Water Resour Manage (2011) 25:21–39
DOI 10.1007/s11269-010-9685-x

Statistical Changes of Lake Stages in Two Rapidly Urbanizing Watersheds

Shayne Paynter · Mahmood Nachabe · George Yanev

Received: 6 April 2009 / Accepted: 11 June 2010 /
Published online: 30 June 2010

Abstract Understanding changes induced by watershed urbanization is integral to developing an effective long-term management strategy. In this research, the authors study statistical changes of lake water surface levels in two urbanizing watersheds by evaluating serial change in time series parameters, autocorrelation, and variance as well as by developing a regression model to estimate weekly lake level fluctuations. The authors fit a seasonal integrated autoregressive moving average model to lake levels over subperiods of the data record to identify trends in parameter values. The authors fit the regression model with rainfall, lake stage, and temperature components for pre-urbanized and urbanized time periods to identify changes in baseflow. The lakes were located in Pasco County, Florida, USA and have not been significantly influenced by changes in rainfall patterns, pumping, surface water extraction or physical modification. Furthermore, the lakes exhibit consistent watershed urbanization and have sufficiently long and complete records. Based upon the research, the authors reach the following general conclusions about lakes in urbanizing watersheds: (1) the statistical structure of lake level time series is systematically altered, (2) in the absence of other forcing mechanisms, autocorrelation and baseflow 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.

S. Paynter (✉) · M. Nachabe
Department of Civil and Environmental Engineering,
University of South Florida, Tampa, FL 33620, USA
e-mail: pshayne21@gmail.com

G. Yanev
Department of Mathematics, The University of Texas-Pan American,
Edinburg, TX 78539, USA

Keywords Multiple regression · Lake · Time series · Autocorrelation · Urbanization · Basin · Watershed · SARIMA

1 Introduction

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, including lakes. While prolonged droughts or other changes in rainfall patterns present clear impacts to water resources, the impacts of urbanization are more ambiguous since the extent of the effect correlates to multiple stressors that may be linked to urbanization but are largely basin specific. These may include significant watershed or lake or stream alteration, water withdrawal, dredging, filling, diversion, installation of control structures, inflow from upstream water bodies, etc. Lakes are especially important as they provide drinking water, allow an alternate means of transportation and provide for commercial or recreational fishing and tourism among many other uses. Furthermore, lakes are important for environmental and ecological reasons; they serve to moderate climate, provide storage to regulate stream flow, recharge aquifers and provide wildlife habitat. According to estimates from the US National Oceanic and Atmospheric Administration, the Great Lakes alone supply drinking water for 40 million people and 10% of the US population lives within the Great Lakes watershed. As these and other lake watersheds become urbanized, development generally results in increased impervious area, more efficient stormwater collection systems and, hence, increased runoff and decreased infiltration and baseflow. Understanding the extent of these changes in lake dynamics in urbanizing watersheds is integral to effective future water resource management.

Watershed urbanization and the associated increase in impervious cover are expected to continue at a rapid pace throughout the US. Elvidge (2004) found that approximately 112,665 km^2 of impervious area has been created in the US as of 2000. Beach (2002) predicts that impervious cover may nearly double to 213,837 km^2 by 2025. Much of the literature regarding the impacts of this urbanization on water resources has focused on streams rather than lakes as stream data is often more available. Researchers have found the general hydrologic changes observed in urbanizing stream watersheds to include decreases in baseflow, decreases in recession times and increases in peak flow due to increased impervious area and rapid storm runoff from efficient collection systems (McMahon et al. 2003; Meyer and Wilson 2001; Rose and Peters 2001; Smith and Baeck 2002). 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. The techniques utilized to separate groundwater recharge (baseflow) from surface runoff in these studies are generally not applicable to lakes, since the smaller lake basins often do not meet the assumptions needed for stream baseflow separation techniques (Mau and Winter 1997). However, the results of the latter study indicate the potential for wetlands or other water-impoundment systems to influence groundwater recharge in lakes. This is consistent with Wang et al. (2010) who found that wetland restoration in urbanizing areas provided significant benefits in regards to flood control and groundwater recharge.

While the concept of urbanization altering the autocorrelation, a measure of how related two subsequent values are, of lakes or streams has not explicitly been researched, the aforementioned studies have addressed this phenomenon indirectly. Although streams have far less autocorrelation than lakes, decreased response time and increased flashiness point towards decreased autocorrelation or shortened memory. A major focus of the current research is to explicitly study changes in autocorrelation in urbanized lake basins. Given the lack of lake data relative to streams and the fact that in many lakes it is difficult to isolate the signal of urbanization due to multiple signal sources such as pumping or surface withdrawal, interconnected upstream lakes with differing degrees of urbanization or major alterations to the lake itself, many studies have invoked water budget models or coarse time scales for time series or regression modeling of lake levels. For this particular research, a finer time scale will be utilized than that of previous studies to investigate more directly the effects of urbanization on lake water levels. Furthermore, the basins selected for study have sufficient data and signals that can largely be isolated.

Nearly all of the literature regarding the impacts of urbanization on water resources that have utilized models has focused on changes in model output, i.e. stream flow, baseflow, recession times, etc. There is a dearth of studies focusing on systematic changes in the models themselves. Cheng et al. (2010) investigated the effect of urbanization on the parameters of a runoff hydrograph and found some systematic changes. In this research, the authors investigate changes in the statistical structure of lake stages in urbanizing watersheds by evaluating serial changes in time series parameters, autocorrelation and variance as well as by developing a regression model to estimate weekly lake level fluctuations. The authors fit a seasonal integrated autoregressive moving average (SARIMA) model to lake level subperiods of approximately equal length over the data record and identified trends in parameter values. The authors fit the regression model with rainfall, lake stage, and temperature components to weekly fluctuations in water surface levels. These analyses were performed for two lakes in Pasco County, Florida, USA, that demonstrate consistent urbanization but have not been substantially anthropogenically altered to minimize the amount of conflicting signals.

The objectives of this research are to: (1) determine changes in the structure of lake level time series, (2) analyze serial changes in lake level autocorrelation and variance, (3) identify changes in the runoff/baseflow relationship. The expected outcome of this study is the development of general lake stage trends corresponding to urbanization that water resource managers can use for effective planning.

2 Materials and Methods

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 (2009) found that west-central Florida rainfall patterns are spatially homogenous and temporally stationary and no significant shifts that would correlate to trends in lake levels were evidenced. For the present study, the authors selected two lakes

Table 1 Lake characteristics summary

Lake	Basin area (km^2)	Lake area (km^2)	Basin/lake ratio	Adjacent wetland area[a] (km^2)	Adjacent wetland percent of basin area
Moon	0.78	0.43	1.8	0.07	9.0
Cow	1.55	0.40	3.9	0.00	0.0

[a] Area defined as contiguous with the lake surface and within 0.5 km of the lake

undergoing rapid urbanization within the west-central Florida region based on the availability of sufficient data, lack of direct lake withdrawals and lack of proximity to wellfields. These lakes have been relatively unaltered other than urbanization of the basin. Because the lakes are located within a small geographic area, they exhibit similar geologic characteristics, including being situated in a silty-sandy environment above a limestone formation. Table 1 summarizes the lake characteristics. Lake stage data available to the general public were obtained from the US Geological Survey (USGS) and the Southwest Florida Water Management District (SWFWMD). The data are daily but as there are several small gaps of a week or more, the time series of each lake was converted to weekly increments by adding 7 days consecutively to the first day of the data set and selecting the lake level that corresponds to the weekly date (Table 2).

Data from the US census for 1980, 1990 and 2000 at the census tract level were utilized to determine the degree of watershed urbanization. Census data was used rather than developing land-use type changes from historic aerials or land use GIS data as the data were more readily available for the time periods of interest. Land-use GIS data from SWFWMD with sufficient detail to estimate impervious area was found for the period from 1995 through 2010. Pasco County was largely rural through the 1970s and urbanization in cities and around lakes began to accelerate during the 1970s through the present time. Land use data would be more readily available in areas with longer histories of urbanization, however lakes in these areas tend to be greatly impacted and not as useful for isolating the effects of urbanization. Population data for both Moon Lake and Cow Lake is available closer to the onset of basin urbanization. Population data has been shown to be a good indicator of impervious cover (Bird et al. 2002; Hicks and Woods 2000; Stankowski 1972), generally has small errors when compared to actual impervious area and has been found to be more accurate than estimating impervious area from satellite imagery. Population data prior to 1980 are only available at the county level for Pasco County. As the census tracts were substantially larger than the lake basins studied, population

Table 2 Lake stage data summary

Lake	Period of record	Average (m) (NGVD)	Maximum (m) (NGVD)	Minimum (m) (NGVD)	Percent complete[a]	Standard deviation (m)	Variance (m)
Moon	1965–2007	11.63	12.58	10.24	96.0	0.47	0.22
Cow	1976–2007	23.56	24.11	23.03	82.3	0.17	0.29

[a] Percent complete of the available data; excludes years with too few data points to use for analysis
NGVD US National Geodetic Vertical Datum

was distributed evenly across the tract and proportioned to each basin. Population density, which has a more meaningful implication for estimating basin urbanization, was developed from these basin-specific population values. For each particular lake, rainfall data from nearby gages maintained by the USGS and SWFWMD were

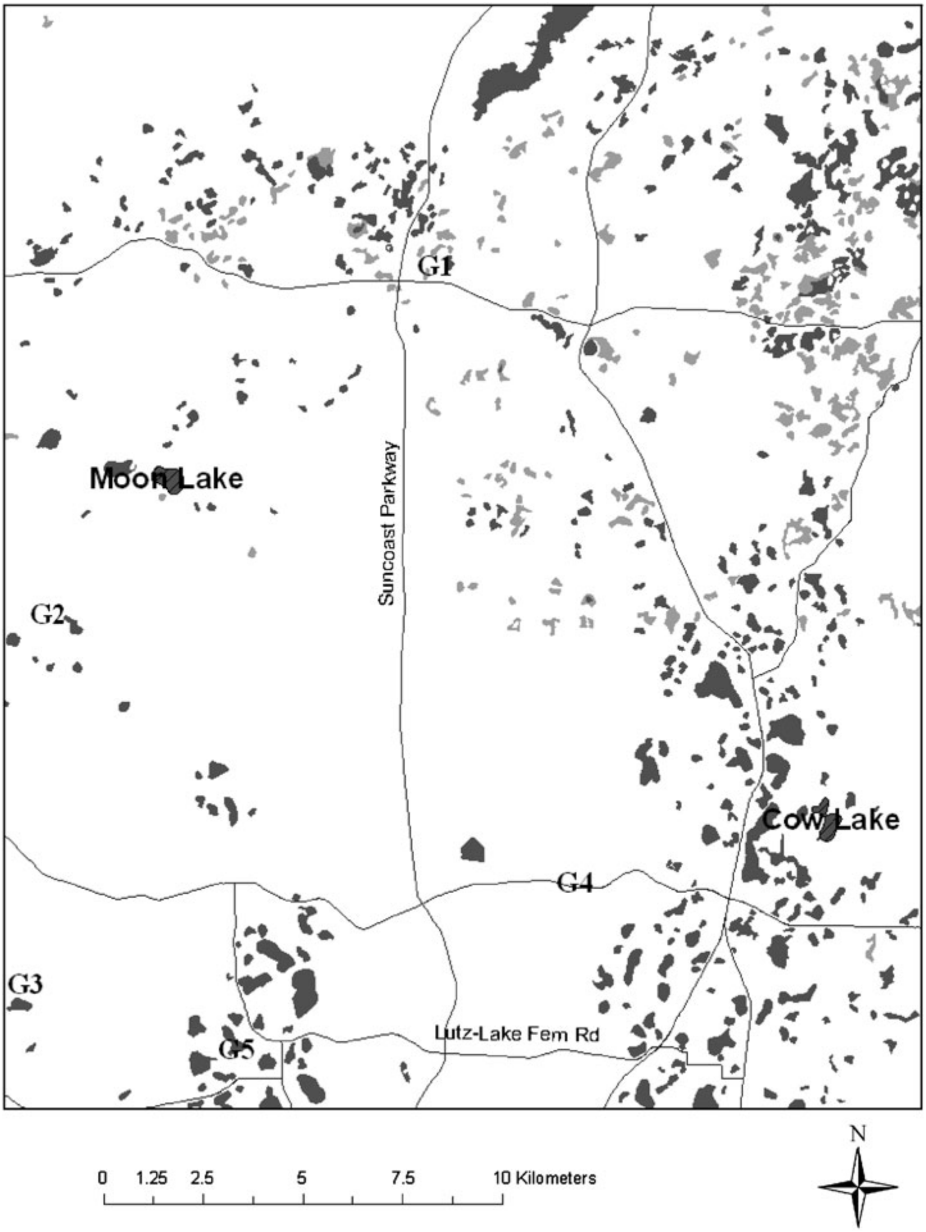

Fig. 1 Lake and rainfall gage locations

utilized. Figure 1 provides an aerial view of the lakes and available rainfall gages. Figures 2 and 3 depict aerial views of Moon Lake and Cow Lake.

Based upon the GIS census data, population in the vicinity of each lake has substantially increased over the time periods studied. As the population grows,

Fig. 2 Moon Lake

Fig. 3 Cow Lake

the land use in the watershed changes from rural to residential development with significant increases in impervious area, channelized drainage and infill due to the raising of lots for home construction. The population density growth around each lake is summarized in Table 3. The Moon Lake watershed exhibits the greatest overall gains with well over 100% density growth from both 1980 to 1990 and 1990 to

Table 3 Lake watershed population growth

Lake	1980 density (pop./km^2)	1990 density (pop./km^2)	2000 density (pop./km^2)	Percent density growth 1980–1990	Percent density growth 1990–2000
Moon	19.9	92.7	238.4	366.2	157.2
Cow	181.1	367.9	645.3	103.2	75.4

2000. Cow Lake population density was considerably higher than that of Moon Lake, however the growth in density was slower.

2.2 Time Series Analysis

The authors first explored changes in lake levels with time series analysis, a methodology for extracting statistics from sequential data points. 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 cycles, such as those induced by the approximately 10-year El Nino phenomenon, may also be present in hydrologic time series. Time series analysis assumes stationarity; any systematic changes in the mean (trend) and any periodic variations (seasonality) were removed through differencing by subtracting adjacent values or values from one seasonal period prior (Chatfield 2004).

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 systematic parameter changes. Time series analysis also requires constant or nearly constant variance. Some of the time series subunits were shifted slightly by 1 or 2 years to achieve homoscedasticity in subseries that exhibited heteroscedasticity, indicating statistically significant differences in variance.

Based on the literature (Altunkaynak 2007; Irvine and Eberhardt 1992; Yin and Nicholson 2002), autoregressive moving average models (ARMA) are often found to fit lake data well. An ARMA model utilizes the inherent memory of the data (autoregressive) as well as values of error terms (moving average) to determine subsequent data values. Although there are more robust methods for forecasting lake levels, the focus of this research is on changes in the structure of lake stages, 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 with mean zero is given by:

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

where x_t is the current value of the lake level, x_{t-p} is the time series value at lag p, ϕ_i, $i = 1, 2 \ldots p$, are unique constant parameters for each lagged value and w_t is Gaussian white noise with mean zero. The order of an autoregressive model is the number of previous data points upon which the estimate of the current data point is regressed. A moving average process is based completely on previous values of white noise. A moving average process of order q with mean zero is given by:

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

where θ_i, i = 1, 2...q, are unique constant parameters for each white-noise value. The order of a moving average model is the number of previous values of white noise the current data point is regressed upon. An ARMA process combines Eqs. 1 and 2 and is said to be of order (p, q). An ARMA model with simple differencing yields an integrated autoregressive moving average (ARIMA) model of order (p, d, q), where d is the number of times the series needs to be differenced to achieve stationarity. Once the series is stationary, ARMA parameters were determined using maximum likelihood estimation. Maximum likelihood estimation is a distribution based method to determine the set of parameters that maximizes the likelihood of the sample data. The likelihood of the model is given by:

$$L\left(\beta, \sigma_w^2\right) = \left(2\pi\sigma_w^2\right)^{-n/2} \left[r_1^0(\beta) r_2^1(\beta) ....r_n^{n-1}(\beta)\right]^{-1/2} \exp\left[-\frac{S(\beta)}{2\sigma_w^2}\right] \tag{3}$$

where

$$S(\beta) = \sum_{t=1}^{n} \left[\frac{\left(x_t - x_t^{t-1}(\beta)\right)^2}{r_t^{t-1}(\beta)}\right] \tag{4}$$

and β is the vector of model parameters $\phi_1.....\phi_p$, $\theta_1.....\theta_q$, σ_w^2 is the variance of the white noise and $r(\beta) = r_1^0(\beta), r_2^1(\beta), \ldots\ldots\ldots r_n^{n-1}(\beta)$ is the mean squared error vector of the one step ahead prediction, $x_t - x_t^{t-1}$. Parameter estimates were obtained by maximizing Eq. 3 with respect to β and σ_w^2 (Shumway and Stoffer 2006). A seasonal component was added to the ARIMA model to adequately capture periodic fluctuations. A seasonal ARIMA (SARIMA) is given by:

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

where

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

and

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

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 rather than the immediately preceding value. SARIMA models are noted as ARIMA (p, d, q) × (P, D, Q) where D is the number of seasonal differences. Once a model was fitted to the data, the authors performed diagnostics to ensure randomness of the residuals, including inspecting the autocorrelations of the residuals, $r_e^2(h)$, where h is the lag, and utilizing the Ljung–Box statistic, given by:

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

where n is the sample size and H is arbitrarily chosen, typically near lag 20 (Shumway and Stoffer 2006). To achieve parsimony, the authors sought the simplest model that adequately fit the data for all time series subsets for each lake and parameters were compared to identify trends. In ranking the models, the authors utilized adjusted R^2

and the Bayesian information criterion (BIC). 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. 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.

2.3 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:

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

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. 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. 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:

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

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, the authors employed an F test for significantly different variances.

2.4 Regression

Any changes in lake level statistical signatures such as a consistent change in parameters found with time series modeling or autocorrelation analysis were further explored with regression. Regression has been used to characterize lake levels throughout the literature (Gibson et al. 2006; Lall et al. 2006; McBean and Motiee 2008; Mendoza et al. 2006). Instead of focusing on the water level itself, the authors modeled the differences between weekly stages. The time scale of overland flow to lakes is in hours while that of baseflow recharge to lakes is in weeks. For the regression model, the authors regressed the difference (Y_t) of the current week's water level from that of the previous week 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 was used to capture evaporation as pan data was not available. The starting water level captures the effects of lake morphology. In general, lakes exhibit increasing area from the bottom to the top. Subsequently, for each unit increase in stage, larger volumes of runoff and baseflow are required. This variable has an inverse relationship with lake stages since the higher the initial stage, the less impact a given volume of rainfall ultimately has on fluctuations. The regression equation is given by:

$$Y_t = \beta_0 + \beta_1 R_w + \beta_2 R_m + \beta_3 T + \beta_4 W + \varepsilon_t \tag{5}$$

where ε_t is the random error term. In most cases, a longer period of lake level data was available than rainfall data and lake level records had to be truncated for the regression analysis. In order to assure parsimony, the authors utilized adjusted R^2 and Akaike's information criterion (AIC) on the entire data set to verify each of the four variables substantially contributes to explaining water stage fluctuation. AIC takes into account the number of parameters required to achieve a particular fit. The model selected for each lake consisted of the four variables in Eq. 5 or a subset thereof. The authors calculated the correlation matrix for each model to ensure there was no substantial correlation between regressors. Cook's distance was utilized to measure the influence of specific data points and identify outliers. As with the time series and autocorrelation analysis, sequential subunits of time were analyzed to identify any systematic changes in parameters. However, since population in the region began to substantially increase in the 1970s, 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 4 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.

3 Results and Discussion

Table 2 provides a summary of the lake stage data. The difference between the maximum and minimum for Moon and Cow Lakes is 2.34 and 1.08 m, respectively. Given the very flat topography of west-central Florida, relatively small differences of less than 0.3 m in water level fluctuations can inundate large areas and houses are routinely set as low as 0.3 m above expected high water marks. Moon Lake exhibited a larger standard deviation while variance at the two lakes was fairly consistent.

3.1 Time Series Modeling

Plots of the lake level time series (Fig. 4) do not display any obvious trends. However, in cases where the autocorrelation of the raw data slowly decays to insignificance, indicating a potential trend, the authors utilized differencing; a single simple difference in which adjacent values are subtracted from each other sufficiently removed any trend at each of the lake time series. Differencing essentially centers the time series around a constant mean so time series modeling or forecasting techniques can be utilized. The model order for each lake as well as a summary of the parameters for each model can be found in Table 4. Figure 5 indicates that errors are random for the Cow Lake 1976–1980 model since none of residuals are significant and there are no clear patterns in the residuals. Similar results were achieved for other time series periods as well as Moon Lake. In a single subseries case for both lakes,

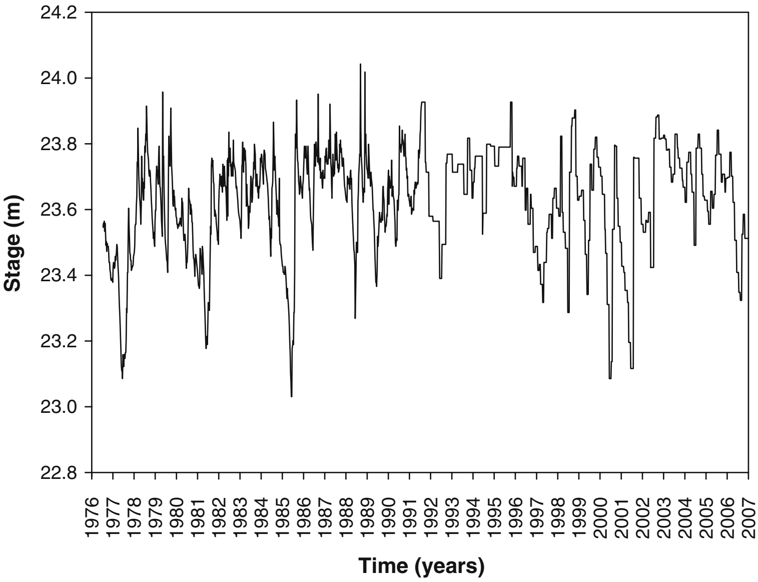

Fig. 4 Cow Lake stages (1976–2007)

Ljung–Box values were slightly outside 95% 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. Both lakes required a single autoregressive term, a single differencing and a single seasonal term. Moon Lake required two moving average terms. As shown in Table 4, Moon Lake demonstrates a consistent increase in the first autoregressive parameter and the first moving average parameter and no consistent change in the seasonal term; Cow Lake demonstrates a consistent decrease in both the autoregressive and seasonal autoregressive parameters. An increase in the autocorrelation parameters should point towards increasing lake memory while a decrease indicates the opposite.

Table 4 SARIMA model parameters

Lake/model order	Subseries period	Parameters φ_1	φ_2	θ_1	θ_2	Φ_1	Ljung–Box
Moon $(1,1,2) \times (1,0,0)$	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
Cow $(1,1,0) \times (1,0,0)$	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

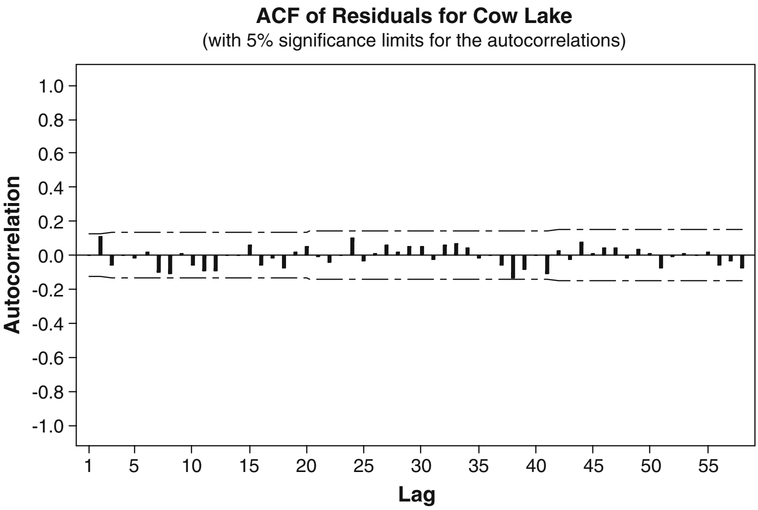

Fig. 5 Autocorrelation of residuals for Cow Lake (1976–1980 with lag in weeks)

3.2 Autocorrelation and Variance

Whereas the time series modeling included terms for moving average and seasonal autoregressive terms, the autocorrelation analysis focuses on lake memory alone. The results of the autocorrelation analysis are found in Table 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 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, the time it takes runoff to reach the watershed outfall from the most hydrologically remote part of the basin. This intuitively would lead to a larger fraction of runoff from a given rainfall event reaching a lake more quickly and reducing the fraction that contributes to the much slower process of groundwater recharge to the lake. This change would

Table 5 Autocorrelation and variance

Lake	Subseries period	Exponential parameter	Variance	Significantly different[a]
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
Cow	1976–1980	0.077	0.03	
	1981–1985	0.107	0.03	N
	1986–1991	0.219	0.01	Y

[a] Indicates if the current time period is significantly different from the previous at the 95% confidence level

conceivably result in higher peak stages due to the larger fraction of runoff and lower low stages as 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. The physical implication of less autocorrelation in a lake is that the lake is beginning to behave more like a stream in that the lake stage reacts quickly to a precipitation event, both rising and falling more rapidly with less memory of the rainfall event left behind in the form of infiltration and recharge. This is the case for Cow Lake, which demonstrates a consistently shorter memory, as shown in Fig. 6.

However, Moon Lake shows a consistent trend towards longer memory. The major difference between the two lakes is the presence of wetlands abutting Moon Lake. As a larger fraction of runoff associated with increasing urbanization flows into these adjacent wetlands, the average wetland water levels may increase and gradually release 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 both Meyer (2005) and Wang et al. (2010) 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 or restored wetland areas demonstrated increases in baseflow. Another difference between the two lakes is that Moon Lake discharges directly into wetlands. Higher antecedent tailwaters due to increased overall runoff volume into the downstream wetland storage areas may also contribute to increases in lake memory. The results of the autocorrelation analysis are consistent with the time series modeling; similar patterns of autoregressive parameter increase or decrease through time are observed.

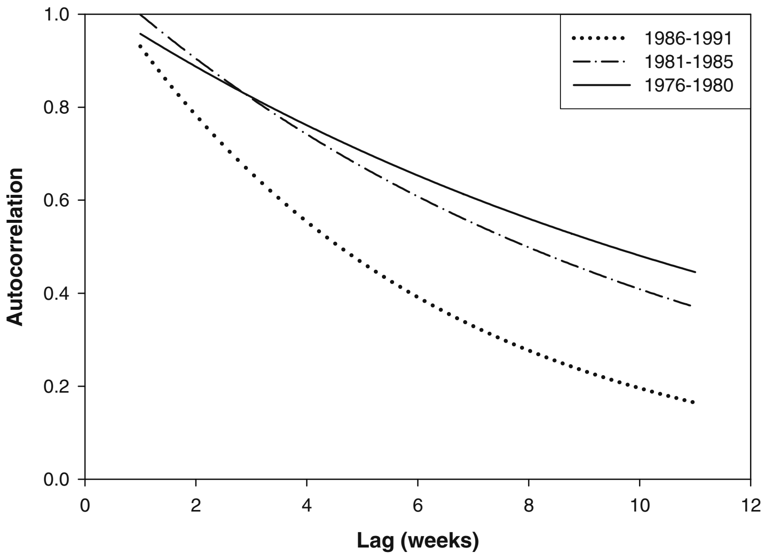

Fig. 6 Cow Lake autocorrelation with exponential fit lines (1976–1991)

3.3 Regression

Table 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 Akaike Information Criterion (AIC) for the overall model which includes the entire data set. There was a large spread in the multiple R^2 values for the subperiod models, from 36.3 to 68.6. 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 such as:

- 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% complete for the time periods analyzed, however, none of the lakes had gages located immediately at the lakes themselves but were within 2 to 6 km of the lakes analyzed. This is well within an acceptable range for this analysis (Paynter and Nachabe 2009) but there will be slight variability between gage site and lake rainfall amounts.
- 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), transpiration data for the lakes studied were not available.
- Although the lakes are located in a geologically similar region and generally have silty and sandy soils, local differences in soil types, including those associated with wetlands, can have an influence on the rainfall–baseflow interaction.

Given these factors, the R^2 values indicate that the models explain a significant portion of lake level variability. For both lakes, the regression model inclusive of all four independent variables was deemed most appropriate.

Figure 7 demonstrates the fit of the modeled lake level fluctuations based on all four variables versus the actual fluctuations for Cow Lake. If the actual values match the modeled values exactly, they would fall on the 45° line. The clustering around this line indicates a good fit. Figure 8 demonstrates the distribution of the residuals for Cow Lake. Data points atop or close to the dotted match line are distributed normally while points further away exhibit less normality. Residuals for all regression models exhibited normality and homoscedasticity with some deviation

Table 6 Regression model parameters

Lake	Subseries period	Parameters[a]				R^2	AIC
		Pre-week rain	Pre-month rain	Starting lake stage	Temp.		
	All values					44.16	2.92
Moon	1973–1976	0.0088	0.0009	N/A	−0.0005	48.8	
	1977–2007	0.0084	0.0012	−0.0092	−0.0009	45.7	
	All Values					44.2	1.34
Cow	1976–1979	0.0129	0.0010	−0.0410	−0.0005	68.6	
	1980–1991	0.0097	N/A	−0.0794	N/A	36.3	

[a]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

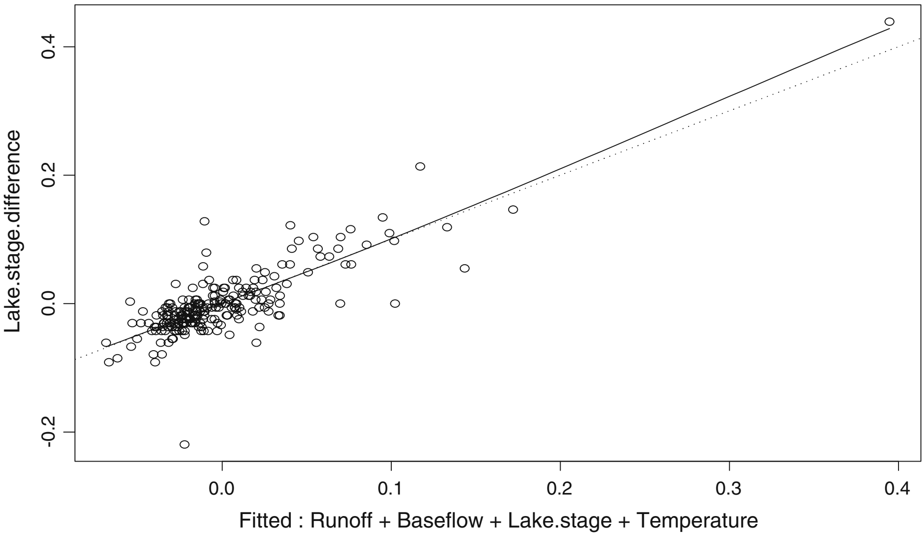

Fig. 7 Cow Lake response versus fit (1976–1980)

from normality at the extremes. This deviation is likely due to events such as hurricanes or extreme drought that are not distributed as normal rainfall events (Paynter and Nachabe 2010) All values for the correlation coefficients of the regressors at each lake were equal to or less than 0.4, indicating no significant correlation and that the regressors are sufficiently independent of each other. Both lakes exhibited Cook's distances of less than 0.5, indicating no presence of outliers that may skew the estimates of the regression coefficients.

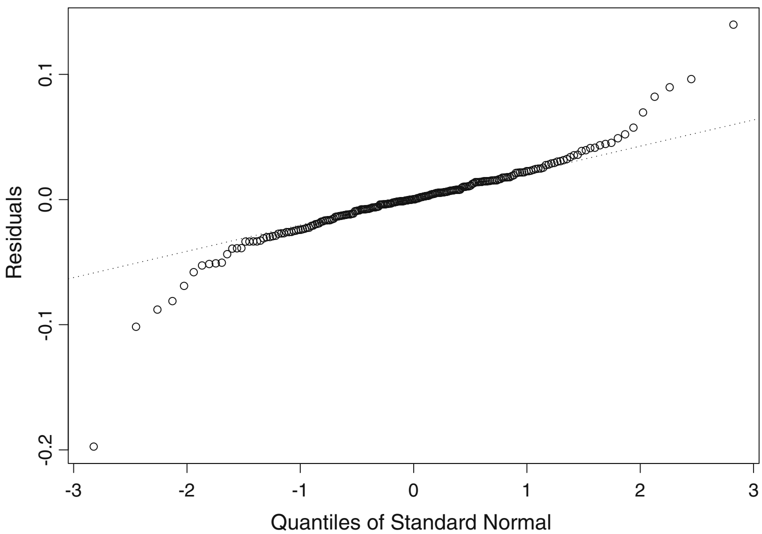

Fig. 8 Cow Lake normal quantile–quantile plot of the residuals (1976–1980)

The previous-week rainfall variable is highly significant for all models. Cow Lake demonstrates 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. However, given that runoff has a time scale of hours and baseflow of days to weeks, it is likely that the majority of this variable consists of baseflow. It is expected that it would decrease for Cow Lake given the trends in the regression baseflow parameter and the time series and autocorrelation analysis. The Moon Lake parameter decrease is somewhat at odds with the regression baseflow parameter trend and the time series and autocorrelation analysis, however it may be that in the timescale of a week, adjacent wetlands have not fully begun to pass flow back to the lake and a reduction in baseflow due to urbanization is occurring in the short-term. Daily data would likely be required to reach definitive conclusions as to urbanization-induced changes in this parameter. Cow Lake demonstrated a trend towards decreased baseflow based upon the pre-month rainfall parameter; baseflow is significant in the pre-urbanized period but not thereafter. Moon Lake exhibited increases in baseflow; the baseflow parameter is 0.0009 in the first time period and 0.0012 thereafter, a 33% increase. As previously noted, Moon Lake has a large 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. The baseflow trends were consistent with the autocorrelation and time series analysis; increasing baseflow correlated to larger autoregressive terms or longer autocorrelation while decreasing baseflow correlated to smaller autoregressive terms or shorter autocorrelation.

Starting lake levels were significant for all but one time period for which lake levels were generally higher. This is consistent with the morphology of these 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 all but one case as well, most likely due to factors such as wind or humidity exerting a greater relative impact on evaporation. Cow Lake provides the best case to examine any effects watershed urbanization may incur as it does not have adjacent wetlands or discharge to a wetland and there is a high degree of dense urbanization around the lake.

4 Conclusions

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. With regard to the time series modeling, a significant increase in basin population density appears to systematically alter the time series signature. 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 an adjacent wetland, it was not true in the Moon Lake watershed. 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 gradually releasing it back to the lake over time. At Moon Lake, variance increased with time while at Cow Lake it was nearly constant. For the regression analysis, Cow Lake demonstrated a decrease in baseflow contribution while Moon Lake demonstrated the opposite, consistent with the time series and autocorrelation analysis. 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, (2) In the absence of other forcing mechanisms, autocorrelation and baseflow 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.

References

Altunkaynak A (2007) Forecasting surface water level fluctuations of Lake Van by artificial neural networks. Water Resour Manag 21:399–408
Beach D (2002) Coastal sprawl. The effects of urban design on aquatic ecosystems in the United States. Pew Oceans Commission, Arlington
Bird S, Harrsion J, Exum L, Alberty S, Perkins C (2002) Screening to prevent urban storm water problems: estimating impervious area accurately and inexpensively. In: Proceedings for national water quality monitoring conference
Chatfield C (2004) The analysis of time series. CRC, Boca Raton
Cheng S, Lee C, Lee J (2010) Effects of urbanization on model parameters. Water Resour Manag 24:775–794
Elvidge C (2004) U.S. constructed area approaches the size of Ohio. EOS (Washington, DC) 85(24):233–240
Gibson J, Prowse T, Peters D (2006) Hydroclimatic controls on water balance and water level variability in Great Slave Lake. Hydrol Process 20:4155–4172
Hicks R, Woods S (2000) Pollutant load, population growth and land use. Progress: Water Environment Research Foundation 11:10
Irvine K, Eberhardt A (1992) Multiplicative seasonal ARIMA models for Lake Erie and Lake Ontario water levels. J Am Water Resour Assoc 28:385–396
Lall U, Moon Y, Kwon H, Bosworth K (2006) Locally weighted polynomial regression: parameter choice and application to forecasts of the Great Salt Lake. Water Resour Res 42:1–11
Mau D, Winter T (1997) Estimating ground-water recharge from streamflow hydrographs for a small mountain watershed in a temperate humid climate, New Hampshire, USA. Ground Water 35:297–304
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. Hydrol Earth Syst Sci 12:239–255
McMahon G, Bales J, Coles J, Giddings E, Zappia H (2003) Use of stage data to characterize hydrologic conditions in an urbanizing environment. J Am Water Resour Assoc 39:1529–1546
Mendoza M, Bocco G, Bravo M, Granados E, Osterkamp W (2006) Predicting water-surface fluctuation of continental lakes: a RS and GIS based approach in Central Mexico. Water Resour Manag 20:291–311
Meyer S (2005) Analysis of base flow trends in urban streams, Northeast Illinois, USA. Hydrogeol J 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
Nachabe M, Shah N, Ross M, Vomacka J (2005) Evapotranspiration of two vegetation covers in a shallow water table environment. Soil Sci Soc Am J 69:492–499
Paynter S, Nachabe M (2009) Regional scale spatio-temporal consistency of precipitation variables related to water resource management and planning. Meteorol Appl 16:413–423
Paynter S, Nachabe M (2010) Use of generalized extreme value covariates to improve estimation of trends and return frequencies for lake levels. J Hydroinform. doi:10.2166/hydro.2010.077

Rose S, Peters N (2001) Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): comparative hydrological approach. Hydrol Process 15:1441–1457
Shumway R, Stoffer D (2006) Time series analysis and its applications. Springer, New York
Smith J, Baeck M (2002) The regional hydrology of extreme floods in an urbanizing drainage basin. Am Meteorol Soc 3:267–282
Stankowski S (1972) Population density as an indirect indicator of urban and suburban land-surface modifications. Geol Surv Res 800B:219–224
Wang M, Qin D, Lu C, Li Y (2010) Modeling anthropogenic impacts and hydrological processes on a wetland in China. Water Resour Manage. doi:10.1007/s11269-010-9577-0
Yin X, Nicholson E (2002) Interpreting annual rainfall from the levels of Lake Victoria. J Hydrometeorol 3:406–416