Journal of Great Lakes Research 36 (2010) 172–178

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

Lake Superior water level fluctuation and climatic factors: A dynamic linear model analysis

E.C. Lamon [a], C.A. Stow [b,*]

[a] Statistical Ecology Associates LLC, 578 Pecan St., Canyon Lake, TX 78133, USA
[b] NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, MI 48108, USA

ARTICLE INFO

Article history:
Received 27 July 2009
Accepted 30 October 2009

Communicated by William M. Shertzer

Index words:
Lake Superior
Water levels
Bayesian
Dynamic linear models

ABSTRACT

We use Dynamic Linear Models (DLM) to analyze the time series of annual average Lake Superior water levels from 1860 to 2007, as well as annual averages of climate drivers including precipitation (1900–2007), evaporation and net precipitation (1951–2007). Our results indicate strong evidence favoring the presence of a systematic trend over a random walk for Lake Superior water levels, and this trend has been negative in recent decades. We then show decisive evidence, in terms of improved predictive performance, favoring a model in which the trend component is replaced with regression components consisting of climatic drivers as predictor variables. Because these models use lagged values of precipitation or net precipitation as predictors, the models can be used to forecast water levels, with the associated uncertainty, several years into the future. We use several of the best fit models and compare one (2008) and two step-ahead (2009) forecasts. The 2008 forecasts compare very well with the observed 2008 water level; the two step-ahead 2009 forecasts are offered as testable hypotheses. The Bayesian context in which these models are developed provides a rigorous framework for data assimilation and regular model updating.

Published by Elsevier B.V.

Introduction

The Laurentian Great Lakes provide numerous ecosystem services that confer benefits throughout Canada and the United States (Costanza et al., 1997). Many of these services are influenced by fluctuating water levels. Water level extremes cause particular problems; high water in the 1980s resulted in coastal erosion and damage to structures near the shoreline; more recently low water has impeded shipping, water extraction, and recreational activities (Sellinger, 2008). Though current low levels may seem dire, it has long been recognized that water levels have, at times, been much lower than the present norm (Wilson, 1931). A recent analysis revealed a period when water levels were so low that the lakes became disconnected (Lewis et al., 2008). Because water levels have been documented to vary greatly, it is important for communities reliant on the services provided by these lakes to anticipate, and adapt to, future water level changes.

Great Lakes water levels fluctuate on several characteristic frequencies possibly related to recurrent, large-scale teleconnection patterns (Cohn and Robinson, 1976; Ghanbari and Bravo, 2008; Hanrahan et al., 2009; Wiles et al., 2009). As early as 1874 a positive relationship with sunspots was identified (Dawson, 1874), and this association was periodically debated until the 1940s (Hubbard, 1887; Nassau and Koski, 1933; Brunt, 1937; Wilson, 1946), when the discussion faltered. More recently Sellinger et al. (2008) noted a sunspot relationship in Lakes Michigan and Huron, but observed that the relationship flipped from positive to negative in ~1940. Such reversals have been documented elsewhere (Lawrence, 2002); this reversal may explain the apparent lapse of literature discussion regarding sunspots and Great Lakes water levels from the 1940s until 2008.

The highs and lows of cyclic fluctuations may impart temporary stress to human populations as well as resident ecological communities; however native populations that have evolved over thousands of years have become resilient to these regular changes and, in principle, humans can identify cycles and plan for them (Stager et al., 2007). Sustained trends present a different challenge, possibly requiring continuous adaptation. Additionally, trends can be hard to differentiate from periodic variability, leaving room for debate as to whether resources should be allocated for adaptation when a trend is suspected.

Following the high levels of the 1980s Great Lakes water levels experienced an overall decline that was particularly pronounced from 1997 to 2000 (Assel et al., 2004). This decline was apparent even in Lake Superior, although water levels have been regulated by structures in the St. Marys River since the 1920s. The lower levels attained since then have generally been sustained; however, Lake Superior has rebounded from near-record low levels experienced in late 2007 (Holden, 2007). Ongoing low levels are worrisome because lower Great Lakes water levels are consistent with most climate

* Corresponding author.
E-mail addresses: eclamon@gmail.com (E.C. Lamon), craig.stow@noaa.gov (C.A. Stow).

0380-1330/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.jglr.2009.11.009

change forecasts (Magnusen et al., 1997, Lofgren et al., 2002) raising concern that recent declines are early indications of a long-term pattern. The fact that this decline is mirrored in nearby seepage lakes, suggests a common regional driver, with climate the most logical perpetrator (Stow et al., 2008). Although Redway (1924) argued that lake water level fluctuation is an unreliable indicator of “progressive” climate change, more recently Williamson et al. (2009) proposed lakes as ideal climate sentinels because they integrate and reveal climatic signals occurring throughout their watersheds. The Laurentian Great Lakes may be particularly informative in this regard because their large watersheds and long hydraulic retention times integrate spatially and temporally thus damping small-scale noise and making large-scale climatic patterns more apparent.

To explore the climatic signals that are reflected in Lake Superior water level fluctuations and evaluate if climatic data can be used to make useful near-term water level forecasts, we investigated annual average water level data, available from 1860 to 2007, in its entirety (Model 1), and then the relationship between annual average water level and potential climatic predictor variables (Model 2) which are only available for shorter periods (precipitation from 1900 to 2007, evaporation from 1948 to 2007, and net precipitation from 1948 to 2007). Because we evaluated up to a 3-year lag in predictor variable relationships, Model 2 spans the period 1951–2007. Model 1 provides a basis to differentiate random from progressive patterns while Model 2 explores whether the relationship with climatic drivers may better explain the variability in Lake Superior water level than the simple trends evaluated for Model 1, albeit for a more limited time span. Additionally, to facilitate discussion, we analyzed trends in the climatic predictors over the periods of record for which they are available applying the same approach used to develop Model 1.

Although there are many approaches available to analyze time series data, researchers commonly use linear regression models to estimate time trends. With modern software linear regression is extremely easy and in many applications a linear approximation is adequate for the intended inference. However, time series data often deviate from the assumptions that support inference using linear regression (such as linearity and homoscedastic, uncorrelated residuals) in which case summary statistics including slope coefficients, confidence intervals, and p-values may be misleading. Ultimately, this situation can result in a poorly founded decision of “statistical significance.” As an alternative, we analyzed Lake Superior annual average water levels using Bayesian Analysis of Time Series (BATS) software to estimate Dynamic Linear Models (DLMs) (Pole et al., 1994). DLMs are similar in concept to linear regression but allow model parameters to evolve, systematically, with time, capturing the nonlinearity in the series. Additionally, the Bayesian framework provides a probabilistic uncertainty estimate for each model parameter, which is not automatically conflated with a decision of “significance” or “non-significance.” Instead, parameter estimates and their associated uncertainty can be evaluated and decisions that are appropriate for the problem under consideration can be made (Zhang and Arhonditsis, 2008).

Methods

Data

Water level data for Lake Superior (1860–2007) are available from the National Oceanic and Atmospheric Administration's (NOAA) National Ocean Services database (NOAA, 2007). NOAA's Geodetic Survey computes an International Great Lakes Datum (IGLD) every 25 years to account for iso-static rebound; presently, all the Great Lakes are referenced to IGLD85. Monthly precipitation (1900–2007) and evaporation (1948–2007) data were obtained from NOAA's Great Lakes Environmental Research Laboratory's Hydrologic Database (Croley and Hunter, 1994). Precipitation data were synthesized using a Thiessen weighting approach to obtain a value for the watershed.

Dynamic linear models (DLMS)

DLMs partition variation in the response variable (water level) into three components: a trend, regression coefficients that describe the relationship with predictor variables, and a random component (West and Harrison, 1989; Lamon et al., 1998). DLMs are similar to linear regression models; however, DLMs allow model coefficients to change, systematically, with time whereas linear regression models are based on the assumption that model coefficients are static. With DLMs, information from earlier time periods is discounted by adding uncertainty as time progresses, based on the idea that newer information is more relevant than older information to predict current or near-future conditions. The discount factor is δ which is equal to $1+\lambda$, where λ is the discount rate. For a discount factor δ between 0 and 1, the information loss for each time interval is $V_t = \delta^{-1} V_{t-1}$. With $\delta = 1$ the relationship becomes static with time like a linear regression model; for a 5% information loss with each time increment, δ is about 0.95 (Pole et al., 1994). As δ approaches zero there is effectively no memory from one time step to the next and the method is similar to “connecting the dots.” Useful discounts are typically >0.8; smaller discounts lead to models that make predictions based on only the two or three most recent observations (West and Harrison, 1989).

DLMs can be used for both “online” forecasting (Lamon et al., 1998), in which successive forecasts are based only on the data preceding each forecast time period, as well as for “retrospective” analysis (Lamon et al., 1999), which is based on all the data in the time series. In contrast, many methods used to examine time series data, such as linear regression, are strictly retrospective.

In the online analysis using DLMs, at any time t, posterior information regarding the model parameters (θ) is obtained by combining prior information with information in the current observation (the likelihood) via Bayes' theorem:

$$p(\theta_t | D_{t-1}, y_t) = \frac{p(Y_t = y_t | \theta_t) p(\theta_t | D_{t-1})}{p(Y_t = y_t)} \quad (1)$$

where D_{t-1} denotes the state of knowledge at time t-1, the first term in the numerator is the likelihood, the second is the prior distribution (commonly referred to as the prior), and θ_t is the state (parameter) vector. The denominator is a normalizing constant that can be dropped, allowing the equality from eq. (1) to be re-expressed as a proportionality:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}. \quad (2)$$

Retrospective analyses using DLMs, alternatively referred to as filtering or smoothing, are useful to ask: “where has the process under study been?” or “what happened?” (West and Harrison, 1989; Pole et al., 1994). In retrospective analysis, we modify the posterior distributions from the online analysis (eq. 1), considering later information to obtain filtered distributions $p(\theta_t|D_{t+k})$, where k is some positive constant. This notation indicates that information arising after time t informs our knowledge about θ_t.

Evaluating how well alternative models capture data dynamics requires quantitative criteria for model assessment. Several criteria are available for model comparison and selection; many have a similar basis, differing mainly in the degree to which model complexity is penalized. We calculate the Akaike Information Criterion (AIC, Akaike, 1973) and the Bayesian Information Criterion (BIC, Schwarz, 1978) to aid in selection between competing models and between discount factors for a given model structure. Because AIC and BIC are deviance based measures (a generalization of the variance or sum of squares), models with lower values fit better than those with higher values. We