Climatic Change (2019) 154:125–141
https://doi.org/10.1007/s10584-019-02400-1

Modeling seasonal onset of coastal ice

Xialong Ji[1] · Andrew D. Gronewold[2] · Houraa Daher[3] · Richard B. Rood[1]

Received: 10 April 2018 / Accepted: 25 February 2019 / Published online: 1 April 2019

Abstract
To support regional management planning decisions, and to protect human health and safety, we developed a new statistical model that simulates the onset of seasonal ice cover along the shoreline of a US National Park (the Apostle Islands National Lakeshore, or APIS). Our model encodes relationships between different modes of climate variability and regional ice cover from 1972 to 2015, and successfully simulates both the timing of ice onset and the probability that ice cover might form at all in a particular winter season. We simulate both of these endpoints using a novel combination of statistical hazard (or survival) and beta regression models. Our analysis of coastal ice cover along the APIS reinforces findings from previous research suggesting that the late 1990s signified a regime shift in climate conditions across North America. Before this period, coastal ice cover conditions at the APIS were often suitable for pedestrian access, while after this period coastal ice cover at the APIS has been highly variable. Our new model accommodates this regime shift, and provides a stepping stone towards a broad range of applications of similar models for supporting regional management decisions in light of evolving climate conditions.

Keywords Coastal ice · Climate variability · Statistical model · Decision-making

1 Introduction

The maximum areal extent, and the timing of onset and retreat of seasonal ice cover across Earth's marine and fresh water bodies has been changing over the past century (Magnuson et al. 2000; Comiso et al. 2008; Wang et al. 2012). These changes pose challenges to coastal

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10584-019-02400-1) contains supplementary material, which is available to authorized users.

✉ Andrew D. Gronewold
drewgron@umich.edu

Xialong Ji
xiaoloji@umich.edu

1 University of Michigan, Ann Arbor, MI, USA

2 NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, MI, USA

3 University of Miami, Coral Gables, FL, USA

and off-shore management agencies for which ice cover facilitates, or impedes, commercial and recreational activities (Laidler et al. 2009; Smith and Stephenson 2013). Across the Laurentian Great Lakes (Fig. 1), Earth's largest lake system, winter ice cover provides a migratory route for wildlife (Hebert 1998; Mlot 2015), can pose hazards for the shipping industry (Millerd 2010), and is a critical component of the lake-atmosphere energy and water flux dynamics that dictate lake effect snow severity (Notaro et al. 2015; Gronewold et al. 2015; Fujisaki-Manome et al. 2017). Ice formation and retreat along the Great Lakes shoreline is of particular importance to regional infrastructure planning and, given the length of Great Lakes coastline (commonly referred to as the "third coast" of the USA), serves as an ideal case study for research on relationships between modes of climate variability and coastal physical processes.

Previously developed models for simulating Great Lakes ice cover range from simple to complex, and span a variety of space and time scales. One-dimensional lakewide-average thermodynamics models, for example, have been employed in operational seasonal water supply forecasting for decades (Croley II and Assel 1994; Gronewold et al. 2011). Statistical models (Assel 1991; Assel et al. 2004; Bai et al. 2015) and three-dimensional hydrodynamic models (Fujisaki-Manome et al. 2013) have been used to simulate and forecast ice cover across shorter time horizons. Other research has focused on multi-decadal projections and simulations of ice cover change using regional climate models (Goyette et al. 2000; Xiao et al. 2016). Importantly, most of the statistical models used to simulate Great Lakes ice cover have been built on empirical relationships between modes of climate variability and basin-scale seasonal maximum ice cover (Assel et al. 2000; Rodionov and Assel 2003; Ghanbari and Bravo 2008; Bai et al. 2015). Recent studies, however, suggest that seasonal

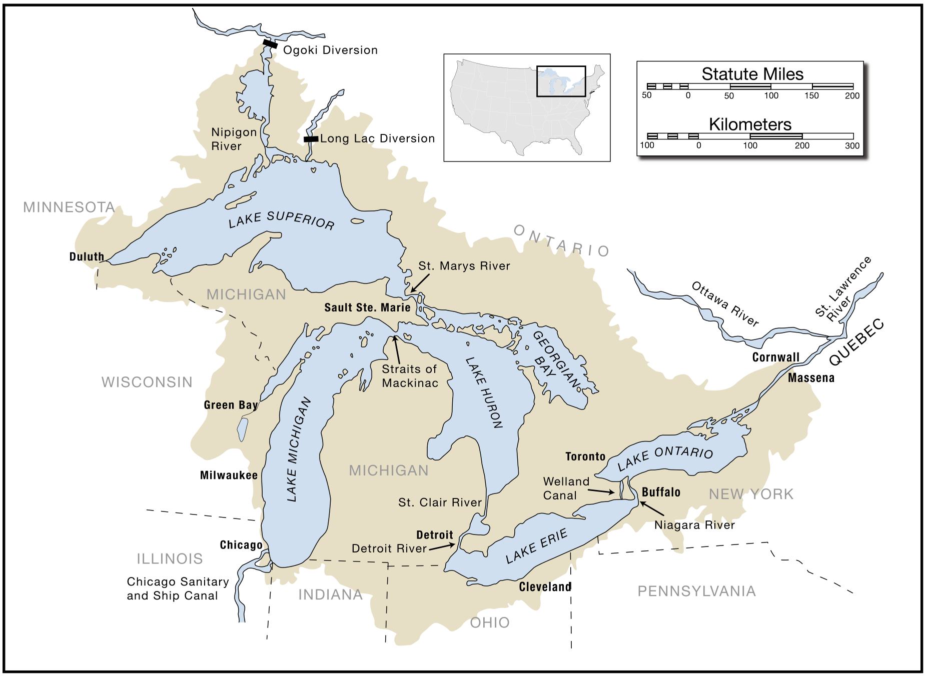

Fig. 1 Map of the North American Great Lakes drainage basin (brown-shaded region) including major cities, political boundaries, interbasin diversions, and interconnecting channels

and interannual ice cover dynamics, and their relationship to climate patterns, can be highly variable when assessed at finer spatial and temporal scales (Magnuson et al. 2000; Mason et al. 2016).

Here, we collectively advance coastal ice modeling, statistical modeling, and climate change research by developing a new model to simulate the seasonal onset of coastal ice cover at regional scales, and apply it to an area managed by the National Park Service—the Apostle Islands National Lakeshore (Krumenaker 2005, 2016). The Lakeshore (hereafter referred to as APIS) is located in Wisconsin along the southwest shoreline of Lake Superior (Fig. 2), the largest freshwater surface on Earth. The APIS includes 21 islands and a 20-km stretch of mainland coastline, while covering a total area of roughly 280 km^2.

The seasonal formation of ice within and along the caves in this area is a popular tourist attraction, and there are periods in some winters when visitors can directly access the caves by walking along the ice-covered lake shoreline. Changes in coastal ice cover areal extent and thickness over the past two decades, however, have periodically limited access to the caves while also making it challenging for managers to anticipate both "low-risk" ice conditions (defined here as ice cover that can sustain high volumes of pedestrian traffic) and the number of tourists that might visit the APIS; when ice caves form, there is a need for additional staff.

More specifically, ice cover on Lake Superior and near the APIS formed with some regularity prior to 1997 (Wang et al. 2012; Van Cleave et al. 2014). After 1997, however, ice cover has been more sporadic. There have been several winters where ice cover was not at any time a low risk for pedestrian access, and others (especially the winter of 2013–2014) when ice onset and retreat were much earlier or later than usual (Assel et al. 2003; Clites et al. 2014). Importantly, in 2014, prominent lake ice brought an unusually high number of visitors to the APIS that were concentrated along the small portion of the shoreline where the ice caves are located, posing a management challenge to APIS personnel.

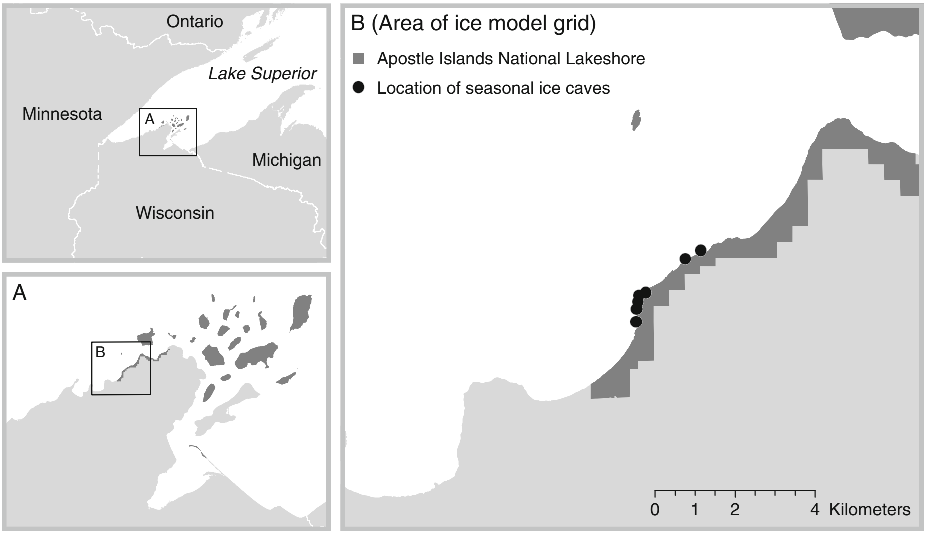

Fig. 2 Location map of the Apostle Islands National Lakeshore. Panel B represents the spatial domain of the gridded ice cover data used in our study and indicates locations of the caves frequently accessed by tourists along the mainland shoreline

Previous research indicates that the exceptionally cold North American winters of 2014 and 2015 were related to the Arctic Oscillation (AO), or its regional manifestation, the North Atlantic Oscillation (NAO). Much attention has been brought to the AO through research (Liu et al. 2012; Francis and Vavrus 2012; Francis et al. 2017) suggesting its behavior might be changing due to the rapid decline of Arctic sea ice and snow cover and, more generally, to rapid changes in regional land and water surface characteristics (for further reading, see Barnston and Livezey 1987).

This study addresses not only a widespread need for improved understanding of connections between climatological drivers and coastal ice cover variability but also for propagating those relationships into regional decision support tools. The latter need catalyzed during recent interactions and planning workshops with APIS management and staff (Star et al. 2015). These interactions led to subsequent questions about the potential to use knowledge of modes of weather and climate variability to forecast the state of APIS shoreline lake ice with sufficient lead time and skill, and to forewarn the need for additional winter Park staff.

The primary goal of our study, then, was the development of a new model for potential application by coastal management agencies (including the APIS) in planning for seasonal ice cover variability that encodes relationships between monthly teleconnection indices each fall (August through December), and ice cover dynamics in the following winter and early spring months (January through June). A secondary goal was the introduction of a robust framework for developing and testing models that is broadly applicable to areas around the world with a need for more accurate simulations of seasonal coastal ice cover dynamics, particularly in light of climate change. These models could be useful not only for supporting safe pedestrian traffic on ice in coastal areas of the Great Lakes, but for large and small-craft vessel navigation, fishing and hunting, and other human activities around the world (Laidler et al. 2009).

2 Methods

2.1 Data

We began by collecting daily gridded ice cover data for all of the Great Lakes for the period December 1972 (the beginning of the available data record) through May 2015 from the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) Great Lakes Ice Atlas, and related suite of products (Assel 2003, 2005). Daily ice cover at each grid point is represented as an areal fraction, and we calculated daily average ice cover for the area around the APIS ice caves (Fig. 2) using the arithmetic average of the grids bounded by latitudes 46.9704° N and 46.8485° N and longitudes 90.9392° W and 91.1363° W.

We then calculated the date of APIS ice onset in each year of the historical record, defined as the first day on which the rolling 10-day average of ice cover, based on the NOAA-GLERL data, exceeds 90%. This criterion is derived from discussions with representatives from the APIS and their historical knowledge of the relationship between duration of surface ice cover area, ice thickness and stability, and the suitability for pedestrian access. We select 90% as a threshold for stable ice cover in the NOAA-GLERL gridded data as a reasonable reflection of the APIS management practices. Because solid ice cover rarely forms before December in most regions of the Great Lakes, we represent the timing of ice onset as the number of days (y^n) from December 1 to the day of ice onset. Hereafter, we refer to this event as the onset of low-risk ice cover.

We obtained average monthly values of the AO, NAO, Pacific-North American Pattern (PNA), Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), and El Niño-Southern Oscillation (NINO3.4) from NOAA's National Centers for Environmental Information (NCEI) for each year from 1972 to 2015 (to coincide with the historical ice cover record). From a meteorological perspective, these teleconnections divide into three quasi-distinct time scales of variability. The AO, NAO, and PNA patterns vary on time scales of days to weeks. The SOI and NINO3.4 vary on times scales of 2 to 7 years, and the PDO is decadal (Perlwitz et al. 2017). From a spatial perspective, the AO and the NAO are more likely than the other indices to indicate conditions local to APIS.

On an annual basis, for example, the AO is the strongest measure of temperature variability in the Northern Hemisphere; it is stronger in the winter than summer (see Hurrell and Deser 2010). When the AO or NAO is in its negative phase, the polar vortex is relatively weak, and the cold air isolated within the polar vortex is likely to be displaced away from the North Pole towards the Great Lakes. That is, a negative phase of the AO or the NAO is often associated with a cool temperature anomaly in the Great Lakes region. The PNA pattern acts more as a guide, loosely focusing the paths of, especially, winter storms as they move from the Pacific Ocean and across the high mountains of western North America.

We then assessed serial autocorrelation within and correlation between each of these indices prior to model calibration. If we found two indices were highly correlated in a given month (with a correlation coefficient greater than 0.65; for details, see Weisberg 2005), we included only one of them in our model calibration. Similarly, if we found that a teleconnection index was autocorrelated across two or more months, we used only the earliest of the autocorrelated monthly values for that index.

Other climate and environmental variables are potential predictors of seasonal ice cover dynamics near the APIS (and other areas in the Great Lakes region) including, for example, surface water temperatures, heat content, winter severity index, and cooling degree days in the months preceding ice onset (Assel 1998; Rodionov and Assel 2003; Bai et al. 2015). The goal of this study, however, was to assess the extent to which continental-scale modes of climate variability alone serve as an adequate predictor of coastal ice cover at local and regional scales. Findings of a significant relationship between climate indices and ice conditions have important management implications not only for the Lake Superior shoreline but for other regions as well. We view analysis of alternative predictor variables throughout the Great Lakes, and other domains, as an area for future research.

2.2 Model description

Our model for simulating seasonal ice cover at the APIS was designed to meet three important criteria closely aligned with the needs of regional stakeholders. First, because of the increased variability in lake ice cover conditions over the past two decades, the model needed to indicate the probability of any low-risk ice cover at the APIS in an upcoming winter season. Second, the model needed to provide an estimation of the first day of low-risk ice onset. Third, the model needed to allow APIS managers to make forecasts at different times throughout the fall months preceding an upcoming winter tourist season.

To meet these criteria, we used statistical hazard (or survival) models because they are typically applied to model duration (i.e., survival time) until an event (for related applications, see Therneau and Grambsch 2000; Read and Vogel 2016). Survival models are particularly useful to our study because they can simulate both the probability of ice onset on each day of an ice season as well as the probability of ice cover occurring *at all* by the end of an upcoming ice season. We modeled the number of days (y_i^n) from December 1 until the onset of ice cover in season i using a Cox proportional hazards regression model and a Weibull (parametric) survival regression model (Anderson and Gill 1982; Kalbfleisch and Prentice 2011). We set September 1 (after the end of the ice season) as our censored date and interpreted the hazard function $h(t \mid \mathbf{x})$ as the instantaneous probability of solid ice at time t (in days) between December 1 and September 1 of the following calendar year. For the Cox proportional hazards model (h^c), this function is as follows:

$$h^c(t \mid \mathbf{x}) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j\} \tag{1}$$

where β_1, β_2, ..., β_j are regression coefficients and $x_1, x_2, \ldots, x_j$ are teleconnection indices. Similarly, the hazard function for the parametric model with Weibull response variable (h^w) is as follows:

$$h^w(t \mid \mathrm{x}) = \gamma\theta t^{\gamma-1} \exp\{\beta_1' x_1 + \beta_2' x_2 + \ldots + \beta_j' x_j\} \tag{2}$$

with scale parameter γ and shape parameter θ.

In addition to the survival models, we developed a beta regression model for simulating the probability of low-risk ice cover before the end of a winter season in which the maximum 10-day average ice areal extent in a season, when expressed as a fraction, is a beta random variable $y^{\max} \sim \mathsf{Be}(\mu, \phi)$ with seasonal varying mean μ and constant (across all seasons) precision ϕ. We modeled μ for each season through a logit link to a linear model with seasonal teleconnection indices $\mathbf{x}$ as predictors:

$$\mu = \frac{\exp\{\beta_1'' x_1 + \beta_2'' x_2 + \ldots + \beta_j'' x_j\}}{1 + \exp\{\beta_1'' x_1 + \beta_2'' x_2 + \ldots + \beta_j'' x_j\}}$$

We then used the beta model (described further in the next section) to simulate the probability distribution of a maximum 10-day average ice cover for each season, and to calculate the probability that it exceeds 90%.

2.3 Model calibration and verification

After eliminating monthly teleconnection indices with significant serial auto- and cross-correlation, we calibrated each model using conventional stepwise regression analyses (Weisberg 2005) in the `R` statistical software environment (R core team 2017). All software packages and functions referenced hereafter are used in `R`.

For the beta model, we initialized the regression procedure using a generalized additive model with a beta link in the `gamlss` function. We then conducted an automated stepwise regression (searching both forwards and backwards) with the `stepGAIC` function, which uses generalized Akaike information criterion (GAIC) for model selection (Stasinopoulos et al. 2017). We then further refined the model using a single manual selection procedure in which we eliminated model coefficients from the `stepGAIC` model using conventional p values as a selection criterion (rather than GAIC), and recalibrated the remaining coefficients using the `betareg` function (Cribari-Neto and Zeileis 2010). This approach yielded a calibrated model that simulates deterministic values of maximum ice cover areal extent for each season, along with estimates of parameters μ and ϕ. We used these results to calculate the probability that ice cover areal extent in a given season exceeds the 90% threshold using the `pbeta`(α, β) function with first shape parameter $\alpha = \mu * \phi$ and second shape parameter $\beta = (1 - \mu) * \phi$ (see Supporting Information for details).

We acknowledge that, for a given data set, manual and automated stepwise regression routines often lead to disparate model results depending on the stopping rules employed and other variations in model and variable selection (Bendel and Afifi 1977). The methods we employed here represent one particular approach to model selection, and we believe that alternate model selection procedures should be explored in future research.

We conducted a similar calibration for the Cox and Weibull survival models using the `coxph` and `survreg` functions, respectively, in the `survival` package (Therneau and Grambsch 2000). The data for the calibration of the survival models includes the first day (in days after December 1) of low-risk ice cover, and a binary variable indicating whether or not any low-risk ice had formed by the censored date (i.e., September 1 of the following calendar year). The `coxph` and `survreg` functions transformed these inputs into a hazard ratio suitable for calibration of Eqs. 1 and 2.

For the Cox and Weibull models, we conducted a manual backwards stepwise regression procedure. We began with all potential model coefficients and, after each step, eliminated those with a p value greater than 0.05 (for further reading, see Wasserstein and Lazar 2016). We continued this procedure until all remaining coefficients had a p value lower than 0.05. The resulting models provide a probabilistic (i.e., with an expression of uncertainty) simulation of the earliest date of ice onset. We used the lower 10% quantile of these simulations as a conservative estimate of the earliest possible date of ice onset in a given season, and recognize that different thresholds could be used depending on the preferences of the model user.

We repeated the calibration routines for all three models using two separate ice cover periods: one includes winter seasons from 1972–1973 through 1995–1996 and the other includes winter seasons from 1996–1997 through 2014–2015. We based the timing of the separation between these two periods on findings from previous research (Van Cleave et al. 2014), and on our own change point analysis (following Mason et al. 2016), that collectively provide strong evidence for a significant difference in mean ice cover before and after 1997 (for details of our change point analysis, see the Supporting Information).

Our analysis therefore explicitly tests the hypothesis that the change in ice cover in the late 1990s (Van Cleave et al. 2014; Mason et al. 2016) might warrant separate models for pre- and post-1997 periods. For these six "split" models (two time periods for each of the three models), we used the model structures (i.e., teleconnection indices) derived from the regression analysis across the entire historical period, but updated coefficient values via calibration using separate pre- and post-1997 time periods.

3 Results and discussion

Daily areal ice extent along the shoreline near the APIS caves during each ice season from 1973 to 2015 (Fig. 3) indicates significant interseasonal variability in the onset date and duration of ice cover, and suggests that management actions, including restrictions on pedestrian access to the ice caves, might also differ depending on the definition of low-risk ice cover. For example, we found, as noted in previous Great Lakes ice cover studies (Magnuson et al. 2000), that while the date of ice onset (Fig. 4) appears to be gradually progressing later in each season from 1973 to 1997, there is low-risk ice cover at some point in each of those seasons. Between 1997 and 2015, however, low-risk ice developed in only 11 seasons and, for years when there was low-risk ice cover, the onset date appears to progress earlier in each season over time. These ice cover onset changes, and the interseasonal variability

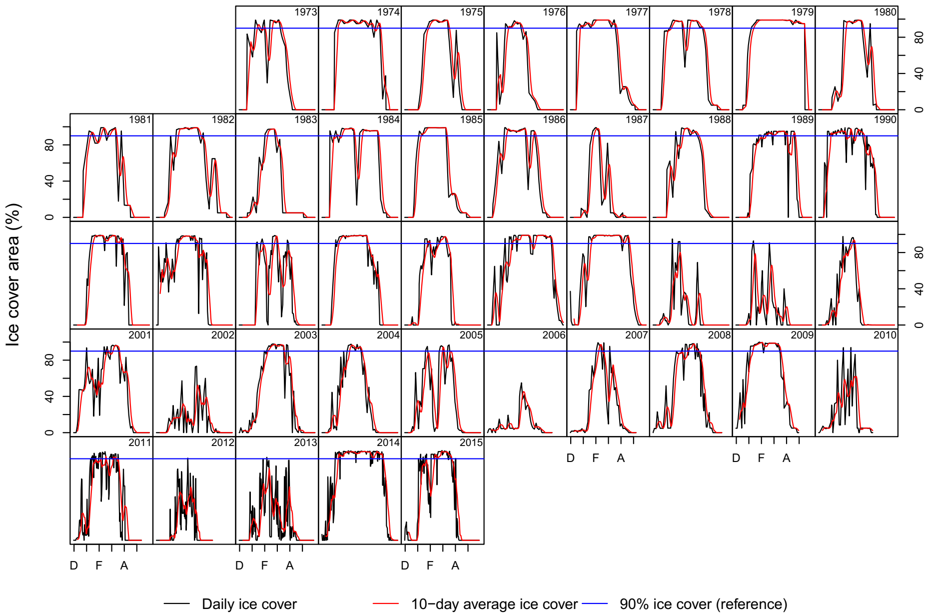

Fig. 3 Observed daily ice cover (black line) and 10-day rolling average of daily ice cover (red) along the APIS coastline (i.e., area in Fig. 2, panel B) for each ice season from 1973 through 2015

before and after the late 1990s, are consistent with previous studies indicating changes in environmental and physical processes not only in the Great Lakes, but across other regions as well (Assel 1998; Chavez et al. 2002; Navarrete et al. 2002; Scott and Marshall 2010; Van Cleave et al. 2014; McCarthy et al. 2015).

We also found that, for several ice seasons (1998, 1999, 2010, 2012, 2013, and 2015), daily ice cover areal extent exceeded or came very close to the 90% threshold, while the 10-day rolling average in those years either exceeded the 90% threshold later in the season, or

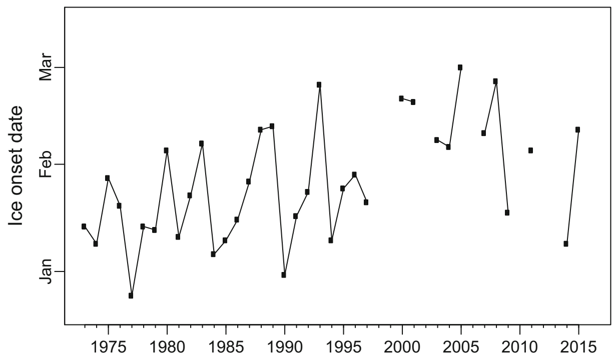

Fig. 4 Date of onset of ice cover suitable for pedestrian access to the APIS ice caves (i.e., "low-risk" ice cover) for each winter ice season. No point is shown for a season if low-risk ice cover was never present. For clarity, lines connect only consecutive years in which there was low-risk ice cover

not at all. This finding underscores the sensitivity of APIS decision-making protocols, and our model results, to the 90% threshold criteria.

Visual inspection of historical monthly teleconnections from August through December (Fig. 5) indicates a noticeable shift in PDO in months preceding the winter ice season starting in the late 1990s and persisting through 2014. Interestingly, the PDO also shifts in 2014 and 2015, suggesting it might be a strong predictor of the unusually high ice cover observed across the Great Lakes region in those years (Clites et al. 2014).

Our analysis of correlation among teleconnection indices (see figures in Supporting Information) indicates that PDO, SOI, and NINO3.4 are each strongly autocorrelated across time, and therefore, only one of the monthly values for each of these three teleconnections in the months preceding the ice season was evaluated in our regression analysis. This finding is significant because it may allow decision-makers to use observed teleconnection information as early as September in seasonal forecasting. This finding is also significant because it represents a departure from previous studies on Great Lakes seasonal ice forecasting in which annual (rather than monthly) average teleconnection indices are used (Bai et al. 2015). The

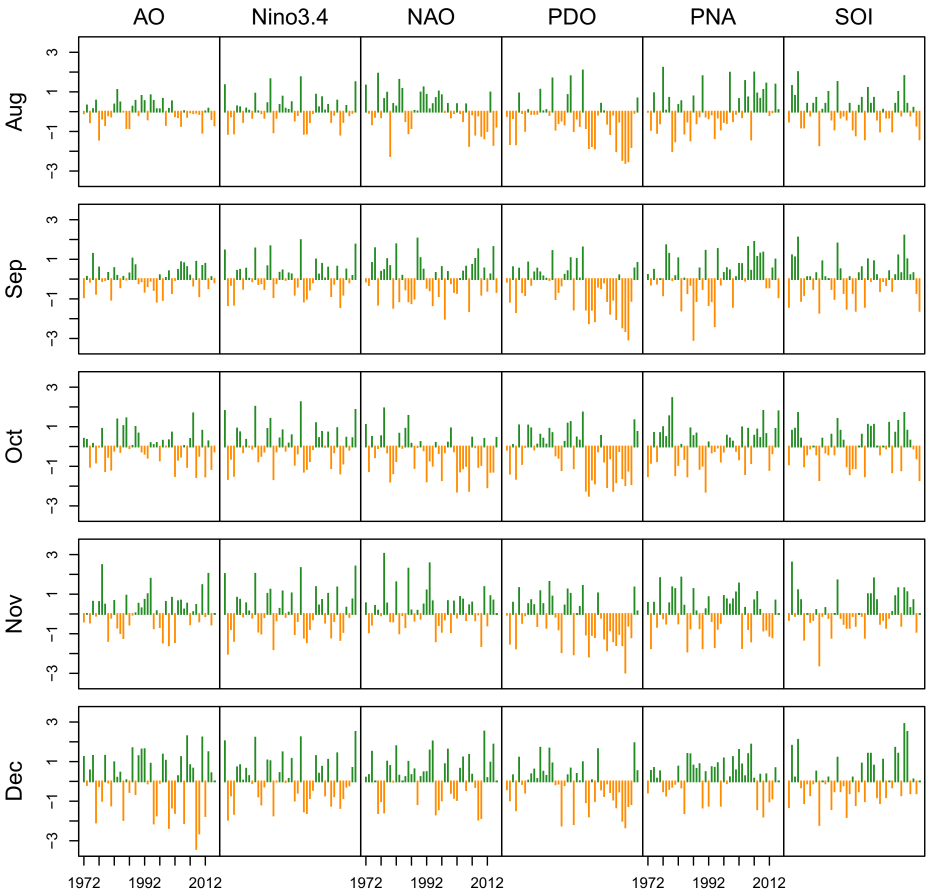

Fig. 5 Monthly average teleconnection indices for August through December (rows) for each year from 1972 (the year preceding the first ice cover records) through 2015. Because our study only extended through the 2015 ice season, teleconnection indices in 2015 (though shown here) were not evaluated as predictors

ability to make forecasts from monthly teleconnection indices prior to the ice season has the potential to allow Park managers to better anticipate the extent of ice cover and improve planning decisions.

Similarly, our analysis of cross-correlation between teleconnection indices for each month from August through December (see figures in Supporting Information) indicates that SOI and NINO3.4 are highly correlated in each month, and that NAO and AO are either highly correlated (November and December) or moderately correlated (August, September, and October) throughout the fall.

It is informative to note that while SOI and NINO3.4 are the strongest measures of global variability on the 2- to 7-year scale (Perlwitz et al. 2017), the signals of seasonal weather conditions associated with the SOI and NINO3.4 are weaker in the Great Lakes region than in other parts of North America. Further, these indices are both measures of variability in the tropical Pacific and influence the geographic distribution of the transport of heat and humidity from the Tropics to middle and high latitudes; it is therefore not particularly surprising that they are highly correlated.

The relationships between these teleconnection patterns are probabilistic and in no cases deterministic. Therefore, there is an implicit challenge in extracting a signal from potentially large amounts of variability. This challenge is made more difficult as we are now in a time of rapidly warming temperatures. Indeed, our entire data record is from a time when global surface air temperatures are increasing. The modes of variability may, themselves, be changing. Even if the modes of variability remain stationary with time, the spatial and temporal existence of air that is, historically, colder than average is decreasing.

A year-by-year analysis of simulations of ice cover onset date from the Cox and Weibull models calibrated to the entire historical record indicates (Fig. 6) that there is much more uncertainty in the Weibull model. Furthermore, in almost every year of our analysis, the Weibull model indicates a nonzero probability of ice onset almost immediately after December 1. The Cox model, however, tends to have narrower and more accurate uncertainty bounds around the actual ice onset date. In 2007, for example (Fig. 6), the Cox model indicated that it would be very unlikely for ice to begin forming before early January, whereas the Weibull model indicated ice might begin forming very soon after December 1. Similarly, in 2007, the Cox model indicated that ice cover was almost certain to form by early March, but the Weibull model indicated there was a small possibility it might not form until early April. From a management perspective, the Weibull model (in 2007 and in nearly ever other year) therefore provides an estimate of ice onset much earlier than both the Cox model and the actual ice onset date.

Neither the Cox nor the Weibull model, however, accurately estimated the overall probability of ice cover in a given season relative to the beta model. In the seven seasons without any low-risk ice cover (1998, 1999, 2002, 2006, 2010, 2012, and 2013), the beta probability model estimated a higher probability of ice absence (i.e., lower probability of low-risk ice cover) than the Cox and Weibull models and, in many of those years, estimated a probability of ice absence very close to 1.0. In 2010, for example (bottom right panel Fig. 6), a year in which there was never low-risk ice cover, the Cox model estimated a probability of a season with no ice cover close to 0.6, and the Weibull model estimated a probability close to 0.2. The beta model, in 2010, estimated a probability of no low-risk ice cover very close to 1.0. Similarly, in many years when low-risk ice cover was observed, the beta regression model estimated a lower probability of ice absence than the Cox and Weibull models. This finding suggests that the beta regression is a potentially suitable predictor of the probability of ice cover in a given season, while the Cox model is most suitable for estimating when that ice cover might first begin to form.

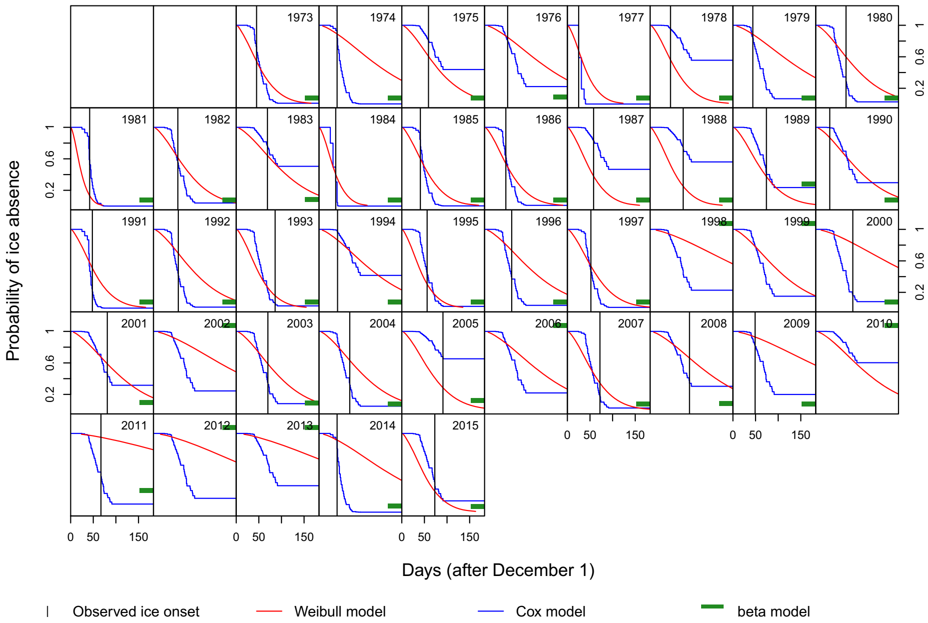

Fig. 6 Seasonal hazard functions (curved lines) for date of ice onset derived from the Cox (blue line) and Weibull (red line) survival models along with the observed date of ice onset (vertical black line), and the probability of ice absence (green bar along right-hand axis) derived from the beta regression model. Panels without a vertical black line indicate a season in which there was never low-risk ice cover

However, additional insight is needed before these models should be considered for forecasting, and that insight can be derived from our assessment of the impacts of calibrating the models separately to periods before and after the late 1990s regime shift. We found, more specifically, that doing so led to a significant improvement in skill for simulating both the date of ice onset and the probability of a season without any low-risk ice cover (Fig. 7). For example, the Cox model, when calibrated to the entire period of record (upper-left panel, Fig. 7), simulated ice onset dates that do not vary significantly from season to season. When calibrated separately to pre- and post-1997 periods, however, the ice onset dates for the post-1997 period (upper-right panel, Fig. 7) are quite different; the "split" Cox model simulated an ice onset date still earlier than, but closer to, the actual onset date in 8 of the 11 years after 1997 with low-risk seasonal ice cover.

For the other three post-1997 years, the "split" Cox model simulated ice onset dates earlier than the non-split model (but still before the actual onset date) in one (2007), after the observed onset date in one (2009) and, for 1 year (2006) did not yield a valid ice onset date simulation. We suspect that the missing simulated ice onset date in 2006 is an artifact of the model attempting to simulate a very late ice onset date, but not being able to do so because of the parameters we set for censored data. Here (Fig. 7), we present results only for the Cox model because it provided more accurate simulations than the Weibull model. However, we did find (results not shown) that relative to the non-split model, the "split" Weibull model also showed considerable improvements.

Differences between the split and non-split versions of the beta model for simulating absence of ice cover across an entire season (bottom panels, Fig. 7) are more profound. The

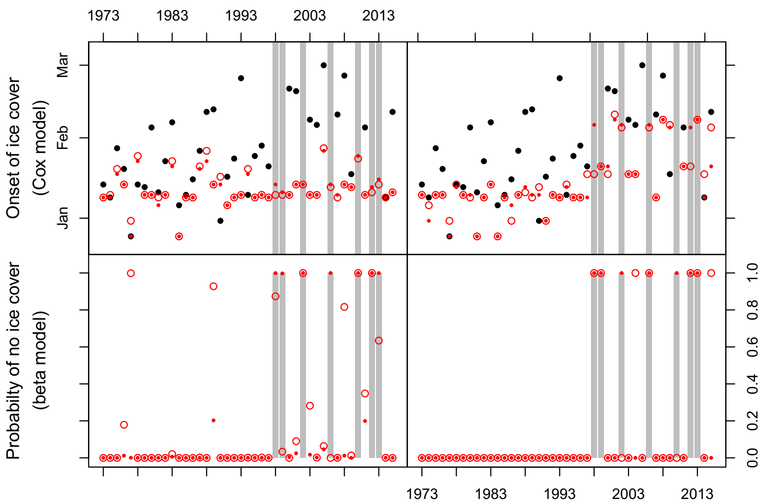

Fig. 7 (Top row) Simulated (red) and observed (black) ice cover onset dates for each ice season using the Cox hazard model calibrated to the entire period of record (left column) and calibrated separately to pre- and post-1997 periods (right column). (Bottom) Simulated (red) probability of ice absence in each season using the beta model. Hollow red circles in all panels represent leave-one-out cross-validation simulations. Gray vertical shaded regions are added for reference, and are aligned with years in which there was no low-risk ice cover

"split" model, for example (bottom right panel, Fig. 7), simulated a probability of seasonal ice absence very close to 1.0 for each of the years with no low-risk ice, and a probability of seasonal ice absence very close to 0.0 for each of the years in which there was low-risk ice. These results represent a significant improvement over the model when calibrated to the entire period (bottom left, Fig. 7).

To address potential problems associated with overfitting, we conducted a leave-one-out cross-validation procedure for all three models for both the entire time period and the post-1997 period. These results, presented for the Cox and beta models (Fig. 7), indicate that the "split" beta regression model is robust; in each year after 1997, the results of the model validation are nearly identical to those from the model calibration. These findings suggest that the post-1997 ice cover regime is explained well by our beta model, and that it could potentially be used in real-world forecasting by the APIS. The validation results for the Cox model are also robust, but not so much as the beta model; in the 11 seasons after 1997 in which low-risk ice cover formed, the validation results for the Cox model are very close to the calibration results in all but two.

Finally, an assessment of the model coefficients selected for each of the three models (Table 1) indicates that August NAO is the most consistent predictor of ice cover response. This is not particularly surprising, given that August NAO appears to have shifted in the late 1990s (Fig. 5), coinciding with the shift in ice cover at the APIS. Our results also indicate that the Cox model relies on three seasonal indices, and that simulations could potentially be made earlier in the fall season using only August indices—understanding the extent to which doing so might lead to loss of skill is an area for future research. While the beta regression model requires a relatively high number of predictor variables, the results of our cross-validation analysis indicate that, at least at seasonal time scales, overfitting is

Table 1 Summary of model variables and coefficients in our regression analysis

	Cox					Beta				
	Aug	Sep	Oct	Nov	Dec	Aug	Sep	Oct	Nov	Dec
AO						·	⊡	·		□
NINO3.4	·					⊡				
NAO	⊡					⊡	⊡			
PDO						□				
PNA				□		⊡	⊡	·		⊡
SOI										

Gray cells correspond to teleconnection indices not included in our model due to cross- or autocorrelation. Empty white cells correspond to variables that were evaluated, but for which the corresponding regression model coefficient was determined to be insignificant. White cells with a dot or square indicate that a coefficient was significant for the corresponding variable (dot) or the squared value of the variable (square), or both

not a significant problem. For further details on the model coefficients, see the Supporting Information.

4 Conclusions

Through an analysis of coastal ice cover variability at a National Park, we have developed a new statistical model that serves as a stepping stone towards improving seasonal ice cover forecasting and an increased level of preparedness for APIS management and staff. Novel aspects of our approach include the use of hazard (survival) models to simulate the date of seasonal ice onset, and the use of a beta regression model to simulate the probability that low-risk ice cover might occur (or not occur) at any point in a season. We have been informed through several iterations of this research by APIS management and staff. This iterative process helped define and refine model parameters and predictive variables. Though the model has yet to be exercised in real-world applications, our results suggest it has predictive skill. Hence, we look forward to evaluating the model's usability, with usability ultimately defined by the decision-makers at the APIS. The engagement process between the partners in this research will be described elsewhere.

A representative simulation from our final composite model (Fig. 8) underscores the benefits of our approach. The advancements represented by our work are particularly important given that ice cover in coastal regions of the Great Lakes, and elsewhere, has been highly variable over the past two decades, and that many conventional statistical models do not carefully align the selection of a probability density function with the corresponding model response variable. More importantly, our results may help serve as a basis for improving management decisions at the APIS; in nearly every year in the study, and particularly in years after 1998, our model (had it been used over the past few decades) could have added value to the decision-making process by indicating that the onset of ice might have either been delayed relative to the long-term average date, or that it might not have occurred at all.

Our findings also indicate that a beta regression model using teleconnection indices as predictors can estimate the probability of seasonal ice cover with high skill, and that a Cox survival model (with similar predictors) can estimate the timing of ice onset with reasonable

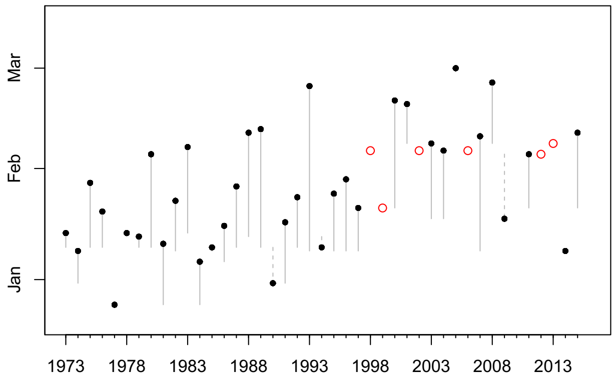

Fig. 8 Model results combining the simulated probability of low-risk ice cover in a given season (from the beta model), and the simulated earliest possible date of low-risk ice (from the Cox model). Black dots indicate the date of low-risk ice onset based on historical NOAA-GLERL data; absence of a black dot for a particular year indicates that there was no ice cover suitable for pedestrian access according to NOAA-GLERL data. Gray vertical segments are drawn between the simulated first date of low-risk ice and the historical date of low-risk ice onset. If the simulated onset date is before the historical onset date, the segment is solid; if after, the segment is dashed. Seasons with a simulated low probability of low-risk ice cover are represented as hollow red circles, and are positioned vertically at the date (if low-risk ice were to form) of ice onset based on the Cox model

skill. Dividing the model into pre- and post-1997 periods led to improvements in skill, and we suggest users of our model utilize this finding while also continuously updating model coefficients to determine if other significant change points evolve. These findings indicate promise for use of regression-based models in other ice onset simulation problems and warrant further research using alternate predictor variables that might provide a more localized representation of drivers of ice cover. We also note that separate models could be used to improve simulation and forecasting of ice break-up at the end of each season; APIS management and staff have indicated that these events (and the meteorological and hydrodynamic phenomena the drive them) are a cause for more urgent safety considerations, though are perhaps less predictable.

Our statistical analysis has brought renewed attention to regional impacts of the NAO and PDO, and underscores the controversy surrounding research on changes in these modes of variability as the planet warms (Trenberth and Fasullo 2013; England et al. 2014). Of special focus in recent research is the influence of sea ice and snow cover decline in the high Arctic (Francis and Vavrus 2012; Liu et al. 2012; Francis et al. 2017).

The PDO, specifically, is less well described and less well understood than the other indices used here (Deser et al. 2012). However, as with the other indices, the phase of the PDO suggests preferential geographical focusing of atmospheric wave patterns that influence both temperature anomalies on seasonal time scales, and the propagation of winter storms as they transport heat from the west and south to the north and east. Our results suggest that information from each of these time scales (i.e., days to weeks, 2 to 7 years, and decadal) have quantitative information about ice cover at the APIS. While placing our statistical results into broader geographic contexts is beyond the scope of this study, the results do suggest that scenarios of plausible futures to benefit planning for changes in lake ice (and other parameters) need to consider these modes of behavior and how they might change.

Finally, rather than presenting climate change in the Great Lakes as either a persistent trend or even a sudden incremental change, the observed non-stationarity of surface parameters both locally (i.e., on the Great Lakes) and regionally (i.e., in the Arctic) suggests a future punctuated by high variability (Briley et al. 2017; Gronewold and Rood 2019).

Acknowledgements Anne Clites, Craig Stow, Song Qian, and Jia Wang provided valuable comments on the technical aspects of the manuscript and use of ice cover data. Kaye LaFond, Becky Bolinger, Lacey Mason, and Nicole Rice provided data management, editorial, and graphical support. The authors thank Robert Krumenaker (Superintendent), David Cooper, and Neil Howk from the APIS for helpful discussions related to this project. This is NOAA-GLERL contribution number 1914.

Funding information Funding for this study was provided by the University of Michigan, the Great Lakes Integrated Science and Assessment (GLISA) center, and NOAA-GLERL.

References

Anderson P, Gill R (1982) Cox's regression model for counting processes. Ann Stat 10:1100–1120
Assel RA (1991) Implications of CO_2 global warming on Great Lakes ice cover. Clim Chang 18(4):377–395
Assel RA (1998) The 1997 ENSO event and implications for North American Laurentian Great Lakes winter severity and ice cover. Geophys Res Lett 25(7):1031–1033
Assel RA (2003) An electronic atlas of Great Lakes ice cover, winters 1973-2002. Tech rep
Assel RA (2005) Classification of annual Great Lakes ice cycles: winters of 1973–2002. J Clim 18(22):4895–4905
Assel RA, Janowiak JE, Boyce D, O'Connors C, Quinn FH, Norton DC (2000) Laurentian Great Lakes ice and weather conditions for the 1998 El Niño winter. Bull Am Meteorol Soc 81(4):703–717
Assel RA, Cronk K, Norton DC (2003) Recent trends in Laurentian Great Lakes ice cover. Clim Chang 57(1–2):185–204
Assel RA, Drobot S, Croley II TE (2004) Improving 30-day Great Lakes ice cover outlooks. J Hydrometeorol 5:713–717
Bai X, Wang J, Austin JA, Schwab DJ, Assel RA, Clites AH, Bratton JF, Colton M, Lenters JD, Lofgren BM, Wohlleben T, Helfrich S, Vanderploeg H, Luo L, Leshkevich GA (2015) A record-breaking low ice cover over the Great Lakes during winter 2011/2012: combined effects of a strong positive NAO and La Niña. Climate Dynam 44(5–6):1187–1213
Barnston AG, Livezey RE (1987) Classification, seasonality, and persistence of low-frequency atmospheric circulation patterns. Mon Weather Rev 115(6):1083–1126
Bendel RB, Afifi AA (1977) Comparison of stopping rules in forward 'stepwise' regression. J Am Stat Assoc 72(357):46–53
Briley LJ, Ashley WS, Rood RB, Krmenec A (2017) The role of meteorological processes in the description of uncertainty for climate change decision-making. Theor Appl Climatol 127(3–4):643–654
Chavez FP, Pennington JT, Castro CG, Ryan JP, Michisaki RP, Schlining B, Walz P, Buck KR, McFadyen A, Collins CA (2002) Biological and chemical consequences of the 1997-1998 El Niño in Central California waters. Prog Oceanogr 54(1):205–232
Clites AH, Wang J, Campbell KB, Gronewold AD, Assel RA, Bai X, Leshkevich GA (2014) Cold water and high ice cover on Great Lakes in spring 2014. Eos Trans AGU 95(34):305–306
Comiso JC, Parkinson CL, Gersten R, Stock L (2008) Accelerated decline in the Arctic sea ice cover. Geophys Res Lett 35(1):L01703
Cribari-Neto F, Zeileis A (2010) Beta regression in R. J Stat Softw 34(2):1–24
Croley II TE, Assel RA (1994) A one-dimensional ice thermodynamics model for the Laurentian Great Lakes. Water Resour Res 30(3):625–639
Deser C, Phillips AS, Tomas RA, Okumura YM, Alexander MA, Capotondi A, Scott JD, Kwon YO, Ohba M (2012) ENSO and Pacific decadal variability in the community climate system model version 4. J Climate 25(8):2622–2651
England MH, McGregor S, Spence P, Meehl GA, Timmermann A, Cai W, Gupta AS, McPhaden MJ, Purich A, Santoso A (2014) Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. Nat Clim Chang 4(3):222–227

Francis JA, Vavrus SJ (2012) Evidence linking Arctic amplification to extreme weather in mid-latitudes. Geophys Res Lett 39:L06801
Francis JA, Vavrus SJ, Cohen J (2017) Amplified Arctic warming and mid-latitude weather: new perspectives on emerging connectoins. WIREs: Clim Change 8(5):e474
Fujisaki-Manome A, Wang J, Bai X, Leshkevich GA, Lofgren BM (2013) Model-simulated interannual variability of Lake Erie ice cover, circulation, and thermal structure in response to atmospheric forcing, 2003–2012. J Geophys Res Oceans 118(9):4286–4304
Fujisaki-Manome A, Fitzpatrick L, Gronewold AD, Anderson EJ, Lofgren BM, Spence C, Chen J, Shao C, Wright DM, Xiao C (2017) Turbulent heat fluxes during an extreme lake effect snow event. J Hydrometeorol 18(2):3145–3163
Ghanbari RN, Bravo HR (2008) Coherence between atmospheric teleconnections, Great Lakes water levels, and regional climate. Adv Water Resour 31(10):1284–1298
Goyette S, McFarlane NA, Flato GM (2000) Application of the Canadian regional climate model to the Laurentian Great Lakes region: implementation of a lake model. Atmosphere-Ocean 38(3):481–503
Gronewold AD, Rood RB (2019) Recent water level changes across Earth's largest lake system and implications for future variability. J Great Lakes Res 45(1):1–3
Gronewold AD, Clites AH, Hunter TS, Stow CA (2011) An appraisal of the Great Lakes advanced hydrologic prediction system. J Great Lakes Res 37(3):577–583
Gronewold AD, Anderson EJ, Lofgren BM, Blanken PD, Wang J, Smith JP, Hunter TS, Lang GA, Stow CA, Beletsky D, Bratton JF (2015) Impact of extreme 2013–2014 winter conditions on Lake Michigan's fall heat content, surface temperature, and evaporation. Geophys Res Lett 42(9):3364–3370
Hebert CE (1998) Winter severity affects migration and contaminant accumulation in Northern Great Lakes herring gulls. Ecol Appl 8(3):669–679
Hurrell JW, Deser C (2010) North Atlantic climate variability: the role of the North Atlantic Oscillation. J Mar Syst 79(3-4):231–244
Kalbfleisch JD, Prentice RL (2011) The statistical analysis of failure time data. Wiley, New York
Krumenaker R (2005) New wilderness 'can' be created: a personal history of the Gaylord Nelson wilderness at Apostle Islands National Lakeshore. The George Wright Forum 22(3):35–49
Krumenaker R (2016) The view from the Apostle Islands. Wisconsin people and ideas 62(2)
Laidler GJ, Ford JD, Gough WA, Ikummaq T, Gagnon AS, Kowal S, Qrunnut K, Irngaut C (2009) Travelling and hunting in a changing Arctic: assessing unit vulnerability to sea ice change in Igloolik, Nunavit. Clim Change 94(3–4):363–397
Liu J, Curry JA, Wang J, Song M, Horton RM (2012) Impact of declining Arctic sea ice on winter snowfall. Proc Natl Acad Sci USA 109(17):4074–4079
Magnuson JJ, Robertson DM, Benson BJ, Wynne RH, Livingstone DM, Arai T, Assel RA, Barry RG, Card V, Kuusisto E, Granin NG, Prowse TD, Stewart KM, Vuglinski VS (2000) Historical trends in lake and river ice cover in the Northern Hemisphere. Science 289(5485):1743–1746
Mason LA, Riseng CM, Gronewold AD, Rutherford ES, Wang J, Clites AH, Smith SDP, McIntyre PB (2016) Fine-scale spatial variation in ice cover and surface temperature trends across the surface of the Laurentian Great Lakes. Clim Change 138(1–2):71–83
McCarthy GD, Haigh ID, Hirschi JJM, Grist JP, Smeed DA (2015) Ocean impact on decadal Atlantic climate variability revealed by sea-level observatoins. Nature 521(7553):508–510
Millerd F (2010) The potential impact of climate change on Great Lakes international shipping. Clim Chang 104:629–652
Mlot C (2015) Inbred wolf population on Isle Royale collapses. Science 348(6233):383–383
Navarrete SA, Broitman B, Wieters EA, Finke GR, Venegas RM, Sotomayor A (2002) Recruitment of intertidal invertebrates in the southeast Pacific: interannual variability and the 1997-1998 Niño. Limnol Oceanogr 47(3):791–802
Notaro M, Bennington V, Lofgren BM (2015) Dynamical downscaling-based projections of Great Lakes water levels. J Clim 28(24):9721–9745
Perlwitz J, Knutson T, Kossin J (2017) Large-scale circulation and climate variability. In: Wuebbles D, Fahey DW, Hibbard KA, Dokken DJ, Stewart BC, Maycock TK, Report ClimateScienceSpecial (eds) Climate science special report: a sustained assessment of the U.S. Global change research program, U.S. global change research program, chap 5, pp 228-266
R core team (2017) R: a language and environment for statistical computing. Vienna, Austria. http://www.r-project.org
Read LK, Vogel RM (2016) Hazard function analysis for flood planning under nonstationarity. Water Resour Res 52(5):4116–4131
Rodionov S, Assel RA (2003) Winter severity in the Great Lakes region: a tale of two oscillations. Clim Res 24(1):19–31

Scott JBT, Marshall GJ (2010) A step-change in the date of sea-ice breakup in Western Hudson Bay. Arctic 63(2):155–164

Smith LC, Stephenson SR (2013) New trans-Atlantic shipping routes navigable by midcentury. Proc Natl Acad Sci U S A 110(13):E1191–E1195

Star J, Fisichelli N, Schuurman G, Welling L, Rood RB, Briley LJ, Baule W (2015) Climate change scenario planning workshop summary. Tech rep, Apostle Islands National Lakeshore

Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, De Bastiani F (2017) Flexible regression and smoothing: using GAMLSS in R. Chapman & Hall/CRC, London

Therneau TM, Grambsch PM (2000) Modeling survival data: extending the Cox model. Springer, New York

Trenberth KE, Fasullo JT (2013) An apparent hiatus in global warming? Earth's Future 1(1):19–32

Van Cleave K, Lenters JD, Wang J, Verhamme EM (2014) A regime shift in Lake Superior ice cover, evaporation, and water temperature following the warm El Niño winter of 1997–1998. Limnol Oceanogr 59(6):1889–1898

Wang J, Bai X, Hu H, Clites AH, Colton M, Lofgren BM (2012) Temporal and spatial variability of Great Lakes ice cover, 1973–2010. J Clim 25(4):1318–1329

Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: context, process, and purpose. Am Stat 70(2):129–133

Weisberg S (2005) Applied linear regression. Wiley series in probability and statistics, 3rd edn. Wiley-Interscience, Hoboken

Xiao C, Lofgren BM, Wang J, Chu PY (2016) Improving the lake scheme within a coupled WRF-lake model in the Laurentian Great Lakes. J Adv Model Earth Syst 8(4):1969–1985

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.