2009

Statistical changes in lakes in urbanizing watersheds and lake return frequencies adjusted for trend and initial stage utilizing generalized extreme value theory

Shayne Paynter
University of South Florida

Statistical Changes in Lakes in Urbanizing Watersheds and Lake Return Frequencies Adjusted for Trend and Initial Stage Utilizing Generalized Extreme Value Theory

by

Shayne Paynter

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Civil and Environmental Engineering
College of Engineering
University of South Florida

Major Professor: Mahmood Nachabe, Ph.D.
Mark Ross, Ph.D.
George Yanev, Ph.D.
Mark Stewart, Ph.D.
Jayajit Chakraborty, Ph.D.

Date of Approval:
January 5, 2009

Keywords: regression, time series, autocorrelation, flood, drought

Table of Contents

List of Tables

List of Figures

Statistical Changes in Lakes in Urbanizing Watersheds and Lake Return Frequencies Adjusted for Trend and Initial Stage Utilizing Generalized Extreme Value Theory

Shayne Paynter

ABSTRACT

Many water resources throughout the world are demonstrating changes in historic water levels. Potential reasons for these changes include climate shifts, anthropogenic alterations or basin urbanization. The focus of this research was threefold: 1) to determine the extent of spatio-temporal changes in regional precipitation patterns 2) to determine the statistical changes that occur in lakes with urbanizing watersheds and 3) to develop accurate prediction of trends and lake level return frequencies.

To investigate rainfall patterns regionally, appropriate distributions, either gamma or generalized extreme value (GEV), were fitted to variables at a number of rainfall gages utilizing maximum likelihood estimation. The spatial distribution of rainfall variables was found to be quite homogenous within the region in terms of an average annual expectation. Furthermore, the temporal distribution of rainfall variables was found to be stationary with only one gage evidencing a significant trend.

In order to study statistical changes of lake water surface levels in urbanizing watersheds, serial changes in time series parameters, autocorrelation and variance were evaluated and a regression model to estimate weekly lake level fluctuations was developed. The following general conclusions about lakes in urbanizing watersheds were reached: 1) The statistical structure of lake level time series is systematically altered and is related to the extent of urbanization 2) in the absence of other forcing mechanisms, autocorrelation and baseflow appear to decrease and 3) the presence of wetlands adjacent to lakes can offset the reduction in baseflow.

In regards to the third objective, the direction and magnitude of trends in flood and drought stages were estimated and both long-term and short-term flood and drought stage return frequencies were predicted utilizing the generalized extreme value (GEV) distribution with time and starting stage covariates. All of the lakes researched evidenced either no trend or very small trends unlikely to significantly alter prediction of future flood or drought return levels. However, for all of the lakes, significant improvement in the prediction of extremes was obtained with the inclusion of starting lake stage as a covariate.

1.0 Regional Scale Spatio-temporal Consistency of Precipitation Variables Related to Water Resource Management and Planning

1.1 Background

This dissertation consists of three main sections which correspond to three papers submitted to technical journals and represent three distinct but interrelated subjects. The first section investigates the spatial homogeneity and temporal stationarity of rainfall in a given region. As lakes and the changes in lake levels are the major focus of this research, it is necessary to first identify any spatial or temporal trends in the major input to lake levels so that any further analysis can take these trends into account. The second section investigates changes in lake levels induced by urbanization. The focus was to remove or account for signals other than urbanization, such as rainfall or control structures, as much as possible. Once lakes with sufficiently isolated urbanization signals were identified, changes in time series model parameters, autocorrelation and baseflow were investigated. The third section, utilizing the methods developed in the first section as well as taking advantage of the autocorrelation in lake levels identified in the second section, sought to significantly improve the prediction of lake level return periods.

During the last century, the planet has experienced tremendous growth in population and development. In addition, many researchers world wide have concluded that climate change is occurring in the form of increases or decreases in temperature, rainfall, drought, evaporation and other climatological variables. Specifically, in regards to precipitation, Kunkel and Andsager (1999) and Karl and Knight (1998) investigated extreme rainfall events and found upward trends utilizing the nonparametric Kendall statistical test in both the magnitude and number of events in some parts of the United States. However, Dahamsheh and Aksoy (2007) found no such trends in rainfall in other areas of the world using the nonparametric Spearman rank order correlation statistical test. Garcia et al (2007) has identified positive, negative and absence of trends in extreme rainfall in various regions of Spain with nonparametric tests as well as by utilizing time-dependent parameters of the Generalized Extreme Value (GEV) distribution. Zolina et al (2004) applied the gamma distribution to daily rainfall from various gages across Europe and found substantial variability from region to region in terms of both the presence and direction of trends and the shape and scale parameters of the distribution.

At the same time that rainfall patterns are changing, many changes in the general trends of lake, stream and other surface water levels have occurred. Many anthropogenic factors, such as watershed urbanization, water supply pumping and structural changes to the water body itself or climatic changes could be responsible for these trends. Water resources are vital for many reasons including recreation, tourism, environment, ecology and water supply. Understanding what is responsible for impacts to water levels is key to determining an effective future management plan. As rainfall, or in the case of drought the lack thereof, is the major contributing factor influencing water stages, precipitation trends must be identified before investigating other potential factors. Much of the literature evaluates different statistical methods for determining trend since trend detection in hydro-climatological data is complicated by the time scale of data, non-normal distributions, seasonality, autocorrelation, data collection methods, censored or missing data, non-stationarity and other difficulties. Hirsch et al (1982) developed a seasonal Kendall nonparametric test that overcomes some of the difficulties of traditional methods of trend detection. Using Monte Carlo simulations, Zhang et al (2004) compared ordinary least squares regression, the nonparametric Kendall test, and allowing the parameters of the GEV distribution to vary with time. The study found that while the nonparametric test is more robust than ordinary least squares, varying the parameters of the GEV distribution significantly outperforms the other two methods and increases the chances of correctly identifying a trend. Recent research shows wide application of GEV theory to hydro-climatological data. Katz et al (2002) provides a discussion of the statistics of hydro-climatological extremes and the application of the generalized extreme value distribution. Nadarajah and Shiau (2005), Zhang et al (2004) and Garcia et al (2007) utilized maximum likelihood estimation of generalized extreme value parameters, allowed parameter covariates to vary with time or other phenomena, and employed a likelihood ratio test to determine if the parameter covariates improve the fit.

Identifying trends in both space and time is integral to effective water resource management and planning. Spatial trends are important at the regional scale because often, insufficient climatological data at a particular lake, stream, reservoir or other water resource are available. Establishing regional precipitation homogeneity allows data from regional gages to be utilized at a particular water resource anywhere within that region.

Temporal trends are equally important to water resource management, as robust identification of a downward trend could allow for early implementation of mitigation plans for water supply, wetland or lake restoration and other water resource issues. There is substantial research regarding both spatial and temporal variability of rainfall. Zolina et al (2004) found substantial regional variation in the parameters of the gamma distribution across Europe. In a similar study, Groisman et al (1999) found spatial and temporal regional stability in the shape parameter of the gamma distribution when applied to daily rainfall. The spatial distribution of rainfall trends was investigated by Cannarozzo et al (2006) and found to be spatially homogenous across Sicily. Further research evaluated spatial and temporal dependence of rainfall data across South America and found evidence of regional dependence. Furthermore, it was found that the regional dependence extended further in the latitudinal direction (Kuhn et al, 2007).

Spatial and temporal changes in rainfall are investigated in this research by analyzing the probability distribution of a set of precipitation variables that are likely to influence regional lake, stream or other water resource levels. These variables include annual rainfall, annual rainfall days per year, the maximum rainfall per week in any given year, and the number of days between events. These variables are determined for rainfall gages surrounding Moon Lake, located in Pasco County, Florida, United States. This lake was chosen because there are a sufficient number of rainfall gages within the region with complete data sets and long records. Furthermore, the lake itself has not been anthropogenically altered by pumping or other means and rainfall is the major influence on lake levels. To determine spatial homogeneity, an average set of distribution parameters are estimated and confidence limits established to determine if the variables at each gage fall within these limits. To determine temporal stationarity, distribution parameters are allowed to vary with time to ascertain if doing so provides a better fit than constant parameters.

Much of the referenced literature regarding climate has focused on identifying large-scale global or continental changes in weather patterns. Identifying trends utilizing GEV distribution parameter covariates rather than ordinary least squares or Mann-Kendall statistical tests is a recent development in the literature; the papers cited have applied this method to flood peaks, daily precipitation and temperature. This research applies GEV distribution covariates to rainfall variables that have not yet been analyzed in such a way. Furthermore, the utilization of parameter confidence limits applied in such a way to the variables analyzed was not found in the literature review. The objectives of this research are: 1) Determine if the rainfall variables analyzed exhibit spatial homogeneity and temporal stationarity utilizing methods that can be adapted to other regions 2) Develop a representative distribution for each variable that can be utilized for water resource management and planning within a given region.

1.2 Materials and Methods

1.2.1 Data

Because rainfall is highly variable from gage to gage even within a limited area, several rainfall gages within a 40-kilometer radius of Moon Lake were analyzed. Several gages within 8, 16, 24, 32 and 40 kilometer radii of the lake were selected based on achieving adequate coverage of the region. Furthermore, gages with at least 25 years of daily rainfall data that were at least 95 percent complete were selected. In some cases, gage records were extended by joining the records of two immediately adjacent gages. Rainfall data available to the general public was obtained from both Southwest Florida Water Management District and the National Climatic Data Center. Figure 1-1 gives a graphic overview of gage locations. Table 1-1 gives a summary of the data associated with each gage.

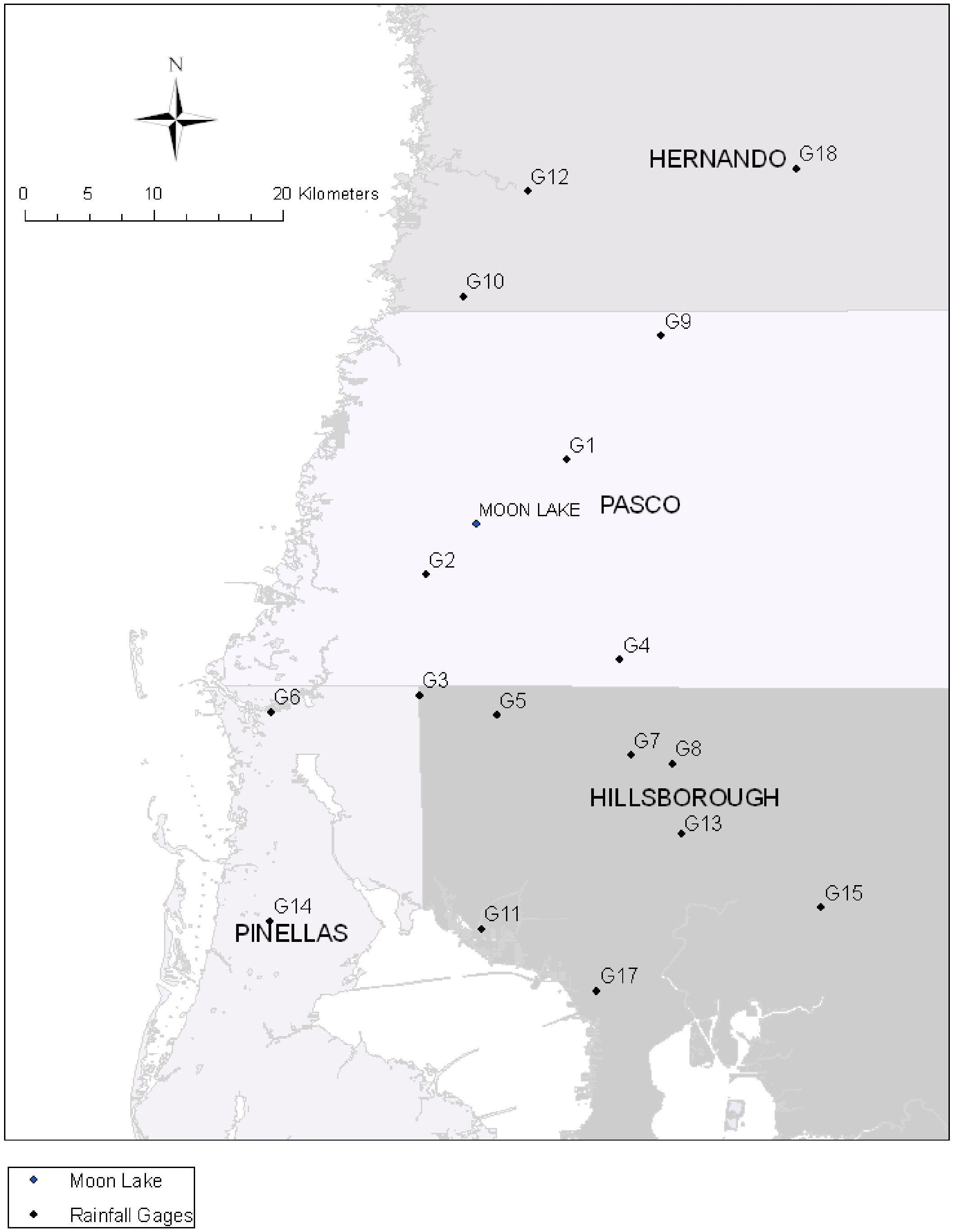

Figure 1-1: Rainfall gage location

Table 1-1: Rainfall gage data summary

Gage No.	Station Name	Year Data Begins	Year Data Ends	Percent Complete	Kilometers from Moon Lake
G1	Growers Kent	1973	2007	99.8	8
G2	Starkey	1983	2007	97.5	8
G3	Eldridge Wilde	1973	2007	99.0	16
G4	South Pasco	1976	2007	98.8	16
G5	Island Ford	1973	2007	98.3	16
G6	Tarpon Springs	1901	2004	98.7	24
G7	Lutz	1965	2005	100.0	24
G8	Whalen	1975	2005	100.0	24
G9	Crews Lake	1976	2007	98.9	24
G10	Hunter Lake	1976	2006	99.1	24
G11	Imperial Key	1974	2007	99.2	32
G12	Weeki Wachee	1971	2007	99.6	32
G13	Bay Lake	1970	2007	99.1	32
G14	Dunedin	1970	2005	99.9	40
G15	Temple Terrace	1975	2007	99.7	40
G16	St. Leo	1901	2007	98.3	40
G17	Tampa Int.	1901	2007	99.8	40
G18	Horse Lake	1981	2007	98.6	40

1.2.2 Precipitation Variables

Because the focus of this research is on water resource planning and management, the rainfall/runoff/water body stage relationship is key. As such, rainfall variables to be analyzed for the presence of trends need to be selected carefully. Variables that will likely have a significant impact on water levels include: total annual rainfall, rainfall days per year, annual maximum event rainfall, and annual maximum interevent days. Because lake levels exhibit significant autocorrelation, variables too close in time cannot be considered independent and therefore cannot be used to analyze trends. Before selecting the aforementioned variables, Moon Lake level data was analyzed to determine the extent of autocorrelation via the autocorrelation function. Before the autocorrelation could be applied, the lake data was detrended and any seasonality removed by differencing. The autocorrelation function, a dimensionless measure of linear dependence of time series values at lag k, r_k, is given by:

$$\frac{\{\sum_{t=1}^{N-K}(x_t-\bar{x})(x_{t+k}-\bar{x})\}}{\sum_{t=1}^{N}(x_t-\bar{x})^2} \tag{1}$$

where the numerator of the equation is the autocovariance, a measure of how related the variance from the mean adjacent time series values are and the denominator is the variance at lag 0. Moon Lake exhibits statistically significant autocorrelation up to nearly one year; therefore, annual variables are the focus of this research.

Annual rainfall is the major water resource input and is directly related to water stages. Any significant upwards or downwards trend will correspond to increasing or decreasing future water levels. Lakes, streams and other water resources require a regular replenishment of rainfall to maintain water levels throughout the year. The distribution of rainfall over a given year will contribute to changes in water levels as a few heavy storms per year would generate a different water surface level signature than many small events. The number of rainfall days per year, in combination with the event and interevent annual maximum, will capture this dynamic. Because there is unlikely to be any effect on stages from very small amounts of rainfall due to interception, pooling, transpiration and other extractions, rainfall less than 0.5 cm for any given day is filtered out. For purposes of this research, rainfall days are defined as any day with rainfall greater than 0.5 cm.

Changes in the magnitude and frequency of extreme events will be identified by the maximum annual event rainfall variable. For purposes of this variable and its relation to water stages, an event is defined as the sum total of a week of rainfall; a moving weekly window was applied throughout each year. Increasing trends in interevent times, which correspond to drought, will correlate to future lowered water levels. Interevent time is defined as any number of consecutive days with daily rainfall less than 0.5 cm. In order to evaluate spatial homogeneity, each variable at each gage location was fitted with the most applicable distribution. In the case of the number of total annual rainfall and rainfall days per year, the gamma distribution was utilized. In the case of extreme variables, the annual maximum event rainfall and the annual maximum interevent time, the GEV distribution was used.

1.2.3 Gamma Distribution

The use of the gamma distribution for precipitation data has been established in the literature (Zolina et al, 2004; Semenov and Bengstsson, 2002; Watterson and Dix, 2003; Groisman et al, 1999), although other distributions such as the Weibull and Poisson have also been used (Sharda and Das, 2005; Burgueno et al, 2004). The gamma distribution is positively skewed and has good flexibility by allowing for variability in both mean and variance with its shape and scale parameters. The gamma distribution function is given by:

$$f(x,\alpha,\beta) = \frac{(x/\beta)^{\alpha-1} e^{-x/\beta}}{\beta\Gamma(\alpha)} \quad \text{where } x \ge 0, \alpha, \beta > 0 \qquad (2)$$

and

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \tag{3}$$

where x is the random variable, i.e. total annual rainfall, α is the shape parameter, β is the scale parameter and Γ(α) is the gamma function. The parameters of the gamma function are estimated using maximum likelihood estimation. Maximum likelihood has been found to generally provide better estimates of parameters for the gamma distribution when compared to other methods (Choi and Wette, 1969; Wilks, 1990). Maximum likelihood estimation determines the parameters that maximize the probability of the sample data by maximizing the likelihood function, either by differentiating the log-likelihood function and equating it to zero or, if this does not yield explicit solutions, by using numerical techniques such as the Newton-Raphson method. The log-likelihood function for the gamma distribution developed by Choi and Wette (1969) is given as follows:

$$l(\alpha, \beta) = n\left\{\alpha \log \beta - \log \Gamma(\alpha)\right\} + (\alpha - 1)\sum_{i=1}^{n} \log x_i - \beta \sum_{i=1}^{n} x_i \tag{4}$$

where x_i………x_n represent a random sample of the gamma distribution random variable.

1.2.4 GEV Distribution

The use of the GEV distribution has gained widespread application in recent literature because of its flexibility and ability to capture the frequency of extremes (Martins and Stedinger, 2000; Nadarajah and Shiau, 2005; Morrison and Smith, 2002). The GEV is the generalized form of three commonly applied extreme value distributions: the Gumbel, the Frechet and the Weibull. The GEV is applicable to variables of block maxima, where the blocks are equal divisions of time. The GEV cumulative distribution function is given by:

$$F(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{5}$$

where x is the random variable, μ is the location parameter, σ is the scale parameter and ξ is the shape parameter and 1+ ξ(x- μ)/ σ > 0. It readily follows that the sub-distributions are:

$$\text{Gumbel:}\quad F(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, -\infty < x < \infty \tag{6}$$

$$\text{Frechet:}\quad F(x) = \begin{cases} 0 & x \le \mu \\ \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\} & x > \mu \end{cases} \tag{7}$$

$$\text{Weibull:}\quad F(x) = \begin{cases} \exp\left\{-\left(-\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\} & x < \mu \\ 1 & x \ge \mu \end{cases} \tag{8}$$

In a similar fashion to the gamma distribution parameters, GEV distribution parameters are determined using maximum likelihood estimation. The log-likelihood function, for ξ ≠ 0, is given by:

$$l(\mu,\sigma,\xi) = -m\log\sigma - (1+1/\xi)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi}$$

given that $1+\xi\left(\frac{x_i-\mu}{\sigma}\right) > 0$ for i = 1,……, m (Coles, 2004) (9)

Also similar to the gamma distribution, there is no explicit solution for the log-likelihood function and it must be solved using numerical methods.

1.2.5 Spatial Analysis

Once the distribution parameters were estimated for each variable at each of the rainfall gages, the fits were confirmed with the Kolmogorov-Smirnov test statistic. The Kolmogorov-Smirnov test statistic was applied to the gamma and GEV distributions at a given significance level to determine the goodness of fit. If the test statistic, D, given by $D = \max\left|P_x(x) - S_n(x)\right|$, where $P_x(x)$ is the complete theoretical cumulative distribution function and $S_n(x)$ is the cumulative density function based on n observations, was greater than the test statistic at a given significance level, the hypothesis that the sample data fits a given distribution was rejected (Haan, 2002). Fits were then averaged with 99-percent confidence limits to create a representative distribution for the region. Given the substantial variation of rainfall across large regions of Florida, if the gage-specific distributions fall within the confidence limits of the average, spatial homogeneity is reasonably established for the region in which the gages are located. The focus is to establish an average expectation for each variable in any given year; therefore, it is especially important that variable distribution fits are contained by the confidence limits near the 0.5 percentile.

1.2.6 Temporal Analysis

In order to analyze the temporal variation, distribution parameters were allowed to vary with time and then compared to the original distribution model to determine if a statistically significant better fit was achieved. If there is an upward or downward trend in a particular hydro-climatological variable, the extreme values themselves are generally getting larger or smaller over time and the distribution itself is potentially changing. Changing distribution parameters with time allows for the distribution to be non-stationary and also gives an estimate on the rate of change. The use of a GEV parameter covariate, such as time, to identify trends in hydro-climatological data has been well established (Katz et al, 2002; Nadarajah and Shiau, 2005; Garcia et al, 2007). For consistency in temporal trend analysis and evaluation of parameter rates of change, the annual rainfall and annual rainfall days variables were fitted with the GEV distribution similar to the block maxima variables and then parameters were allowed to vary with time. It should be noted that as a check on the GEV parameters, the original gamma fitted variable parameters were allowed to vary with time to identify any differences from the GEV models in trend detection.

For purposes of this research, model 1 was the GEV distribution with parameters μ, σ and ξ held constant; model 2 is a submodel of model 1 with

$$\mu = a+bt \tag{10}$$

$$\sigma = c+dt \tag{11}$$

$$\xi = e+ft \tag{12}$$

where t is the time in years and a, b, c, d, e and f are constants of a linear trend evaluated at each year. The log-likelihood for the GEV distribution with parameters that are a function of time is given by:

$$l(\mu,\sigma,\xi) = -\sum_{i=1}^{m}\left\{\log\sigma(t) + (1+1/\xi(t))\log\left[1+\xi(t)\left(\frac{x_t-\mu(t)}{\sigma(t)}\right)\right] + \left[1+\xi(t)\left(\frac{x_t-\mu(t)}{\sigma(t)}\right)\right]^{-1/\xi(t)}\right\}$$

$$\text{given that } 1+\xi(t)\left(\frac{x_t-\mu(t)}{\sigma(t)}\right) > 0 \text{ for all t = 1,……..,m (Coles, 2004)} \qquad (13)$$

Once parameters were estimated for both cases, the models were compared to determine if the time covariate gives a statistically significant better fit and parameters are indeed changing with time. In order to test model 1 against model 2, the likelihood ratio test was utilized. If l_1 and l_2 represent the maximized log-likelihoods of model 1 and model 2, respectively, then a deviance statistic is given by $D = 2\{l_2 - l_1\}$. Assuming a chi-square distribution, a quantile, c_α, at significance α can be determined and if D> c_α, the submodel explains significantly more of the variation in the data (Coles, 2004). In cases where a model with time-dependent parameters shows a significantly better fit based on the likelihood ratio fit, fits were further investigated by examining standard quantile plots. However, because model 2 is non-stationary and parameters are varying at each observation, the random variable X should be transformed to a new variable Z for the quantile plot. A transform to the standard Gumbel distribution is given by (Coles, 2004):

$$Z_t = \frac{1}{\xi(t)}\log\left\{1+\xi(t)\left(\frac{X_t-\mu(t)}{\sigma(t)}\right)\right\} \qquad (14)$$

1.3 Results and Discussion

In order to examine the general bounds and consistency of the data, average, maximum and minimum values were calculated for variables at each gage. Table 1-2 gives a summary of the variables determined at each rainfall gage as well as the standard deviation, variance and average for all gages. From Table 1-2, it is apparent that the data is generally quite consistent. The standard deviation for each of the four variables, annual rainfall, annual rainfall days, annual event maximum and annual interevent days indicates minimal variation around the average. One exception is gage G17; the values for this gage represent the low average value for annual rainfall and annual event rainfall. Although the values for both variables are within three standard deviations of the mean, indicating the gage may not be a complete outlier, there are some potential physical reasons this gage may not be consistent with the others. This is the southernmost gage, located 40 kilometers from Moon Lake, and is near the northernmost part of the South Tampa peninsula. It is possible that precipitation dynamics at this gage are influenced by proximity to Tampa Bay and Little Tampa Bay. Furthermore, this gage is in the most urbanized area of all the gages, which may have an effect on rainfall patterns.

Table 1-2: Precipitation variable summary

Gage No.	Avg. annual rainfall (cm)	Max	Min	Avg. annual rainfall days	Max	Min	Avg. annual max. event rainfall (cm)	Max	Min	Avg. annual inter-event days	Max	Min
G1	145.3	236.	88.4	61.1	103.	41.0	19.8	42.2	9.9	39.4	71.0	19.0
G2	127.0	199.	92.2	55.9	71.0	38.0	18.0	43.7	6.6	40.1	67.0	25.0
G3	130.3	175.	95.3	54.9	74.0	39.0	19.1	38.6	9.4	42.7	84.0	24.0
G4	130.3	172.	78.2	57.2	70.0	41.0	17.3	34.5	10.2	40.4	82.0	22.0
G5	142.0	197.	83.1	60.6	81.0	32.0	20.1	44.7	11.4	40.4	76.0	22.0
G6	130.0	190.	87.9	55.0	80.0	34.0	19.3	40.1	11.2	42.7	76.0	24.0
G7	128.5	190.	87.6	54.8	72.0	33.0	17.8	33.3	8.4	44.3	94.0	21.0
G8	142.2	209.	95.8	58.5	73.0	43.0	18.8	30.0	9.4	39.3	67.0	21.0
G9	139.2	203.	65.8	55.7	72.0	35.0	19.1	40.4	7.6	42.5	79.0	24.0
G10	131.8	187.	89.2	54.9	73.0	42.0	19.3	46.5	9.9	41.4	66.0	24.0
G11	150.4	225.	101.	55.5	76.0	35.0	18.0	30.0	10.4	36.5	76.0	19.0
G12	148.1	255.	96.3	58.6	87.0	42.0	21.3	38.4	10.2	41.5	73.0	24.0
G13	148.8	225.	62.2	64.0	86.0	27.0	19.6	32.5	9.4	39.0	105.	19.0
G14	136.7	212.	75.2	59.6	79.0	38.0	20.1	65.0	11.2	40.9	81.0	24.0
G15	139.4	222.	85.6	59.8	71.0	41.0	18.3	34.5	8.6	40.2	66.0	22.0
G16	138.2	192.	92.5	59.7	79.0	45.0	17.8	39.6	9.9	41.7	78.0	23.0
G17	115.6	172.	75.9	51.9	68.0	37.0	16.0	35.1	7.9	41.3	69.0	23.0
G18	142.2	186.	93.2	59.3	82.0	41.0	18.0	36.3	9.9	39.9	70.0	23.0
Std. Dev.	9.1	23.1	10.7	3.0	8.4	4.6	1.3	8.1	1.3	1.8	10.2	1.9
Var.	32.5	209.	44.7	9.0	70.7	21.2	0.5	25.4	0.8	3.1	104.	3.7
Total	136.9	202.	85.9	57.6	77.6	38.0	18.8	39.1	9.7	40.8	76.7	22.4

1.3.1 Spatial Analysis

Both the gamma and GEV distributions gave generally good fits to the respective variables they were applied to, based upon the Kolmogorov-Smirnov 99-percent test statistic. The parameters for the gamma distribution based upon maximum likelihood estimation as well as goodness-of-fit estimations are summarized in Table 1-3. Since gage G16 has the longest record and is generally representative of the fit dynamics at other gages, quantile-quantile plots for each variable at this gage are displayed for visual inspection (Figures 1-2, 1-4 and 1-6).

Table 1-3: Gamma variable parameter summary

Gage No.	Annual Rainfall				Annual Rainfall Days			
	Shape α	Scale β	K-S α*	K-S D**	Shape α	Scale β	K-S α	K-S D
G1	16.49	8.81	0.28	0.11	20.98	2.91	0.28	0.06
G2	28.47	4.46	0.34	0.14	57.84	0.97	0.34	0.16
G3	32.76	3.98	0.27	0.14	47.77	1.15	0.27	0.1
G4	34.22	3.81	0.29	0.10	48.85	1.17	0.29	0.14
G5	33.77	4.21	0.29	0.07	28.47	2.13	0.29	0.17
G6	34.06	3.82	0.24	0.08	44.6	1.23	0.24	0.12
G7	25.12	5.12	0.26	0.10	34.34	1.6	0.26	0.14
G8	29.47	4.82	0.29	0.12	52.53	1.11	0.29	0.08
G9	18.61	7.47	0.29	0.13	45.93	1.21	0.29	0.15
G10	31.60	4.17	0.29	0.10	47.55	1.15	0.29	0.09
G11	21.30	7.06	0.28	0.11	40.59	1.37	0.28	0.12
G12	17.34	8.54	0.27	0.12	27.97	2.09	0.27	0.14
G13	18.20	8.19	0.27	0.12	22.64	2.83	0.27	0.11
G14	20.60	6.63	0.27	0.09	26.24	2.27	0.27	0.12
G15	22.30	6.26	0.28	0.06	77.83	0.77	0.28	0.12
G16	32.71	4.22	0.23	0.11	53.54	1.12	0.23	0.1
G17	22.93	5.04	0.24	0.11	60.03	0.86	0.24	0.11
G18	25.39	5.60	0.32	0.11	35.79	1.66	0.32	0.07

*K-S α refers to Kolmogorov-Smirnov statistic at 99 percent significance
**K-S D refers to Kolmogorov-Smirnov test statistic, D for the gamma distribution

1.3.1.1 Annual Rainfall

Examining the quantile-quantile plot (Figure 1-2) of the observed annual rainfall and the gamma predictions provides a visual confirmation of the fit; observed and fitted values roughly follow a 45-degree line indicating agreement. Below approximately 110 cm and upwards of 160 cm of rainfall, Figure 1-2 data points begin to show some deviance from the 45-degree match line. The cumulative distribution for each gage, with the average distribution and corresponding 99-percent confidence limits superimposed, is shown in Figure 1-3. The 99-percent lower confidence limit for the average distribution contains the fits for all gages with the exception of gage G17, which is well outside this limit for all frequencies. Also observed in Figure 1-3, and as observed in the quantile plot, for frequencies near the 0.7 percentile, corresponding to nearly of 165 cm of rainfall, the fits for gages G1, G11, G12 and G13 begin to fall slightly outside of the upper confidence limit.

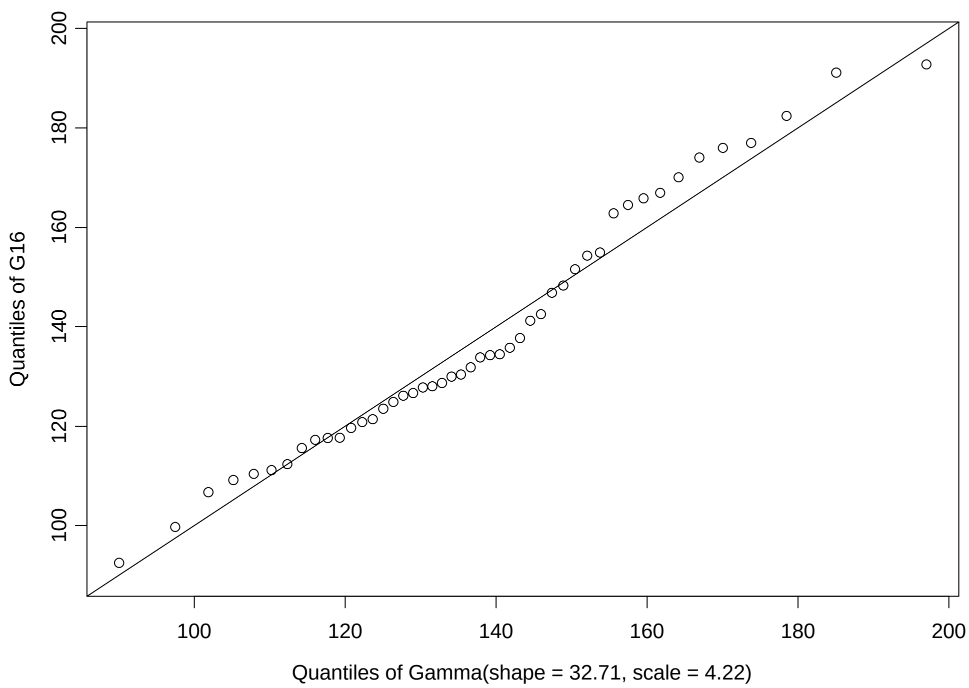

Figure 1-2: Gage G16 quantile plot of annual rainfall

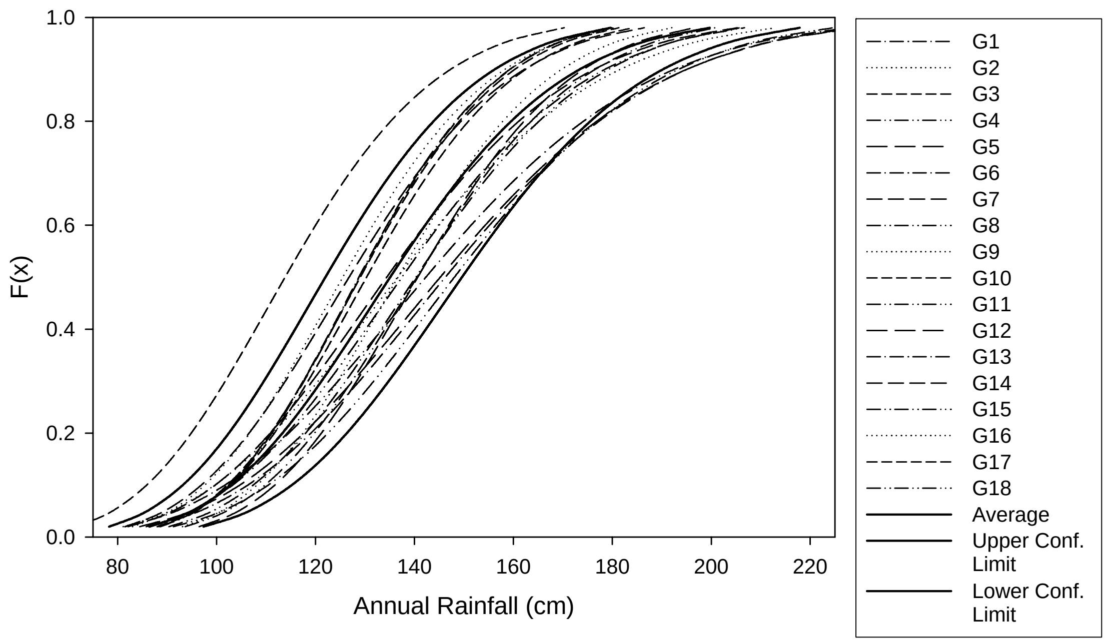

Figure 1-3: Annual rainfall cumulative distribution

1.3.1.2 Rainfall Days per Year

The quantile-quantile plot for gage G16 rainfall days per year (Figure 1-4) indicates a fit with slight variation around the match line near the lower percentiles and above 60 days. The cumulative distribution (Figure 1-5) shows that while only gage G15 falls outside of the confidence limits at a low percentile, several fits fall outside the confidence limits at high percentiles. Gage G17 falls outside the lower confidence limit above the 0.18 percentile and all other gage fits are contained by the lower limit. Along the upper confidence limit, gage G13 falls outside above the 0.38 percentile and gages G1, G5 and G14 fall out approximately above the 0.7 percentile, corresponding to just above 60 days.

The parameters for the GEV distribution based upon maximum likelihood estimation as well as goodness-of-fit estimations are summarized in Table 1-4. It can be seen from the table that the GEV distribution generally gives a good fit based upon the Kolmogorov-Smirnov test statistic. All but two gage fits, including G17, are contained at the 0.5 percentile.

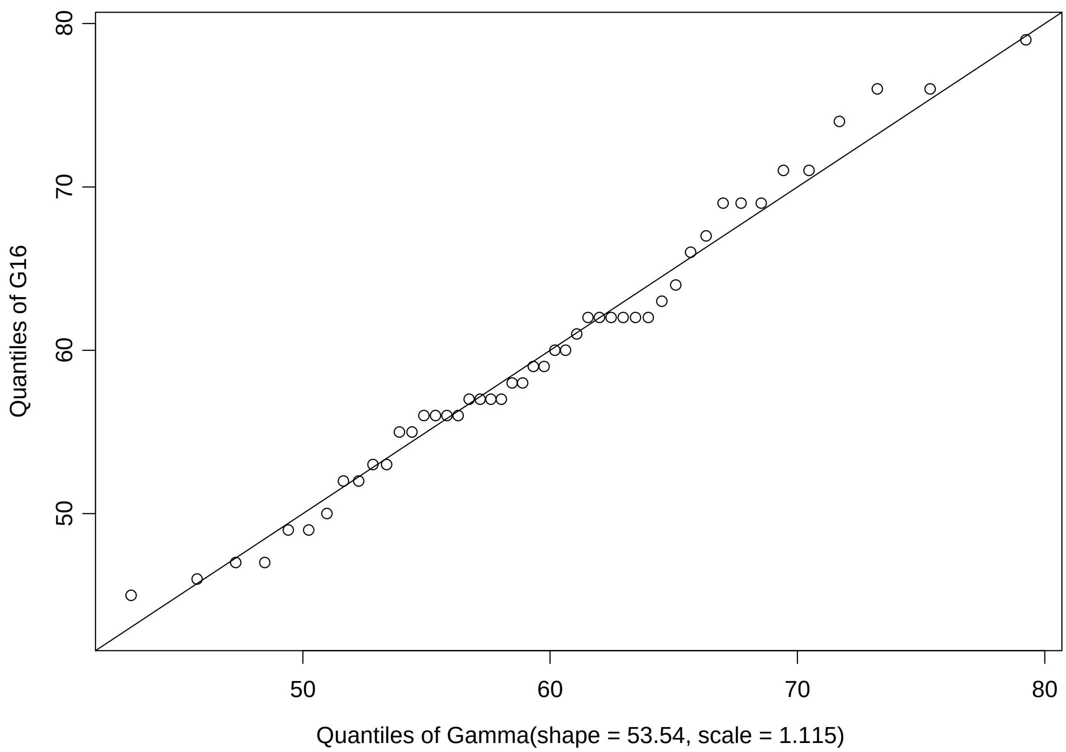

Figure 1-4: Gage G16 quantile plot of the rainfall days per year

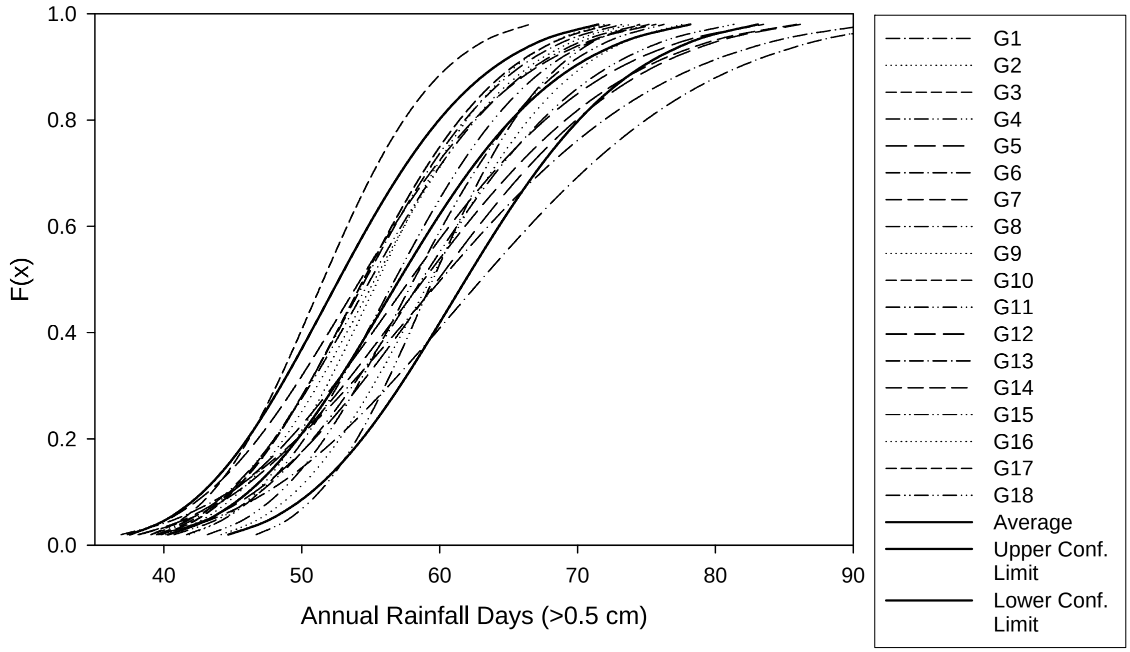

Figure 1-5: Rainfall days per year cumulative distribution

Table 1-4: GEV variable parameter summary

Gage No.	Annual Maximum Event Rainfall					Annual Maximum Interevent Days				
	Location μ	Scale σ	Shape ξ	K-S α*	K-S D**	Location μ	Scale σ	Shape ξ	K-S α*	K-S D**
G1	16.32	5.16	-0.10	0.28	0.10	36.27	9.20	0.03	0.28	0.10
G2	14.90	5.65	0.00	0.34	0.12	36.97	11.47	0.04	0.34	0.16
G3	15.91	4.67	-0.10	0.27	0.12	35.03	9.37	0.22	0.27	0.11
G4	13.84	3.81	-0.28	0.29	0.06	35.82	9.03	0.15	0.29	0.08
G5	16.93	4.33	-0.14	0.29	0.11	33.66	9.80	-0.10	0.29	0.07
G6	16.01	4.07	-0.23	0.27	0.08	40.79	10.89	0.03	0.24	0.06
G7	15.04	4.69	-0.01	0.26	0.07	35.72	8.85	0.09	0.26	0.07
G8	16.55	4.82	0.14	0.29	0.07	33.67	9.80	0.10	0.29	0.09
G9	15.93	5.90	0.05	0.29	0.10	33.70	9.16	0.14	0.29	0.10
G10	15.61	4.66	-0.17	0.29	0.10	33.76	11.81	0.12	0.29	0.09
G11	16.11	4.12	0.13	0.28	0.08	30.85	8.59	-0.08	0.28	0.11
G12	18.35	5.55	0.05	0.27	0.10	34.62	8.39	-0.22	0.27	0.08
G13	16.91	5.27	0.09	0.27	0.09	32.04	8.99	-0.16	0.27	0.11
G14	16.06	4.30	-0.25	0.27	0.12	34.66	7.92	-0.18	0.27	0.11
G15	15.62	4.83	0.00	0.28	0.10	35.95	9.15	0.13	0.28	0.10
G16	14.97	3.71	-0.17	0.23	0.10	34.55	9.44	-0.17	0.23	0.11
G17	12.98	4.21	-0.15	0.24	0.08	36.75	8.40	0.04	0.24	0.08
G18	15.76	3.84	-0.03	0.32	0.09	34.72	8.49	-0.03	0.32	0.12

*K-S α refers to Kolmogorov-Smirnov statistic at 99 percent significance
**K-S D refers to Kolmogorov-Smirnov test statistic, D for the GEV distribution

1.3.1.3 Event Rainfall Annual Maximum

The GEV fitted variables demonstrated somewhat more variability at the extreme end of the scale. Examining the quantile plot for the gage G16 fit (Figure 1-6), a good fit is evidenced up until approximately 20 cm of rainfall when a large deviance from the match line is observed. This deviance is likely due to the fact that some of these event rainfall maximums are a result of hurricanes or tropical storms, which are not part of normal rainfall mechanisms or distributions. The cumulative distribution for all gages for the event rainfall annual maximum (Figure 1-7) indicates that several GEV fits fall outside the confidence limits near the 0.8 percentile, corresponding to approximately 24 cm of rainfall. Gages G6, G14 and G16 fall outside of the lower confidence limit above the 0.86 percentile while gages G2, G8, G9, G11 and G13 fall outside of the upper confidence limit above the 0.76 percentile. The fits of gages G4, G12 and G17, located 16, 32 and 40 kilometers from Moon Lake, respectively, fall out well before the 0.8 percentile. Gage G17 falls entirely outside of the lower confidence and gage G4 falls outside the lower confidence limit above the 0.64 percentile. Gage G12 falls outside the upper confidence limit above the 0.42 percentile. All but two gage fits, including G17, are contained at the 0.5 percentile.

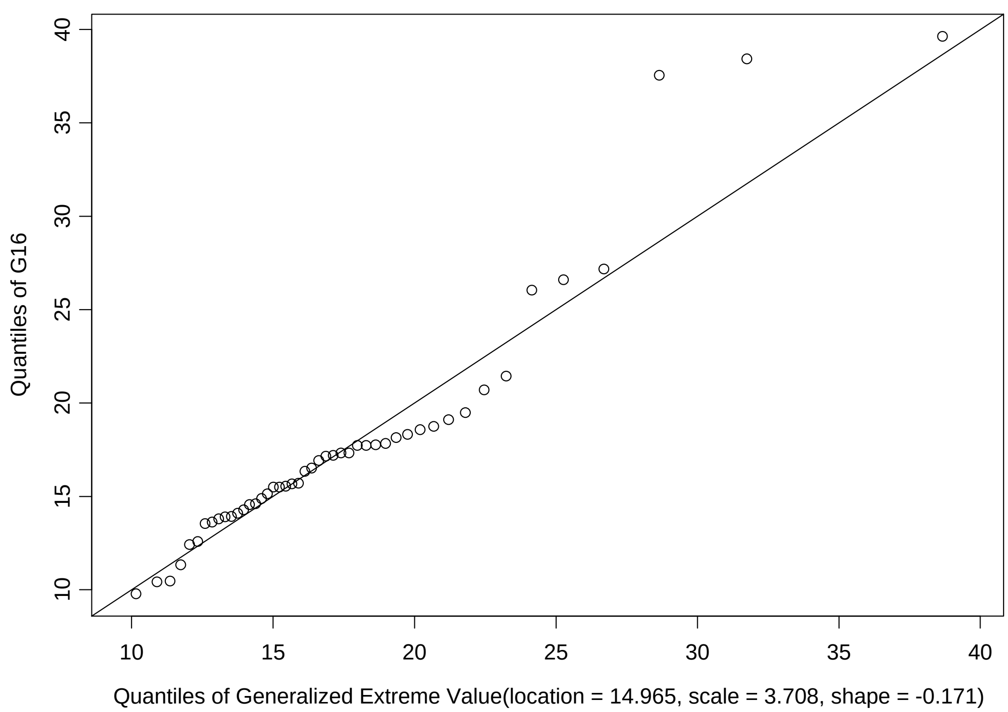

Figure 1-6: Gage G16 quantile plot of event rainfall annual maximum

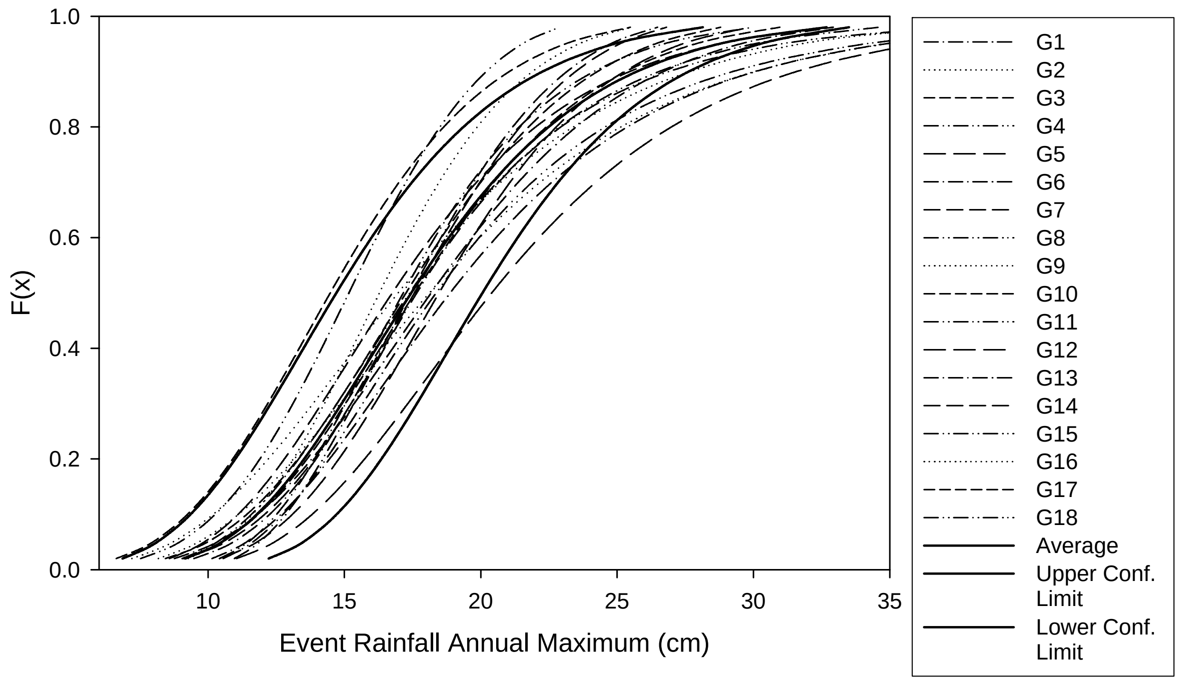

Figure 1-7: Event rainfall annual maximum cumulative distribution

1.3.1.4 Annual Maximum Interevent Days

As shown in Figure 1-8, the quantile plot for gage G16 shows similar dynamics to other gages for the annual maximum interevent days. A good fit is evidenced up until approximately 65 days of interevent time. From this point, some deviance from the match line is observed. The cumulative distribution (Figure 1-9) exhibits substantial variability at higher percentiles. Nearly all gage fits are contained within the confidence limits up to approximately the 0.74 percentile with the exception of gage G6, which falls just outside the upper confidence limit from the 0.30 percentile. From the 0.74 percentile upwards, gages, G11, G12, G13, G14 and G16 fall slightly outside the lower confidence limit. Although the interevent annual maximum exhibits more variability outside the 99-percent confidence limits associated with the average distribution, a case for regional homogeneity can be made. Six of the gages that fall outside the confidence limits are greater than 32 kilometers from Moon Lake. Gage G6 is the only gage that falls outside the limits prior to the 0.72 percentile. All other fits are contained at the 0.5 percentile.

In terms of establishing spatial homogeneity in regards to water resource management, it is important to establish average expected conditions. Although for most of the variable fits analyzed there were areas at the upper and lower percentiles where a few individual gage fits exceeded the 99-percent upper or lower confidence interval of the average, nearly all gage fits were contained at the 0.5 percentile, representing the average annual variable value. Gage G17 was the only fit that consistently fell outside the confidence limits for multiple variables. As previously discussed, this gage is located the furthest from Moon Lake and may be subject to different rainfall forcing mechanisms. Most of the other fits that exceeded the confidence bounds at high or low percentiles exceeded them at locations where the fits themselves break down. Exceeding the confidence limits slightly at high or low percentiles is most likely a function of the fit variability and not an indication of spatial non-homogeneity. Thus, the precipitation variables evaluated appear to exhibit spatial homogeneity within the given confidence limits for the region analyzed. An average distribution of rainfall variables that could be used for water resources management and planning within the region studied was developed. Furthermore, the methods utilized to evaluate spatial variability and develop representative distributions could easily be applied to other regions.

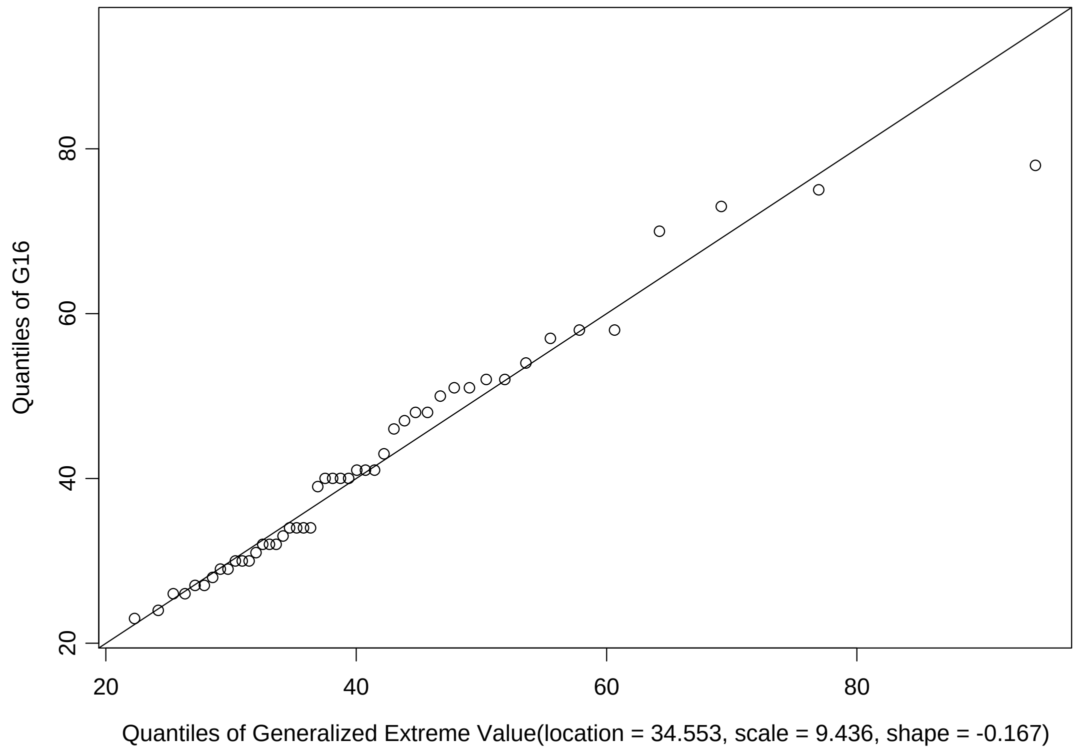

Figure 1-8: Gage G16 quantile plot of the annual maximum interevent days

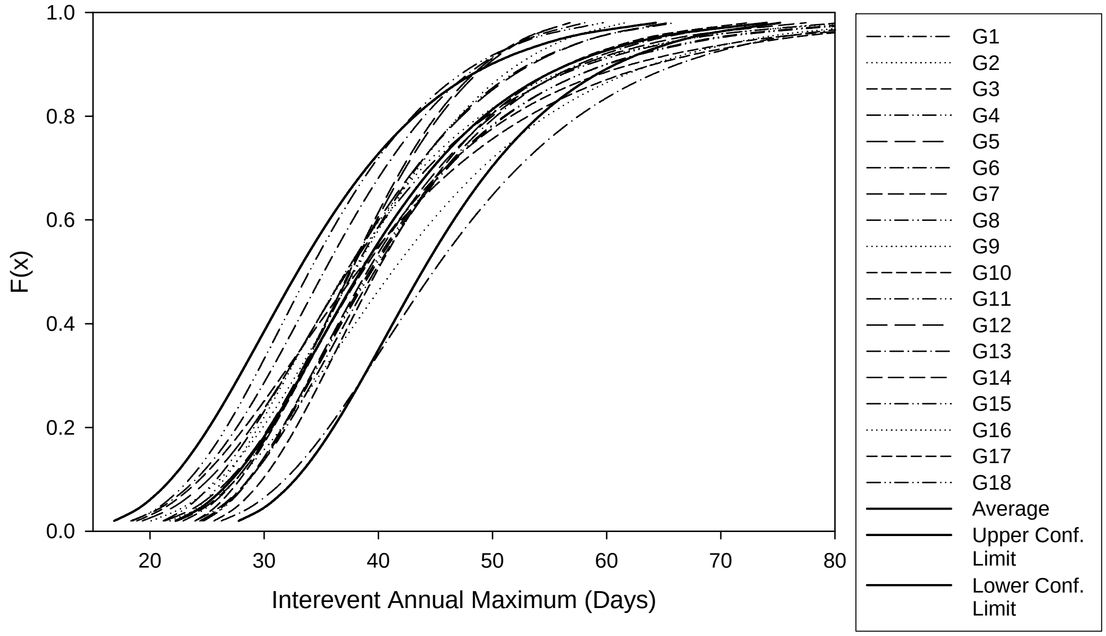

Figure 1-9: Interevent days annual maximum cumulative distribution

1.3.2 Temporal Analysis

Although the total annual rainfall and rainfall days per year were fitted with the gamma distribution for purposes of spatial analysis, they were fitted with the GEV distribution in order to investigate temporal variability in a consistent manner with the other variables and to compare relative changes in parameters with covariates. As such, the Kolmogorov-Smirnov test statistic was applied to these fits as well; nearly all of the variables at each gage gave as good or better a fit with the GEV distribution. This is largely due to the Weibull distribution, which is also often applied to rainfall data, being a subdistribution of the GEV. The use of the Weibull distribution is consistent with the findings of Burgueno et al (2004), which found the Weibull distribution well suited to fit time intervals between rainfall and Sharda and Das (2005) which found the Weibull distribution fit weekly rainfall data better than the gamma distribution as well as other rainfall research. As a check, the gamma parameters of the original fits were also allowed to vary with time with very similar results in terms of trend identification.

All parameters (μ, σ, ξ) were allowed to vary with time; however, no significant improvement was evidenced by using one parameter over another. Therefore, for purposes of model comparison, the location parameter, μ, was allowed to vary with time in model 2 and this was compared with the constant parameter model 1. Table 1-5 gives a summary of the likelihood ratio for the two models at the 99-percent significance level.

Table 1-5: GEV model comparison

Gage No.	Annual Rainfall		AnnualRainfall Days		Annual Maximum Event Rainfall		Annual Maximum Interevent Days	
	M2/M1*	α**	M2/M1*	α**	M2/M1*	α**	M2/M1*	α**
G1	9.05	6.63	17.06	6.63	0.24	6.63	11.08	6.63
G2	0.49	6.63	1.70	6.63	.08	6.63	2.73	6.63
G3	0.75	6.63	0.42	6.63	0.07	6.63	0.49	6.63
G4	0.79	6.63	6.92	6.63	3.73	6.63	0.20	6.63
G5	1.26	6.63	1.81	6.63	1.06	6.63	3.39	6.63
G6	0.34	6.63	0.02	6.63	0.95	6.63	0.03	6.63
G7	0.02	6.63	2.75	6.63	0.00	6.63	0.17	6.63
G8	0.00	6.63	0.18	6.63	0.47	6.63	0.04	6.63
G9	0.03	6.63	0.00	6.63	2.18	6.63	0.30	6.63
G10	0.29	6.63	1.17	6.63	0.26	6.63	2.21	6.63
G11	1.11	6.63	3.51	6.63	2.38	6.63	0.61	6.63
G12	5.26	6.63	13.72	6.63	1.11	6.63	0.57	6.63
G13	0.00	6.63	14.60	6.63	0.00	6.63	2.47	6.63
G14	6.46	6.63	3.55	6.63	0.40	6.63	0.75	6.63
G15	0.76	6.63	0.21	6.63	0.00	6.63	0.93	6.63
G16	0.01	6.63	2.42	6.63	4.14	6.63	0.99	6.63
G17	0.27	6.63	0.02	6.63	1.35	6.63	0.65	6.63
G18	1.01	6.63	0.35	6.63	6.19	6.63	0.05	6.63

* Model 2/Model 1 likelihood ratio
** 99-percent significance (likelihood ratio should be greater to indicate a trend)

From the table it can be seen that gages G4, G12 and G13 show a weak trend in the rainfall days per year. In the case of G4, this is an upward trend; G12 and G13 demonstrate a downward trend. Because a similar trend in the other variables is not evidenced, a possible increase or decrease in rainfall days is not corresponding to changes in annual rainfall nor to changes in event maximums. Furthermore, an increase or decrease in rainfall days would be expected to correlate to a decrease or increase, respectively, in annual maximum interevent days, which has not occurred at these three gages. As such, the trend for these gages does not appear to be significant. Gage G1, however, shows a significant trend in rainfall days, interevent days and annual rainfall. Looking at the trend equations for the location parameter for each variable:

$$\text{Rainfall Days: } \mu = 70.90 - 0.88t \tag{15}$$

$$\text{Interevent Days: } \mu = 23.53 + 0.64t \tag{16}$$

$$\text{Annual Rainfall: } \mu = 156.05 - 1.72t \tag{17}$$

It can be seen that there is a decrease in the number of rainfall days GEV fit location parameter of approximately 0.88 per year. This statistically and logically corresponds to an 0.64 per year increase in the interevent days GEV fit location parameter and a 1.72 decrease in the total annual rainfall per year GEV fit location parameter. Because annual rainfall days exhibited the strongest trend, quantile and probability plots for model 2 (Figure 1-10) were visually investigated to confirm the model fit.

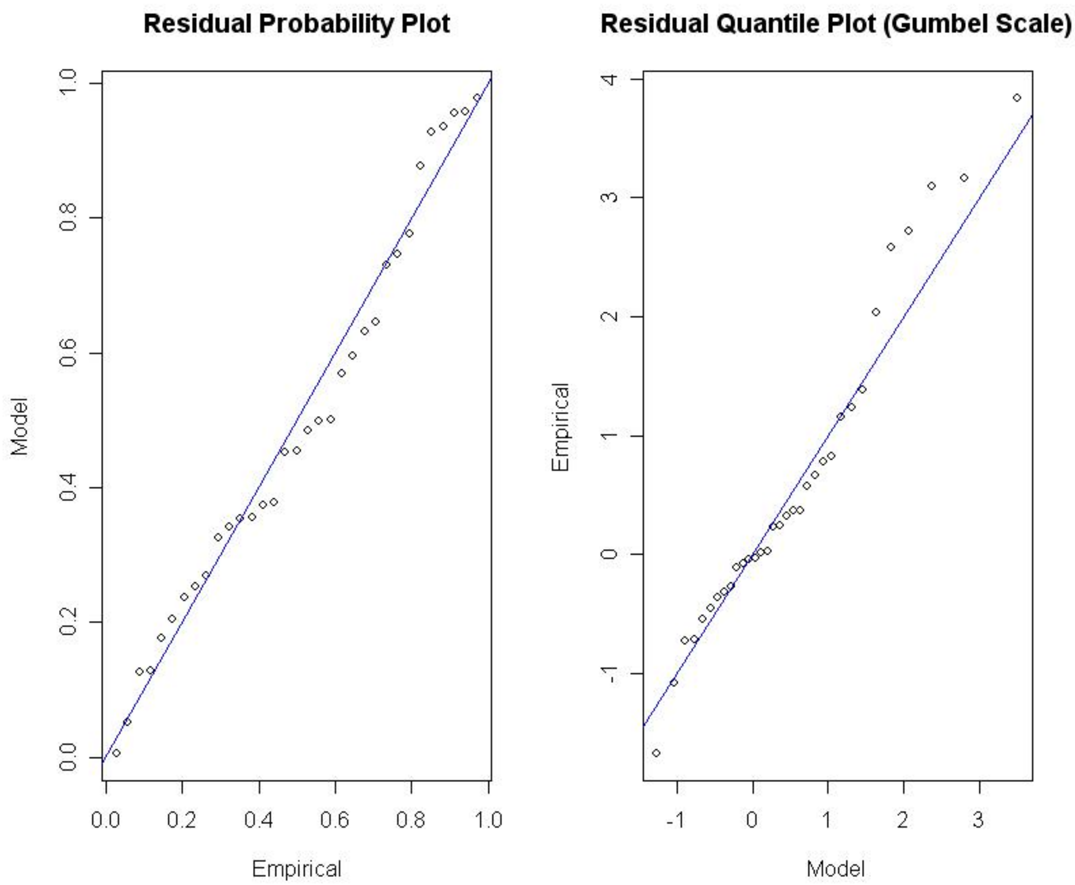

Figure 1-10: Quantile plot for gage G1 annual rainfall days - model 2

From the plot, the fit is adequate. Other gage G1 variables were investigated visually and similar results were found. However, from the graph of the observed variable data below (Figure 1-11), it does appear that the trends indicated by model 2 are apparent. It is possible that, given the lack of similar trends for all other gages analyzed, the trends observed at this particular gage are due to time scale and a longer record would weaken the apparent trend. Focusing on the annual rainfall variable, which has been divided by three to scale with the other variables in Figure 1-11, it appears that there is a downward trend up until 2002 and 2004 when two of the highest rainfall totals on record are achieved. The annual rainfall days follow a similar pattern. The maximum interevent days variable is nearly a mirror image of the annual rainfall where an upward trend is apparent until 2002, after which several low extremes are recorded.

Given that only one of the gages analyzed demonstrates a significant trend based on time-dependent parameters of the GEV distribution, a very strong case for temporal stationarity in rainfall patterns in this region can be made. This method of allowing GEV parameters to vary with time is a robust method of analyzing trends and can be applied to rainfall data in other regions.

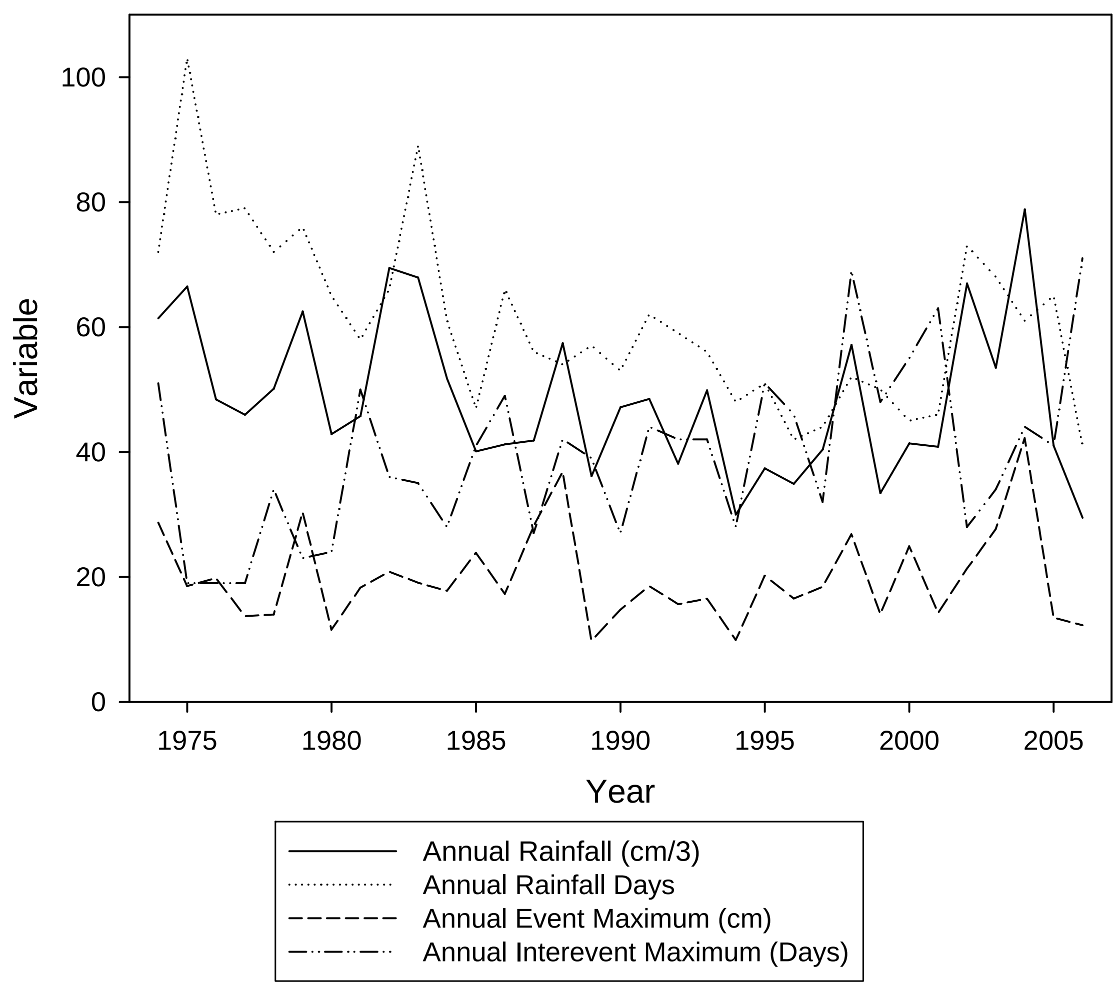

Figure 1-11: Gage G1 observed variable data

1.4 Conclusions

The focus of this research was to determine the extent of spatio-temporal changes in precipitation patterns in a particular region utilizing methods that can be generalized and applied at the regional scale. Furthermore, the aim was to develop average distributions for each rainfall variable that can be applied to any water resource within the region analyzed, regardless of the proximity of gage data. Given the documented changes in spatial and temporal rainfall patterns in many parts of the world, it was expected that some changes in rainfall patterns would be evident in the region analyzed. However, based upon the spatial analysis, the vast majority of variables analyzed at each gage were confined to a 99-percent confidence band associated with the average fit, gamma or GEV, of the data. There were some exceptions; however most of these were at gages at the outer fringes of the area analyzed and at percentiles near the high or low end. Nearly all of the fits were contained at the 0.5 percentile, representing the average annual variable a particular water resource can expect to experience. The method utilized to establish spatial homogeneity can easily be applied to other areas. Furthermore, developing an average, representative fit for a given region can be a powerful tool when forecasting return levels of the various variables and managing and planning water resources.

In regards to temporal variability, it was also somewhat surprising that almost no significant trends were detected. Many of the gages investigated would have demonstrated a trend if analyzed with traditional methods such as ordinary least squares or non-parametric Mann-Kendall. However, varying parameters with time to detect a trend is a robust method to overcome traditional difficulties inherent in real-world hydro-climatological data, including limited or incomplete data, autocorrelation, and changes in collection methods. Furthermore, in cases where trends are detected, this method gives an estimate on how much distributions are changing with time. In essence, since the distribution itself is changing with time, return frequencies are a function of time and the trend in parameters gives an estimate of this change.

2.0 Statistical Changes of Lake Stages in Urbanizing Watersheds

2.1 Background

During the last century, the planet has experienced tremendous growth in population and development. In order to support this growth, an ever-increasing strain is being placed on water resources. While prolonged droughts or other changes in rainfall patterns present clear impacts to water resources, the urbanization of a basin changes the rainfall runoff relationship in such a way that impacts to water resources management are not so clear. Many researchers have studied some of the impacts of urbanization on water resources in various watersheds. Meyer and Wilson (2001) found that streams in urbanized basins exhibited a reduction in baseflow. In a similar study, Rose and Peters (2001) found that streams in urbanized watersheds demonstrated a decrease in baseflow, an increase in peak flows and a decrease in recession times for both baseflow and peak flow. Both studies attribute their findings to an increase in impervious area and rapid storm runoff from efficient collection systems and the corresponding decrease in infiltration. However, Meyer (2005) found no trend in baseflow changes in several streams with urbanized watersheds and attributed this to the low permeability of near-surface soils and presence of stormwater detention systems. Smith and Baeck (2002) found that the increased efficiency of the drainage network in a rapidly urbanizing stream watershed is the major factor in a positive trend in flood magnitude primarily due to a shortened response time. Changnon and Demissie (1996) found that in two urban and two rural stream basins undergoing changes in rainfall and land use, the majority of the increase in mean flows was due to the land use changes. Furthermore, mean and peak flows in the urban watersheds demonstrated considerably more response to rainfall shifts. McMahon et al (2003) investigated changes in stream watersheds undergoing urbanization in three different locales. The study found that increased urbanization was related to increases in stream flashiness and variability but found less relation to the duration of high or low stage conditions.

Much of the literature has focused on accurate modeling and prediction of water levels in order to predict future levels, identify the contribution of individual forcing mechanisms such as land use change, or explore changes in the rainfall or runoff and water level response. Altunkaynak (2007) employed artificial neural networks to model increases in Van Lake, Turkey, water levels and compared the results to traditional autoregressive moving average models (ARMA) and found that while the neural network outperformed the time series models, both had low error. In a similar study, Khan and Coulibaly (2006) compared a support vector machine and a seasonal autoregressive (SAR) model in long-term prediction of lake water levels and found that while the support vector machine outperforms the SAR model, both also had low error. Privalsky (1992) utilized a SAR model in combination with spectral analysis to study the statistical properties of Lake Erie, United States, water levels. Irvine and Eberhardt (1992) utilized an integrated autoregressive moving average (ARIMA) model to characterize water levels at Lake Erie and Lake Ontario. Montanari et al (1997) applied a fractionally differenced ARIMA model to Lake Maggiore, Italy, and found the model outperformed traditional ARIMA models with some limitations in accounting for seasonality. Yin and Nicholson (2002) utilized an autoregressive model coupled with rainfall inputs to predict Lake Victoria, Tanzania, Uganda and Kenya, levels.

Several studies focused on water budget models or physically based models to characterize changes in lake levels. Lenters (2004) found that changes in Lake Superior, United States and Canada, were primarily due to climatic and land use changes based upon a water budget model. Li et al (2007) and Jones et al (2001) also utilized a water budget model to determine that declines in water levels at Lake Qinghai, China, and at several lakes in Africa, respectively, were primarily due to climatic changes rather than land-use or other changes. Elias and Ierotheos (2006) utilized a transfer function model, a dynamic linear relationship model and a physically based model to describe the relationship between precipitation and lake levels.

In addition to time series modeling, some of the research literature has applied regression to lake stages. Gao (2004) utilized multiple linear regression to develop quantiles of lake level fluctuations. McBean and Motiee (2008) utilized regression to identify long-term trends in precipitation, temperature and inflow to the Great Lakes of North America. Gibson et al (2006) found rainfall variability to be the primary driving force on Great Slave Lake, Canada, levels and developed a regression model for water level fluctuations based on this variable alone. However, Mendoza et al (2006) determined that lake level fluctuations can be estimated by regressing monthly mean precipitation as well as temperature. Lall et al (2006) developed a locally weighted polynomial regression model to forecast the Great Salt Lake, United States, biweekly volume.

Because lakes are vital for tourism, recreation, ecology, the environment and water supply, understanding statistical changes of lake dynamics in urbanizing watersheds is integral to effective future water resource management. In this research, changes in the statistical structure of lakes in urbanizing watersheds are investigated by evaluating serial differences in time series parameters, autocorrelation and variance as well as by developing a regression model to estimate weekly lake level fluctuations. The focus of the research was to develop general expectations for lake levels in urbanizing areas that can be applied globally utilizing methods that can be applied to other locations. The time series modeling involves fitting a seasonal integrated autoregressive moving average (SARIMA) model to lake levels over subperiods of approximately equal length over the data record available and identifying trends in parameter values. The regression model was fit to weekly fluctuations in water surface levels. The regression model independent variables consist of rainfall components as well as lake stage and temperature components. These analyses were performed for six lakes in Pasco County, Florida, United States, that demonstrate consistent urbanization and have not been substantially anthropogenically altered other than the addition of control structures or culverts at two lakes as part of lake management for urban planning.

While some studies have focused on modeling stream stages or the effect of urbanization on stream stages, very few have focused on lake levels. For both streams and lakes, it is difficult to separate out the signal of urbanization from the multitude of forcing mechanisms inherent in an urbanized watershed, including pumping, surface withdrawal, dredging, filling, diversion, installation of control structures, etc. Furthermore, lake stage data are typically less available than stream data, hence the previously cited lake regression analyses utilized a coarser time scale than that employed in this research. The focus on isolating the urbanization signal on lake levels and the methods utilized to identify statistical changes in lake levels in urbanizing basins as applied in this research were not found in the literature review. The objectives of this research in regard to urbanizing lake watersheds are to: 1) Determine changes in the statistical structure of lake level time series. 2) Analyze serial changes in lake level autocorrelation and variance. 3) Identify changes in the runoff/baseflow relationship. As a result of researching these objectives, general expectations for water resource managers will be developed as these are requisite tools for effective planning.

2.2 Materials and Methods

2.2.1 Lake Information and Data

In order to sufficiently isolate the impacts of watershed urbanization, other change mechanisms, including anthropogenic alterations, pumping and climate change must be accounted for or eliminated as much as possible. Paynter and Nachabe (2008) found that rainfall patterns within west-central Florida are spatially homogenous and temporally stationary and no significant shifts that would correlate to trends in lake levels were evidenced. For the present study, six lakes were selected within the west-central Florida region based on the availability of sufficient data, lack of direct lake withdrawals and lack of proximity to well fields. Some of the lakes selected have had control structures added while others have been relatively unaltered other than urbanization of the basin. Because changes such as adding a weir or a culvert to a lake are inherent to urbanization and management of water resources, these lakes were included in this analysis. Some of the lakes analyzed are flow-through lakes in which the lakes receive substantial flow from upstream water bodies and discharge to downstream lakes or wetlands. Other lakes are simply drainage lakes with no flow-through. Because all of the lakes are located within a fairly small geographic area, they generally evidence similar geologic characteristics, including being located in a silty-sandy environment above a limestone formation. A summary of the characteristics of the six lakes utilized can be found in Table 2-1. The presence of other potential signal sources in addition to urbanization, including flow from upstream lakes, control structures and adjacent wetlands are also included in the table. Lake data available to the general public were obtained from the United States Geological Survey and the Southwest Florida Water Management District. The data are generally daily but because there are several small gaps of a week or more throughout the data, the time series of each lake was converted to weekly increments by selecting the first day of available stage data, adding seven days to this date consecutively throughout the record and selecting the lake level that corresponds to the weekly date.

In order to determine the degree of watershed urbanization, data was utilized from the United States census for 1980, 1990 and 2000 population counts at the census tract level. Population data prior to 1980 are only available at the county level for Pasco County.

Census tracts were substantially larger than the lake basins studied. For each lake basin, population was distributed evenly across the tract and proportioned to each basin. Population density within each watershed, which has a more meaningful implication for estimating basin urbanization, was developed from these basin-specific population values. Lake Thomas and King Lake fall within the same census tract as do Cow Lake and Lake Padgett. As such, these pairs of lakes will exhibit similar densities although the actual population numbers will be different. For each particular lake, rainfall data from nearby gages were utilized. Figure 2-1 provides an aerial view of the lakes and available rainfall gages. Figures 2-2 and 2-3 depict aerial views of Moon Lake and Cow Lake.

Table 2-1: Lake characteristics summary

Lake	Basin Area (km^2)	Lake Area (km^2)	Basin/Lake Ratio	Flow Through	Weir Structure	Adjacent Wetland Area (km^2)	Adjacent Wetland Percent of Basin Area
Moon	0.78	0.43	1.8	N	N	0.07	9.0
Padgett	17.09	0.79	21.6	Y	N	0.12	0.7
Thomas	2.59	0.66	3.9	N	N	0.12	4.6
Ann Parker	8.00	0.38	21.1	Y	Y	0.31	3.9
King	4.40	0.55	8.0	Y	Y	0.25	5.7
Cow	1.55	0.40	3.9	N	N	0.00	0.0

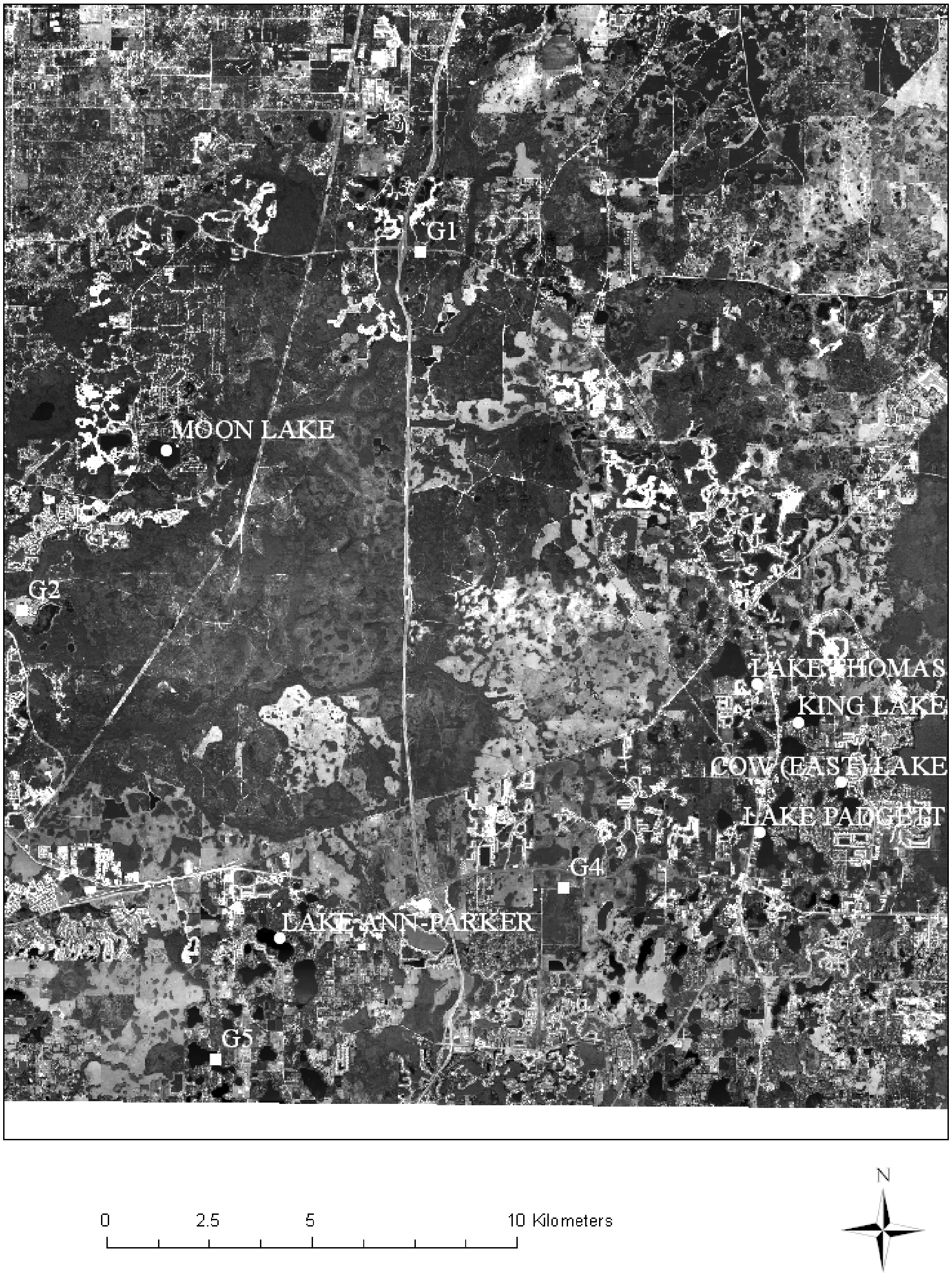

Figure 2-1: Lake and rainfall gage location

Figure 2-2: Moon Lake

Figure 2-3: Cow Lake

2.2.2 Time Series Analysis

Statistical changes in lake levels were first explored with time series analysis. For each lake, the total stage record was split into approximately equivalent units of less than 10 years and fit with a time series model to investigate any systematic changes in parameters. Some of these subunits were shifted slightly by one or two years to achieve nearly constant variance in time subseries that exhibited heteroscedasticity. Time series analysis assumes a stationary time series; any systematic change in the mean (trend) and variance and any periodic variations (seasonality) were accounted for. Surface water levels throughout the world have evidenced trends related to pumping, climate change or other factors and seasonal variation is found as water levels rise in the rainy season and fall in the dry season. In addition, longer seasonal trends, such as those induced by the approximately 10-year El Nino phenomenon, may also be present in hydrologic time series. For this data, any trends were removed by differencing, i.e, by subtracting adjacent values. Seasonality was also removed by differencing, i.e., by subtracting values one period away. For purposes of this research, the period utilized was 12 months.

Based on the literature (Irvine and Eberhardt, 1992; Altunkaynak, 2007), autoregressive moving average models (ARMA) are often found to fit lake data well. Although there are more robust methods for forecasting lake levels, such as neural networks, the focus of this research is on changes in the statistical structure of lake levels, for which time series analysis is aptly suited. An autoregressive process is based completely on previous values of the time series. An autoregressive process of order p is given by:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2+\ldots\ldots\ldots} \phi p x_{t-p} + w_t \tag{18}$$

where x_t is the current value of the lake level time series, x_{t-p} is the time series value at lag p, ϕ_p is a unique constant parameter for each lagged value and w_t is Gaussian white noise with mean zero. A moving average process is based completely on previous values of white noise. A moving average process of order q is given by:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ...\theta_q w_{t-q} \tag{19}$$

where θ_q is a unique constant parameter for each white-noise value. An ARMA process combines equations 18 and 19 and is said to be of order p, q. For most of the lakes studied, an ARMA model with simple differencing was required, yielding an integrated autoregressive moving average (ARIMA) model of order p, d, q where d is the number of time the series needs to be differenced to remove trends and achieve stationarity. Once the series is stationary, ARMA parameters were determined using maximum likelihood estimation. The likelihood of the model is given by:

$$L(\beta, \sigma_w^2) = (2\pi\sigma_w^2)^{-n/2} \left[r_1^0(\beta) r_2^1(\beta) ....r_n^{n-1}(\beta) \right]^{-1/2} \exp\left[-\frac{S(\beta)}{2\sigma_w^2} \right] \tag{20}$$

where

$$S(\beta) = \sum_{t=1}^{n} \left[\frac{(x_t - x_t^{t-1}(\beta))^2}{r_t^{t-1}(\beta)} \right] \tag{21}$$

and β is the vector of model parameters $\phi_1 .....\phi_p, \theta_1 .....\theta_q$, σ_w^2 is the variance of white noise and r is the mean squared error of the one step ahead prediction, $x_t - x_t^{t-1}$.

Parameter estimates were obtained by maximizing (20) with respect to β and σ_w^2 (Shumway and Stoffer, 2006). A seasonal component was added to the ARIMA model in some cases to adequately capture periodic fluctuations. A seasonal ARIMA is given by:

$$\Phi_P(\mathrm{B}^s)\phi(\mathrm{B})x_t = \Theta_Q(\mathrm{B}^s)\theta(\mathrm{B})w_t \tag{22}$$

where

$$\Phi_P(\mathrm{B}^s) = 1 - \Phi_1\mathrm{B}^s - \Phi_2\mathrm{B}^s - ..... - \Phi_P\mathrm{B}^{Ps} \tag{23}$$

and

$$\Theta_Q(\mathrm{B}^s) = 1 + \Theta_1\mathrm{B}^s + \Theta_2\mathrm{B}^s + ... + \Theta_Q\mathrm{B}^{Qs} \tag{24}$$

are the seasonal autoregressive and seasonal moving average operators, respectively, of order P and Q with seasonal period s. In essence, the seasonal part of the model estimates current time series values from values one seasonal period or more in the past. SARIMA models are noted as ARIMA (p, d, q) x (P, D, Q) where D is the number of seasonal differences. Once a model was fitted to the data, diagnostics were performed to ensure randomness of the residuals, including inspecting the autocorrelations of the residuals, $r_e^2(h)$, where h is the lag, and the Ljung-Box statistic, given by:

$$Q = n(n+2)\sum_{h=1}^{H}\frac{r_e^2(h)}{n-h} \tag{25}$$

where n is the sample size and H is arbitrarily chosen, typically near lag 20 (Shumway and Stoffer, 2006). The test statistic Q is chi-square distributed and the null hypothesis of randomness is rejected if Q falls above the selected significance limit quantile. To achieve parsimony, the simplest model that adequately fit the data for all time series subsets was sought for each lake and model parameters were compared to identify any trends. In ranking the models, adjusted R^2 and the Bayesian information criterion (BIC) were utilized. Whereas R^2 is a common measure of the goodness-of-fit, BIC takes into account the number of parameters required to achieve the fit; lower BIC values indicate a preferred model. The general approach for each lake was to increase the number of parameters from a first order autoregressive model and note any improvements in the R^2, Ljung-Box statistic or BIC criteria.

2.2.3 Autocorrelation and Variance

Lake levels exhibit significant autocorrelation, a measure of how related adjacent values are. The autocorrelation function, a dimensionless measure of linear dependence of time series values at lag k, is given by:

$$\frac{\{\sum_{t=1}^{N-K}(x_t-\bar{x})(x_{t+k}-\bar{x})\}}{\sum_{t=1}^{N}(x_t-\bar{x})^2} \tag{26}$$

The potential change in lake level autocorrelation was evaluated by analyzing any serial changes in the autocorrelation of the lake levels. A time scale of weeks is too long to capture any changes in autocorrelation due to shifts in the rainfall/runoff response from urbanization. However, changes in the slower process of infiltration and base flow into a lake, including a hypothesized reduction due to basin urbanization, should be evident. Furthermore, effects of changing mechanisms, such as a control structure, inherent to urbanization may also alter lake memory. Baseflow, for purposes of this research, is defined as the fraction of watershed rainfall that infiltrates the ground and subsequently over weeks and months enters the lake. This includes potential spill-over from wetlands that may enter the lake well after a precipitation event has occurred. Statistically significant autocorrelation values for each time series subunit were approximated with an exponential fit so that they could be characterized by a single parameter and compared to other subperiods within each lake. The exponential fit is of the form:

$$r_k = e^{-\lambda k} \tag{27}$$

where λ is a constant and k is the lag. Because some of the time subseries appear to exhibit heteroscedasticity when compared to one another, an F test for significantly different variances was employed.

2.2.4 Regression

Any changes in lake level statistical signatures found with time series modeling were further explored with regression. Instead of focusing on the water level itself, the differences between weekly stages were modeled. The time scale of overland flow to lakes is in hours or days while the time scale of baseflow recharge to lakes is in weeks. Hence, lake level fluctuations over a week contain both a runoff and baseflow component; insufficient data are available to reduce the time scale and separate these components. As such, evaluating changes over time between the runoff fraction and baseflow fraction may be difficult unless these changes are prominent. However, changes over time in the baseflow recharge to lakes derived from rainfall that has fallen more than a week in the past can be evaluated. For the regression model, the difference (Y_t) of the current week's water level from that of the previous week was regressed against both the total rainfall for the current week (R_w), representing a combination of rainfall runoff and baseflow, and the month's rainfall total previous to the current week (R_m), representing solely baseflow. In order to improve the goodness-of-fit of the model, terms for temperature (T) and starting water level (W) were also included. The average temperature helps to capture evaporation, which, contrary to rainfall, has an inverse relationship with lake levels. Although evaporation pan data for the lakes analyzed were not available, temperature is highly correlated to evaporation. The starting water level helps capture lake morphology since, in general, when a lake is at a lower level, less volume is available at a given elevation difference; as lake stage increases, generally the area of the lake also increases and larger amounts of runoff and baseflow are required to make a unit change in elevation. This variable also has an inverse relationship with lake stages since the higher the initial stage, the less impact rainfall ultimately has on fluctuations.

The regression equation is given by:

$$Y_t = \beta_0 + \beta_1 R_w + \beta_2 R_m + \beta_3 T + \beta_4 W + \varepsilon_t \tag{28}$$

where ε_t is the random error term. The parameters $\beta_{0\text{-}4}$ were estimated with the method of least squares. In most cases, a longer period of lake level data were available than rainfall data and lake level records had to be truncated for the regression analysis. In order to assure parsimony, adjusted R^2 and Akaike's information criterion (AIC) was utilized on the entire data set to verify each of the four variables substantially contributes to explaining water stage fluctuation for each lake. AIC takes into account the number of parameters required to achieve a particular fit; lower AIC values indicate a preferred model. The model selected for each lake consisted of the four variables in equation (28) or a subset thereof to be applied to each time subperiod. Several common diagnostics were run to verify there were no substantial correlations among the regressors and no overly influential outliers. In particular, the correlation matrix was calculated for each model while Cook's distance was utilized to measure the influence of specific data points and identify outliers. Data points with Cook's distances of near one are considered to have significant influence and merit further investigation. As with the time series analysis, sequential subunits of time were analyzed to identify any systematic changes in parameters for the independent variables. However in the case of regression, since population in the region began to substantially increase in the 1970's, regression values for a time period as close to 1970 as possible was desired to represent pre-urbanization in each basin. As such, the first four years of data for each lake were modeled to represent the pre-urbanized lake dynamics and the remaining portion of the data was utilized to represent the urbanized lake basin dynamics; the two subperiods were compared to identify changes.

2.3 Results and Discussion

Table 2-2 provides a summary of the lake data. The average difference between the maximum and minimum for the lakes analyzed is 1.9 m. Given the very flat topography of west-central Florida, relatively small differences in water level fluctuations can inundate large areas and houses are routinely set as low as 0.3 m above expected high water marks. The standard deviation and variance are fairly consistent with ranges of 0.30 m and 0.23 m, respectively. All but two lake data sets were greater than 95 percent complete when the number of weekly data points available are divided by the total number of weeks for the time period analyzed. The two lakes with less data had evenly spaced gaps usually less than 2 to 3 weeks apart for which weekly data could be interpolated easily from adjacent values. In the cases of King Lake and Cow Lake, data had to be truncated after 1991 as data points thereafter were too sparse to achieve meaningful results. Data for Lake Thomas had to be truncated after 1994 and the period from 1992 to 1999 had to be excised from Lake Ann Parker for similar reasons.

Based upon the GIS census data, population in the vicinity of each lake has measurably increased over the time periods studied. As the population grows, the watershed morphs from rural to residential development with significant increases in impervious area, channelized drainage provisions and possible infill due to the raising of lots for home construction. The population density growth around each lake is summarized in Table 2-3. The Moon Lake watershed exhibits the greatest overall gains with well over 100 percent density growth from both 1980 to 1990 and 1990 to 2000. Lake Ann Parker, although heavily urbanized, showed the smallest percent gains in population density while Lake Thomas showed the smallest densities overall. Cow Lake population density was considerably higher than that of other lakes.

Table 2-2: Regional lake data summary

Lake	Period of Record	Average (m) (NGV	Maximum (m) (NGVD	Minimum (m) (NGVD	Percent Complete*	Standard Deviation (m)	Variance (m)
Moon	1965-2007	11.63	12.58	10.24	96.0	0.47	0.22
Padgett	1970-2003	21.18	21.9	20.23	95.5	0.25	0.06
Thomas	1968-2003	22.34	23.01	21.04	98.6	0.34	0.12
Ann Parker	1969-2007	15.91	16.8	14.75	72.4	0.36	0.13
King	1976-2007	21.85	22.53	20.39	100.0	0.27	0.07
Cow	1976-2007	23.56	24.11	23.03	82.3	0.17	0.29

*Note: Percent complete excludes years with too few data points to use for analysis

Table 2-3: Lake watershed population growth

Lake	1980 Density (pop./km^2)	1990 Density (pop./km^2)	2000 Density (pop./km^2)	Percent Density Growth 80-90	Percent Density Growth 90-00
Moon	19.9	92.7	238.4	366.2	157.2
Padgett	39.5	80.2	140.7	103.2	75.4
Thomas	27.5	55.9	59.5	103.2	6.4
Ann Parker	137.6	181.7	203.5	32.1	12.0
King	58.0	117.9	125.5	103.2	6.4
Cow	181.1	367.9	645.3	103.2	75.4

2.3.1 Time Series Modeling

Plots of the lake level time series do not display any obvious trends, refer to Figure 2-4 for Cow Lake stages. However, in cases where the autocorrelation of the raw data indicated that a trend is likely present since the correlations slowly decay to insignificance, differencing was utilized. Because lake levels exhibit a high degree of autocorrelation, subtracting an adjacent value is not substantially different than subtracting a value one seasonal period ago; a single simple difference sufficiently removed any trend or seasonality at each of the lake time series. The model order for each lake as well as a summary of the parameters for each model can be found in Table 2-4. The Ljung-Box values were sufficiently low to ensure randomness of residuals. In a single subseries case for Moon Lake, Lake Padgett and Cow Lake, Ljung-Box values were slightly outside 95-percent significance limits for randomness of the residual. In order to bring Ljung-Box values within these limits, several additional parameters would be required, overfitting the other subseries for each lake. As such, the simpler overall model was chosen and is sufficient for the analysis herein. All lakes required at most two autoregressive terms. Moon Lake, Lake Thomas, King Lake and Cow Lake required a single differencing as well as a single seasonal term. Moon Lake, Lake Padgett and Lake Ann Parker required at most two moving average terms.

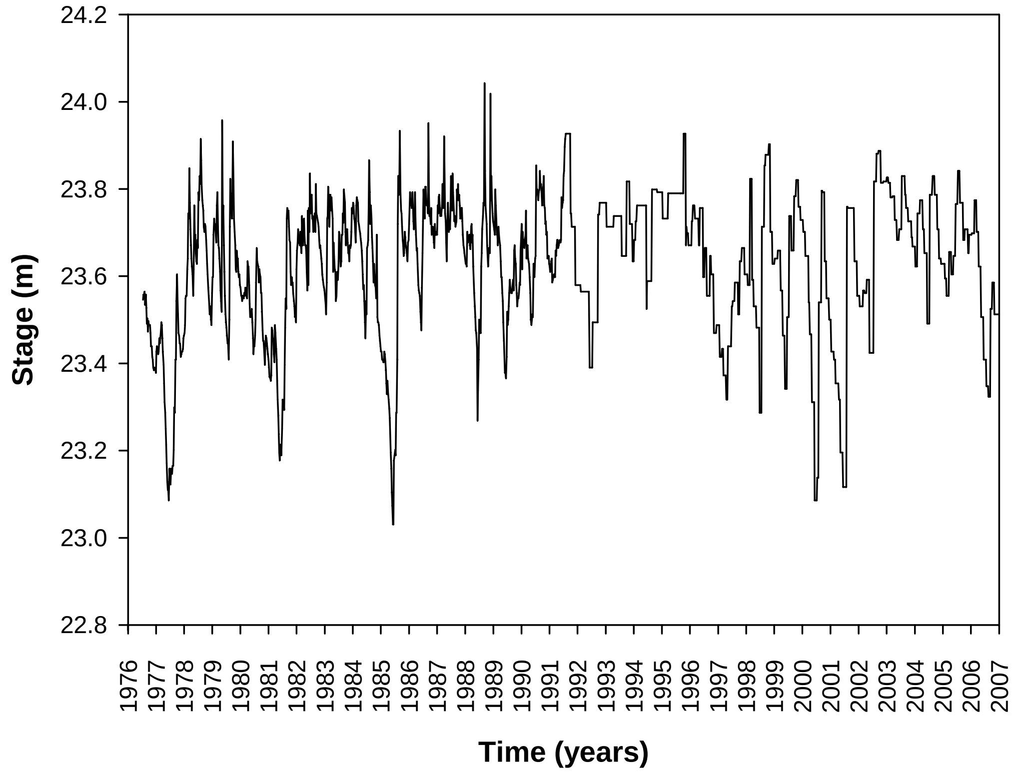

Figure 2-4: Cow Lake stages (1976-2007)

Table 2-4: SARIMA model parameters

Lake/model order	Subseries Period	φ_1	φ_2	θ_1	θ_2	Φ_1	Ljung-Box
Moon (1,1,2)x (1,0,0)	1965-1976	0.68	N/A	0.29	0.07	0.02	15.1
	1977-1986	0.72	N/A	0.49	0.00	0.03	23.2
	1987-1996	0.78	N/A	0.56	0.06	0.13	24.6
	1997-2006	0.81	N/A	0.68	-0.09	0.01	35.0
Padgett (2,0,1)	1970-1975	1.15	-0.20	-0.11	N/A	N/A	21.5
	1976-1983	1.40	-0.43	0.31	N/A	N/A	14.1
	1984-1991	1.594	-0.62	0.45	N/A	N/A	20.3
	1992-2000	1.44	-0.46	0.30	N/A	N/A	36.7
Thomas (2,1,0)x(1,0,0)	1968-1976	0.20	0.03	N/A	N/A	0.01	28.9
	1977-1985	0.09	0.10	N/A	N/A	0.12	18.4
	1986-1994	0.09	0.03	N/A	N/A	0.11	28.3
Ann Parker (2,0,1)	1969-1976	1.12	-0.15	-0.10	N/A	N/A	26.6
	1977-1984	1.20	-0.23	0.02	N/A	N/A	14.3
	1985-1991	1.24	-0.28	0.08	N/A	N/A	23.8
	2000-2007	1.71	-0.72	0.46	N/A	N/A	15.9
King (1,1,0)x(1,0,0)	1976-1980	-0.01	N/A	N/A	N/A	0.12	18.7
	1981-1985	0.17	N/A	N/A	N/A	0.18	23.7
	1986-1991	0.12	N/A	N/A	N/A	0.13	13.6
Cow (1,1,0)x(1,0,0)	1976-1980	-0.03	N/A	N/A	N/A	0.13	17.2
	1981-1985	-0.05	N/A	N/A	N/A	0.06	40.7
	1986-1991	-0.22	N/A	N/A	N/A	0.07	17.7

From inspection of Table 2-4, it can be seen that Lakes Padgett and Ann Parker are the only lakes to exhibit a consistent pattern for all parameters. For every subseries, Lake Ann Parker shows serial increases in the first autoregressive parameter and the first moving average parameter and a serial decrease in the second autoregressive parameter. Lake Padgett shows a nearly identical pattern with the exception of the most recent subseries. Lakes Padgett and Ann Parker have a substantially larger basin to lake area ratio than the other lakes; it is surmised that the effects of watershed urbanization are widespread enough compared to the lake itself to overcome the addition of outlet culverts and control structures and other potential anthropogenical influences on the statistical structure of the time series. While they do not reflect serial changes in all parameters, Moon Lake does demonstrate a consistent increase in the first autoregressive parameter and the first moving average parameter while Cow Lake demonstrates a consistent decrease in the autoregressive parameter. Moon Lake demonstrates the largest percent population density gains, representing the greatest relative change in urbanization, while Lake Padgett and Cow Lake demonstrate the largest population gain in overall numbers. Lake Ann Parker has consistently high density for the time periods analyzed and although its percent growth has been small, it has added substantial population.

King Lake and Lake Thomas do not demonstrate appreciable patterns in parameters for consecutive subseries. These two lakes exhibit the lowest population density as well as the lowest percent growth for the most recent time period. It is possible that an insufficient level of urbanization was achieved within these watersheds to overcome other signals and cause serial changes in lake level time series. From the research, the degree and extent of urbanization appears to have influence on changes in the statistical structure of lake level time series. Furthermore, these changes are apparent despite the multitude of other signals present in the basins. Figure 2-5 indicates that errors are random for the Cow Lake time series model. Similar results were achieved for other lakes.

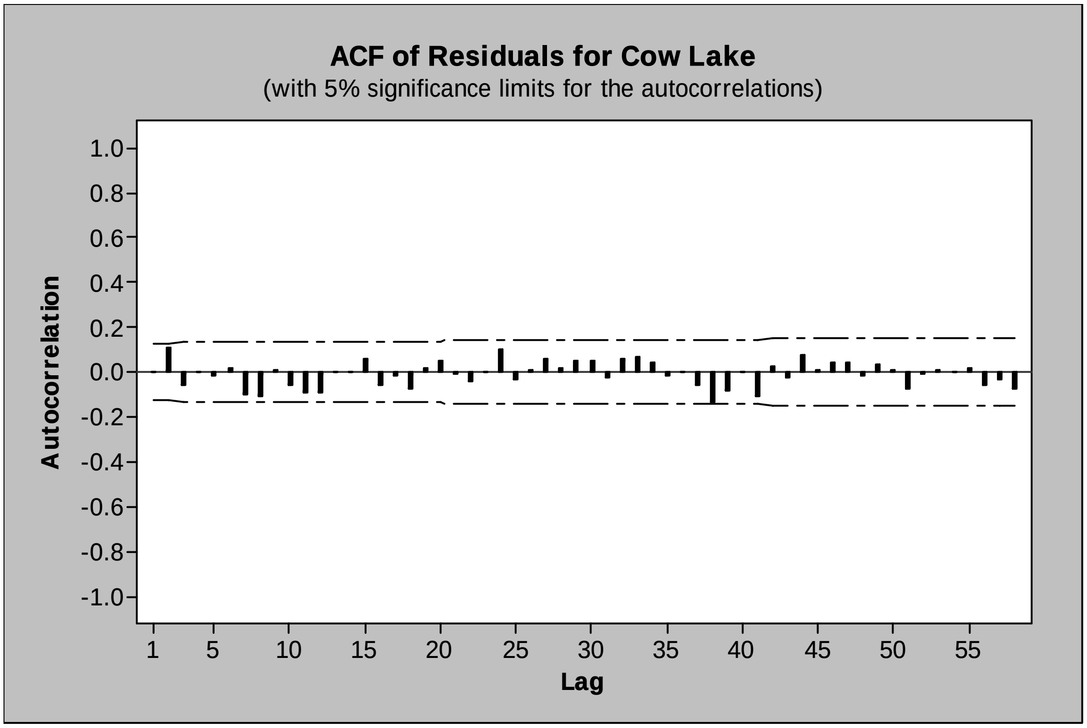

Figure 2-5: Autocorrelation of residuals for Cow Lake (1976-1980, lag in weeks)

2.3.2 Autocorrelation and Variance

The results of the autocorrelation analysis for the time subseries are found in Table 2-5, which includes the exponential parameter that characterizes the autocorrelation from lag 0 until the autocorrelation drops to insignificance. The table also displays the variance in lake levels for each time subperiod. As a basin becomes more and more urbanized, the increase in basin imperviousness and more efficient runoff collection systems should reduce the infiltration within the basin and decrease the time of concentration. This intuitively would lead to a larger fraction of runoff from a given rainfall event reaching a lake more quickly and reducing the much slower process of groundwater recharge to the lake. This change would conceivably result in higher peak stages and lower low stages, since in times of drought, baseflow replenishment would be reduced. It is surmised that a decrease in baseflow will translate to a reduction in autocorrelation in an urbanizing lake watershed with no other signals. A reduction in autocorrelation translates to a steeper autocorrelation curve with a larger exponential parameter. This is the case for Cow Lake and Lake Ann Parker; however, Moon Lake and Lake Padgett show a consistent trend towards longer memory and Lake Thomas and King Lake display a general trend towards longer memory. One potential reason for increased autocorrelation may be wetlands adjacent to the lakes. Lake Thomas, King Lake and Moon Lake have the largest adjacent wetland area as a percent of the total basin. As a larger fraction of runoff from increasing urbanization flows into these adjacent wetlands, the average wetland water levels may increase and pass a larger amount of baseflow into the lake, increasing memory. It is also possible that wetlands are inherently more efficient at recharging the lakes than watershed infiltration due to their proximity. This is consistent with Meyer (2005) who found that although most streams in urbanized watersheds demonstrate a decrease in baseflow, streams with low-permeability near-surface soils and a substantial number of detention basins which can serve as a recharge mechanism demonstrated increases in baseflow. For all lakes that discharge into wetlands, all lakes except Cow and King Lakes, higher antecedent tailwaters for the lakes due to increased overall runoff volume into the wetland storage areas may also contribute to increases in lake memory. In cases where a control structure helps regulate lake stages, including Lake Ann Parker and King Lake, effects on autocorrelation are difficult to isolate; it is expected that if lake stages are lowered by the control structure, autocorrelation would generally decrease as less water is stored in the lake while the opposite should occur if the control structure increases stages. From the data available, it is unclear when the control structures in these lakes were installed or modified. Differences in autocorrelation might also be difficult to filter out at Lake Padgett and Lake Ann Parker, which are flow-through lakes and receive flow from upstream lakes with their own sets of basin and lake alterations.

Table 2-5: Autocorrelation and variance

Lake	Subseries Period	Exponential Parameter	Variance	Significantly Different*
Moon	1965-1976	0.034	0.13	
	1977-1986	0.035	0.11	N
	1987-1996	0.022	0.17	Y
	1997-2006	0.017	0.38	Y
Padgett	1970-1975	0.261	0.04	
	1976-1983	0.064	0.06	Y
	1984-1991	0.097	0.06	N
	1992-2000	0.099	0.07	Y
Thomas	1968-1976	0.063	0.05	
	1977-1985	0.040	0.08	Y
	1986-1994	0.033	0.06	Y
Ann Parker	1969-1976	0.060	0.13	
	1977-1984	0.104	0.06	Y
	1985-1991	0.110	0.08	Y
	2000-2007	0.237	0.20	Y
King	1976-1980	0.089	0.04	
	1981-1985	0.091	0.07	Y
	1986-1991	0.033	0.08	N
Cow	1976-1980	0.077	0.03	
	1981-1985	0.107	0.03	N
	1986-1991	0.219	0.01	Y

* Indicates if the current time period is significantly different from the previous at the 95-percent confidence level

As should be expected, the results of autocorrelation analysis are consistent with the time series modeling; similar patterns of autoregressive parameter increase or decrease through time are observed. However, for lakes that required two autoregressive terms, including Lake Padgett, Lake Thomas and Lake Ann Parker, the trend in parameters for these terms was usually opposite of one another. Most of the lakes demonstrated an increase in variance over time, although the variance for Cow Lake was nearly constant. While Lake Ann Parker had the highest variance in the most recent time period, it did not demonstrate a serial increase in variance as did the other lakes. With the exception of Moon Lake, the increases in variance were marginal and, in many cases, insignificant at the 95-percent confidence level.

Cow Lake appears to give the best representation of the effects of urbanization on lake level autocorrelation (Figure 2-6) as it is the only lake of the six studied that does not have an adjacent wetland or discharge to a wetland, is not a flow-through lake and does not have a control structure. Furthermore, there is a high degree of dense urbanization around the lake. Cow Lake exhibits a significant decrease in autocorrelation over time. However, as it is the only lake studied with the aforementioned characteristics, a strong trend cannot be established.

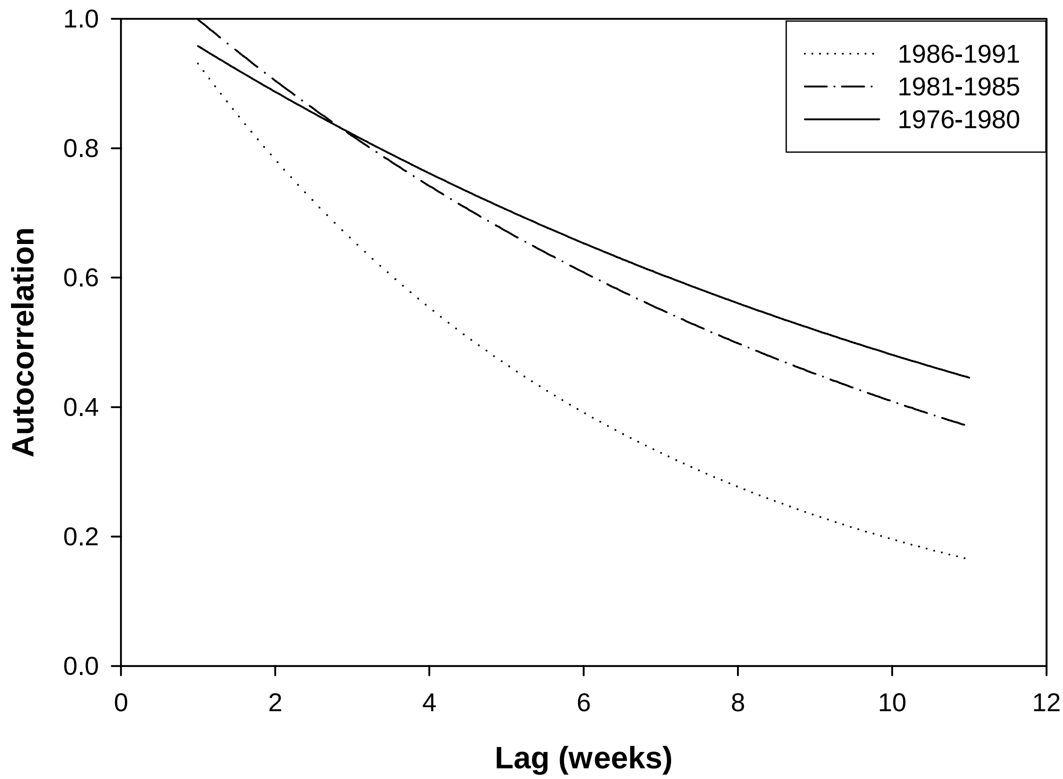

Figure 2-6: Cow Lake autocorrelation with exponential fit lines (1976-1991)

2.3.3 Regression

Table 2-6 gives a summary of the independent variables and associated R^2 for the pre-urbanized and urbanized time periods at each lake as well as the R^2 and AIC for the overall model which includes the entire data set.

Table 2-6: Regression model parameters

Lake	Subseries Period	Parameters*				R^2	AIC
		Pre-week Rain	Pre-month rain	Starting Lake Stage	Temp.		
Moon	All Values					44.16	2.92
	1973-1976	0.0088	0.0009	N/A	-0.0005	48.8	
	1977-2007	0.0084	0.0012	-0.0092	-0.0009	45.7	
Padgett	All Values					52.9	2.62
	1976-1979	0.0152	N/A	-0.0334	-0.0011	73.9	
	1980-2000	0.0133	N/A	-0.0292	-0.0005	48.0	
Thomas	All Values					53.0	1.30
	1976-1979	0.0096	0.0008	N/A	-0.0009	62.3	
	1980-2000	0.0120	N/A	-0.0218	-0.0008	51.5	
Ann Parker	All Values					39.4	3.09
	1972-1976	0.0078	0.0017	-0.0581	N/A	30.7	
	1977-2007	0.0097	0.0010	-0.0214	-0.0011	43.4	
King	All Values					52.8	1.61
	1976-1979	0.0155	N/A	-0.0356	-0.0009	66.5	
	1980-1991	0.0128	0.0006	-	-0.0010	48.1	
Cow	All Values					44.2	1.34
	1976-1979	0.0129	0.0010	-0.0410	-0.0005	68.6	
	1980-1991	0.0097	N/A	-0.0794	N/A	36.3	

*N/A indicates the parameter was not significant at the 0.05 level of significance. AIC was run for the model selection on the entire data set only

There was a large spread in the multiple R^2 values for the subperiod models, from 30.7 to 73.9. However, a fair amount of uncertainty is expected due to the multitude of variables that contribute to lake levels as well as the availability of data. In time periods in which there were more data gaps, i.e., one or more weeks in which missing values had to be imputated by interpolation, R^2 values decreased. Rainfall records were nearly 100-

percent complete for the time periods analyzed, however, there is substantial regional variability in rainfall and none of the lakes had gages located immediately at the lakes themselves. Rainfall gages were within two to six km of the lakes analyzed. Although evaporation is highly correlated with temperature, it is also dependent on wind, humidity and other factors for which data were not available. While transpiration can be a significant fraction of the water budget in a shallow water table environment (Nachabe et al 2005), data for the lakes studied was not available and a transpiration variable term was not included in the regression. Although the lakes are located in a geologically similar region and generally have silty and sandy soils, local differences in soil types, including the presence of wetlands adjacent to a lake, can have an influence on the rainfall-baseflow interaction within individual lakes. Given these factors, the obtained R^2 values are generally acceptable. The independent variables are significant in explaining the changes in lake levels. For all lakes, the regression model inclusive of all four independent variables was deemed most appropriate. Figure 2-7 demonstrates the fit of the model versus the actuals for Cow Lake. Figure 2-8 demonstrates the normality of the residuals for Cow Lake. Residuals for all regression models exhibited normality and homoscedasticity with some deviation from normality at the extremes. All values for the correlation coefficients of the regressors at each lake were equal to or less than 0.4, indicating there is no significant correlation. All lakes with the exception of King Lake exhibited Cook's distances of less than 0.5, indicating no presence of outliers. One data point for King Lake exhibited a Cook's distance of near one, indicating a value of significant influence. However, from inspection of the data, this appears to be a valid data point.

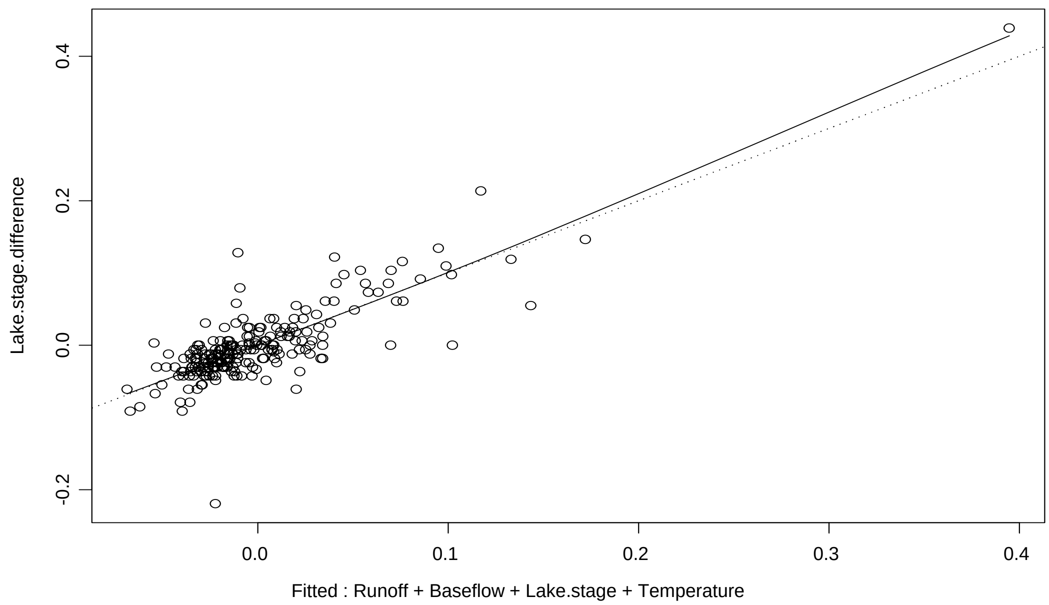

Figure 2-7: Cow Lake response versus fit (1976-1980)

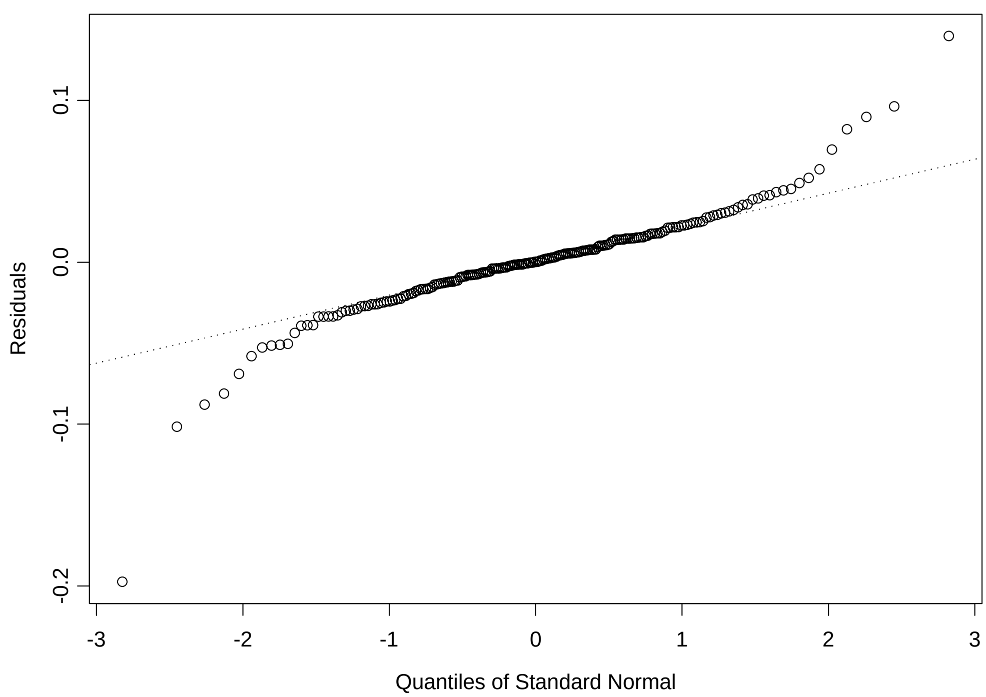

Figure 2-8: Cow Lake quantile-quantile plot of the residuals (1976-1980)

Because of the unique nature of each lake studied, including the presence of adjacent wetlands, control structures, degree of urbanization and flow-through characteristics, it is important to ascertain the individual aspects of each lake that contribute to changes in the regression parameters. For most of the lakes analyzed, the previous-week rainfall variable is fairly consistent with time and is always highly significant. Lake Thomas and Lake Ann Parker are the only lakes to demonstrate an increase in this parameter. Lake Padgett, King Lake and Cow Lake demonstrate a significant decrease in this parameter while Moon Lake shows a marginal decrease. Because this parameter includes the effects of both runoff and baseflow, it is difficult to draw significant conclusions as any increase in runoff may be offset by decreases in baseflow. Daily data would likely be required to reach definitive conclusions as to urbanization-induced changes in this parameter.

Most lakes demonstrated a trend towards decreased baseflow based upon analysis of the pre-month rainfall parameter. At Lake Thomas and Cow Lake baseflow is significant in the pre-urbanized period but not thereafter, representing a reduction in baseflow. At Lake Ann Parker there is a significant decrease of 41 percent in this parameter. At Lake Padgett, the baseflow parameter is not significant for either time period. However, since the runoff/baseflow pre-week rainfall parameter is highly significant for both subperiods and due to the extremely large basin-to-lake-area ratio, this is likely more representative of a reduction in baseflow than a reduction in runoff. Two sites exhibited increases in baseflow; at Moon Lake and King Lake, the baseflow parameter is either low or insignificant in the first time period and larger or significant thereafter. These two lakes have the highest adjacent wetland percentage of the lake basin, giving credence to wetlands being a mechanism of increased baseflow and offsetting some of the impacts of urbanization. Furthermore, despite Moon Lake demonstrating substantial gains in population density and having a higher density that all but Cow Lake in the most recent time period, baseflow still declined. The baseflow trends were consistent with the autocorrelation analysis; increasing baseflow correlated to longer autocorrelation while decreasing baseflow correlated to shorter autocorrelation. Lake Thomas and Lake Padgett were exceptions to this consistency.

Starting lake levels were significant for nearly all lakes. In the few cases where starting lake levels were not significant, lake levels were generally higher than in periods in which starting levels were significant. This is consistent with the morphology of most lakes in which lake surface area increases with depth and a similar volume of runoff makes for a smaller increase in stages as lake levels rise. The temperature variable is significant in most cases. However, in cases where it is not significant it is probably due to factors such as wind or humidity exerting a greater relative impact on evaporation.

Based on the fact that Lake Padgett, Lake Ann Parker and King Lake are flow-through lakes and Lake Ann Parker and King Lake have added control structures, it is difficult to isolate the signal of urbanization on the relative values of the regression parameters. Although Moon Lake, Lake Thomas and Cow Lake do not have the aforementioned complications, the presence of wetlands adjacent to Moon Lake and Lake Thomas also cloud the results due to possible increases in basin runoff storing in the wetlands and recharging the lake in lieu of entering the lake as baseflow from the watershed. As was the case for the autocorrelation analysis, Cow Lake provides the best case to examine any effects watershed urbanization may incur. Furthermore, despite the presence of many other signals unique to each lake, most lakes demonstrated an overall decrease in baseflow.

2.4 Conclusions

Separating the signal of lake-basin urbanization from the multitude of signals inherent in an urbanizing watershed is problematic. The particular lakes chosen for this study were not substantially influenced by pumping, surface water extraction or precipitation trends, helping isolate the effects of urbanization. Many of the lakes did exhibit other sources of influence on water levels, including the addition of control structures or culverts, presence of adjacent wetlands and inflow from upstream lakes. With regard to the time series modeling, lakes with a large basin to lake area ratio demonstrated definite trends in model parameters. As the basin/lake ratio increases, the urbanization signal is likely increased enough to be detected by the time series modeling. Furthermore, a significant increase in basin population density appears to systematically alter the time series signature, despite the presence of other conflicting signals. It was hypothesized that urbanization would shorten the autocorrelation of lakes as the baseflow fraction was decreased due to more efficient drainage and increased impervious area. While this was certainly true in the Cow Lake basin, which is the most heavily urbanized lake and does not have a control structure, inflow from an upstream lake or an adjacent wetland, it was not true in several other lake watersheds. Because all other watersheds have wetlands immediately adjacent to the lakes studied or the lakes discharge into wetlands, it is surmised that wetlands serve as an efficient recharge mechanism and can compensate for effects of urbanization on baseflow by storing increased runoff volume associated with urbanization and slowly passing it back to the lake over time. In nearly all the lakes studied, variance increased with time; in most cases, however, the increase was negligible. For the regression analysis, four lakes demonstrated a decrease in baseflow contribution while the two lakes with the largest relative adjacent wetland areas demonstrated the opposite. Based upon the research, the following general conclusions about lakes in urbanizing watersheds can be reached: 1) The statistical structure of lake level time series is systematically altered and is related to the extent of urbanization. 2) In the absence of other forcing mechanisms, autocorrelation and baseflow appear to decrease. 3) The presence of wetlands adjacent to lakes can offset the reduction in baseflow. These conclusions can be applied globally to similar regions that consist of lakes undergoing urbanization in flat, humid, shallow water table environments with wetlands. Furthermore, the methodology utilized can be applied at lakes in both similar and dissimilar environments to those studied in this research.

3.0 Use of Generalized Extreme Value Covariates to Improve Estimation of Trends and Return Frequencies for Lake Levels

3.1 Background

One of the most important tools in effective water management is the accurate forecast of both long-term and short-term extreme values for both flood and drought conditions. High water stages associated with flood can cause extensive erosion or property damage while low stages associated with drought affect wildlife, ecology, recreation, and water supply. Frequency return periods for both peak highs and lows are often utilized to gage risk and evaluate mitigation methods to minimize this risk. Accurately identifying trends in lake levels can affect long-term decision making such as forecasting water supply, while improving the prediction of near-term frequency return periods can affect short-term planning such as the determination of evacuation zones in the face of an approaching hurricane. Another significant benefit of more accurate short-term forecasts is giving resource managers adequate tools in January to determine how much water to let out of a lake to prepare for flood stages that often occur in August or September. Changes in the general trends of lake, stream and other surface water bodies have been observed in many parts of the world. These trends may be due to factors such as watershed urbanization, water supply pumping and morphological changes to the water body itself or climatic changes. Traditional methods of trend detection, such as ordinary least squares (OLS) or the Mann-Kendall test, are not aptly suited for hydrologic systems since these systems often exhibit time scale issues, non-normal distributions, seasonality, autocorrelation, inconsistent data collection, missing data other complications that render these traditional methods unreliable. In a similar fashion, traditional methods of predicting extreme flood and drought frequencies, such as distribution fitting without parameter covariates, may be highly inaccurate in lake-type systems, especially in the short-term. In the case of lakes, traditional frequency return estimates assume extremes are independent of trend or starting lake stages. However, due to the significant autocorrelation of lake levels, the initial stage can have a significant influence on the severity of a given event. If a 100-year precipitation event occurs at a low lake stage, the peak stage will be much lower than if the initial lake stage is high due to the additional storage available. In Florida, with the annual threat of hurricanes and flat topography where small differences in extreme stages can have significant impacts, utilizing flood or drought predictions that take starting lake stage and future trends into account will allow for more accurate appraisals of both short-term and long-term risk.

One of the objectives of this research was to evaluate trends in a robust manner that can accommodate the autocorrelation, missing data, non-stationarity, etc. previously noted. Many studies have attempted to identify appropriate methods of trend detection in hydrologic data. Hirsch et al (1982) presents the seasonal Mann-Kendall test to improve upon traditional methods of trend detection to accommodate some of the aforementioned complications in hydrologic data. Katz et al (2002) describes the use of extreme distribution parameter covariates in combination with maximum likelihood estimation as a more rigorous methodology to identify trends in hydrologic data. Zhang et al (2004) utilized Monte Carlo simulations to compare OLS, the nonparametric Kendall test, and allowing the parameters of the GEV distribution to vary with time. According to the study, while the nonparametric test is more effective at identifying trends than OLS, allowing a GEV parameter covariate significantly outperforms both OLS and the Kendall test.

The GEV distribution has recently been widely applied to hydrologic studies. Nadarajah and Shiau (2005) utilized the distribution to model flood events for 39 years of data at the Pachang River, Taiwan, and employed parameter covariates of flood volume, duration and time to peak to both identify trends and improve the fit. Morrison and Smith (2002) found the GEV distribution to adequately fit flood peaks in streams with at least 30 years of data in the Appalachian Mountains, United States. Garcia et al (2007) found the GEV distribution with a time covariate to adequately fit and identify trends in daily extreme rainfall in the Iberian Peninsula at gages with 40 years of data.

Another objective of this research was to investigate methods to improve estimation of extreme lake stages by incorporating variables, including starting stage and time, in addition to lake stage. Several studies have analyzed the relation between initial stages, antecedent conditions and flood return periods in various hydrologic systems. Other studies have attempted to quantify the multivariate nature of flooding in streams and other water bodies by incorporating terms such as time to flood peak, initial stage, flood volume, etc. into models and predictions. Buchberger (1995) developed near-term flood risk estimates for Lake Erie, United States, based on an autoregressive time series model and the joint occurrence of a normally distributed storm surge and found that conventional frequency analysis underestimates flood risk when starting lake stages are high and overestimates flood risk when starting lake stages are low. Struthers and Sivapalan (2007) developed flood return periods dependent upon thresholds of evaporation, rainfall frequency, catchment response time, field capacity storage and catchment storage capacity. In a similar study, Kusumastuti et al (2007) developed lake-specific flood frequency return period curves based on field capacity storage and total storage thresholds. Kusumastuti et al (2008) developed lake flood frequency return periods based on several catchment and lake thresholds including antecedent storage, catchment to lake area ratio and magnitude of storm depths. The antecedent storage in the lake was found to be a dominant control on flood frequency and magnitude. Goel et al (1998) developed flood frequency curves based on the joint probability of flood volume and flood peak for the Narmada River, India.

Lake level trends in both flood and drought were investigated in this research utilizing the GEV distribution with a time parameter covariate. Lake flood and drought stages were also modeled with the GEV distribution utilizing covariates of starting lake stage and time. If the addition of time or lake stage covariates offered a significant improvement of the fit, frequency return period curves were developed for these cases. Lakes studied are located in Florida, United States, and have at least 50 years of data that are not significantly anthropogenically altered.

Trend identification in lake levels utilizing the GEV distribution as well as the development of variable return periods based on starting lake stages are a practical application of GEV distribution theory that has not yet been applied to lakes. Estimates of trend that are more accurate than those derived from traditional methods such as OLS as well as more accurate flood and drought frequencies based on starting water level will be of significant use in water resource management in terms of hurricane evacuation decisions, lake management decisions including letting an adequate amount of water out of a lake to minimize flooding impacts from an approaching hurricane or tropical storm, development of appropriate average water levels to maintain throughout the year based upon return curves that can be adjusted to the average water levels selected and preparation for increases or decreases in future flooding or drought. The objectives of this research in regards to lake levels were to 1) accurately identify the direction and magnitude of trends in flood and drought stages and 2) provide more accurate predictions of both long-term and short-term flood and drought stage return frequencies utilizing GEV with time and starting stage covariates.

3.2 Materials and Methods

3.2.1 Lake Information and Data

Lakes with at least 50 years of data were selected across the southwestern portion of Florida that were mostly anthropogenically unaltered, i.e., from significant dredging, placement of berms, pumping, installation of major control structures, etc. in such a way that would significantly change the time series signature and, hence, the underlying distribution. Given the degree of urbanization across Florida, it is not possible to find completely unaltered lakes with sufficient data. However, four lakes, including Lake Arbuckle, Lake Carroll, Lake Trafford and Lake Weohyakapka (Figure 3-1) that are relatively unaltered were utilized.

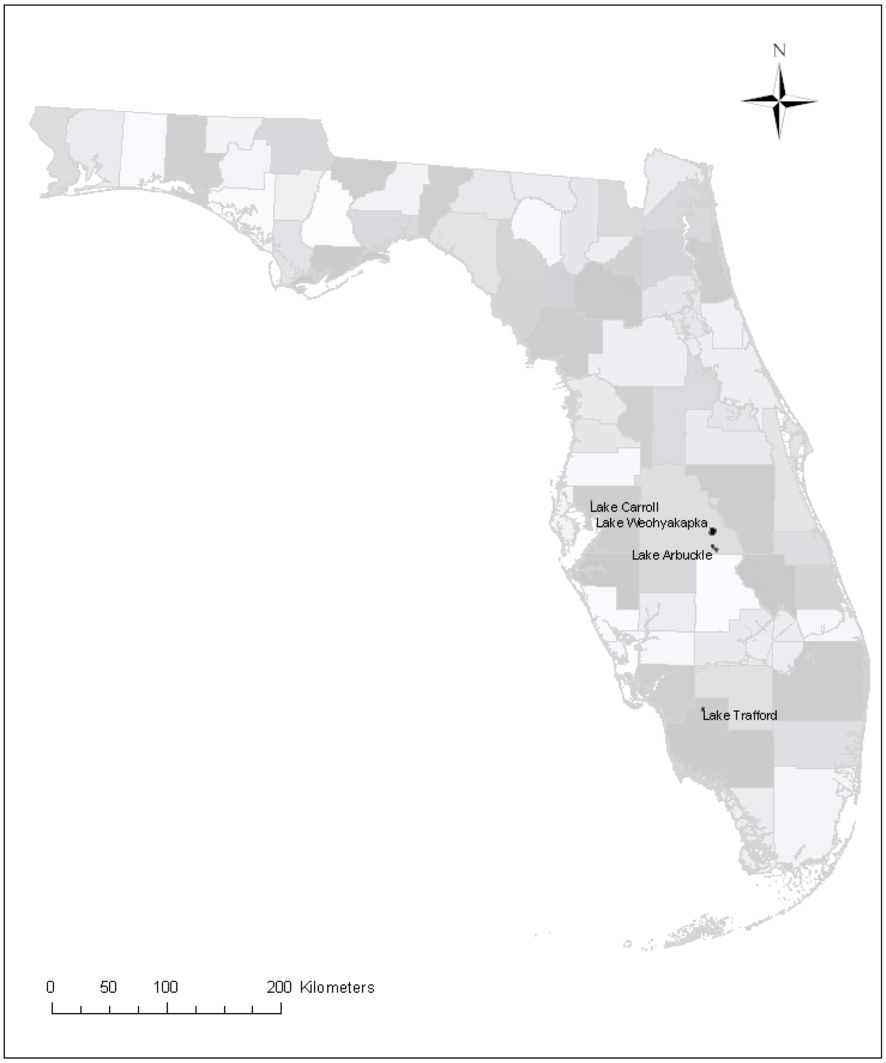

Figure 3-1: Location map of study lakes

3.2.2 GEV Distribution

Trends in lake levels and return level frequencies were identified utilizing extreme value models. The main variables modeled were the annual maximum and minimum lake levels, the flood and drought stages. In order to analyze any trends, distribution parameters were allowed to vary with time. Because lake levels exhibit substantial autocorrelation, it is surmised that annual starting lake levels have a significant impact on the distribution of annual extremes; therefore, the GEV parameters also were allowed to vary with initial stage. The starting lake stage was taken as the water level on January 1st of any given year. The time and starting lake stage covariate models were compared to the original distribution model to determine if a statistically significant better fit was achieved. If covariates do significantly improve the fit, the distribution itself is potentially changing as these covariates change. Changing distribution parameters with time or starting stage allows for the distribution to be non-stationary and also gives an estimate on the rate of change.

The GEV is the generalized form of three commonly applied extreme value distributions: the Gumbel, the Frechet and the Weibull. The GEV is applicable to variables of block maxima, where the blocks are equal divisions of time. The GEV cumulative distribution function is given by:

$$F(x) = \exp\left\{ -\left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \tag{29}$$

where x is the random variable, μ is the location parameter, σ is the scale parameter and ξ is the shape parameter and $1+\xi(x-\mu)/\sigma > 0$. It readily follows that the sub-distributions are:

$$\text{Gumbel:}\quad F(x)=\exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, -\infty < x < \infty \tag{30}$$

$$\text{Frechet:}\quad F(x)=\begin{cases} 0 & x \le \mu \\ \exp\left\{-\left(\dfrac{x-\mu}{\sigma}\right)^{-1/\xi}\right\} & x > \mu \end{cases} \tag{31}$$

$$\text{Weibull:}\quad F(x)=\begin{cases} \exp\left\{-\left(-\dfrac{x-\mu}{\sigma}\right)^{-1/\xi}\right\} & x < \mu \\ 1 & x \ge \mu \end{cases} \tag{32}$$

GEV distribution parameters are determined using maximum likelihood estimation. The log-likelihood function, for $\xi \neq 0$, is given by:

$$l(\mu,\sigma,\xi) = -m\log\sigma - (1+1/\xi)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi}$$

$$\text{given that } 1+\xi\left(\frac{x_i-\mu}{\sigma}\right) > 0 \text{ for i = 1,……, m (Coles, 2004)} \tag{33}$$

The log-likelihood for the GEV distribution with parameters that are a function of time t or starting lake stage s is given by:

$$l(\mu,\sigma,\xi) = -\sum_{i=1}^{m}\left\{\begin{array}{l} \log\sigma(t,s) + (1+1/\xi(t,s))\log\left[1+\xi(t,s)\left(\dfrac{x_{t,s}-\mu(t,s)}{\sigma(t,s)}\right)\right] + \\ \left[1+\xi(t,s)\left(\dfrac{x_t-\mu(t,s)}{\sigma(t,s)}\right)\right]^{-1/\xi(t,s)} \end{array}\right\}$$

$$\text{given that } 1+\xi(t,s)\left(\frac{x_{t,s}-\mu(t,s)}{\sigma(t,s)}\right) > 0 \text{ for all t = 1,……..,m (Coles, 2004)} \tag{34}$$

For purposes of this research, model 1 is the GEV distribution with parameters μ, σ and ξ held constant. The distribution parameters for model 1 for each lake were estimated and the goodness-of-fit was evaluated with the Kolmogorov-Smirnov test statistic at the 95-percent significance level. For model 2, the location parameter of model 1 was allowed to vary with time or starting stage or both to investigate the presence of trends and determine if model 1 could be improved. Model 2 is therefore a submodel of model 1 with

$$\mu = a + by \tag{35}$$

where y is either the time in years or the starting lake stage and a and b are constants. Model 3 is a submodel of model 2 with

$$\mu = c + dt + es \tag{36}$$

where t is the time in years, s is the starting lake stage and c, d and e are constants. Once parameters were estimated for all three cases, the models were compared to determine if the time and/or starting lake stage covariate give a statistically significant better fit. In order to test one model against another, the likelihood ratio test was utilized. If l_1 and l_2 represent the maximized log-likelihoods of the models to be compared, then a deviance statistic is given by:

$$D = 2\{l_2 - l_1\} \tag{37}$$

Assuming a chi-square distribution, a quantile, c_α, at significance α can be determined and if D> c_α, the submodel explains significantly more of the variation in the data (Coles, 2004). Model 2 will be compared to model 1 while model 3 will be compared to both model 1 and model 2. In cases where a model with parameter covariates demonstrated a significantly better fit, fits were further investigated by examining standard quantile plots for visual confirmation of the fit improvement. However, because models 2 and 3 are non-stationary and parameters are varying at each observation, the random variable X should be transformed to a new variable Z for the quantile plot. A transform to the standard Gumbel distribution is given by (Coles, 2004):

$$Z_t = \frac{1}{\xi(t,s)} \log\left\{1 + \xi(t,s)\left(\frac{X_t - \mu(t,s)}{\sigma(t,s)}\right)\right\} \tag{38}$$

Quantile-quantile plots were developed for these transformed standardized variables. If models 2 or 3 demonstrate an improved fit, it means estimated frequency return periods are changing with time or starting lake stage. Although the maximum likelihood ratio test is given more weight than the quantile-quantile plots, the test compares the fit of all actual data points to the model and gives even weight to all frequency events. Because low frequency events are of main interest, quantile-quantile plots were utilized to focus on the fit in the extreme end of the distribution. If an adequate fit in this region was not confirmed via the plots, the simplest model that adequately predicted extremes was selected. Return level plots for the most appropriate model for both flood and drought were developed at each lake. Estimates of quantiles for the return level plots are given by:

$$q_{x,p} = \begin{cases} \mu(t,s) + \dfrac{\sigma}{\gamma}\left[(-\ln(1-p))^{-\gamma}\right], & \gamma \neq 0 \\ \mu(t,s) - \sigma \ln(-\ln(1-p)), & \gamma = 0 \end{cases} \tag{39}$$

where q is the quantile estimate for lake stage x at frequency p (Beirlant et al ,2004).

3.3 Results and Discussion

A summary of the data utilized is given in Table 3-1. Specifically, the number of years of record, maximum, minimum and average stages and variance are provided. Plots of the maximum, minimum and starting stage for Lakes Carroll and Weohyakapka are provided in Figures 3-2 and 3-3. From the table, the standard deviation in lake levels is consistently near 0.3 m. The average difference between the maximum and minimum for the lakes analyzed is 2.13 m. Given the flat topography of west-central Florida and other similar regions, relatively small differences in water level fluctuations can inundate large areas and impact structures that are routinely set as low as 0.3 m above expected high water marks. From inspection of the figures, it appears likely that annual starting stage is correlated with both annual maximum flood and minimum drought stages as the starting stage approximately parallels both the flood and drought stages. The fits of the lake stages for flood and drought are given in Tables 3-2 and 3-3, respectively, for all GEV models. The Kolmogorov-Smirnov values were well within the 95-percent test statistic for the no-covariate fits, indicating the fits are acceptable.

Table 3-1: Lake data summary

Lake	Period of Record	Average (m) (NGVD)	Maximum (m) (NGVD)	Minimum (m) (NGVD)	Standard Deviation (m)
Arbuckle	1942-2008	16.35	17.79	15.59	0.36
Carroll	1946-2003	10.76	12.10	9.41	0.34
Trafford	1941-2007	5.98	6.95	4.85	0.29
Weohyakapka	1958-2008	18.64	19.46	17.95	0.24

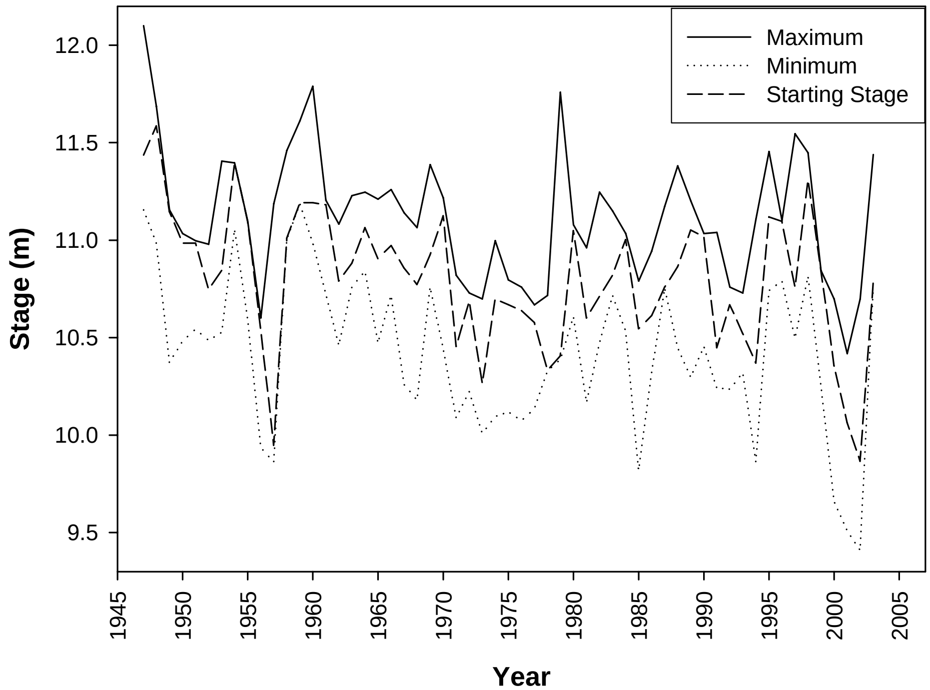

Figure 3-2: Lake Carroll stage data

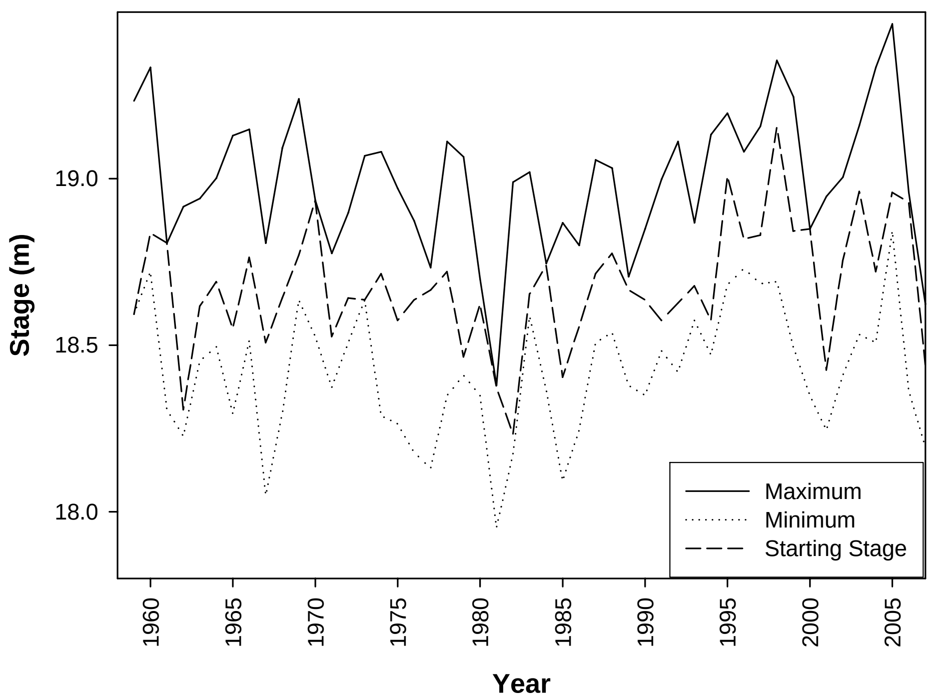

Figure 3-3: Lake Weohyakapka stage data

Table 3-2: GEV flood parameter summary

Lake	Loc. μ	μ-time cov.	μ-start stage cov.	Scale σ	Shape ξ	Likelihood ratio*
			No covariate (model 1)			
Arbuckle	16.831	N/A	N/A	0.347	-0.286	
Carroll	10.984	N/A	N/A	0.300	-0.147	
Trafford	6.285	N/A	N/A	0.193	-0.191	
Weohyakapka	18.928	N/A	N/A	0.214	-0.335	
			Time covariate (model 2)			Model 2/ Model 1
Arbuckle	17.085	-0.007	N/A	0.336	-0.385	**10.301**
Carroll	11.176	-0.006	N/A	0.282	-0.149	**6.567**
Trafford	6.301	-0.000	N/A	0.193	-0.192	0.090
Weohyakapka	18.876	0.002	N/A	0.216	-0.386	1.204
			Starting stage covariate (model 2)			Model 2/ Model 1
Arbuckle	8.954	N/A	0.484	0.343	-0.387	**8.203**
Carroll	3.299	N/A	0.712	0.165	0.194	**45.895**
Trafford	4.015	N/A	0.376	0.179	-0.115	**4.457**
Weohyakapka	5.691	N/A	0.710	0.192	-0.464	**16.431**
			Time and starting stage covariates (model 3)			Model 3/Model 1:Model
Arbuckle	13.178	-0.007	0.240	0.335	-0.488	16.626/8.422
Carroll	3.470	-0.000	0.697	0.169	0.169	45.782/0.113
Trafford	3.858	-0.001	0.407	0.177	-0.109	4.898/0.441
Weohyakapka	6.341	-0.000	0.676	0.190	-0.449	17.081/0.649

* At the 95-percent confidence interval, a maximum likelihood ratio of greater than 3.842 for model 2/model 1 or model 3/model 2 and 5.992 for model 3/model 1 indicates a significantly better fit. The model 2/model 1 and model 3/model 1 ratios were also compared to determine if the additional degree of freedom improves the fit. Selected models are bolded.

Table 3-3: GEV drought parameter summary

Lake	Loc. μ	μ-time cov.	μ-start stage cov.	Scale σ	Shape ξ	Likeli-hood ratio*
			No covariate (model 1)			
Arbuckle	15.844	N/A	N/A	0.182	-0.150	
Carroll	10.304	N/A	N/A	0.409	-0.395	
Trafford	5.554	N/A	N/A	0.301	-0.504	
Weohyakapka	18.354	N/A	N/A	0.194	-0.326	
			Time covariate (model 2)			Model 2/ Model 1
Arbuckle	15.915	-0.002	N/A	0.181	-0.169	2.484
Carroll	10.561	-0.008	N/A	0.380	-0.455	**11.005**
Trafford	5.498	0.001	N/A	0.295	-0.465	0.549
Weohyakapka	18.293	0.0026	N/A	0.193	-0.363	2.023
			Starting stage covariate (model 2)			Model 2/ Model 1
Arbuckle	10.414	N/A	0.335	0.154	-0.383	7.536
Carroll	0.117	N/A	0.947	0.200	-0.264	**71.923**
Trafford	-0.804	N/A	1.055	0.200	-0.323	**39.937**
Weohyakapka	3.855	N/A	0.778	0.149	-0.445	**32.044**
			Time and starting stage covariates (model 3)			Model 3/Model 1:Model
Arbuckle	9.441	-0.001	0.396	0.166	-0.276	**21.753/14.217**
Carroll	0.439	-0.001	0.921	0.198	-0.259	72.413/0.489
Trafford	-1.519	0.000	1.177	0.177	-0.329	34.432/5.505
Weohyakapka	15.839	0.002	0.134	0.176	-0.430	8.160/23.884

* At the 95-percent confidence interval, a maximum likelihood ratio of greater than 3.842 for model 2/model 1 or model 3/model 2 and 5.992 for model 3/model 1 indicates a significantly better fit. The model 2/model 1 and model 3/model 1 ratios were also compared to determine if the additional degree of freedom improves the fit. Selected models are bolded.

3.3.1 Trend Analysis

In regards to modeling both lake flood and drought stages with the GEV distribution and a time covariate, only Lake Carroll demonstrated a statistically significant improvement in the model 2 fit over the model 1 fit with the GEV distribution alone. Lake Arbuckle exhibited a trend in annual flood stages but not drought stages. For Lake Carroll, the model 2 location parameter for flood stages, which yields an estimate of the relation between lake stage and time t in years, is given by:

$$11.176 - 0.006t \tag{40}$$

And the model 2 location parameter for drought stages is given by:

$$10.561 - 0.008t \tag{41}$$

Although the maximum likelihood ratio for both trends is substantially larger than the 95-percent confidence limit threshold, the actual change in flood or drought stage is relatively small, 0.006 m and 0.008 m of decrease per year that the trend is extended into the future. This slight trend is visually confirmed in Figure 3-2. For Lake Arbuckle, the model 2 location parameter for flood stages is given by:

$$17.085 - 0.007t \tag{42}$$

The trend is again downward and of similar order, a decrease of 0.007 m per year. The lakes studied are relatively unaltered in regards to excessive pumping, dredging, management or other mechanisms that may induce dramatic trends. Furthermore, Paynter and Nachabe (2008) determined that the rainfall patterns in the southwest Florida region do not exhibit significant trends that would correlate to changes in lake levels. Lakes Arbuckle, Trafford and Weohyakapka are fairly undeveloped when compared to Lake Carroll, which is highly urbanized. Although many lakes in Florida have demonstrated significant trends due to pumping or anthropogenic change, it appears lakes left in a fairly natural state such as the four studied for this research exhibit slight but statistically significant trends in the case of Lakes Carroll and Arbuckle or no trends in the cases of Lakes Trafford and Weohyakapka.

3.3.2 Starting Stage Analysis

3.3.2.1 Flood Return Period

According to the maximum likelihood ratios, model 2, with a starting stage covariate, is most appropriate for Lakes Carroll, Trafford and Weohyakapka while model 3 is most appropriate for Lake Arbuckle. It should be noted that the magnitude of the likelihood ratio is proportional to the degree of improvement of the fit; the model 2 ratios are generally high. Only Lake Arbuckle demonstrated a statistically significant improvement in fit when covariates for both time and starting stage are included. However, as the trend component of model 3 is negligible, the simpler model 2 was selected. The Lake Arbuckle (Figure 3-4) and Lake Carroll (Figure 3-5) quantile-quantile plots demonstrate an adequate fit for model 2. Quantile-quantile plots for Lakes Trafford and Weohyakapka also demonstrated adequate model 2 fits. With the exception of Lake Carroll, in most of the quantile-quantile plots for flood, the fit breaks down at the extreme end for all models. This is partly due to these points representing hurricanes or tropical storms that are not part of the same distribution as normal rainfall events and partly due to extrapolating extreme events with 50 years of data. After evaluating both the maximum likelihood ratios and the quantile-quantile plots, model 2 was selected for all four lakes in terms of flood stage. The location parameter, which yields an estimate of the relation between starting stage and flood stage, is given by the following for Lakes Arbuckle, Carroll, Trafford and Weohyakapka, respectively, for starting stage s:

$$8.954 + 0.484s \tag{43}$$

$$3.299 + 0.712s \tag{44}$$

$$4.015 + 0.376s \tag{45}$$

$$5.691 + 0.710s \tag{46}$$

For every unit change in starting stage, there is a substantial change ranging from 0.376 to 0.712 m in the flood stage for a given year, indicating a very high degree of correlation.

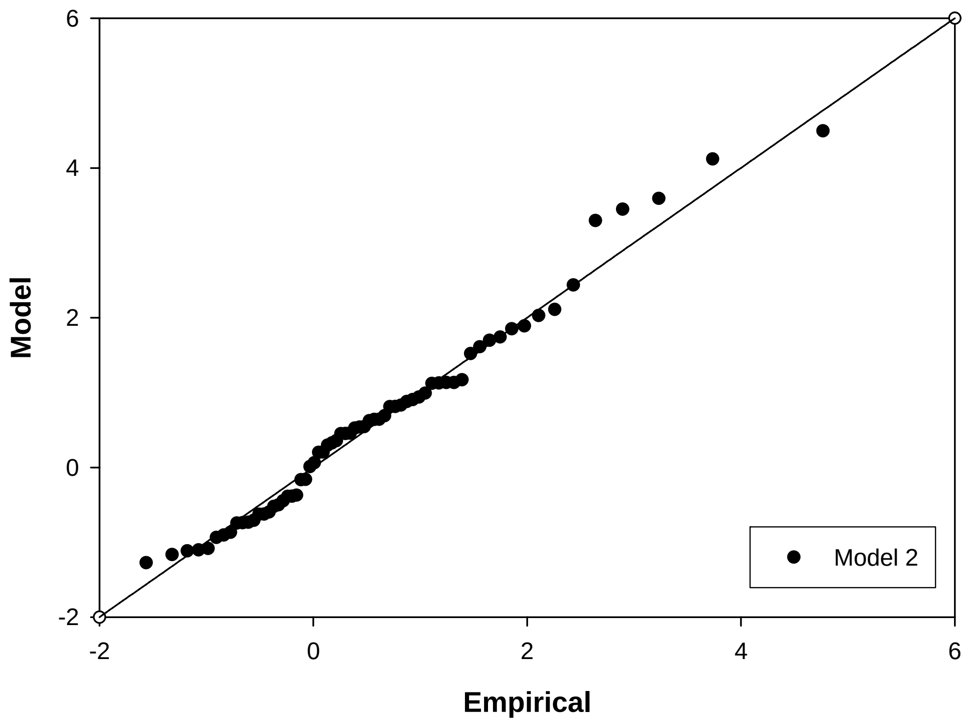

Figure 3-4: Lake Arbuckle flood stage standardized residual quantiles

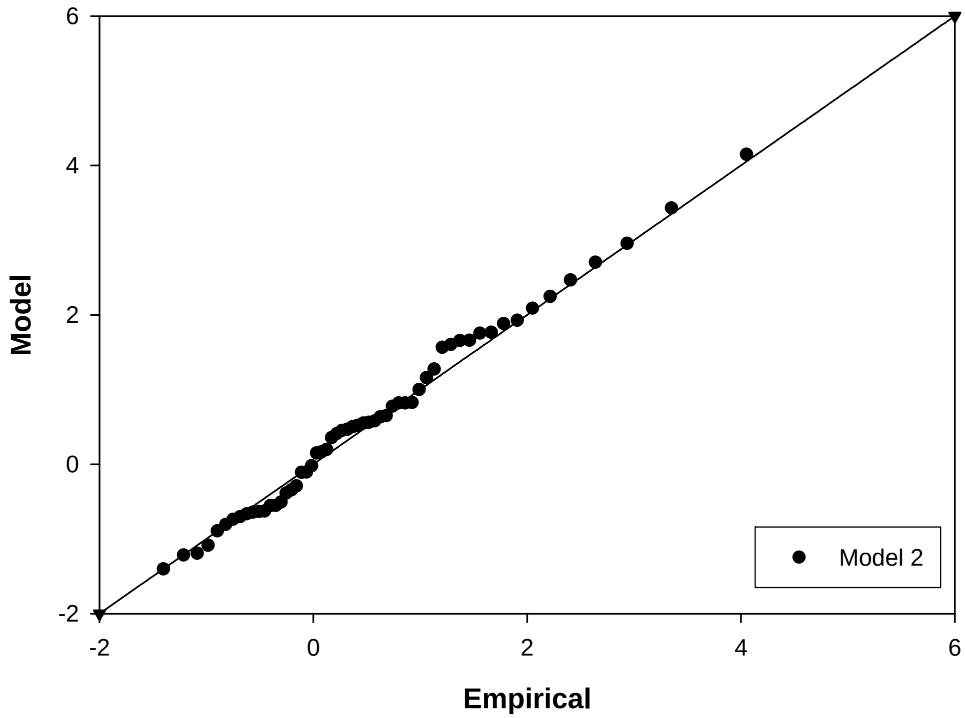

Figure 3-5: Lake Carroll flood stage standardized residual quantiles

Figures 3-6 and 3-7 give the model 2 flood return period for Lakes Arbuckle and Trafford associated with the maximum, minimum and average starting stage as well as the return period associated with no covariate. For each lake, the return period associated with the average starting stage is fairly close to the return period associated with no covariate. For Lakes Arbuckle, Carroll and Weohyakapka, there is some divergence between these two curves towards the larger return periods. At Lakes Arbuckle and Carroll, this is likely due to the fact that these lakes exhibit some trends and since the starting stage should correlate to any trends, the inclusion of the starting stage covariate improves the fit and causes divergence from the fit without a covariate. The flood return period associated with no covariate is bounded by that associated with the maximum and minimum starting stage. In years with a low starting stage, traditional frequency analysis overpredicts the 100-year flood by 108.3, 129.4, 75.9 and 179.2 percent of standard deviation for Lakes Arbuckle, Carroll, Trafford and Weohyakapka, respectively. In years with a high starting stage, traditional frequency analysis underpredicts the 100-year flood by 50, 232.4, 69.0, and 91.7 percent of standard deviation for the same lakes. As such there is a 0.57m, 1.22m, 0.42m and 0.65m difference, respectively, between the 100-year return period stage for the maximum and minimum starting lake stage covariate. Given the flat topography in Florida and other similar regions, a difference of as much as 1.22m can mean a substantial increase in the extent of flooding and potential number of structures flooded.

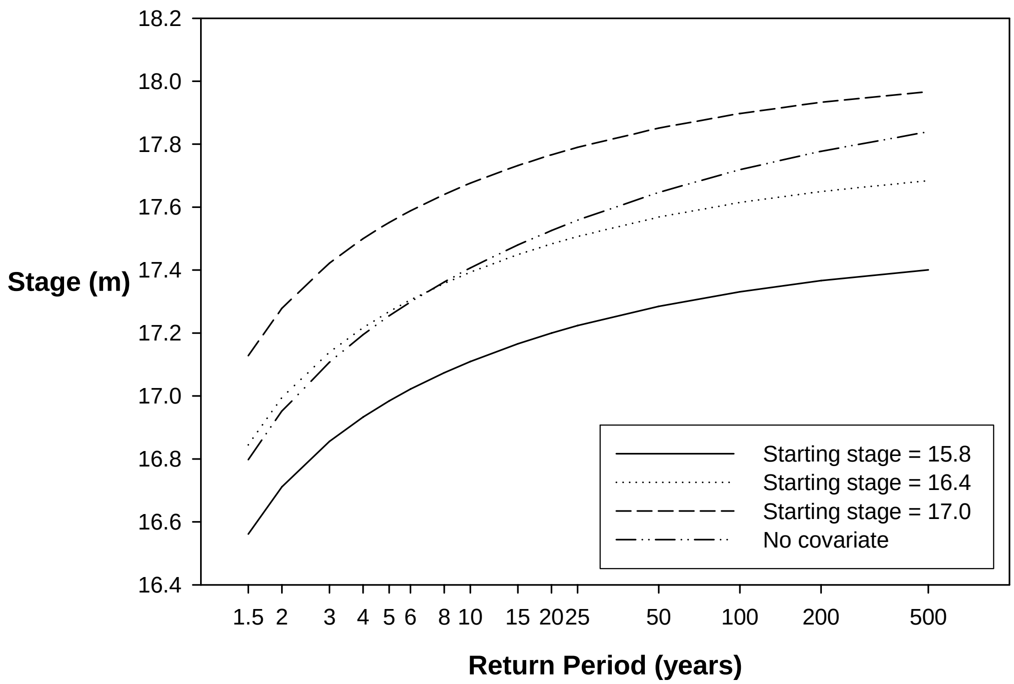

Figure 3-6: Lake Arbuckle flood frequencies with and without covariates

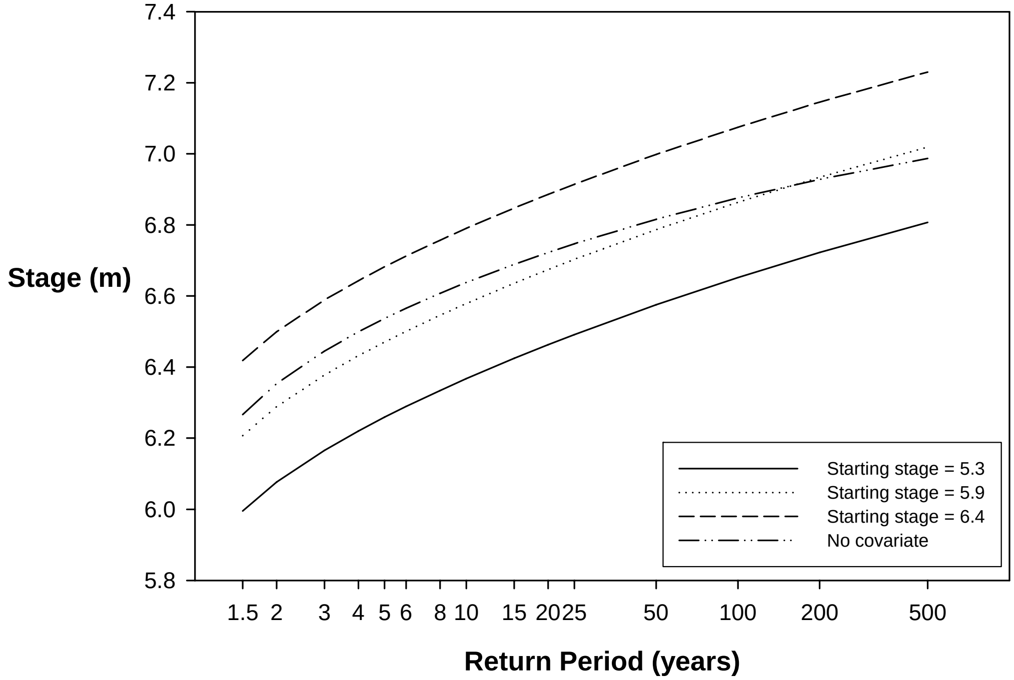

Figure 3-7: Lake Trafford flood frequencies with and without covariates

Since more area is available at consistently higher elevations of a lake, it takes more runoff or baseflow volume to cause a unit rise in stage at higher lake elevations. Because of this it would be expected that in a lake left in its natural stage, return period curves would flatten out at more extreme frequencies. However, once a lake basin is urbanized, the watershed infilled with construction and management structures installed, it is difficult to consistently predict the shape of these curves in a general sense. Lakes Arbuckle, Trafford and Weohyakapka are relatively undeveloped and they demonstrate the expected flattening of the return period curves at higher frequencies. Lake Carroll is the most urbanized and it shows some steepening of the return period curves at extreme events.

3.3.2.2 Drought Return Period

According to the maximum likelihood ratios, model 2 (with a starting stage covariate) is most appropriate for Lakes Carroll, Trafford and Weohyakapka while model 3 is most appropriate for Lake Arbuckle. As in the flood analysis, the trend component is quite small and the simpler model 2 was deemed appropriate. Similar to the flooding case, the likelihood ratios for model 2 are quite high, indicating that model 2 explains substantially more of the variation. The quantile-quantile plots for Lakes Arbuckle, Carroll, Trafford (Figure 3-8) and Weohyakapka (Figure 3-9) indicate an adequate fit for model 2. As with the flood quantiles, there is divergence between the model and empirical data at the extremes. This is likely due to longer time-scale cycles, such as La Nina, that cause excessively dry years and are not explicitly included in the models; model 2 should capture some, but not all, of these longer cycles with the inclusion of starting stage. Some of the fit breakdown is also due to extrapolating events greater than the 50-year from 50 years of data. After evaluating both the maximum likelihood ratios and the quantile-quantile plots, model 2 was selected for all four lakes.

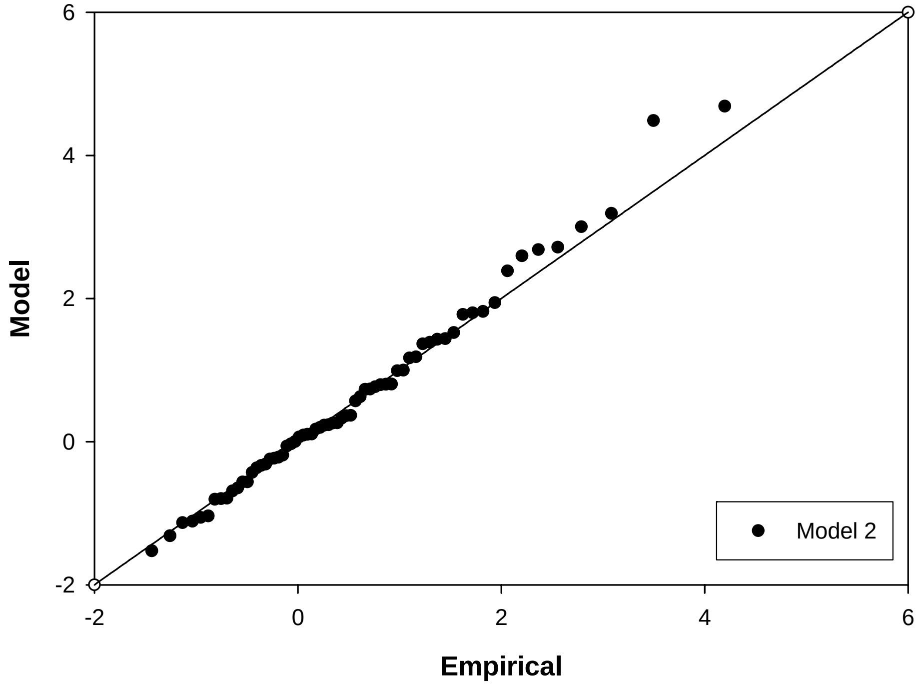

Figure 3-8: Lake Trafford drought stage standardized residual quantiles

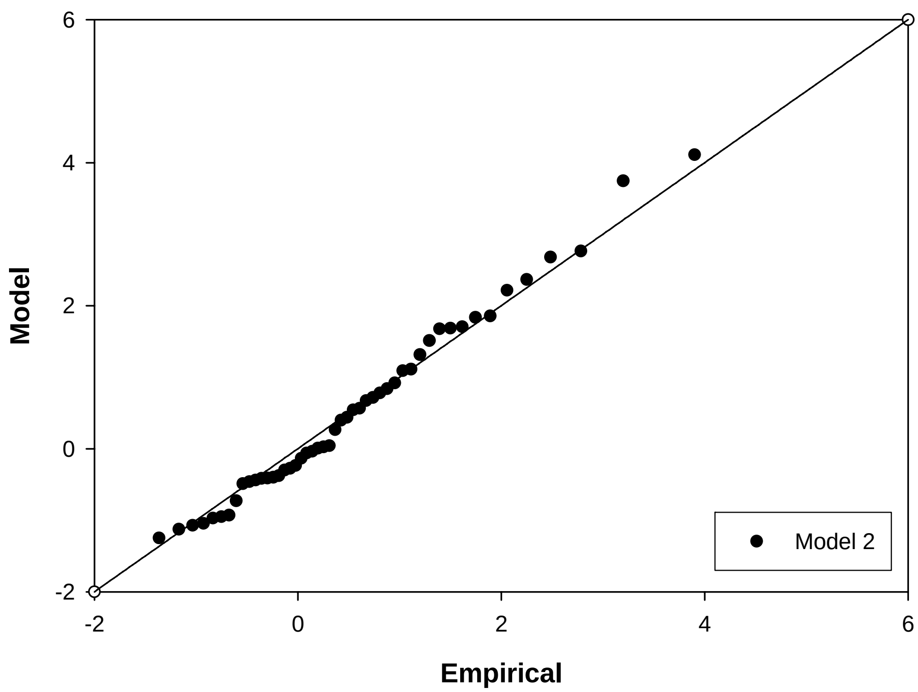

Figure 3-9: Lake Weohyakapka drought stage standardized residual quantiles

The location parameter associated with the most appropriate model for each lake is given by the following for Lakes Arbuckle, Carroll, Trafford and Weohyakapka, respectively, for starting stage s:

$$10.414 + 0.335s \tag{47}$$

$$0.117 + 0.947s \tag{48}$$

$$-0.804 + 1.055s \tag{49}$$

$$3.855 + 0.778s \tag{50}$$

As in the flood case, for every unit change in starting stage, there is a substantial change in the drought stage for that year, in this case ranging from 0.335m to 1.055m. Figures 3-10 and 3-11 give the drought return period for Lakes Arbuckle and Carroll associated with the maximum, minimum and average starting stage as well as the return period associated with no covariate. Similar to the flood return period case, the return period associated with no covariate is bounded by that associated with the maximum and minimum starting stage and nearly parallels the return period associated with the average starting stage covariate. One exception is Lake Carroll where the no-covariate return period curves deviate significantly from the average starting stage covariate curves towards the extreme end. Lake Carroll was the only lake to exhibit a significant drought trend and, as in the flood case, including the starting stage as a covariate captures some of this trend and provides a better fit. In years with a low starting stage, traditional frequency analysis overpredicts the 100-year drought by 41.4, 105.9, 158.6 and 104.2 percent of standard deviation for Lakes Arbuckle, Carroll, Trafford and Weohyakapka, respectively. In years with a high starting stage, traditional frequency analysis underpredicts the 100-year drought by 66.7, 373.5, 251.7 and 191.7 percent of standard deviation for the same lakes. As such there is a 0.39m, 1.63m, 1.19m and 0.72m difference, respectively, between the 100-year return period stage for the maximum and minimum starting lake stage covariate. In similar fashion to flood stages, it is expected that drought return period curves would flatten at more extreme return periods since there are more water loss mechanisms at higher lake stages. At lower stages, the only method of water loss may be evapotranspiration or recharge to the ground. All four lake drought return curves follow this general pattern.

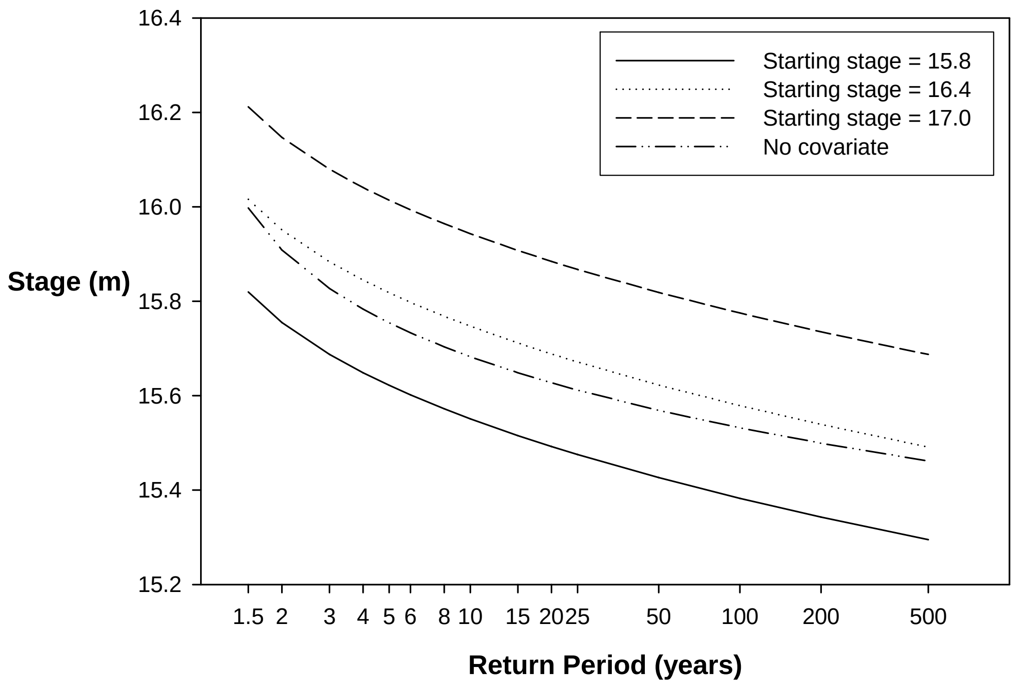

Figure 3-10: Lake Arbuckle drought frequencies with and without covariates

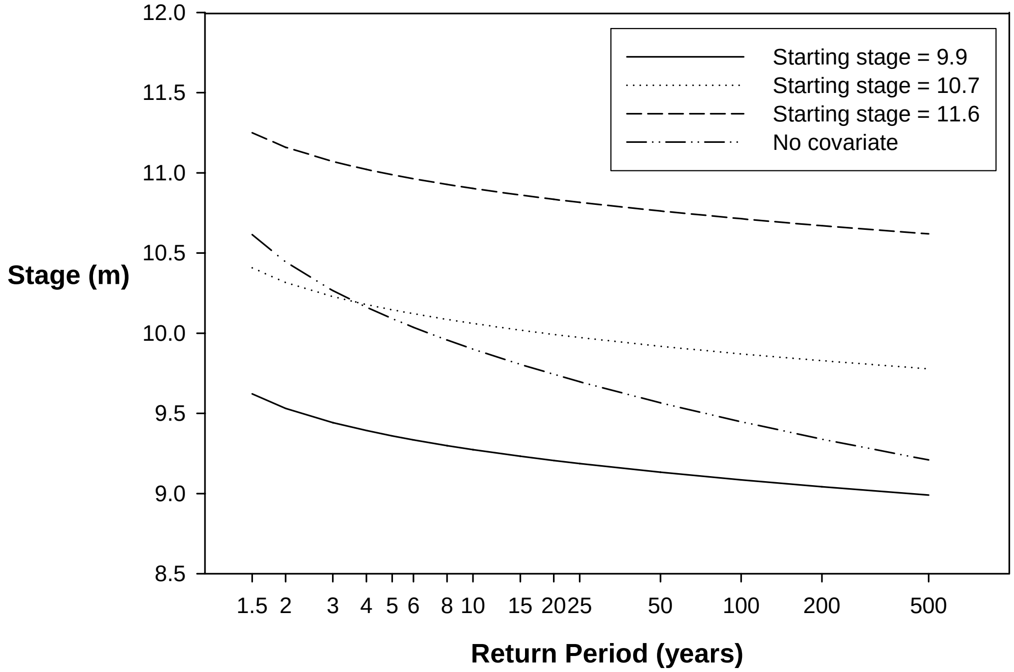

Figure 3-11: Lake Carroll drought frequencies with and without covariates

There appears to be no correlation in the difference between flood stages and drought stages for the minimum and maximum starting lake stages within each lake, i.e, a small difference in the Lake Carroll flood stage associated with the maximum and minimum starting stage does not indicate a small difference in the drought stage associated with the maximum and minimum starting stage. This is likely due to different physical dynamics operating in the flood and drought cases. Flood stages are generally controlled by some management mechanism, i.e., a weir, culvert, gate, etc. while drought stages are largely uncontrolled other than natural losses such as evapotranspiration or seepage to the ground. Furthermore, at extreme flood stages, the basin morphology may change relative to the lake at lower stages, i.e., higher stages may be flatter than at lower stages, a basin popoff to another basin may be reached or housing construction may have significantly altered the historic basin by infill.

In all cases for both flood and drought, adding covariates for both trend and starting stage offer little improvement over a starting stage alone. This is likely because any monotonic trend in time should be captured in the starting stage variable and because the trends identified were very small. Potential scenarios for the inclusion of both time and starting stage covariates improving a fit include situations in which overall trends may not be reflected in the January 1st stage. One possibility may be seasonal trends such as an increase in summer floods due to hurricanes or tropical storms followed by periods of low rainfall whereby the annual starting stage returns to normality.

3.4 Conclusions

The lakes studied were relatively unaltered in terms of extensive pumping, dredging, filling or other measures that would significantly alter the underlying lake level distribution. All of the lakes researched evidenced either no trend or very small trends unlikely to significantly alter prediction of future flood or drought return levels. However, for all of the lakes, significant improvement in the fits was obtained with the inclusion of starting lake stage as a covariate. This is likely because any monotonic trends are captured in the starting stage itself and the trends identified were negligible. Traditional methods of estimating flood or drought stages significantly overpredict stages when starting lake stages are low and underpredict stages when starting stages are high. The difference between these predictions can be substantially more than one meter, a significant amount in urbanized watersheds in areas of the world with flat topography. Flood differences of over one meter can mean significant alterations in evacuation or other water management decisions. In addition to improving prediction of extreme events, utilizing GEV with time or starting stage covariates can provide guidance in lake management decisions in regards to how much water to release from a lake in preparation for an approaching hurricane, appropriate lake levels to maintain throughout the year or determining minimum structure flood elevations in the watershed. Although there is less that can be done from a management standpoint in regards to drought, utilizing GEV with covariates provides a more accurate estimate of expected drought return periods, which can be useful in forecasting future water supply or impacts to tourism. The methodology employed in this research provides a means to estimate the direction and magnitude of lake trends that is robust despite the inherent difficulties in determining trends in hydrologic data. The methodology also allows for more accurate prediction of flood and drought return frequencies that can be applied to nearly any region globally.

4.0 Conclusion

The focus of this research was threefold: 1) to determine the extent of spatio-temporal changes in precipitation patterns utilizing methods that can be regionalized and applied to various water resources such as lakes, streams, reservoirs or other water bodies 2) to determine the statistical changes that occur in lakes with urbanizing watersheds and 3) to develop accurate prediction of trends and lake level return frequencies.

In terms of spatial changes in precipitation, the vast majority of variables analyzed at each gage were confined to a 99-percent confidence band associated with the average fit, gamma or GEV, of the data. There were some exceptions; however most of these were at gages at the outer fringes of the area analyzed and at percentiles near the high or low end. Nearly all of the fits were contained at the 0.5 percentile, representing the average annual variable a particular water resource can expect to experience. In regards to temporal variability, it was also somewhat surprising that almost no significant trends were detected. Many of the gages investigated would have demonstrated a trend if analyzed with traditional methods such as ordinary least squares or non-parametric Mann-Kendall.

Separating the signal of lake-basin urbanization from the multitude of signals inherent in an urbanizing watershed is problematic. With regard to the time series modeling, lakes with a large basin to lake area ratio demonstrated definite trends in model parameters. As the basin/lake ratio increases, the urbanization signal is likely increased enough to be detected by the time series modeling. Furthermore, a significant increase in basin population density appears to systematically alter the time series signature, despite the presence of other conflicting signals. It was hypothesized that urbanization would shorten the autocorrelation of lakes as the baseflow fraction was decreased due to more efficient drainage and increased impervious area. While this was certainly true in the Cow Lake basin, which is the most heavily urbanized lake and does not have a control structure, inflow from an upstream lake or an adjacent wetland, it was not true in several other lake watersheds. Because all other watersheds have wetlands immediately adjacent to the lakes studied or the lakes discharge into wetlands, it is surmised that wetlands serve as an efficient recharge mechanism and can compensate for effects of urbanization on baseflow by storing increased runoff volume associated with urbanization and slowly passing it back to the lake over time. For the regression analysis, four lakes demonstrated a decrease in baseflow contribution while the two lakes with the largest relative adjacent wetland areas demonstrated the opposite. Based upon the research, the following general conclusions about lakes in urbanizing watersheds can be reached: 1) The statistical structure of lake level time series is systematically altered and is related to the extent of urbanization. 2) In the absence of other forcing mechanisms, autocorrelation and baseflow appear to decrease. 3) The presence of wetlands adjacent to lakes can offset the reduction in baseflow.

In regards to utilizing the GEV distribution with covariates to identify trends, all of the lakes researched evidenced either no trend or very small trends unlikely to significantly alter prediction of future flood or drought return levels. However, for all of the lakes, significant improvement in the fits was obtained with the inclusion of starting lake stage as a covariate. This is likely because any monotonic trends are captured in the starting stage itself and the trends identified were negligible. Traditional methods of estimating flood or drought stages significantly overpredict stages when starting lake stages are low and underpredict stages when starting stages are high. The difference between these predictions can be nearly two meters, a significant amount in urbanized watersheds in areas of the world with flat topography. Differences of near two meters can mean significant alterations in evacuation or other water management decisions. In addition to improving prediction of extreme events, utilizing GEV with time or starting stage covariates can provide guidance in lake management decisions in regards to how much water to release from a lake in preparation for an approaching hurricane, appropriate lake levels to maintain throughout the year, determine minimum structure floor elevations in the watershed and allow more accurate forecasting of future water supply or impacts to tourism.

The methodology utilized for each of the three focus areas of the research can be applied to other regions globally. Furthermore, the results can likely also be applied to similar regions with flat topography and shallow water table environments. The focus of this research is on water management and engineering and there are several implications and applications that can be derived from this research. Developing regional rainfall patterns that take into account potential trends allows water managers to develop realistic expectations for future water supply, water levels, etc. at ungaged water resources in a given region. Understanding the impacts of urbanization allows for better management and engineering decisions in regards to lakes and other water resources. For example, the implication of wetlands mitigating some of the effects of urbanization may imply constructing or preserving wetlands adjacent to a lake as part of a regional development plan should be a future consideration. The benefits and implications of more accurate short-term flood and drought return period predictions are many. Improving these predictions by approximately one meter or more can mean very different flood evacuation zones in areas with flat topography and may alter the design of control structures as to how much water to release and appropriate lake levels to maintain throughout the year.

References

Abraham B, Ledolter J (2006) Introduction to regression modeling. Thompson Brooks/Cole Belmont

Altunkaynak A (2007) Forecasting surface water level fluctuations of Lake Van by artificial neural networks. Water Resources Management 21: 399-408

Beirlant J, et al (2004) Statistics of Extremes. John Wiley &Sons, Limited

Buchberger S (1995) Conditional frequency analysis of autocorrelated lake levels. Journal of Water Resources Planning and Management 121: 158-170

Burgueno A, Serra C, et al (2004) Monthly annual statistical distributions of daily rainfall at the Fabra observatory (Barcelona, NE Spain) for the years 1917-1999. Theoretical and Applied Climatology 77: 57-75

Cannarozzo M, Noto LV, et al (2006) Spatial distribution of rainfall trends in Sicily (1921-2000). Physics and Chemistry of the Earth 31: 1201-1211

Changnon S, Demissie M (1996) Detection of changes in streamflow patterns and floods resulting from climate fluctuations and land-use drainage changes. Climatic Change 32: 411-421

Chatfield C (2004) The analysis of time series. CRC Press LLC Boca Raton

Choi S, Wette R (1969) Maximum likelihood estimation of the parameters of the gamma distribution and their bias. Tecnometrics 11: 683-690

Coles, Stuart (2004) An introduction to statistical modeling of extreme values. Springer-Verlag London Limited

Dahamsheh A, Aksoy H (2007) Structural characteristics of annual precipitation data in Jordan. Theoretical and Applied Climatology 88: 201-212

Elias D, Ierotheos Z (2006) Quantifying the rainfall-water level fluctuation process in a geologically complex lake catchment. Environmental Monitoring and Assessment 119: 491-506

Gao J (2004) Lake stage fluctuation study in West-Central Florida using multiple regression models. Master's thesis University of South Florida.

Garcia J, Gallego M, et al (2007) Trend in block-seasonal extreme rainfall over the Iberian peninsula in the second half of the twentieth century. Journal of Climate 20: 113-130

Gibson J, et al (2006) Hydroclimatic controls on water balance and water level variability in Great Slave Lake. Hydrological Processes 20: 4155-4172

Goel N, et al (1998) Multivariate modeling of flood flows. Journal of Hydraulic Engineering 146-155

Groisman P, Karl T, et al (1999) Changes in the probability of heavy precipitation: important indicators of climate change. Climatic Change 42: 243-283

Haan, Charles T. (2002) Statistical Methods in Hydrology. Iowa State Press

Hirsch R, Slack J, et al (1982) Techniques of trend analysis for monthly water quality data. Water Resources Research 18: 107-121

Irvine K, Eberhardt A (1992) Multiplicative seasonal ARIMA models for Lake Erie and Lake Ontario water levels. Journal of the American Water Resources Association 28: 385-396

Jones R, et al (2001) Modeling historical lake levels and recent climate change at three closed lakes, Western Victoria, Australia (c. 1840-1990). Journal of Hydrology 246: 159-180

Khan M, Coulibaly P (2006) Application of support vector machine in lake water level prediction. Journal of Hydrologic Engineering 11: 199-205

Karl T, Knight R (1998) Secular trends of precipitation amount, frequency, and intensity in the United States. Bulletin of the American Meteorological Society 79: 231-241

Katz R, et al (2002) Statistics of extremes in hydrology. Advances in Water Resources 25: 1287-1304

Kuhn G, Khan S, et al (2007) Geospatial-temporal dependence among weekly precipitation extemes with applications to observations and climate model simulations in South America. Advances in Water Resources 30: 2401-2423

Kunkel K, Andsager K (1999) Long-term trends in extreme precipitation events over the conterminous United States and Canada. American Meteorological Society 12: 2515-2527

Kusumastuti D, et al (2007) Threshold effects in catchment storm response and the occurrence and magnitude of flood events: implications for flood frequency. Hydrology and Earth System Sciences 11: 1515-1528

Kusumastuti D, et al (2008) Thresholds in the storm response of a catchment-lake system and the occurrence and magnitude of lake overflows: implications for flood frequency. Water Resources Research 44: 1-15

Lall U, et al (2006) Locally weighted polynomial regression: parameter choice and application to forecasts of the Great Salt Lake. Water Resources Research 42: 1-11

Lenters J (2004) Trends in Lake Superior water budget since 1948: a weakening seasonal cycle. Journal of Great Lakes Research 30: 20-40

Li X, et al (2007) Lake-level change and water balance analysis at Lake Qinghai, West China during recent decades. Water Resources Management 21: 1505-1516

Martins E, Stedinger J (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research 36: 737-744

McBean, E, Motiee, H (2008) Assessment of impact of climate change on water resources: a long term analysis of the Great Lakes of North America. Hydrology and Earth Systems Science 12: 239-255

McMahon G, et al (2003) Use of stage data to characterize hydrologic conditions in an urbanizing environment. Journal of the American Water Resources Association 39: 1529-1546

Mendoza M, et al (2006) Predicting water-surface fluctuation of continental lakes: a RS and GIS based approach in Central Mexico. Water Resources Management 20: 291-311

Meyer S (2005) Analysis of base flow trends in urban streams, Northeast Illinois, USA. Hydrogeology Journal 13: 871-885

Meyer S, Wilson S (2001) Impacts of urbanization on base flow and recharge rates, Northeast Illinois: summary of year 1 activities. Illinois State Water Survey

Montanari A, et al (1997) Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation, and simulation. Water Resources Research 33: 1035-1044

Morrison J, Smith J (2002) Stochastic modeling of flood peaks using the generalized extreme value distribution. Water Resources Research 38: 41-1 – 41-12

Nachabe M, et al (2005) Evapotranspiration of two vegetation covers in a shallow water table environment. Soil Science Society of America Journal 69: 492-499

Nadarajah S, Shiau J (2005) Analysis of extreme flood events for the Pachang River, Taiwan. Water Resources Management 19: 363-374

Privalsky V (1992) Statistical analysis and predictability of Lake Erie water level variations. Journal of Great Lakes Research 18: 236-243

Rose S, Peters N (2001) Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): comparative hydrological approach. Hydrological Processes 15: 1441-1457

Semenov V, Bengstsson L (2002) Secular trends in daily precipitation characteristics: greenhouse gas simulation with a coupled AOGCM. Climate Dynamics 19: 123-140

Sharda V, Das P (2005) Modeling weekly rainfall data for crop planning in a sub-humid climate of India. Agricultural Water Management 76: 120-138

Shumway R, Stoffer D (2006) Time series analysis and its applications. Springer Science Business Media, LLC New York

Smith J, Baeck M (2002) The regional hydrology of extreme floods in an urbanizing drainage basin. American Meteorological Society 3: 267-282

Struthers I and Sivaplan M (2007) A conceptual investigation of process control upon flood frequency: role of thresholds. Hydrology and Eath Systems Science 11 1405-1416

Watterson I, Dix M (2003) Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. Journal of Geophysical Research 108: ACL 3-1 – ACL 3-20

Wilks D (1990) Maximum likelihood estimation of the gamma distribution using data containing zeros. American Meteorological Society 3: 1495-1501

Yin X, Nicholson E (2002) Interpreting annual rainfall from the levels of Lake Victoria. Journal of Hydrometeorology 3: 406-416

Zhang X, Zwiers F, et al (2004) Monte Carlo experiments on the detection of trends in extreme Values. Journal of Climate 17: 1945-1952

Zolina O, Kapala A, et al (2004) Analysis of extreme precipitation over Europe from different reanalysis: a comparative assessment. Global and Planetary Change 44: 129-161

About the Author

Shayne Paynter received a Bachelor's Degree in Civil Engineering from Florida State University in 1991 and a Master's Degree in Civil Engineering from the University of South Florida in 2002. He currently lives with his wonderful wife, Sunitha, without whose support this research would not have been completed, and daughter, Inara.