0.1195519	0.57602783
#X25	0.169978	0.16579041
#X26	0.1041196	0.75761780
#X27	0.0686000	1.28564331
#X28	0.0700000	1.26168152
#X29	0.1169611	0.60446442
#X30	0.0925000	0.91437988
#X31	0.1502395	0.29613678
#X32	0.1456125	0.33207947
#X33	0.1558299	0.25571930
#X34	0.1552087	0.26004974
#X35	0.1778595	0.12277466
#X36	0.1247218	0.52146423
#X37	0.1018011	0.78749786
#X38	0.0624000	1.39101746
#X39	0.0773000	1.14244995

**Table 6-13. Secchi disk at Tahlequah Using Network Model - III  
after Total Point Source P-reduction from Inputs**

Serial #	Input - P Tahlequah (mg/l)	Secchi Tahlequah (m)
#X1	6.28E-02	1.38481049
#X2	6.87E-02	1.28376679
#X3	8.67E-02	0.99795746
#X4	1.28E-01	0.48840851
#X5	7.06E-02	1.25186585
#X6	8.37E-02	1.04299408
#X7	7.08E-02	1.24854584
#X8	1.26E-01	0.50847290
#X9	7.92E-02	1.11256989
#X10	1.36E-01	0.41334747
#X11	9.11E-02	0.93357819
#X12	7.74E-02	1.14100647
#X13	7.53E-02	1.17463959
#X14	1.02E-01	0.78489960
#X15	8.03E-02	1.09524811
#X16	1.59E-01	0.23406708
#X17	1.67E-01	0.18340088
#X18	6.73E-02	1.30729553
#X19	1.09E-01	0.69713593
#X20	1.11E-01	0.67317413
#X21	1.05E-01	0.74664734
#X22	1.19E-01	0.58194611
#X23	9.51E-02	0.87742676
#X24	1.16E-01	0.61529053
#X25	1.66E-01	0.18931915
#X26	1.00E-01	0.81117096
#X27	6.47E-02	1.35175476
#X28	6.61E-02	1.32764862
#X29	1.13E-01	0.64964539
#X30	8.86E-02	0.96995392
#X31	1.46E-01	0.32890381
#X32	1.42E-01	0.36138214
#X33	1.52E-01	0.28314545
#X34	1.51E-01	0.29050720
#X35	1.74E-01	0.14327210
#X36	1.21E-01	0.56043823
#X37	9.79E-02	0.83917450
#X38	5.85E-02	1.46044891
#X39	7.34E-02	1.20553009



**Table 6-14. Secchi disk at Tahlequah Using Network Model - III  
with 50% Non-Point source Reduction in Phosphorus**

Serial #	Input-P Tahlequah (mg/l)	Secchi Tahlequah (m)
#X1	3.53E-02	1.97476135
#X2	3.83E-02	1.92496124
#X3	4.73E-02	1.77714874
#X4	6.79E-02	1.45149932
#X5	3.92E-02	1.91009338
#X6	4.58E-02	1.80139923
#X7	3.93E-02	1.90836121
#X8	6.69E-02	1.46665588
#X9	4.35E-02	1.83921845
#X10	7.19E-02	1.39087311
#X11	4.95E-02	1.74120606
#X12	4.26E-02	1.85394196
#X13	4.16E-02	1.87039764
#X14	5.49E-02	1.65459717
#X15	4.41E-02	1.82925842
#X16	8.34E-02	1.22342926
#X17	8.74E-02	1.16742218
#X18	3.76E-02	1.93636475
#X19	5.84E-02	1.59887878
#X20	5.94E-02	1.58314484
#X21	5.64E-02	1.63063538
#X22	6.34E-02	1.52064209
#X23	5.15E-02	1.70887207
#X24	6.19E-02	1.54388214
#X25	8.69E-02	1.17435089
#X26	5.39E-02	1.67047546
#X27	3.63E-02	1.95801697
#X28	3.70E-02	1.94646912
#X29	6.04E-02	1.56741089
#X30	4.82E-02	1.76242523
#X31	7.69E-02	1.31682251
#X32	7.49E-02	1.34626953
#X33	7.99E-02	1.27322937
#X34	7.94E-02	1.28059113
#X35	9.09E-02	1.11964294
#X36	6.44E-02	1.50505249
#X37	5.29E-02	1.68635376
#X38	3.32E-02	2.00954926
#X39	4.06E-02	1.88699768

**Table 6-15. Secchi disk at Tahlequah Using Network Model - III  
with 75% Non-Point Source Reduction in Phosphorus**

Serial #	Input - P Tahlequah (mg/l)	Secchi Tahlequah (m)
#X1	1.96E-02	2.23646454
#X2	2.11E-02	2.21134796
#X3	2.56E-02	2.13628693
#X4	3.59E-02	1.96465698
#X5	2.16E-02	2.20297577
#X6	2.49E-02	2.14797913
#X7	2.16E-02	2.20297577
#X8	3.54E-02	1.97302918
#X9	2.37E-02	2.16804352
#X10	3.79E-02	1.93145691
#X11	2.67E-02	2.11795471
#X12	2.33E-02	2.17468353
#X13	2.28E-02	2.18291138
#X14	2.94E-02	2.07291809
#X15	2.40E-02	2.16299133
#X16	4.37E-02	1.83589844
#X17	4.57E-02	1.80313141
#X18	2.08E-02	2.21640015
#X19	3.12E-02	2.04274933
#X20	3.17E-02	2.03452148
#X21	3.02E-02	2.05963806
#X22	3.37E-02	2.00117706
#X23	2.77E-02	2.10135468
#X24	3.29E-02	2.01445709
#X25	4.54E-02	1.80803925
#X26	2.89E-02	2.08129028
#X27	2.01E-02	2.22809235
#X28	2.05E-02	2.22145233
#X29	3.22E-02	2.02614929
#X30	2.61E-02	2.12805908
#X31	4.04E-02	1.89017334
#X32	3.94E-02	1.90662903
#X33	4.19E-02	1.86534546
#X34	4.17E-02	1.86880981
#X35	4.74E-02	1.77541657
#X36	3.42E-02	1.99294922
#X37	2.84E-02	2.08951813
#X38	1.86E-02	2.25320892
#X39	2.23E-02	2.19142792

**Table 6-16. Secchi disk at Tahlequah Using Network Model - III with 80% Non-Point Source Reduction in Phosphorus**

Serial #	Input - P Tahlequah (mg/l)	Secchi Tahlequah (m)
#X1	1.65E-02	2.28871858
#X2	1.77E-02	2.26894287
#X3	2.13E-02	2.20860535
#X4	2.95E-02	2.07118591
#X5	1.81E-02	2.26215851
#X6	2.07E-02	2.21813232
#X7	1.81E-02	2.26215851
#X8	2.91E-02	2.07797028
#X9	1.97E-02	2.23415497
#X10	3.11E-02	2.04448151
#X11	2.21E-02	2.19402618
#X12	1.94E-02	2.23949585
#X13	1.90E-02	2.24613586
#X14	2.43E-02	2.15793915
#X15	2.00E-02	2.23011322
#X16	3.57E-02	1.96725525
#X17	3.73E-02	1.94083954
#X18	1.74E-02	2.27284027
#X19	2.57E-02	2.13412170
#X20	2.61E-02	2.12733734
#X21	2.49E-02	2.14740173
#X22	2.77E-02	2.10063294
#X23	2.29E-02	2.18074616
#X24	2.71E-02	2.11131470
#X25	3.71E-02	1.94473694
#X26	2.39E-02	2.16457916
#X27	1.69E-02	2.28222290
#X28	1.72E-02	2.27688202
#X29	2.65E-02	2.12069733
#X30	2.17E-02	2.20210968
#X31	3.31E-02	2.01128143
#X32	3.23E-02	2.02441711
#X33	3.43E-02	1.99121704
#X34	3.41E-02	1.99381531
#X35	3.87E-02	1.91832123
#X36	2.81E-02	2.09399292
#X37	2.35E-02	2.17136353
#X38	1.57E-02	2.30199860
#X39	1.86E-02	2.25277588

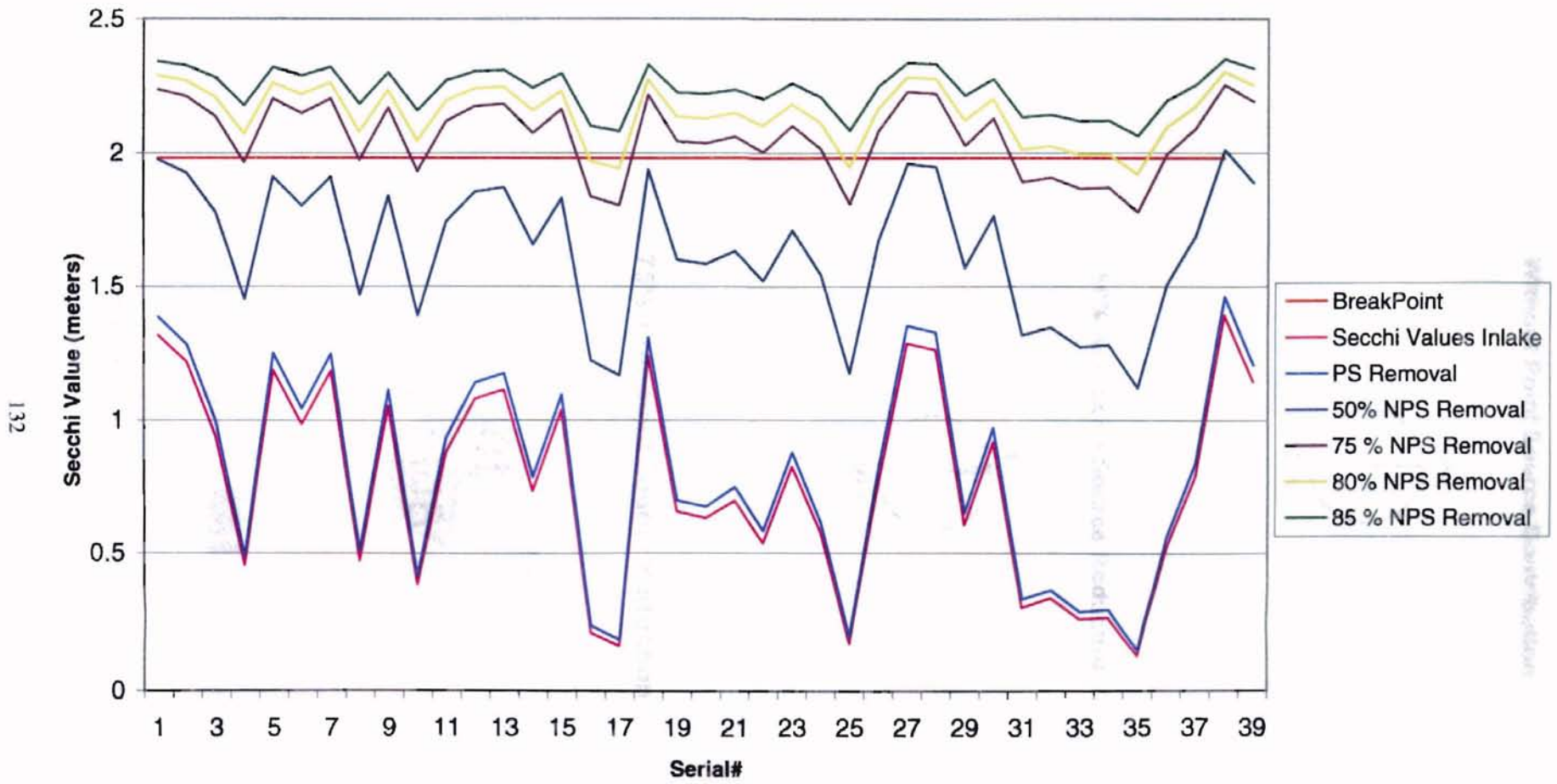
phosphorus reduction for eutrophication control in lake. Compared to this 89% success rate in lowering eutrophication with secchi disk predictions, a similar non-point source phosphorus reduction yielded only a 28% success rate when chlorophyll *a* was used as the metric measurement. To establish a definite and positive management strategy, further reductions in non-point source phosphorus were done for the secchi disk predictions.

Table 6-17 gives the secchi values predicted with 85% reduction in non-point source phosphorus loading (alternative S5). All of the values predicted by P vs. secchi disk model values were above the threshold, indicating a non-eutrophic state of the water body. Only 50% of the corresponding chlorophyll *a* values were positive towards the non-eutrophic state for the same phosphorus reduction level.

Figure 6-9 is a plot of all the secchi values obtained by various phosphorus reductions considered. As observed with chlorophyll values, point source removal alone does not have any effect on the reduction of secchi values. This can be deduced by comparison of point source removal plot against the plot of secchi disk values at Tahlequah. But there is considerable improvement in the eutrophication state of the lake with any of the other alternatives. This is reasoned by observing the movement of plots closer to the breakpoint of secchi values. It is observed that 85% non-point source phosphorus reduction has all secchi values above the breakpoint of 1.98 m, implying a complete remediation of the lake with this alternative. The above deductions can also be made from the pie chart representation of results in Figure 6-10. It is observed that the slice corresponding to the non-eutrophic data range increases from a zero to 100% as phosphorus reductions are applied to the model.

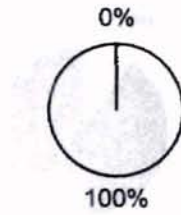
**Table 6-17. Secchi disk at Tahlequah Using Network Model - III  
with 85% Non-Point Source Reduction in Phosphorus**

<b>Serial #</b>	<b>Input - P Tahlequah (mg/l)</b>	<b>Secchi Tahlequah (m)</b>
#X1	1.33E-02	2.34097260
#X2	1.42E-02	2.32610474
#X3	1.69E-02	2.28106812
#X4	2.31E-02	2.17814789
#X5	1.45E-02	2.32105255
#X6	1.65E-02	2.28814117
#X7	1.45E-02	2.32105255
#X8	2.28E-02	2.18291138
#X9	1.58E-02	2.30012207
#X10	2.43E-02	2.15793915
#X11	1.76E-02	2.27024200
#X12	1.55E-02	2.30401947
#X13	1.52E-02	2.30907166
#X14	1.92E-02	2.24324890
#X15	1.60E-02	2.29723511
#X16	2.78E-02	2.09991119
#X17	2.90E-02	2.07999115
#X18	1.40E-02	2.32899170
#X19	2.03E-02	2.22520538
#X20	2.06E-02	2.22015320
#X21	1.97E-02	2.23516541
#X22	2.18E-02	2.20008881
#X23	1.82E-02	2.26013763
#X24	2.13E-02	2.20802795
#X25	2.88E-02	2.08302246
#X26	1.89E-02	2.24815674
#X27	1.36E-02	2.33606476
#X28	1.39E-02	2.33202301
#X29	2.09E-02	2.21510101
#X30	1.72E-02	2.27616028
#X31	2.58E-02	2.13296692
#X32	2.52E-02	2.14307129
#X33	2.67E-02	2.11795471
#X34	2.66E-02	2.11997559
#X35	3.00E-02	2.06295807
#X36	2.21E-02	2.19503662
#X37	1.86E-02	2.25320892
#X38	1.27E-02	2.35093262
#X39	1.49E-02	2.31412384



**Figure 6-6. Secchi Values with Varying Phosphorus Reductions**

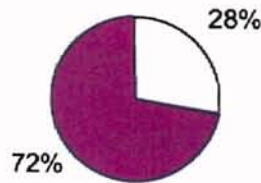
**Without Point Source Contribution**



**50% Non-point Source Reduction**



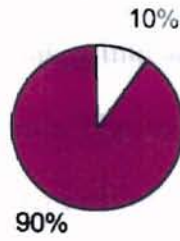
**75% Non-point Source Reduction**



 - Eutrophic       - Non-Eutrophic

**Figure 6-10. Management Alternatives with Secchi Disk Predictions**

### 80% Non-point Source Reduction



### 85% Non-point Source Reduction



 - Eutrophic       - Non-Eutrophic

**Figure 6-10. (cont'd from previous page) Management Alternatives with Secchi Disk Predictions**



Although secchi disk values indicated an improvement in eutrophication with smaller reductions in non-point source phosphorus loading, the accuracy and precision of the secchi disk measurement had to be taken into account. Secchi disk measurements are prone to human errors and inconsistencies. The method of data collection does not follow a particular standard and the accuracy depends on the individual collecting the data as well as specific field conditions, such as a sunny or overcast condition. A listing of some of the factors affecting secchi disk measurements include:

1. eyesight of the viewer,
2. time of the day the readings are taken,
3. the reflectance of the disk,
4. the color of the water, and
5. clay or other particles suspended in water (<http://www.iserv.net/~mlsa/secchi.html>).

A deterministic neural network model provided an advanced technique borrowed from other disciplines for assessing the current state of water quality and evaluating phosphorus reduction goals for controlling the eutrophication of Lake Tenkiller, Oklahoma. The salient features of the model were:

1. few parameter requirements.
2. simple predictions which eliminated the need for excessive sampling, and
3. adaptive learning approach which addressed highly disaggregated data.

In this study the critical factor was model construction. There were two options available for model development, which included model construction without a genetic supervisor and model construction with a genetic supervisor. Though the adaptive learning approach of neural networks helps in modeling disaggregated data, the number of iterations required to develop a robust model would be very large to construct a network with optimum parameter specifications. The genetic supervisor largely reduced the iterative problem by incorporating the strongest links from different network architectures evaluated in developing the optimum network. The ability to cross breed and mutate between different network architectures provided the genetic supervisor with the capabilities which helped in reducing the number of iterations. The gainful application of the genetic supervisor was the high point in this study relative to model environmental data.

Model selection was based on the percentage of “Right” scores made by a model and the consistency exhibited in the learning process. The consistency in learning was projected by the gradual dip in the testing and training error curves of the RMS plots.

Spikes and sudden dips in the curves implied a loss in learning or purely coincidental prediction efficiency. Care was taken to observe the model performance over several epochs before judging the model performance.

As illustrated by the previously collected data and corroborated by the model Lake Tenkiller exhibited highly eutrophic characteristics as indicated by both chlorophyll *a* and secchi disk values. If the phosphorus reduction utilized secchi disk data alone as the indicator of eutrophication, non-point source phosphorus reductions of 80% - 85% were needed to improve the lake water quality. Chlorophyll *a* simulations were more restrictive, suggesting that non-point source reductions of at least 85% were required to stem eutrophication of the Lake Tenkiller. Both parameters were utilized in this effort for suggesting management goals to eliminate eutrophication. The final recommendations from this study are:

1. No additional controls beyond existing NPDES limitations are needed for point source discharges for phosphorus, as they have little or no effect on the eutrophication state of the lake.
2. An overall reduction of 80% to 85% for non-point source loading of phosphorus is needed to reverse eutrophication in Tenkiller. These values are consistent with earlier studies by Harton (1989), and OWRB et al. (1996). Harton suggested between 70% to 90% reduction in total phosphorus loading was needed while the OWRB et al. (1996), study recommended between 70% to 80% reduction in total phosphorus loading. The two studies had also indicated that point sources of phosphorus were of minimal significance in the trophic condition of Lake Tenkiller.

## Bibliography

- American Public Health Association, American Waterworks Association and Water Pollution Control Federation. Standard Methods for the Examination of Water and Wastewater. *Port City Press*, Baltimore, MD, 1992.
- Aziz, A. R. A., and K. V. Wong. A Neural-Network Approach to the Determination of Aquifer Parameters. *Ground Water*, 30(2), 164-166, 1992.
- Basheer, I. A., L. N. Reddi, and Yacoub M. Najjar. Site Characterization by Neuronets: An Application to the Landfill Siting Problem. In Applications of Neural Networks in Environment, Energy and Health. Proceedings of the 1995 Workshop On Environmental and Energy Applications of Neural Networks. P.E. Keller, S. Hashem, L. J. Kangas, R. T. Kouzes (eds.). *World Scientific Publishing Co. Pte. Ltd.*, River Edge, NJ, 1995.
- Chardon, W. J., O. Oenema, P. D. Castilho, R. Vriesema, J. Japenga, and D. Blauuw. Organic Phosphorus in Solutions and Leachates from Soils Treated with Animal Slurries. *Journal of Environmental Quality*, 26, 372-378, 1997.
- Cheshire Engineering Corporation. *Neuralyst v1.41*, 1996.
- Dowd, P. A., and C. Sarac. A Neural Network Approach to Geostatistical Simulation. *Mathematical Geology*, 26(4), 491-503, 1994.
- Dowla, F. U., and L. L. Rogers. Solving Problems in Environmental Engineering and Geosciences with Artificial Neural Networks. *MIT Press*, Cambridge, MA, 1995.
- El-Hawary, F. Applications of Computational Neural Networks in Predicting Atmospheric Pollutant Concentrations due to Fossil-fired Electric Power Generation. In Applications of Neural Networks in Environment, Energy and Health. Proceedings of the 1995 Workshop on Environmental and Energy Applications of Neural Networks. P. E. Keller, S. Hashem, L. J. Kangas, R. T. Kouzes (eds.). *World Scientific Publishing Co. Pte. Ltd.*, River Edge, NJ, 1995.
- Freeze, R.A., J. W. Massmann, J. L. Smith, T. Sperling, and B. James. Hydrogeological Decision Analysis: A Framework. *Groundwater*, 28(5), 738-766, 1990.
- Gakstatter, J. H., M. O. Allum, and J. M. Omernik. Lake Eutrophication: Results from the National Eutrophication Survey. In Water Quality Criteria Research of the U.S. Environmental Protection Agency. Proceedings of EPA-sponsored Symposium on Marine, Estuarine and Freshwater Quality. EPA-600/3-76-079. *U.S. Environmental Protection Agency*, Corvallis, OR, 1974.
- Goldberg, D. E. Genetic Algorithms in Search, Optimization and Machine Learning.

Addison-Wesley, Reading, MA. 1989.

- Gorelick, S. M. Groundwater Contamination: Optimal Capture and Containment. *Lewis Publishers*, Boca Raton, FL. 1993.
- Hagan, M. T., H. B. Demuth and M. Beale. Neural Network Design. *PWS Publishing Company*, Boston, MD. 1996.
- Haraughty, S. The Relationship between Nutrient Limitation and Phytoplankton Community Structure in Lake Tenkiller Ferry. *Masters Thesis, Oklahoma State University*. 1995.
- Haraughty, S. J., and S. L. Burks. Nutrient Limitation in Lake Tenkiller, Oklahoma. *Journal of Freshwater Ecology*. 11(1), 91-100. 1996.
- Harris, G. P. Phytoplankton Ecology: Structure, Function and Fluctuation. *Chapman and Hall Ltd.*, New York, NY. 1986.
- Harton, N. An Analysis of Uncertainty of Point and Non-Point Source Loading on Eutrophication on a Downstream Reservoir. *Master's Thesis, Oklahoma State University*. 1989.
- Hashem, S., P. E. Keller, R. T. Kouzes and L. J. Kangas. Electronic Noses and Their Applications in Environmental Monitoring. In Applications of Neural Networks in Environment, Energy and Health. Proceedings of the 1995 Workshop on Environmental and Energy Applications of Neural Networks. P. E. Keller, S. Hashem, L. J. Kangas, R. T. Kouzes (eds.). *World Scientific Publishing Co. Pte. Ltd.*, River Edge, NJ. 1995.
- Haygarth, P. M., and S. C. Jarvis. Soil Derived Phosphorus in Surface Runoff from Grazed Grassland Lysimeters. *Water Resources*. 31(1), 140-148. 1997.
- Hellman, T. M., and D. A. Hawkins. How Clean is Clean? The Need for Action. *Hazardous Waste Site Management: Water Quality Issues*. National Academy Press, Washington, D. C. 1988.
- <http://vv.carleton.ca/~neil/neural?neuron-a.htm>. The Biological Neuron.
- Johnson, V. M., and L. L. Rogers. Location Analysis in Ground-Water Remediation Using Neural Networks. *Ground Water*. 33(5), 749-758. 1995.
- Karayiannis, N. B., and A. N. Venetsanopoulos. Artificial Neural Networks - Learning Algorithms, Performance Evaluation, and Applications. *Kluwer Academic Publishers*, Boston, MD. 1993.

- Kartalopoulos, S. V. *Understanding Neural Networks and Fuzzy Logic: Basic Concepts and Applications*. *IEEE Press*, New York, NY, 1996.
- Lee, R. E. *Phycology*. *Cambridge University Press*, Cambridge, NY, 1989.
- Metcalf and Eddy, Inc. *Wastewater Engineering, 3<sup>rd</sup> Edition*. McGraw-Hill, New York, NY, 1991.
- Michigan Department of Environmental Quality. <http://www.iserv.net/~mlsa/secchi.html>
- Millhouse, S., and M. Gifford. Net Results. *Civil Engineering-ASCE*, 67(1), 58-60, 1997.
- Mukhopadhyay, A. Spatial Estimation of Transmissivity Using Artificial Neural Network. *Ground Water*, 37(3), 458-464, 1999.
- OWRB (Oklahoma Water Resources Board), Oklahoma State University and EPA. Diagnostic and Feasibility Study on Tenkiller Lake, Oklahoma. Sponsored by US EPA (Region VI). *Environmental Institute – Oklahoma State University*, 1996.
- Palisade Corporation. *@RISK release 3.5d*. Newfield, NY, 1996.
- Ranjithan, S., J. W. Eheart, and J. H. Garrett, Jr. Neural Network-Based Screening for Groundwater Reclamation under Uncertainty. *Ground Water*, 29(3), 563-573, 1993.
- Reynolds, C. S. *The Ecology of Freshwater Phytoplankton*. *Cambridge University Press*, New York, NY, 1984.
- Rich, L. G. *Unit Processes of Sanitary Engineering*. *John Wiley & Sons*, New York, NY, 1963.
- Rizzo, D. M., and D. E. Dougherty, Characterization of Aquifer Properties Using Artificial Neural Networks: Neural Kriging. *Water Resources Research*, 30(2), 483-497, 1994.
- Robinson, D. G. Reliability and Risk Analysis Using Artificial Neural Networks. In *Applications of Neural Networks in Environment, Energy and Health. Proceedings of the 1995 Workshop on Environmental and Energy Applications of Neural Networks*. P. E. Keller, S. Hashem, L. J. Kangas, R. T. Kouzes (eds.), *World Scientific Publishing Co. Pte. Ltd.*, River Edge, NJ, 1995.
- Rogers, L. L., and F. U. Dowla. Optimization of Groundwater Remediation Using Artificial Neural Networks with Parallel Solute Transport Modeling. *Water Resources Research*, 30(2), 457-481, 1994.
- Rogers, L. L., F. U. Dowla, and V. M. Johnson. Optimal Field-Scale Groundwater

- Remediation Using Neural Networks and the Genetic Algorithm. *Environmental Science and Technology*. 29(5). 1145-1155. 1995.
- Rosas, I., A. Velasco, R. Belmont, A. Baez and A. Martinez. The Algal Community as an Indicator of the Trophic Status of Lake Patzcuaro, Mexico. *Environmental Pollution*. 80. 255-264. 1993.
- Sawyer, C. N., and P. L. McCarty. Chemistry for Environmental Engineering. *McGraw-Hill, Inc.*, New York, NY. 1994.
- Schindler, D. Evolution of Phosphorus Limitation in Lakes. *Science*. 195. 260-262. 1977.
- Shannon, E. E., and P. Brezonik. Relationship between Lake Trophic State and Nitrogen and Phosphorus Loading Rates. *Environmental Science and Technology*. 6. 719-725. 1972.
- Thornton, K. W., B. L. Kimmel, and F. E. Payne. Reservoir Limnology: Ecological Perspective. *John Wiley and Sons, Inc.*, New York, NY. 1990.
- United States Environmental Protection Agency (US EPA). Tenkiller Ferry Reservoir Cherokee and Sequoyah Counties, Oklahoma. National Eutrophication Survey, Working Paper No. 593, Corvallis Environmental Research Laboratory, Corvallis, OR. 1977.
- Wang, T. A. A Multilevel Decision Analysis Method for Remediation of Contaminated Sites under Conditions of Uncertainty. *Master's Thesis, Oklahoma State University*. 1995.
- Yip, P. P. C., and Y. H. Pao. A Perfect Integration of Neural Networks and Evolutionary Algorithms. In *Artificial Neural Nets and Genetic Algorithms*. D. W. Pearson, N. C. Steele, R. F. Albrecht (eds.). *Springer-Verlag Wien*. New York, NY. 1995.

## APPENDIX-I

### Example calculation for point source contribution at Tahlequah

The calculation is performed on Osage Creek (Rogers) data.

Known parameters:

Plant Flow	3.5 MGD
Mean Phosphorus Loading	3.829 mg/l
Distance to Tahlequah	75 miles
95 <sup>th</sup> percentile Peak Flow Value at Tahlequah	13000 cfs
50 <sup>th</sup> percentile Channel Phosphorus Mile Loss	.005 (mg/l)/mile

Calculations:

Total Amount Phosphorus Generated by Plant per day =

$$3.829 \text{ (mg/l)} * 3.5 \text{ (MGD)} * 10^6 \text{ (gallons/million gallon)} * 3.781 \text{ (liters/gallon)} \\ = 50724467 \text{ mg}$$

Channel Loss of Phosphorus in Transport to Tahlequah =

$$.005 \text{ ((mg/l)/mile)} * 75 \text{ (miles)} * 3.5 \text{ (MGD)} * 10^6 \text{ (gallons/million gallon)} * \\ 3.781 \text{ (liters/gallon)} = 4967812 \text{ mg}$$

$$\text{Phosphorus Contribution at Tahlequah} = 50724467 \text{ (mg)} - 4967812 \text{ (mg)} \\ = 45703260 \text{ mg}$$

Flow at Tahlequah =

$$13000 \text{ (cfs)} * 28.32 \text{ (l/ft}^3\text{)} * 24 \text{ (hours/day)} * 60 \text{ (min./hour)} * 60 \text{ (sec./min.)} = \\ = 3.18 * 10^{10} \text{ liters per day}$$

Resulting Concentration at Tahlequah =

$$45703260 \text{ (mg)} / 3.18 * 10^{10} \text{ liters per day} = \underline{0.0014 \text{ mg/l (approx.)}}$$



**Vita**  
**Anupam Kumar**  
**Candidate for the Degree of**  
**Master of Science**

Thesis: Neural Network Solution for Assessment of Eutrophication In Lake Tenkiller

Major Field: Environmental Engineering

Biographical:

Personal Data: Born in Calcutta, India. On Dec. 30, 1970, the son of Abinash and Meera Kumar.

Education: Graduated from Loyola High School, Jamshedpur, India; received Bachelor of Technology degree in Civil Engineering from Indian Institute of Technology, Kanpur, India in May, 1993. Completed the requirements for the Master of Science degree with a major in Environmental Engineering at Oklahoma State University in May 2000.

Experience: Employed with Tata Steel, Jamshedpur, India from July 1993 to Dec., 1995 as Senior Officer incharge of construction site supervision, Water and Waste Water treatment plant project management and Stores inventory management; employed by Oklahoma State University, Department of Civil and Environmental Engineering as a graduate research assistant from Jan., 1996 to Aug., 1996; employed by Oklahoma State University, College of Engineering, Architecture and Technology (CEAT) computer laboratories as Lab. Monitor from Aug., 1996 to Sep., 1997; employed by Parsons Engineering Science Inc., as Environmental Engineer/ Environmental Database Systems Developer from Sep., 1997 to present.