Water Quality Assessment of Small Water Supply Reservoir Using Statistical and Analytical Methods

Nicolas A.Gonzalez[1], N. R. Swain[1], O. Obregon[1], G. P. Williams[2], E. J. Nelson[3], Dennis L. Eggett[4]
Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT

Abstract. We evaluated twenty-eight years (1980-2007) of spatial-temporal water quality data from Deer Creek reservoir in Utah. The data came from three sampling points representing the lotic, transitional and lentic zones and included measurements of climatological, hydrological, and surface water quality conditions. The time frame spanned dates before and after the completion of the Jordanelle Reservoir (1987-1992), approximately fourteen miles upstream of Deer Creek. Our analysis showed changed conditions in Deer Creek prior to dam construction. On average chlorophyll-a and phosphorus levels have dropped since 1984 while dissolved oxygen levels have remained steady. We evaluated temporal groupings and found that a traditional month distribution following standard seasons was not effective in characterizing the measured conditions; we developed a more representative seasonal grouping by performing a Tukey-Kramer multiple comparison adjustment . Based on this analysis, we determined the best groupings were Cold (December - April), Transition (May and November) and Warm (June - October). We used Analyses of variance (ANOVA) calculations to determine if the temporal and spatial variations were statistically different. We found significant spatial variation in chlorophyll-*a* and nutrients. In general values were higher in the lotic zone than the lentic zone. We performed a principal component analysis (PCA) to determine principal parameters associated with the water quality of the reservoir which confirmed our seasonal groups showing the Cold, Transition and Warm seasons as separate.

1 Introduction

Deer Creek reservoir was built in 1938 as part of the Provo River Project (PSOMAS, 2002). It is located on the Provo River approximately 20 miles to the north-east of Provo, Utah. Deer Creek is a major source of municipal and agricultural water for Utah and Salt Lake counties. The reservoir is characterized with a dam height of 235 feet and a capacity of 152,700 acre-feet (BOR, 2009; Casbeer, 2009) . Deer Creek is one focus of water quality research at Brigham Young University because of the potential proliferation of late-summer algae blooms (PSOMAS, 2002; Miller, 2008).

We evaluated spatial and temporal water quality trends using data collected from three locations: UpperEnd, MidLake, and NearDam (Figure 1) and common methods (Chapman, 1996; Cunha *et al.*, 2011; Rahman *et al.*, 2005) . The three locations represent

[1] Graduate Student, Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT 84602. Email: oliveroob@byu.edu
[2] Associate Professor, Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT 84602. Email: gus.williams@byu.edu
[3] Professor, Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT 84602. Email: jimn@byu.edu
[4] Associate Research Professor, Department of Statistics, Brigham Young University, Provo, UT 84602. Email: theegg@byu.edu

the lotic, transitional and lentic zones, respectively. Deer Creek is fed by four main inflows: Provo River, Snake Creek, Daniel's Creek and Main Creek all of which are gaged and monitored with the exception of Main Creek. We did not include, Main Creek in the study. The upper location captures the remaining three influents.

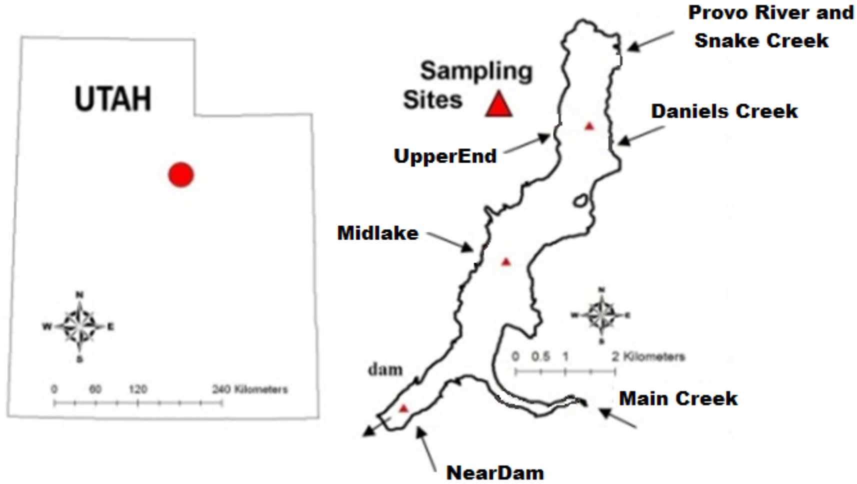

Figure 1. Deer Creek Reservoir and Sampling Locations.

2 Methods

2.1 Data Collection

Data were provided by the Central Utah Water Conservancy District (CUWCD). The data were collected from 1980 to 2007 on average once a month during the sampling season (April-October) and some years when the weather allowed they collected samples throughout the whole year. Subsets of these data were used for different portions of the water quality assessment. Samples were collected at three surface depths; Surface (between 0 and 1 m), Midwater (between the top and bottom of the thermocline) and Bottom (bottom of the reservoir). Most of the water quality parameters used in this study were measurements obtained from the laboratory testing. Some nutrients were considered for this study because of the belief of their direct effect in algae population (Rahman et al., 2005).

Climatological data (air temperature and dew point temperature) were obtained from the National Oceanic and Atmospheric Administration (NOAA). We approximated weather conditions at Deer Creek using data from the Salt Lake City International Airport. Although other sites were closer to our study area, they did not have consistent hourly data for the range of dates needed for our study. Table 1 shows all the parameters considered for this study, not all were used in each analysis.

Table 1. Parameter Names and Descriptions

Model Parameter Name		Parameter Description	Units
Temporal-Spatial Data	MONTH	Month of reading	-
	YEAR	Year of reading	-
	DEPTH	Depth of Reading (Numerical and Categorical)	m
	LOCATION	Sampling site location	-
	TIME	Time rounded to nearest hour	-
Water Quality Data	CHL-a	Chlorophyll-*a*, response variable	μg/L
	DO	Dissolved oxygen	mg/L
	pH	pH	-
	TOT P	Total phosphorus	mg/L
	TW	Water temperature	ºC
Climate Data	TA	Air temperature	ºC
	TD	Dew point temperature	ºC
	REL H	Relative humidity	-
Hydraulic Data	VOLUME	Reservoir volume	$M\ m^3$
	INFLOW	Inflow calculated from outflow and reservoir volume	m^3/s
	OUTFLOW	Outflow measured at the dam	m^3/s
	PROVO UP	Provo river inflow	m^3/s
	PROVO DOWN	Provo river flow downstream of dam	m^3/s
	SNAKE	Snake Creek inflow	m^3/s
	DANIELS	Daniel's Creek inflow	m^3/s

2.2 Water Quality Assessment

We examined chlorophyll-a, dissolved oxygen (DO), and total phosphate as phosphorus (P) over time graphically for any visible trends for the years from 1980 to 2007. We added one unit to each chlorophyll-a value to allow the use of the logarithm transformation.

2.3 Statistical Analyses

The statistical software used to conduct the statistical analyses was JMP version 9.0 (JMP statistical software, SAS Institute, Cary, NC, USA).

2.3.1 Season distribution of month by temperature

We performed a Tukey-Kramer multiple comparison adjustment for the relationship of water temperature versus month. We then grouped the months according to whether their differences were statistically significant or not. We stopped grouping when there was a significant difference between the means of the months.

2.3.2 Analyses of variance

We performed an ANOVA test for DO, chlorophyll-a and P against season group (Cold, Transition or Warm), location (NearDam, MidLake or UpperEnd) and depth (Surface, Midwater or Bottom). We used log transformations for chlorophyll-a and P in order to approach normality. We used a Tukey-Kramer multiple comparison to obtain the p-values and the differences (or ratios when a log transformation was used) for each paired comparison.

2.3.3 Principal components analysis and factor analysis

We used a PCA dimensional reduction to find linear combinations of our dataset that captured most of the variability of Deer Creek. The eigenvalues were used to determine which components were significant to explain the variability; the eigenvectors were used to group the variables into a subset of common factors. We labeled the dataset with different categorical data to assess the temporal and spatial variation by observing how they grouped in a plot of the different principal components (Praus, 2006).

A factor analysis was used to further simplify the dataset by determining the most significant variables at the time of describing the behavior of Deer Creek (Wunderlin *et al.*, 2001).

3 Results And Discussion

3.1 Water quality assessment

3.1.1 Total Phosphorus

Phosphorus concentrations started high in the mid-80s (>0.05) but have since shown a slight decrease throughout the years (See Figure 2). This is consistent with the aims of the Water Quality Management Plan implemented in 1984. The main objective of this plan was to lower the levels of phosphorus in Deer Creek (PSOMAS, 2002). Concentations at the NearDam site are considerably higher than in MidLake and UpperEnd. Howver, the UppderEnd site shows high values in the early 2000s. The inflow from Main Creek (See Figure 1) and storm runoff entering the reservoir may be the cause of these higher concentrations .

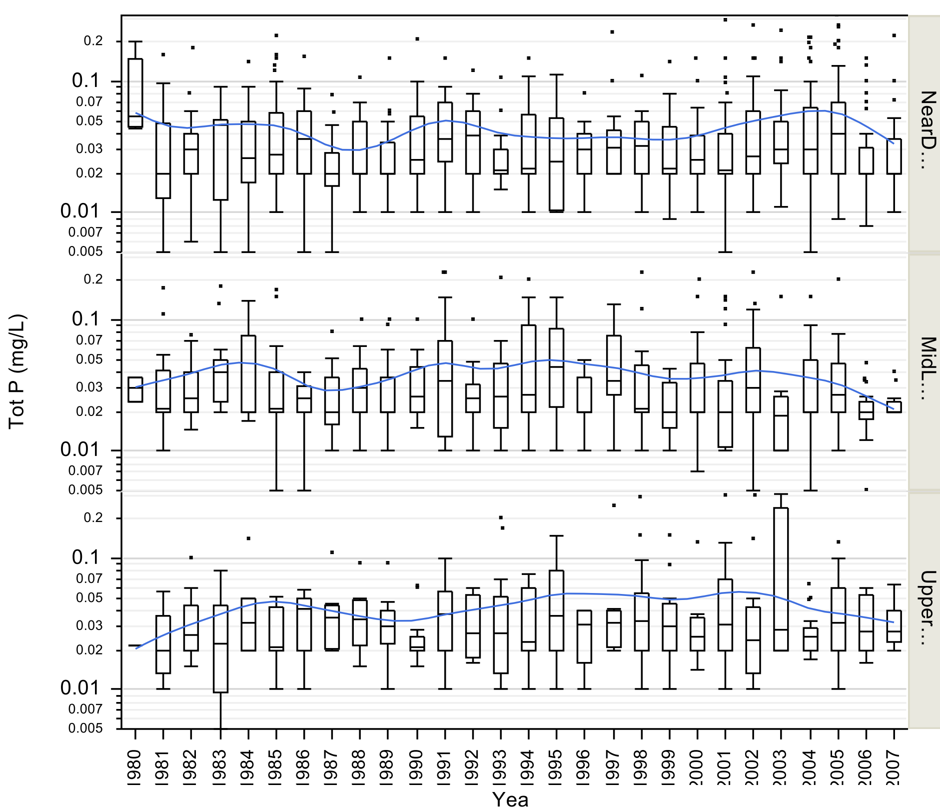

Figure 2. Total Phosphorus versus Year

3.1.2 Dissolved Oxygen

The DO levels do not show any trends over this time period. Figure 3 shows a group of plots of DO versus month for each location. This plot shows lower values in the late summer for all locations. This is probably due to the decay of the high algae concentrations from the summer algal blooms (Stephens *et al.*, 2011). The average DO ranges between 7 and 8 mg/L in all locations sampled. NearDam shows lower DO levels but this may be because deeper locations will have lower DO levels and therefore lower averages.

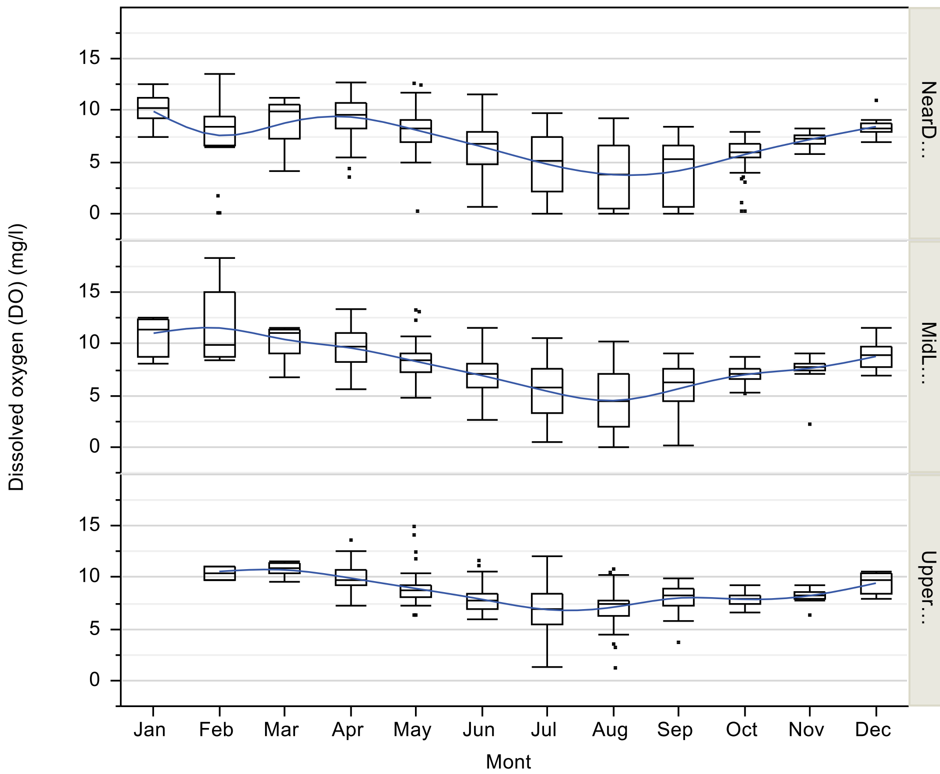

Figure 3. Dissolved Oxigen versus Month

3.1.3 Chlorophyll-a

Chlorophyll-a along the reservoir shows a similar behavior between locations with no apparent spatial trends, this suggests that the parameters affecting the water quality in the lotic zone, affect the transitional and the lentic zones as well.

Figure 4 shows a decrease in chlorophyll-a levels in the mid-90s and early 2000s. This may be due to the completion of the Jordanelle Dam upstream Deer Creek in 1992. In the mid-2000 the average chlorophyll-a increased to approximately the same levels as the 80s. On average across the years analyzed, the chlorophyll-*a* levels have remained the same since 1984, though there have been variations.

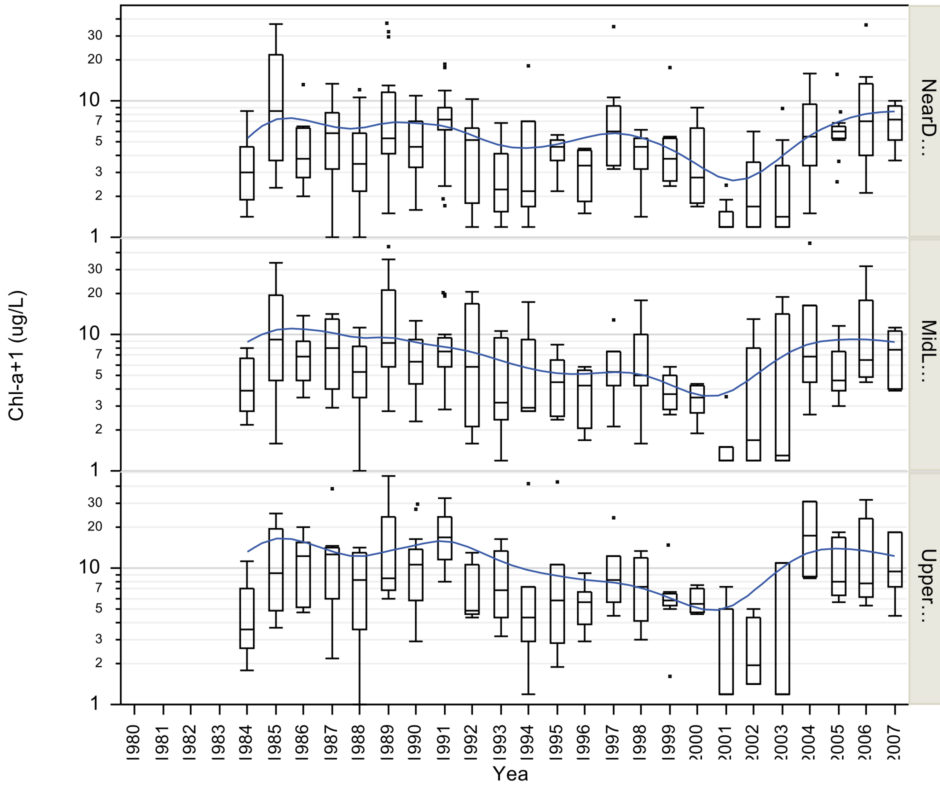

Figure 4. Chlorophyll-a versus Year

3.2 Season distribution

Several methods to characterize temporal data into seasons have been used by different authors. Once these data are group, they are analyzed using PCA, for example Rangel-Peraza *et al.* (2009) used three seasons called Warm dry (March to June), Rainy (November to February) and Cold dry (July to October). Fan *et al.* (2012) described the seasonality by using a continuous wavelet transform and with the results assigning the category of dry season or wet season for values less than 1 and more than 1 respectively.

Reservoirs do not necessarily follow the seasonal pattern established by a calendar or by rainfall, but in order to study them one must impose some sort of a structure where time periods of similar data are grouped together. We formed our seasonal groups using a Tukey-Kramer multiple comparison. The differences of water temperature between months are shown in Table 2. Only the comparisons where there was a significant difference were used to distribute the seasons.

Table 2. Tukey-Kramer Multiple Comparison for Water Temperature versus Month

Month	- Month	Difference	p-Value
Dec	Feb	2.19316	0.1622
Mar	Feb	2.01616	0.2136
Dec	Jan	1.97687	0.6387
Apr	Dec	1.95338	0.0527
Mar	Jan	1.79987	0.7313
May	Nov	0.98652	0.3824
Jun	Oct	0.44631	0.9481
Sep	Jul	0.22630	0.9996
Jan	Feb	0.21629	1.0000
Dec	Mar	0.17701	1.0000

From this grouping we obtained three kinds of seasons. There were called Cold (December, January, February, March, and April), Transition (May and November), and Warm (June, July, August, September and October). See Figure 5.

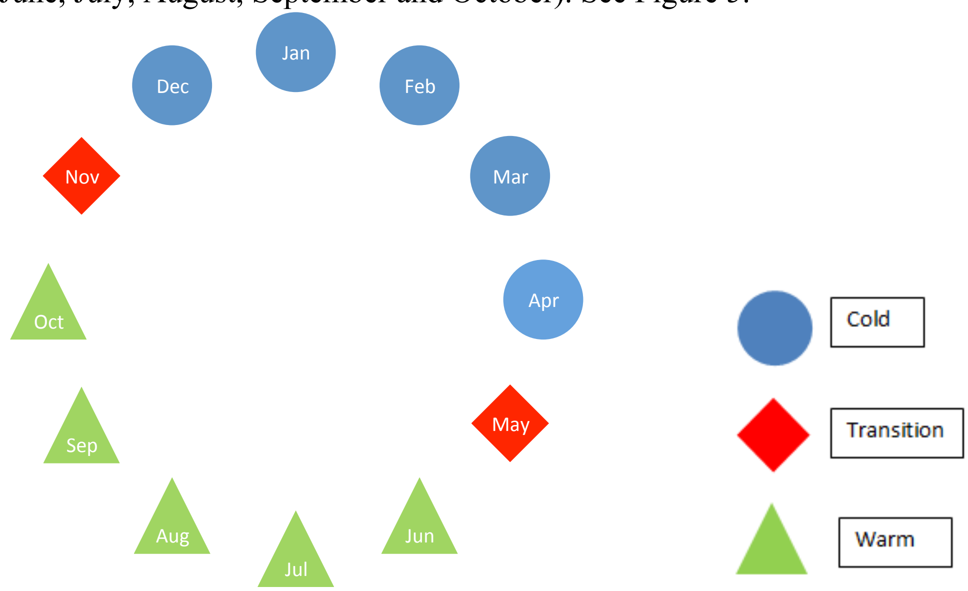

Figure 5. Season Distribution by Month

As it can be seen, the distribution of the Transition months is not continuous but it does not have to be as long as the months show no statistical difference between their means. The data shows some sort of lag in the water conditions compared to air conditions. This suggests that water does not start to get cold until approximately two months into the fall and it does not get warmer until two months into the spring. This approach developed groups better suited to statistical analysis than using traditional seasons.

3.2.1 Analyses of variance

The results for all comparisons are shown in Table 3 along with their p-values and confidence intervals (CI). For DO the results are expressed in differences between the means. Due to the log transformation of phosphates and chlorophyll-a the results are expressed as the ratios between the means (Ramsey *et al.*, 2002).

Table 3. Comparisons for DO, Phosphates and Chlorophyll-a versus Season, Location and Depth

	SEASON						
	Season	- Season	Difference/Ratio	Std Err Dif	Lower CL	Upper CL	p-Value
DO (mg/L)	Cold	Warm	3.764122	0.1602	3.388358	4.139885	<.0001
	Transition	Warm	2.393449	0.146793	2.049133	2.737766	<.0001
	Cold	Transition	1.370672	0.194342	0.914826	1.826518	<.0001
PO4+ (mg/L)	Transition	Warm	0.051413	0.046233	-0.05703	0.15986	0.5067
	Transition	Cold	0.031867	0.061195	-0.11168	0.175409	0.8612
	Cold	Warm	0.019546	0.05017	-0.09814	0.137228	0.9197
CHl-a +1 (mg/L)	Cold	Transition	0.246603	0.101121	0.009127	0.484078	0.0397
	Warm	Transition	0.127692	0.079777	-0.05966	0.315042	0.246
	Cold	Warm	0.11891	0.082337	-0.07445	0.312273	0.3188
	LOCATION						
	Location	- Location	Difference/Ratio	Std Err Dif	Lower CL	Upper CL	p-Value
DO (mg/L)	UpperEnd	NearDam	1.983793	0.167695	1.590448	2.377137	<.0001
	UpperEnd	MidLake	1.197256	0.173207	0.790983	1.603529	<.0001
	MidLake	NearDam	0.786537	0.143182	0.45069	1.122383	<.0001
PO4+ (mg/L)	UpperEnd	MidLake	0.107509	0.048446	-0.00613	0.221146	0.0682
	NearDam	MidLake	0.096475	0.039181	0.004571	0.18838	0.037
	UpperEnd	NearDam	0.011034	0.04699	-0.09919	0.121255	0.9701
CHl-a +1 (mg/L)	UpperEnd	NearDam	0.594902	0.072666	0.424252	0.765551	<.0001
	UpperEnd	MidLake	0.351972	0.075934	0.173647	0.530296	<.0001
	MidLake	NearDam	0.24293	0.069117	0.080613	0.405247	0.0014
	DEPTH						
	Location	- Location	Difference/Ratio	Std Err Dif	Lower CL	Upper CL	p-Value
DO (mg/L)	Surface	Bottom	3.135733	0.153857	2.774817	3.49665	<.0001
	Surface	Midwater	1.74183	0.152858	1.383257	2.100403	<.0001
	Midwater	Bottom	1.393904	0.151869	1.03765	1.750158	<.0001
PO4+ (mg/L)	Bottom	Surface	0.025513	0.048385	-0.08801	0.139036	0.858
	Bottom	Midwater	0.024549	0.043723	-0.07804	0.127134	0.8406
	Midwater	Surface	0.000964	0.048346	-0.11247	0.114396	0.9998
CHl-a +1 (mg/L)	Bottom	Surface	0.433141	0.131992	0.123141	0.74314	0.0031
	Bottom	Midwater	0.32962	0.141238	-0.0021	0.661335	0.0519
	Midwater	Surface	0.103521	0.072328	-0.06635	0.273392	0.3253

3.2.1.1 By season

On average DO is 3.76 mg/L higher in the Cold season than in the Warm season (CI 3.39 and 4.14). Phosphates do not show any statistical difference between seasons and chlorophyll-a showed a differences only in the cold versus transitional months. This indicates that distributing the months in such way helps explain variability related to temperature but is not efficient for explaining variability due to factors like nutrients.

3.2.1.2 By location

Statistical differences were found in all three parameters tested by location. These results indicate higher levels of DO in UpperEnd than in MidLake and NearDam which may have relation to the respiration of the higher levels of algae in UpperEnd (on average 59% higher than NearDam) or from the higher DO concentrations in the inflows. The differences in phosphates are negligible.

3.2.1.3 By depth

DO was significantly different in the three layers representing the water column. No significant differences were found for phosphates. Chlorophyll-a showed differences only between the bottom layer, on average 43% higher than the surface (CI 12% and 74%). A possible reason is that most of the measurements were taken in the surface of the reservoir and chlorophyll-a shows high variability (sample size for surface is ten times greater than the sample size for bottom)

3.3 Principal components analysis

The variables used in the PCA are shown in Table 4 which also shows the eigenvectors for the first three principal components. We chose to use only the first three principal components based on the plot of variance versus component (Figure 6) and picking the location before the curve flattens out (Praus, 2006).

Table 4. Eigenvectors

Parameter	Prin1	Prin2	Prin3
Tot P (mg/L)	-0.03515	-0.22389	-0.15464
pH	-0.11495	0.51470	0.51775
Spec Cond (umho/cm)	-0.39448	-0.22065	0.16839
DO (mg/l)	-0.26643	0.63146	0.05529
Volume (Mcm)	0.23259	0.39628	-0.50558
Air temp (C)	0.47927	-0.00185	0.21057
Release (cms)	0.36519	0.26516	-0.26318
Water Temperature (C)	0.36864	-0.03387	0.54456
Dew Pt (C)	0.45376	-0.09506	0.10418

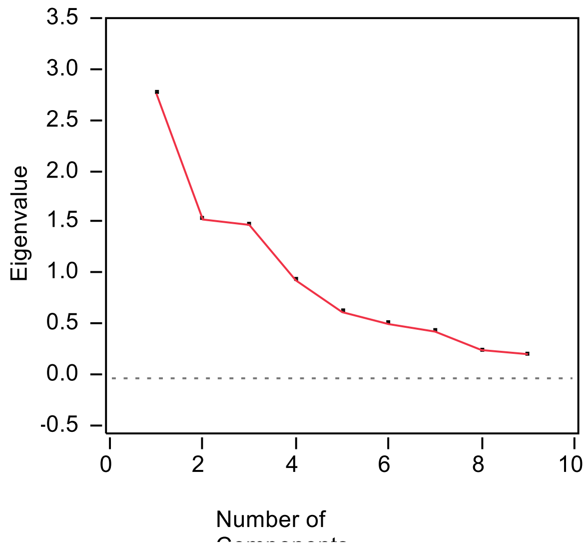

Figure 6. Scree Plot

The eigenvalues are shown in Table 5. The first three eigenvalues explain 65% of the variation in our data set. With a study of 28 years, higher variations are expected, especially when the study is subject to great spatial distances (Ramsey & Schafer, 2002).

Table 5. Eigenvalues and Cumulative Explained Variation

Number	Eigenvalue	Percent	Cum Percent
1	2.7985	31.094	31.094
2	1.5614	17.349	48.443
3	1.5088	16.764	65.207
4	0.9647	10.719	75.926
5	0.6520	7.244	83.170
6	0.5346	5.940	89.110
7	0.4610	5.122	94.232
8	0.2797	3.108	97.340
9	0.2394	2.660	100.000

A scatterplot matrix of the three principal components was created. Season, Location and Depth were used as labels for the plots (Praus, 2006). From the three categorical variables only, Season showed a clear distinction in the cloud created by the principal components (See Figure 7). This shows that there is a moderate transition between the cold and warm season in Deer Creek based on PCA analysis which supports our earlier findings.

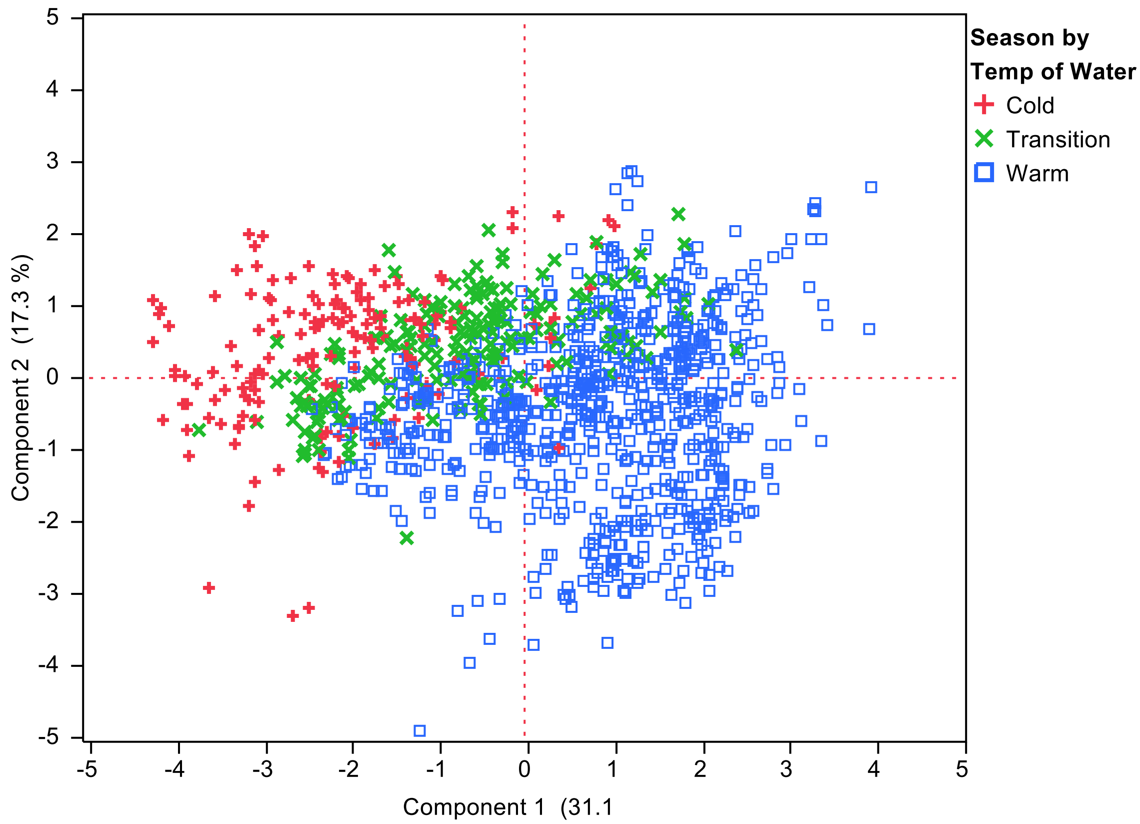

Figure 7. Principal Component 1 versus Principal Component 2

3.4 Factor analysis

Table 6 shows the rotated factor loading with the first three factors obtained from the factor analysis. Correlated factors were underlined (Wunderlin et al., 2001).

Table 6. Rotated Factor Loading

Parameter	**Factor 1**	**Factor 2**	**Factor 3**
Tot P (mg/L)	-0.027520	-0.030049	-0.071707
pH	0.126709	-0.146780	<u>0.674052</u>
Spec Cond (umho/cm)	-0.364387	<u>-0.516087</u>	0.060180
DO (mg/l)	-0.363575	0.134913	<u>0.921632</u>
Volume (Mcm)	-0.077139	<u>0.805347</u>	0.003160
Air temp (C)	<u>0.684650</u>	0.297826	-0.130943
Release (cms)	0.223579	<u>0.585679</u>	-0.036337
Water Temperature (C)	<u>0.928161</u>	-0.098482	0.139794
Dew Pt (C)	<u>0.570706</u>	0.291033	-0.174961

The values underlined follow similar patterns. For factor 1, air temperature, water temperature and dew point temperature were correlated by the factor analysis. For factor 2, Release and Volume were underlined along with Specific Conductivity; this could be due to the high correlation between conductivity and dissolved solids. This factor shows a negative correlation which indicates more dissolved solids at lower volumes in the reservoir or simply less dilution for the same amount dissolved solids. Factor 3 grouped pH and DO which are common characterizers or water chemistry.

4 Conclusions

Time trends of chlorophyll-*a*, DO, and P indicate slight improvement in water quality at Deer Creek since 1980. Lotic, transitional, and lentic zones showed similar or correlated fluctuations throughout the years of the study. The tested season distribution proved to be effective at the moment of showing general variations in the reservoir but it did not show significances when it was correlated to nutrients present in Deer Creek. The first three principal components explained 65% of the variability in the reservoir which is reasonable for a study of 28 years. The factor analysis proved to be effective when grouping different variables by combining variables with similar characteristics.

Acknowledgements. The authors would like to thank the United States Bureau of Reclamation, Upper Colorado Region (USBR) for sponsoring this project and the Central Utah Water Conservancy District (CUWCD) for providing the data used in this study.

References

BOR, 2009. Deer Creek Reservoir. US Bureau of Reclamation.

Casbeer, W., 2009. Phosphorus Fractionation and Distribution across Delta of Deer Creek Reservoir. Brigham Young University, Provo, UT.

Chapman, D. (1996). *Water quality assessments*.

Cunha, M. D. G. F., & do Carmo Calijuri, D. M. (2011). Limiting factors for phytoplankton growth in subtropical reservoirs: the effect of light and nutrient availability in different longitudinal compartments. *Lake and Reservoir Management, 27*(2), 162-172.

Fan, X., Cui, B., Zhang, K., Zhang, Z., & Shao, H. (2012). Water Quality Management Based on Division of Dry and Wet Seasons in Pearl River Delta, China. *CLEAN–Soil, Air, Water*.

Miller, J. B., 2008. East Canyon Reservoir CE-QUAL-W2 Model, 2008 Water Quality Assessment Utah DEQ Phosphorus TMDL. JM Water Quality, LLC, Hooper.

Praus, P. (2006). Water quality assessment using SVD-based principal component analysis of hydrological data. *Water SA, 31*(4), 417-422.

PSOMAS, 2002. Deer Creek reservoir drainage. TMDL Study. Salt Lake City.

Rahman, A., Al Bakri, D., Ford, P., & Church, T. (2005). Limnological characteristics, eutrophication and cyanobacterial blooms in an inland reservoir, Australia. *Lakes & Reservoirs: Research & Management, 10*(4), 211-220.

Ramsey, F. L., & Schafer, D. W. (2002). *The statistical sleuth: a course in methods of data analysis* (Vol. 163): Duxbury/Thomson Learning Australia; Pacific Grove, CA.

Rangel-Peraza, J. G., De Anda, J., González-Farias, F., & Erickson, D. (2009). Statistical assessment of water quality seasonality in large tropical reservoirs. *Lakes & Reservoirs: Research & Management, 14*(4), 315-323.

Stephens, R., Obregon, O., Chilton, R. E., Williams, G. P., & Nelson, E. J. (2011). *Field Algae Measurements Using Empirical Correlations at Deer Creek Reservoir*.

Wunderlin, D. A., Pesce, S. F., & Hued, A. C. (2001). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquia River Basin (Córdoba-Argentina). *Water Research, 35*(12), 2881-2894.