

A Statistical Examination of Water Quality in Two Iowa Lakes

Erin Carlson and Mark D. Ecker
Department of Mathematics
University of Northern Iowa
Cedar Falls, Iowa 50614-0506 USA

Received June 14, 2002

Accepted July 5, 2002

ABSTRACT

Water quality has become an important issue in the state of Iowa as well as across the entire United States. Two Iowa Lakes, Silver Lake and Casey Lake were chosen for study by a team of biologists, chemists, earth scientists and statisticians from the University of Northern Iowa. Our goals are to statistically compare the water quality in the two lakes in each year and examine whether or not each lake has changed, in terms of water quality variables, from 1999 to 2000. In addition, we explore which variables most affect phosphorus levels in each lake in 2000. Lastly, we explore the spatial distribution of phosphorus in the sediment of each lake. Discriminant Analyses and ANCOVA show significant difference between the two lakes in both 1999 and 2000 as well as a change in Silver Lake's water quality data from 1999 to 2000. Regression Analyses show that, in Silver Lake, phosphorus levels increased during the summer of 2000 while they decreased with increasing levels of surface dissolved oxygen and decreased as the water became less clear. The analyses also show that phosphorus levels in Lake Casey decreased as the water became less clear. A significant relationship between phosphorus in the sediment and depth exists in Lake Casey. While a significant 2-dimensional spatial correlation cannot be shown in Silver Lake, spatial analyses do show the existence of a significant 3-dimensional spatial correlation in Lake Casey.

I. INTRODUCTION

Water quality has become an important issue in the state of Iowa as well as across the entire United States. Two Iowa lakes were chosen for study by a team of biologists, chemists, earth scientists and statisticians from the University of Northern Iowa. Silver Lake, in Delhi, Iowa, has a 187-acre watershed with a hog lot and agricultural fields at the lake's shore. Not surprisingly, the water quality is quite poor [1, p. 18]; high levels of phosphorus are a leading factor in making the lake eutrophic. Twenty sites (see Figure 1) were selected in Silver Lake to collect water quality data in each summer of 1999 and 2000. Casey Lake, in Tama County, has a much larger watershed of 738 acres and is buffered by trees and vegetation. The water quality is

much better than Silver Lake; it is one of the better fishing lakes in northeast Iowa [1, p. 2]. Twenty-three sites in Lake Casey were chosen for water quality assessment (see Figure 2).

In 2000, each Silver Lake site in Figure 1 was sampled on May 30, June 27 and August 1. In addition, the ten sites marked with a diamond (♦) in Figure 1 were sampled on May 10, May 17, May 23, May 30, June 6, June 15, June 20, June 27, July 5, July 11, July 18, July 25, August 1 and August 8. In 2000, each site in Lake Casey in Figure 2 (including 3 inflows) was sampled on June 1, June 29 and August 3. Additionally, 14 sites marked by a diamond (♦) in Figure 2 were sampled on May 12, May 28, May 19, May 25, June 1, June 7, June 13, June 22, June 29, July 6, July 13, July 20, July 27, August 3 and August 10.

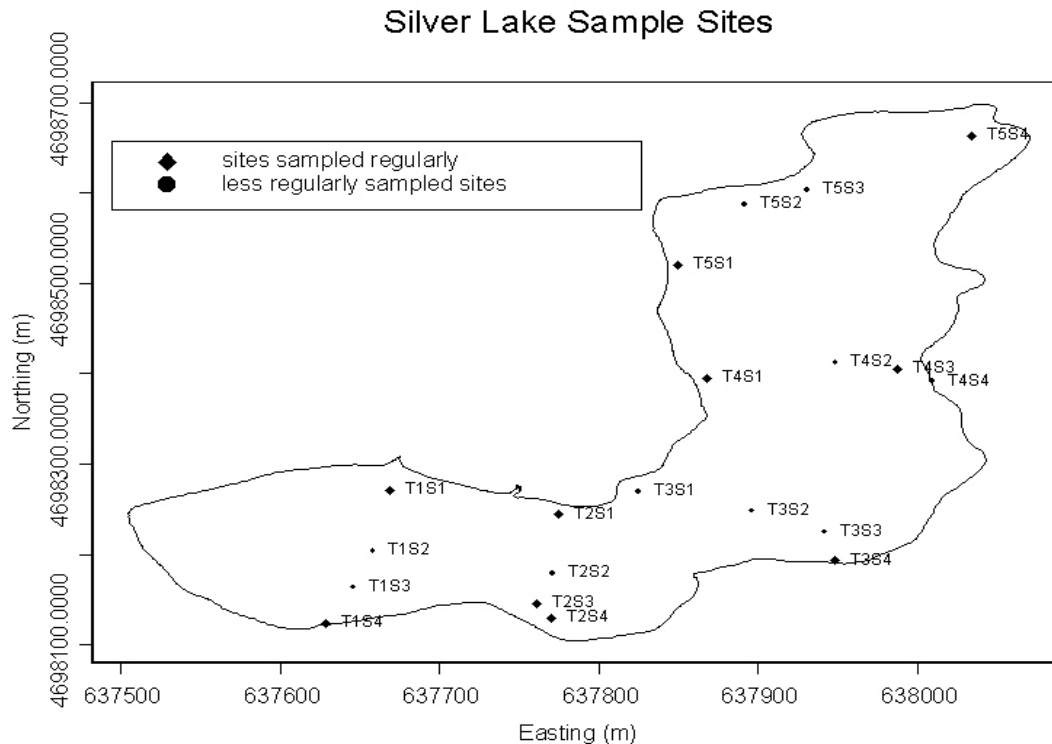


Figure 1. Silver Lake sites sampled in 2000. Filled circles (●) represent sites sampled on May 30, June 27 and August 1. The ten sites marked with a diamond (◆) were sampled on May 10, May 17, May 23, May 30, June 6, June 15, June 20, June 27, July 5, July 11, July 18, July 25, August 1 and August 8.

Water quality variables from Lake Casey and Silver Lake that were collected include: turbidity, secchi, depth, surface temperature, bottom temperature, surface dissolved oxygen, bottom dissolved oxygen, chlorophyll a, total phosphorus, dissolved phosphorus, bacteria and coliform. Secchi, which measures water clarity, is formally defined as “the mean depth of the point where a weighted white disk, 20 cm in diameter, disappears when viewed from the shaded side of a vessel and that point where it reappears upon raising it after it has been lowered beyond visibility” [2]. Overall, the total number of site/time data points in Silver Lake in 2000 was 170, while for Lake Casey 204 were possible. Data from 1999 were also available; see Ecker and Janssen [3] for a detailed description of the sampling regime.

Our goals are to statistically compare the water quality in the two lakes in each year and examine whether or not each

lake has changed, in terms of water quality variables, from 1999 to 2000. In addition, we explore which variables most affect phosphorus levels in each lake in 2000. Lastly, we explore the spatial distribution of phosphorus in the sediment of each lake.

II. WATER QUALITY DATA ANALYSIS - REGRESSION

The purpose of this statistical analysis is to determine which of the covariates are influential in modeling dissolved phosphorus and/or total phosphorus in each lake using the 2000 data. A separate regression analysis was done for each lake. In Silver Lake, data were available on dissolved phosphorus while total phosphorus was analyzed in Lake Casey. Initially, correlations were calculated between all pairs of covariates to determine which of covariates would be used in the multiple regression. Tables 1 and 2 show

Casey Lake Sample Sites

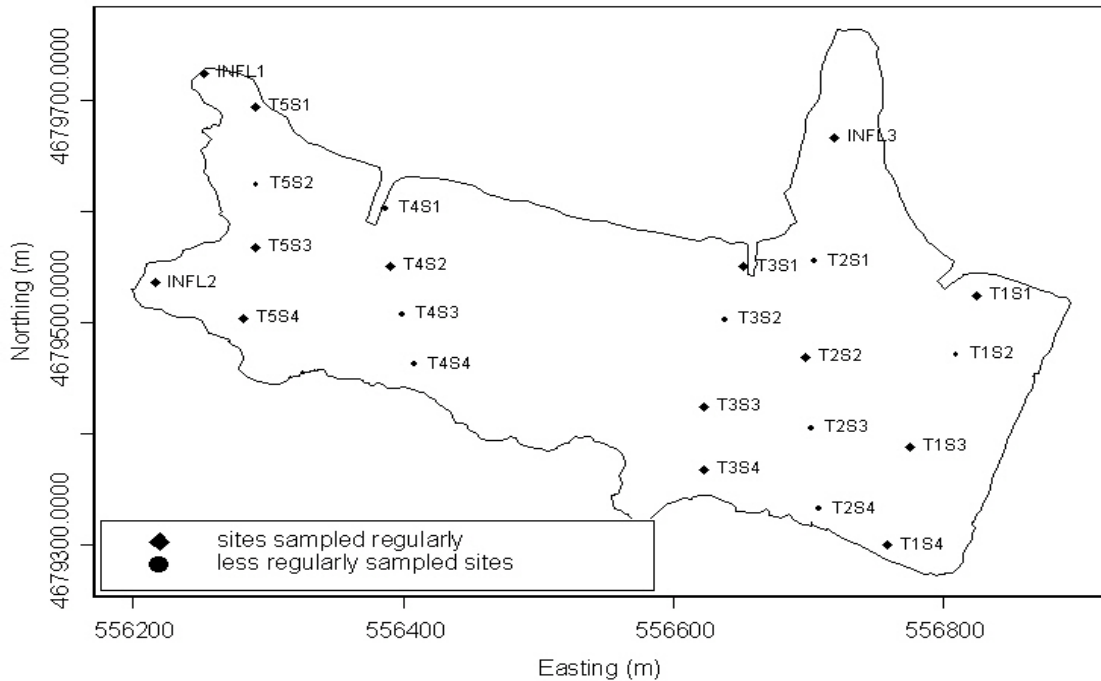


Figure 2. In 2000, each site in Lake Casey (including 3 inflows) was sampled on June 1, June 29 and August 3. Additionally, 14 sites marked by a diamond (◆) were sampled on May 12, May 28, May 19, May 25, June 1, June 7, June 13, June 22, June 29, July 6, July 13, July 20, July 27, August 3 and August 10.

the covariates used for each respective lake, as well as the results of the regression procedure. The regression was performed in SAS. Time series models could be

entertained for the regularly sampled sites in each lake, however, Ecker and Janssen [3] found little temporal correlation in 1999 Silver Lake data.

Variable	Coefficient	p-value
Depth	0.57306	0.8867
Time	0.31347	0.0004
Secchi	-0.76364	0.0057
Surface temperature	0.71108	0.4158
Surface dissolved oxygen	-4.04812	<0.0001

Table 1: Silver Lake, Dependant Variable: Dissolved Phosphorus

Variable	Coefficient	p-value
Depth	-1.03057	0.4531
Time	0.07014	0.6618
Secchi	-0.13651	0.0108
Surface dissolved oxygen	1.43286	0.1481

Table 2: Lake Casey, Dependant Variable: Total Phosphorus

The regression analysis for modeling dissolved phosphorus in Silver Lake reveals three significant variables (using an 0.05 significance level): time, secchi and surface dissolved oxygen. The overall R^2 was 0.40. Time, the number of days after May 1 when the sample was taken, has a direct relationship with dissolved phosphorus. The later in the summer the sample was taken the higher the dissolved phosphorus levels were. Secchi has an inverse relationship with dissolved phosphorus. The deeper the visibility (or smaller the number of particles) the lower the dissolved phosphorus levels were. Surface dissolved oxygen also has an inverse relationship with dissolved phosphorus. As the levels of surface dissolved oxygen increase, the levels of dissolved phosphorus decrease.

The regression analysis measuring total phosphorus in Lake Casey shows only secchi as a significant variable. The overall R^2 was 0.232. (Dissolved phosphorus data were not available for Lake Casey.) Secchi has a negative relationship with dissolved phosphorus implying that as visibility in the water increases, dissolved phosphorus levels decrease. Note that surface temperature was not used in the Lake Casey regression analysis due to its strong correlation with time ($r = 0.765$).

III. WATER QUALITY ANALYSIS - DISCRIMINANT ANALYSIS AND ANALYSIS OF COVARIANCE (ANCOVA)

The goal of the next statistical analysis is to determine if the water quality variables for each respective lake had changed from 1999 to 2000, and also to explore the perception that the water quality in Silver Lake was different from Lake Casey. A discriminant analysis was performed using S-Plus for each lake with data from 1999 and 2000; a graphical plot of the first two discriminant functions was prepared. Discriminant analyses were also performed for each year using data from the two lakes to check for differences by lake. Formal analysis of covariance (ANCOVA) hypothesis tests were also performed. The dependant variable in all ANCOVA models was total phosphorus.

An analysis of covariance (ANCOVA) can be used to determine if two sets of data have common, parallel, or separate regression planes. If the groups have separate regression planes then they are considered to be different. If the two groups have common regression planes, they are considered to be the same. If the groups have parallel planes, they have different intercepts (levels of total phosphorus given small values-potentially extrapolated to zero-of the covariates) but the same underlying trends or rates of change. Fitting the ANCOVA model was done with SAS. Table 3 contains the respective sums of squared error (SSE) and the degrees of freedom (df) to be used in the subsequent ANCOVA hypothesis tests.

IV. 1999 DISCRIMINANT ANALYSIS AND ANCOVA FOR BETWEEN LAKE CASEY AND SILVER LAKE

The following variables were used to discriminate between Lake Casey and Silver Lake in 1999: surface temperature (*stemp*), surface dissolved oxygen (*sDO*), turbidity (*turb*), chlorophyll a (*chla*), and dissolved phosphorus (*dissP*). There were 54 samples from Silver Lake that had all five variables and 54 from Lake Casey. The goal of the discriminant analysis is to find one or two functions (linear combinations) of the observed data (called discriminant functions) that best separate the two known groups (lakes). The two functions (linear combinations) that provide the largest separation between the two lakes are the first and second discriminant functions. The first discriminant function is

$$-0.188(stemp) + 0.094(sDO) + 0.013(turb) + 0.002(chla) + 0.012(DissP); \quad (1)$$

the second discriminant function is

$$+ 0.419(stemp) - 0.388(sDO) - 0.012(turb) + 0.006(chla) + 0.006(dissP). \quad (2)$$

Surface temperature and surface dissolved oxygen contribute the most to these two discriminant functions because of their relative large (in absolute value) coefficients.

Each of the 54 samples from Silver Lake in 1999 and the 54 samples from Lake Casey were transformed using these two discriminant functions. These first two discriminant functions were plotted against each other in Figure 3. Inspecting Figure 3, the data from the two lakes in 1999 appear to have been well separated by the discriminant functions. Thus, water quality variables are distinctly different for the two lakes in 1999.

The ANCOVA formally tests whether the two lakes are similar (common regression planes), have different initial values but change at the same rate (parallel planes), or are completely different (separate planes). Because of the three possible outcomes, a formal hypothesis test first compares common vs. parallel planes, i.e., assume the simplest model. The test H_0 : common planes versus H_A : parallel planes has F-statistic, using the data from Table 3, of

$$F = \frac{(SSE_{\text{common}} - SSE_{\text{parallel}}) / (df_{\text{common}} - df_{\text{parallel}})}{SSE_{\text{parallel}} / df_{\text{parallel}}} = \frac{(721363.608 - 521595) / (104 - 102)}{521595 / 102}$$

$$= \frac{99884.304}{5113.676} = 19.53$$

with a p-value less than 0.0001. Thus, H_0 is rejected in favor of H_A , i.e., the common regression planes model is rejected in favor of the parallel plane model.

Had we failed to reject H_0 , the ANCOVA testing would be complete and the water quality variables in the two lakes would be deemed equivalent. However,

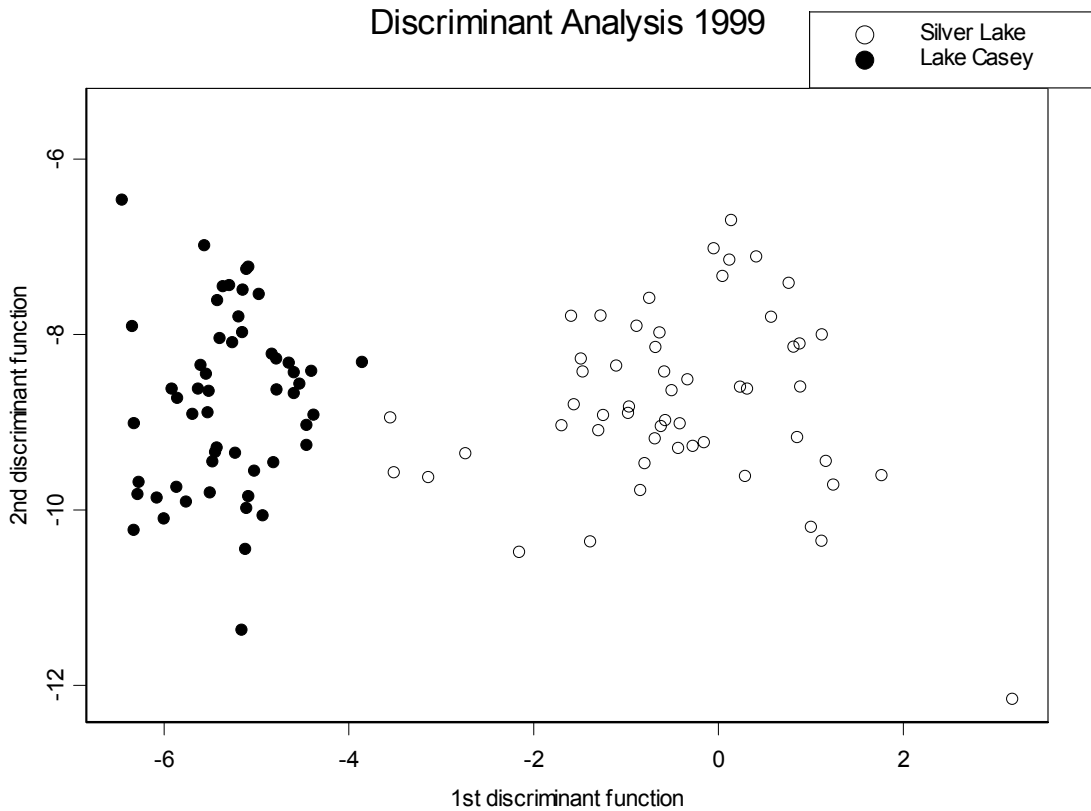


Figure 3. 1999 Discriminant Analysis using the discriminant functions, equations (1) and (2).

since the common regression plane model is not tenable, the next step is to test H_0 :

parallel planes versus H_A : separate planes. The F-statistic for this test is

$$F = \frac{(SSE_{\text{common}} - SSE_{\text{parallel}}) / (df_{\text{common}} - df_{\text{parallel}})}{SSE_{\text{parallel}} / df_{\text{parallel}}} = \frac{(521595 - 390605.677) / (102 - 98)}{390605.677 / 102}$$

$$= \frac{32747.33}{3985.77} = 8.216$$

(p-value < 0.0001). Again, H_0 is rejected in favor of H_A , i.e., the simpler parallel regression planes model is rejected in favor of the separate regression planes model. Since the 1999 Silver Lake data and the

1999 Casey Lake data best support a separate regression planes model, we formally conclude that the water quality variables in the two lakes are quite different in 1999.

1999 Silver vs. 1999 Casey			
	<i>separate</i>	<i>parallel</i>	<i>common</i>
SSE	390605.7	521595.0	721363.6
df	98	102	104
2000 Silver vs. 2000 Casey			
	<i>separate</i>	<i>parallel</i>	<i>common</i>
SSE	1110943.7	1582081.2	601157.0
df	220	226	227
Silver 1999 vs. Silver 2000			
	<i>separate</i>	<i>parallel</i>	<i>common</i>
SSE	1007644.6	1413264.0	19744341.2
df	140	144	145
Casey 1999 vs. Casey 2000			
	<i>separate</i>	<i>parallel</i>	<i>common</i>
SSE	87126.3	108709.2	110605.1
df	164	168	169

Table 3. ANCOVA fittings (sums of squared error = SSE; degrees of freedom = df).

V. 2000 DISCRIMINANT ANALYSIS AND ANCOVA BETWEEN LAKE CASEY AND SILVER LAKE

The variables used in this analysis are surface temperature, bottom temperature (*btemp*), surface dissolved oxygen, bottom dissolved oxygen (*bDO*), turbidity, secchi, and total phosphorus (*totalP*). 109 site/time observations in Silver Lake had all the respective variables while 125 did so in Lake Casey. The two functions (linear combinations) that provide the largest separation between the two lakes are the first and second discriminant functions. The first discriminant function is

$$-0.188(stemp) - 0.157(btemp) - 0.047(sDO) + 0.155(bDO) + 0.030(turb) - 0.014(sec\ chi) + 0.001(totalP); \tag{3}$$

and the second discriminant function is

$$0.045(stemp) - 0.152(btemp) + 0.025(sDO) + 0.114(bDO) + 0.006(turb) + 0.003(sec\ chi) + 0.002(totalP). \tag{4}$$

The variables surface temperature, bottom temperature and bottom dissolved oxygen contribute largely to the separation of the two groups. Figure 4 shows the plot of discriminant functions (3) and (4).

The discriminant functions appear to have separated the 2000 Lake Casey data and the 2000 Silver Lake data fairly well. One observation from Lake Casey, the filled circle in the center of Figure 4 (representing data from Inflow #1 on June 7), is far removed from the bulk of the Lake Casey data and would be misclassified if it were not known from which lake it arose. This particular observation had an elevated total phosphorus value (the highest in Lake Casey for the summer of 2000).

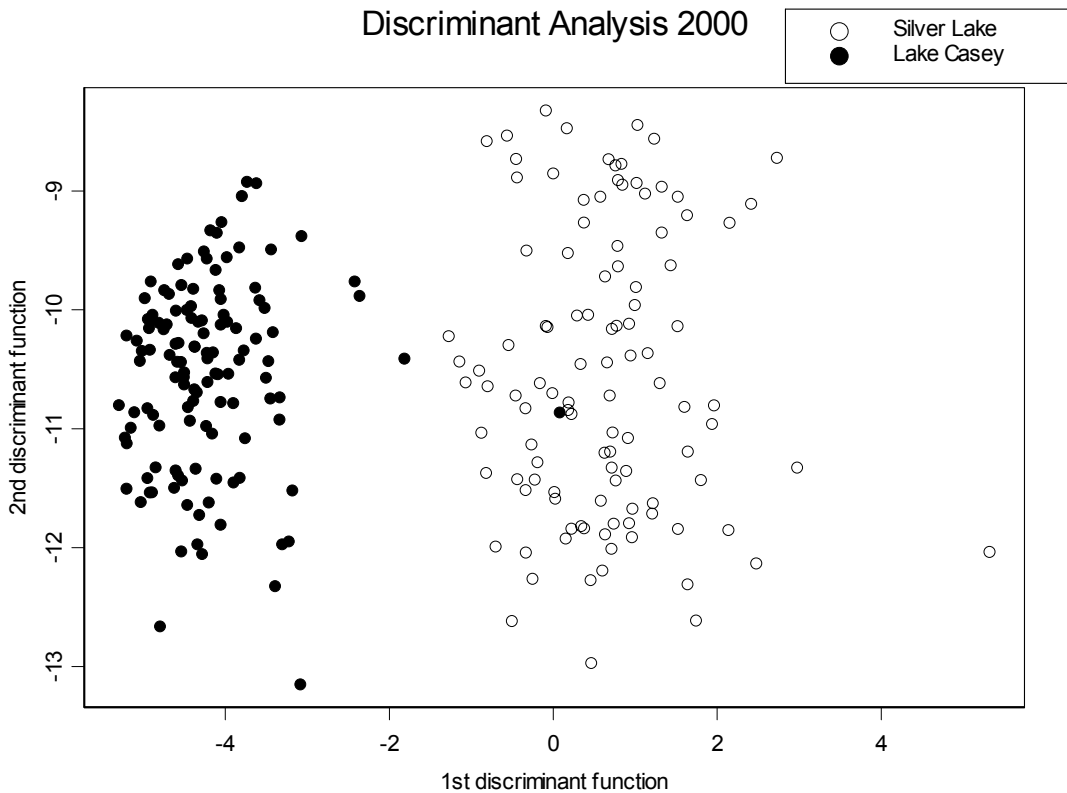


Figure 4. 2000 Discriminant Analysis using the discriminant functions, equations (3) and (4).

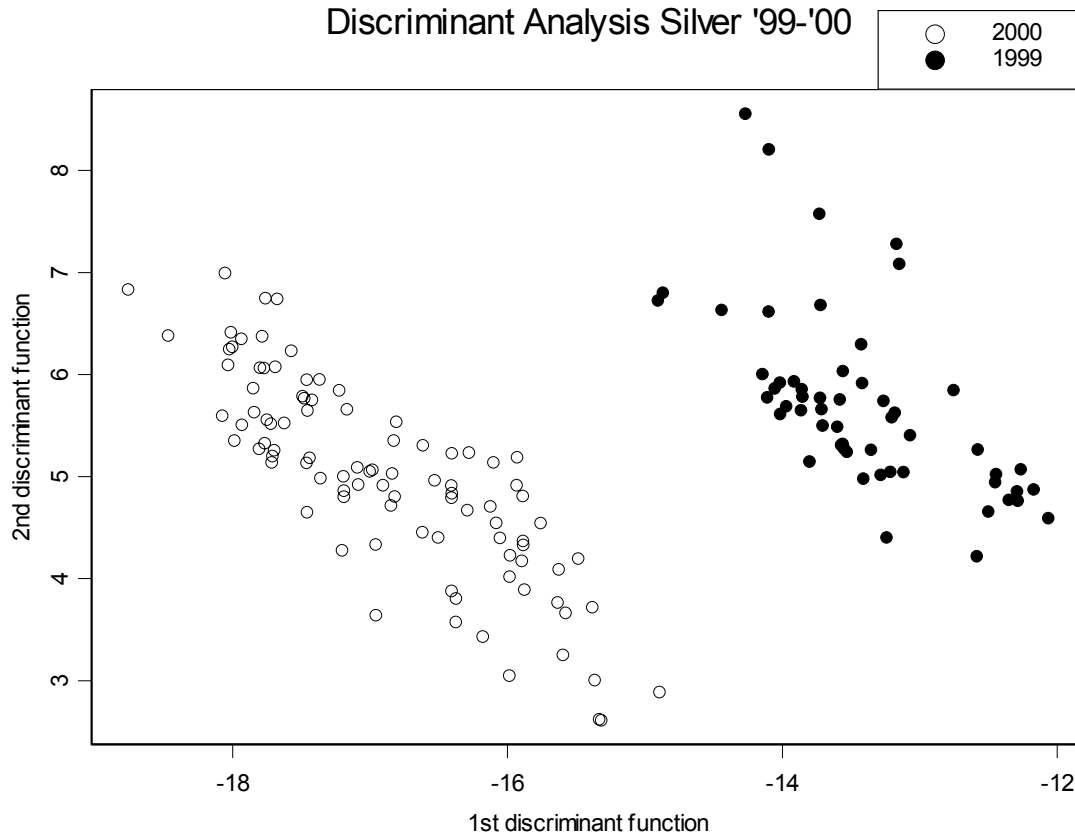


Figure 5. Discriminant Analysis using the 1999 and the 2000 data for Silver Lake and equations (5) and (6).

Overall, the 2000 Lake Casey variables and the 2000 Silver Lake variables appear to be different.

Using the data from Table 3, the ANCOVA hypothesis test H_0 : common planes versus H_A : parallel planes has an F-statistic of 632.7 (p -value < 0.0001). H_0 is rejected in favor of H_A , i.e., the common planes model is rejected in favor of the parallel planes model. Next, H_0 : parallel planes versus H_A : separate planes has an F-statistic of 15.55 (p -value < 0.0001). H_0 is rejected in favor of H_A , i.e., the parallel planes model is rejected in favor of the separate planes model. Since the 2000 Silver Lake data and the 2000 Casey Lake data best support the separate regression lines, the water quality variables between the two lakes are different in 2000.

VI. DISCRIMINANT ANALYSIS AND ANCOVA FOR 1999 AND 2000 SILVER LAKE DATA

The following variables were used to discriminate between the 1999 and 2000 Silver Lake data: surface temperature, bottom temperature, surface dissolved oxygen, bottom dissolved oxygen, turbidity, secchi, and total phosphorus. There are 55 observations from Silver Lake in 1999 and 95 observations in 2000. The two functions that best separate the two known groups are the first two discriminant functions. The first discriminant function is

$$\begin{aligned}
 & -0.188(stemp) - 0.157(btemp) - 0.047(sDO) \\
 & + 0.155(bDO) + 0.030(turb) - 0.014(sec\ chi) \\
 & + 0.006(totalP);
 \end{aligned}
 \tag{5}$$

The second discriminant function is

$$0.0485(stemp) - 0.152(btemp) + 0.025(sDO) + 0.114(bDO) - 0.006(turb) + 0.003(sec\ chi) + 0.002(totalP).$$

(6)

Based on these two equations, surface temperature, bottom temperature and bottom dissolved oxygen are important in separating the two groups. Figure 5 shows the plot of the two discriminant functions (5) and (6). The two discriminant functions have separated the two years of data well. Therefore, the Silver Lake data appears to have changed from 1999 to 2000.

Using the data from Table 3, the ANCOVA hypothesis test H_0 : common lines versus H_A : parallel lines has an F-statistic of 57.178 (p-value <0.0001). H_0 is rejected in favor of H_A , i.e., the common planes model is rejected in favor of the parallel planes model. The hypothesis test H_0 : parallel lines versus H_A : separate lines has an F-statistic of 56.365 (p-value <0.0001). H_0 is rejected in favor of H_A , i.e., the parallel planes model is rejected in favor of the separate planes model. Since the Silver Lake data from 1999 and the Silver Lake data from 2000 best support separate regression lines, the formal ANCOVA verifies the conclusions of the informal discriminant analysis: the Silver Lake water quality data changed from 1999 to 2000.

VII. DISCRIMINANT ANALYSIS AND ANCOVA FOR THE 1999 AND 2000 LAKE CASEY DATA

Analogous procedures to those of the previous section were performed with the Lake Casey data to determine if the water quality variables in 1999 are different than in 2000. The variables used in this analysis were surface temperature, surface dissolved oxygen, turbidity, chlorophyll a, and total phosphorus. There are 54 observations from Lake Casey in 1999 and 120 in 2000. The two functions that best separate the two groups are the first two discriminant functions. The first discriminant function is

$$0.432(stemp) - 0.307(sDO) - 0.016(turb) - 0.001(chla) - 0.007(totalP);$$

(7)

The second discriminant function is

$$-0.247(stemp) - 0.291(sDO) + 0.048(turb) - 0.003(chla) - 0.025(totalP).$$

(8)

Surface temperature and surface dissolved oxygen contribute most to these discriminant functions.

Figure 6 shows the plot of the discriminant functions (7) and (8). There is a great deal of overlap of the 1999 and 2000 data points. This raises questions about whether the water quality data in Lake Casey has changed much.

The ANCOVA hypothesis H_0 : common lines versus H_A : parallel lines has an F-statistic of 2.93 (p-value = 0.088). H_0 cannot be rejected using a 0.05 significance level. Thus, we cannot formally reject the common lines hypothesis (although it should be noted that the p-value of 0.088 is marginally significant). Hence, backed by the graphical discriminant analysis together with the formal ANCOVA, the water quality variables have not changed much in Lake Casey from 1999 to 2000.

VIII. SEDIMENT ANALYSIS

In addition to the water quality analyses performed in 1999 and 2000, sediment samples from Lake Casey and Silver Lake were taken during the summer of 2001. The sampling sites used in the 2000 water quality study were also used in the sediment sampling. (See Figures 1 and 2 for the location of the sites in each respective lake). Samples were taken using a corer and are analyzed in this section. The cored sample was divided into 2-inch intervals and the phosphorus levels were determined for each interval. The midpoint of each interval was used as the depth of the sample (e.g. 1, 3, 5, 7... inches deep). The easting and northing coordinates of the sites were available for each location. The depths in inches were transformed to depths in centimeters. It was of interest to determine if there was a spatial trend in the levels of phosphorus. Regression analyses were performed to determine if there was a relationship between phosphorus levels and depth of the sediment sample. Two- and three-dimensional variograms [4] were constructed to explore for any spatial

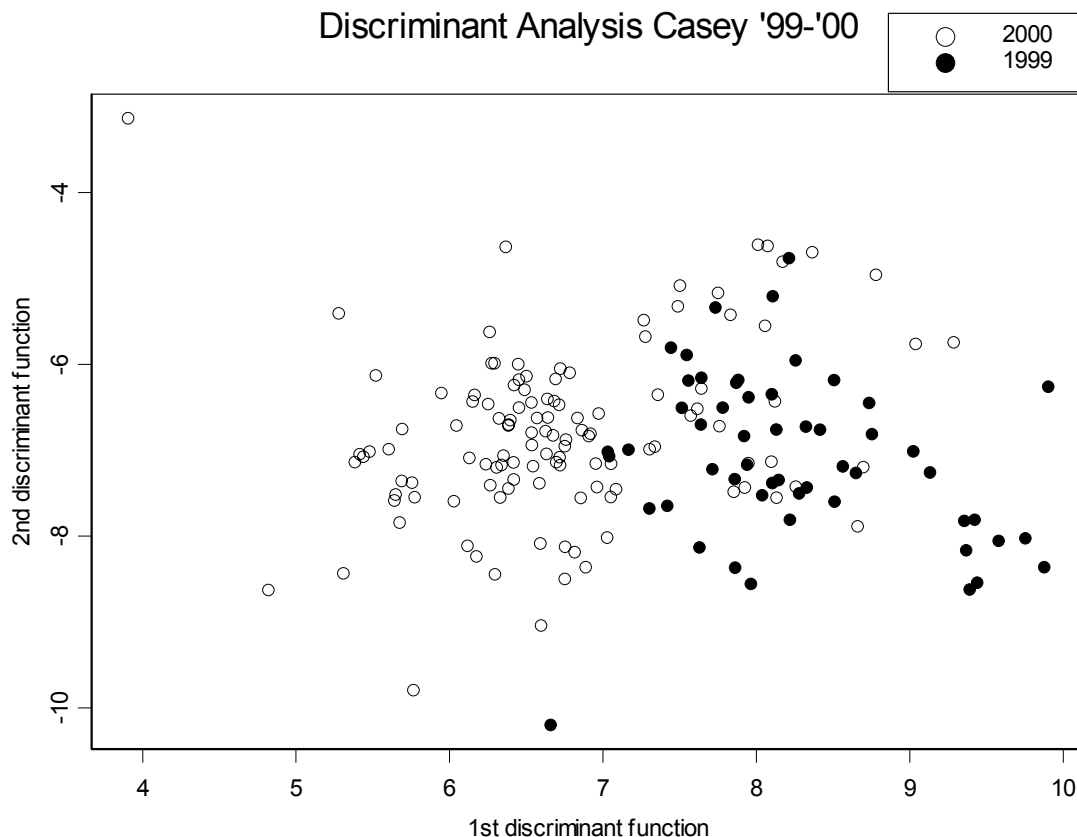


Figure 6. Plot of the discriminant functions (7) and (8) for Lake Casey data.

correlation (the tendency for pairs of observations closer in space to have more similar responses than for pairs further apart) in the data. Kriging or prediction at unsampled locations within each respective lake was also performed to produce a map of the phosphorus levels within the lake [5].

Table 4 contains summary statistics of phosphorus levels for all sediment depths by lake. The sample mean for Silver Lake was over 100 units higher than for Lake Casey, however the Wilcoxon-Summed Rank Test [6] did not find a significant difference (p-value = 0.14).

A regression analysis for each lake was performed in SAS to determine if there is a relationship between the log of the phosphorus levels and the depth of the sample. Because of the right skewed distribution of phosphorus levels, a natural log transformation of these data was performed. There was not a significant relationship between the log of phosphorus and depth for Silver Lake (p-value=0.4742).

However there was a significant relationship between the log of phosphorus and depth for Lake Casey (p-value= 0.0184). The regression equation for Lake Casey is

$$\log(\text{phosphorus}) = 5.66687 + 0.79561(\text{depth})$$

Since, there is a positive relationship between depth and the log of phosphorus, the levels of phosphorus in Lake Casey tend to be higher deeper in the sediment.

	Lake Casey	Silver Lake
Mean	357.70	466.50
Median	324.70	379.2
Standard Deviation	155.424	333.73
Minimum	81.76	11.0
Maximum	925.20	1315.0

Table 4. 2001 Sediment Phosphorus Levels (in units of μ g/gm).

The 2-d spatial analysis of the phosphorus levels in the sediment of these lakes was performed using S-plus, where the two location variables were easting and northing. Figures 7 and 8 are the empirical variograms [4] for Lake Casey and Silver Lake respectively. Figure 8 shows no apparent 2-d spatial association in the phosphorus levels in Silver Lake. A lack of association is evidenced by noise in the Silver Lake variogram, especially at short lags. However, the variogram for Lake Casey in Figure 7 does appear to exhibit spatial correlation; we shall examine this further.

The goal of the sediment analysis was to determine if there is spatial association in the Lake Casey data by including depth as a third location coordinate. The regression analysis showed that there was no significant relationship between depth and phosphorus for Silver Lake. Thus, a 3-d spatial analysis of Silver Lake was not performed. Since there was a significant relationship for Lake Casey, an analysis of Lake Casey was performed. S-plus does not offer a 3-d spatial analysis as part of its standard code; a custom S-Plus

program to compute the empirical 3-d variogram was written (see Appendix I). The variogram obtained from S-plus for Lake Casey is in Figure 9.

The upward trend in the variogram in Figure 9 implies a spatial trend in the data. In particular, the greater the distance between two points in the lake, the greater the variability in phosphorus levels. The theoretical variogram was chosen as an exponential function with nugget = 0.09, sill = 0.25 and range = 350. The theoretical variogram overlaid on the empirical variogram is presented in Figure 10.

This exponential model was chosen for the kriging procedure [5] to create a graph of surface phosphorus concentrations in the Lake Casey sediment (using depth = 0). Figure 11 shows the log scaled concentrations of the sediment surface layer of phosphorus in Lake Casey based on the exponential log scaled 3-d variogram. There is a band of high log phosphorus levels across the middle of the lake. The areas of lowest log phosphorus concentrations are in the northwestern tip and the southeastern corner.

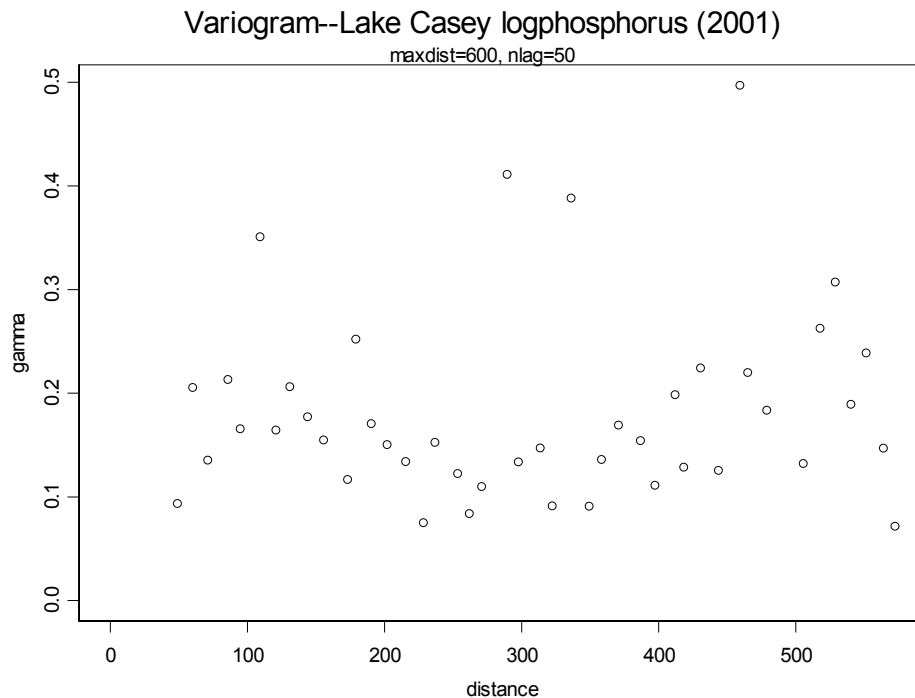


Figure 7. This variogram for Lake Casey appears to exhibit spatial correlation.

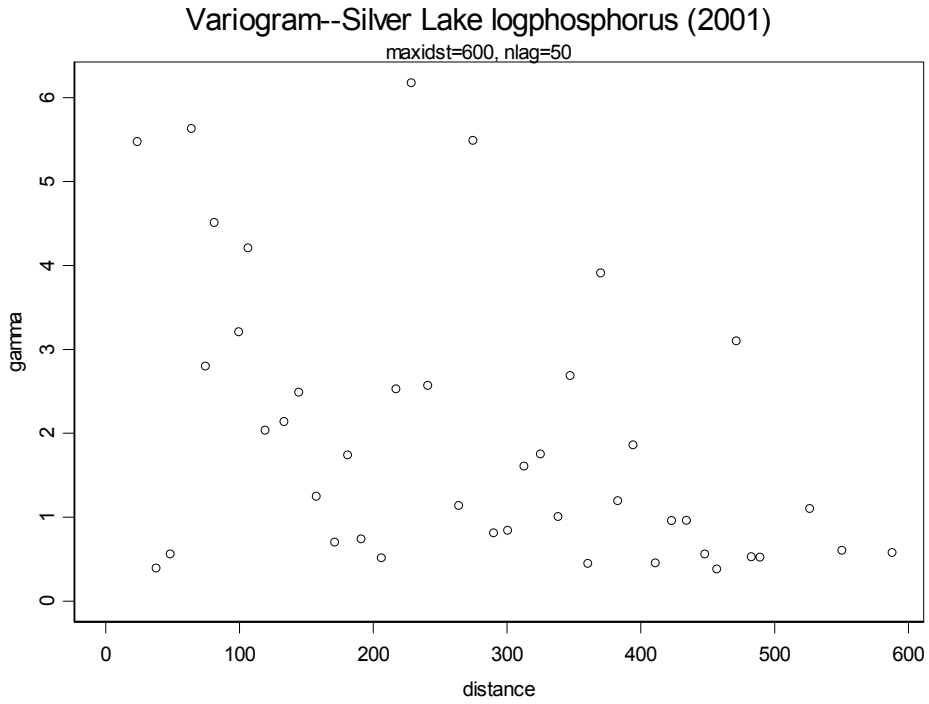


Figure 8. Silver Lake shows no apparent 2-d spatial association in phosphorus levels.

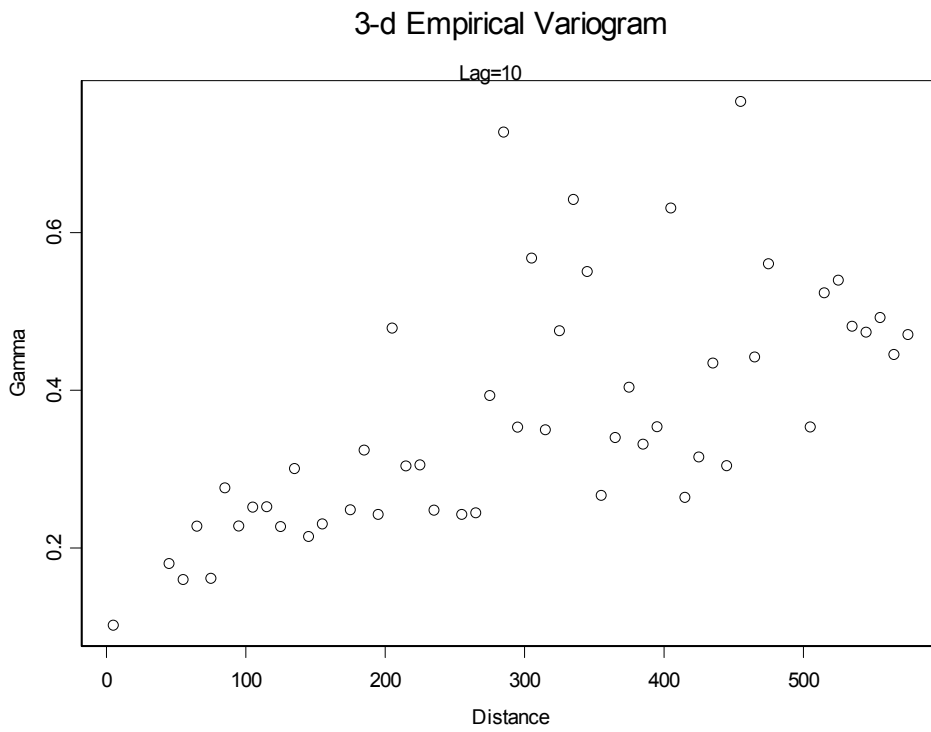


Figure 9. Variogram for Lake Casey obtained from S-plus.

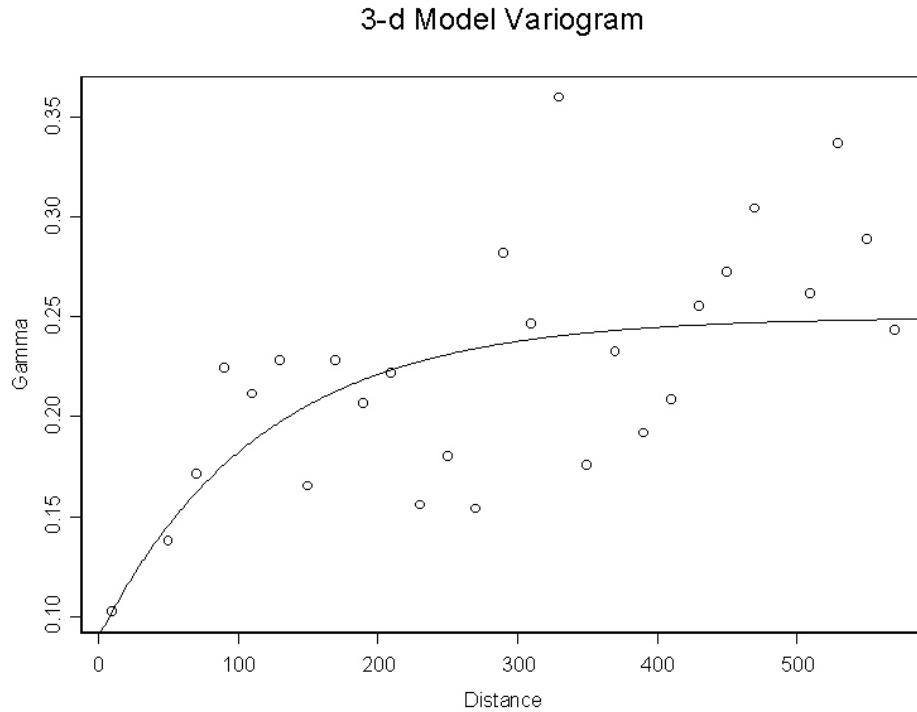


Figure 10. Theoretical variogram (an exponential) for Lake Casey laid over the empirical variogram.

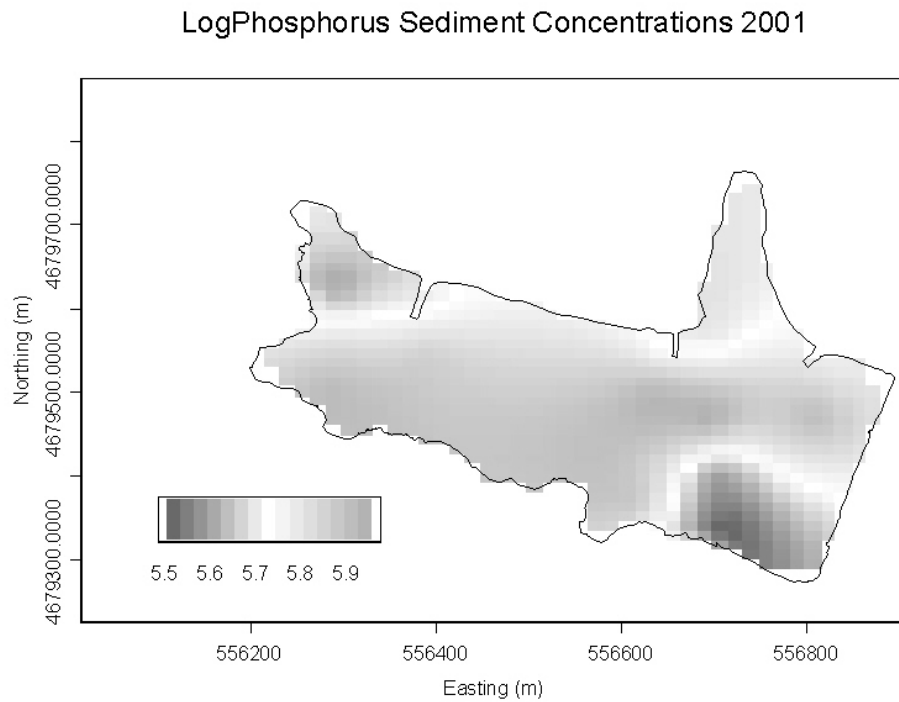


Figure 11. The log scaled concentrations of the sediment surface layer of phosphorus in Lake Casey based on the exponential log scaled 3-d variogram.

IX. CONCLUSIONS

A regression analysis of data from Silver lake found three significant variables for explaining phosphorus levels in the water. Phosphorus levels increased during the summer of 2000 while they decreased with increasing levels of surface dissolved oxygen and decreased as the water became less clear. In Lake Casey the only variable that had a significant relationship with phosphorus was secchi; phosphorus levels in Lake Casey decreased as the water became less clear.

The Discriminant analyses and ANCOVA show there is a significant difference in the water quality data between the two lakes in both 1999 and 2000. In addition, the water quality data from Silver Lake has changed from 1999 to 2000. However, the water quality from Casey Lake has not changed much from 1999 to 2000.

Regression analyses show a significant relationship between the phosphorus in the sediment and depth in Lake Casey but not Silver Lake. A spatial analysis shows that no significant 2-dimensional spatial correlation exists in the levels of phosphorus in Silver Lake; however, a significant 3-dimensional spatial correlation does exist in Lake Casey.

APPENDIX I

S-plus code for 3-d spatial analysis

```
#Program to compute the variogram cloud and
do the subsequent binning for
#three dimensionally spatially referenced data
#
#Written by Erin Carlson 9/19/01
#
#
#Input data should have the form
# X-coordinate : dataset$firstvar
# Y-coordinate : dataset$secondvar
# Z-coordinate : dataset$thirdvar
# Dependent variable : dataset$resp

n<-length(dataset$firstvar)
ncomb<-(n)*(n-1)/2
d<-1:ncomb
gamma<-1:ncomb
k<-1
for(i in 1:n)
{
for(j in i:n)
{
```

```
if(j!=i) d[k]<-sqrt(((dataset$firstvar[i]-
dataset$firstvar[j])**2)+
((dataset$secondvar[i]-
dataset$secondvar[j])**2)+
((dataset$thirdvar[i]-
dataset$thirdvar[j])**2))
if(j!=i) gamma[k]<-
((dataset$respons[i]-
dataset$respons[j])**2)/2
if(j!=i) k<-k+1
}
}
dist<-na.omit(d)
gamma1<-na.omit(gamma)
plot(dist, gamma1, xlab="distance",
ylab="gamma")
title('3-d Variogram Cloud')

#Code to create bins and plot empirical
variogram
#must be used after above variogram program is
used
#must enter number for either nbins (number of
bins) or lag, the other should be 0.
#
nbins<-0
lag<-20
N<-length(d)
maxd<-max(d)
if(nbins==0) nbins<-trunc(maxd/lag)
if(lag==0) lag<-trunc(maxd/nbins)
gammamean<-1:nbins
gammamean1<-1:nbins
mid<-1:nbins
mid1<-1:nbins
bin<-1:nbins
bintotal<-1:nbins
bincount<-1:nbins
for(i in 1:N)
{
binnum<-trunc(d[i]/lag + 1)
bintotal[binnum]<-
(bintotal[binnum]+gamma[i])
bincount[binnum]<-(bincount[binnum]+1)
}
}
for(i in 1:nbins)
{
if(bincount[i]==i) gammamean[i]<-NA
if(bincount[i]!=i) gammamean[i]<-
bintotal[i]/bincount[i]
if(bincount[i]==i) mid[i]<-NA
if(bincount[i]!=i) mid[i]<-(bin[i]-1)*lag+lag/2
if(bincount[i]==i) bincount[i]<-NA
}
}
gammamean1<-na.omit(gammamean)
mid1<-na.omit(mid)
bincount1<-na.omit(bincount)
```

```
plot(mid1, gammamean1, xlab="Distance",
     ylab="Gamma")
title('3-d Empirical Variogram')

#Code to model Empirical variogram with an
#exponential
#must enter values for sill, nugget and range
#

sill<-VALUE
nugget<-VALUE
range<-VALUE

x<-1:maxd
y<-nugget + (sill-nugget)*(1-exp(-3*x/range))

plot(mid1, gammamean1, xlab="Distance",
     ylab="Gamma")
par(new=T, xaxs="d", yaxs="d")
plot(x,y, type='l', xlab="Distance",
     ylab="Gamma")
title('3-d Model Variogram')
```

REFERENCES

1. Stigliani, W. *University of Northern Iowa Summer Lakes Study 2000-2001 Technical Report* (University of Northern Iowa, Cedar Falls, Iowa, 2002).
2. Wetzel, Robert G., *Limnology, Lake and River Ecosystems*. (Academic Press, San Diego, California, 2001).
3. Ecker, M. D. and Janssen, A., "Statistical Analysis of Water Quality Data from Lake Casey and Silver Lake" Iowa Lake Project Report. (University of Northern Iowa, Cedar Falls, Iowa, 2000). [Contact: ecker@math.uni.edu]
4. Cressie, N., *Statistics for Spatial Data*. (John Wiley, New York, 1993).
5. Stein, M. *Interpolation of Spatial Data: Some Theory for Kriging*. (Springer-Verlag, New York, NY, 1999).
6. Conover, W. *Practical Nonparametric Statistics*. (John Wiley, New York, NY, 1980).

