AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE

10.1029/2017WR021731

Linear-Circular Statistical Modeling of Lake Ice-Out Dates

Mussie T. Beyene[1,2], **Shaleen Jain**[1,2,3], **and Ramesh C. Gupta**[4]

[1]Department of Civil and Environmental Engineering, University of Maine, Orono, ME, USA, [2]Mitchell Center for Sustainability Solutions, University of Maine, Orono, ME, USA, [3]Climate Change Institute, University of Maine, Orono, ME, USA, [4]Department of Mathematics and Statistics, University of Maine, Orono, ME, USA

Key Points:

- Timing of lake ice-out critically influences physical and biogeochemical processes
- Winter and spring weather-climate processes lend predictability to lake ice-out dates
- Linear-circular statistical framework allows comprehensive modeling and assessment of lake ice-out

Supporting Information:

- Supporting Information S1
- Data Set S1
- Data Set S2
- Data Set S3
- Data Set S4

Correspondence to:
M. T. Beyene,
mussie.beyene@maine.edu

Citation:
Beyene, M. T., Jain, S., & Gupta, R. C. (2018). Linear-circular statistical modeling of lake ice-out dates. *Water Resources Research*, *54*, 7841–7858. https://doi.org/10.1029/2017WR021731

Received 22 AUG 2017
Accepted 11 SEP 2018
Accepted article online 21 SEP 2018
Published online 16 OCT 2018

Abstract Over the past few decades, lake ice phenology in northern temperate lakes has exhibited increased interannual variability. The resulting increase in the incidence of unusually early spring lake ice-out dates has the potential to affect stability, health, and function of the lake ecosystems. Characterizing the dependency of spring lake ice-out date to winter and/or spring climate variables offers foreknowledge on the annual lake ice cover season, as the spring ice-out date is an integrated response to prevailing weather/climate conditions during winter and spring. Here a circular regression framework is presented where ice-out date regression models, conditioned on a suite of predictor winter and/or spring climate variables (i.e., degree days and snowfall), are developed for 12 Maine lakes to determine the relative import of winter and spring meteorological conditions on year-to-year variability of ice-out dates in Maine lakes. In the circular regression models, ice-out dates are expressed as points on a unit circle instead of real line, as it preserves the periodicity and order of time-of-day variables independent of the choice of reference point. Results show that (a) the magnitude and variance of seasonal spring temperatures explain more than half of the total variability in spring ice-out date for Maine lakes, (b) the modulating efficacy of spring snowfall on the timing of spring ice-out dates is the strongest in northern interior Maine lakes, (c) the role of winter degree days in determining the ice-out dates in Maine lakes is significant across all climate regions, and (d) the effect of winter snowfall on ice-out dates is significant in coastal Maine lakes. Diagnostics suggest that there are other climatic and nonclimatic variables that produce shifts in the lake ice-out dates.

1. Introduction

In temperate regions, there has been an increase in the interannual variability of lake ice phenology over the past three decades (e.g., Kratz et al., 2001; Magnuson et al., 2000; Weyhenmeyer et al., 2011). This has led to the rise in the frequency of unusually short ice cover periods. For instance, Beyene and Jain (2015) found that interannual winter climate variability, linked to northern hemisphere atmospheric teleconnection patterns, promote early ice-out dates in Maine lakes. Winter limnology studies show that the shortening of the ice cover season in lakes has detrimental effect on lake ecology and services (e.g., Hampton et al., 2017; Prowse et al., 2011). The year-to-year variability in the timing of spring lake ice-out dates is primarily a response to prevailing winter and spring meteorological conditions, as they control the surplus/deficit in the energy balance at lake surface determining lake ice growth/melt (e.g., Leppäranta, 2010; Livingstone, 1997). However, aside from spring temperatures, the efficacy of seasonal meteorological variables particularly during winter, in modulating the timing of the spring ice-out dates of lakes, is not well understood. Given that winter climate variability over the northern Temperate and Arctic regions is influenced by large-scale oceanic-atmospheric circulation patterns, determining the role of winter on the variability of ice-out dates would afford seasonal or longer outlook on the ice cover season of lakes, both at local and regional scales.

Winter and spring climate variability affect the timing of spring lake ice-out date differently. Winter meteorological conditions govern ice cover processes related to the characteristics (e.g., type and thickness) of the winter ice that melts in spring. For instance, winter air temperatures (particularly the accumulated freezing and melting degree days—AFDD and AMDD) determine the cold content available at lake surface to cool and thicken the ice cover (e.g., Leppäranta, 2015). Winter snowfall, on the other hand, can alter the composition of the ice cover by promoting snow ice formation, as well as reduce the thickness of lake ice, due to its insulating effect (e.g., Adams, 1976; Vavrus et al., 1996). In contrast, spring climate variables control ice processes that govern the rate of melt. Spring air temperature, for example, influences the thermal energy available in the atmosphere to overcome the freeze content and melt the winter ice cover, while spring snowfall can reduce the melt rate by increasing the surface albedo and cold content of the ice cover (e.g., Williams, 1965). Modeling offers the opportunity to disentangle the dependence of year-to-year variability of lake ice-out dates to winter and spring weather and climate processes.

In empirical lake ice studies, ice phenology models, conditioned on seasonal climate variable(s), are often built using traditional regression method. The underlying assumption in this method is that the response variable (e.g., ice in/out dates) is a linear continuous variable, which has a true start point and magnitude. However, given that day-of-year variable is inherently unbounded (no start and end point) and cyclical, representing time variables as a linear variable results in the (i) loss of the periodic nature of time-of-year (ii) order and rank of ice in/out date variables to change with respect to the choice of origin (e.g., Lee, 2010). For instance, if using the Julian calendar, 31 December and 1 January are always 364/365 magnitudes apart and (say) 1 January and 1 September have a magnitude of 1 and 242, respectively. On the other hand, if using the water year, 31 December and 1 January are 1 magnitude apart and 1 January and 1 September have a magnitude of 124 and 1, respectively. According to Mallows (1998), choosing data appropriate for model is a critical first step in statistical model building, as erroneous representation of the phenomenon under consideration in model produces model uncertainty that is much more than simple statistical inefficiency. This highlights the need for an alternate approach for characterizing day-of-year variables in regression models for ice phenology such that (a) the order of ice in/out dates are insensitive to the choice of reference point and (b) the distributional assumptions employed for analyses take into account cyclical nature of time of year.

One such approach is the use of circular (angular) regression method, where the day-of-year variable is represented as a point on the circumference of a unit circle (Jammalamadaka & Sengupta, 2001; Mardia & Jupp, 2009). On a circle, the beginning coincides with the end, and as such representing day-of-year variables as circular variables captures the periodicity and order of calendar days, independent of the choice in reference point. Furthermore, the circular regression approach employs unimodal circular distributions most notable of which is the standard von Mises distribution (Mardia & Jupp, 2009). In addition, in circular regression models where the covariates are linear variable(s) (e.g., temperature and snowfall), link functions such as $2\tan^{-1}(\,)$ are used to map the covariate variable from the real line onto a unit circle. Consequently, the circular regression method is employed here in developing ice-out models of increasing complexity, to clarify the efficacy of winter and spring temperatures and snowfall in modulating the variability of spring ice dates in Maine lakes.

For this study, the historical ice-out and climate data for 12 Maine lakes are used. The next section discusses source of the ice-out and meteorological data and delineation of winter and spring season for this study. The methodology section provides a concise summary on the theory of circular regression, and the framework applied here for building and assessing circular regression models for ice-out dates. The result section discusses diagnostic results from model outputs and residuals for selected lakes. It also provides an assessment if winter and spring degree days and snowfall are adequate in explaining the variability in the spring ice-out dates in Maine lakes.

2. Data

2.1. Lake Ice-Out Date Data

Lake ice-out date refers to the date in spring, when winter ice completely disappears from the lake surface (Hodgkins et al., 2002). The annual spring ice-out dates from 1950 to 2010 for the 12 studied Maine lakes were obtained from a publication by U. S. Geological Survey (Hodgkins, 2010). Data from this database are selected because of the consistency over site of observation, and ice-out date definition for each lake. Furthermore, the main criterion for selecting these lakes is that they had more than 50 years of ice-out date data. Morphometric data for the 12 lakes are given in Table 1.

2.2. Temperature and Snowfall Data

The 1950 – 2010 daily temperature and snowfall data for each lake are obtained from the nearest U.S. Historical Climatology Network (USHCN) stations. Data from USHCN stations are preferred, because of the long period of serially complete data, genuine quality assurance, and control checks imposed on data. The procedures undertaken by USHCN for data quality control are described in Williams et al. (2007). Appropriate modeling of lake ice-out dates and meteorological variables necessitates that stations have more than 30 years of complete data. In this study, a year is considered complete if it contained 90% of the winter and spring temperature and snowfall data. Climate data from seven USHCN stations are used in this study.

Table 1
Summary Statistics of Ice-Out Dates From 1950 to 2010 for Selected Maine Lakes Using Circular Statistical Approach

	Circular statistics					
	Mean			von Mises distribution		
Lakes	μ_o	Date (Julian day)	Standard deviation	θ (SE)	κ (SE)	Watson's U^2 statistic (p value)
Damariscotta	1.74	11 April (101)	11.37	1.74 (0.03)	24.59 (4.41)	0.07 ($p > 0.1$)
China	1.80	14 April (104)	10.16	1.80 (0.03)	35.32 (6.35)	0.07 ($p > 0.1$)
Maranacook	1.84	17 April (107)	9.55	1.84 (0.02)	39.64 (7.13)	0.06 ($p > 0.1$)
Auburn	1.86	18 April (108)	8.73	1.86 (0.02)	44.3 (8.18)	0.04 ($p > 0.1$)
West Grand	2.00	26 April (116)	8.61	2.00 (0.02)	41.89 (7.54)	0.07 ($p > 0.1$)
Norway	1.92	21 April (111)	7.57	1.92 (0.02)	65.37 (12.42)	0.08 ($p > 0.1$)
Sebec	2.03	28 April (118)	7.28	2.03 (0.02)	65.49 (11.89)	0.08 ($p > 0.1$)
Mooselucmeguntic	2.15	5 May (125)	7.17	2.15 (0.02)	66.34 (12.17)	0.05 ($p > 0.1$)
Rangeley	2.17	6 May (126)	7.16	2.17 (0.02)	66.17 (11.94)	0.05 ($p > 0.1$)
Moosehead	2.16	5 May (125)	7.62	2.16 (0.02)	60.36 (10.88)	0.05 ($p > 0.1$)
Squapan	2.11	3 May (123)	6.05	2.11 (0.01)	88.12 (16.91)	0.04 ($p > 0.1$)
Portage	2.17	6 May (126)	5.78	2.17 (0.01)	100 (18.68)	0.04 ($p > 0.1$)

Note. SE = standard error.

Mohseni et al. (1998) has shown that climate data from meteorological stations, as far as 200 km from site, can be applicable in predicting stream water temperatures. Here the distance between lake and nearby meteorological station is on average about 20 km, while the maximum distance is about 80 km (see Figure 1).

2.3. Seasonal Degree Day Indices

The net energy balance at lake surface determines the formation, growth, and melt of surface ice cover, and air temperature is directly or indirectly related to the net long wave radiation, sensible heat, and latent heat flux. Consequently, seasonal winter temperature indices such as AFDD and AMDD have often been used to approximate the available freeze/thaw energy to form/melt lake ice (e.g., Ashton, 1986; Kirillin et al., 2012; Leppäranta, 2015). When calculating seasonal AFDD/AMDD, lake, and glacial ice studies, different temperature thresholds are employed for freezing/melting of water/ice, to compensate for different atmospheric conditions or sampling problems. However, in this study, the AFDD (AMDD) during the ice cover period is computed, as the daily degree days below (above) freezing of water (0 °C or 32 °F) summed over the total number of days when daily average temperature was below (above) freezing.

2.4. Delineating Winter and Spring Period in Maine

In lake ice studies, the winter season provides the bulk of freezing energy to grow lake ice, and consequently, the winter AFDD have often be used to gage the freezing energy available to form and grow ice. Thus, to delineate the winter months during the lake ice cover period in Maine, the smoothened mean profile of AFDD from 1950 to 2010 over the period between 1 December and 30 April is generated using nonparametric kernel estimators for each station (see Figure 2). Across the six stations, the mean (median) date when 90% of the winter and spring AFDD is attained lies prior to 10 March. Thus, in this study, winter season represents the period between December and February months, and spring season refers to the period between March and April.

3. Methodology

3.1. Circular Data: Lake Ice-Out Dates

Circular/directional data are the data that can be represented as locations (points) on the circumference of a unit circle (e.g., Lee, 2010). They are encountered in various scientific fields and are usually expressed as angles from an arbitrarily selected zero reference and sense of rotation. Examples of this type of data include readings of wind direction or animal orientation, relative to a reference direction. In addition to data that are initially measured as angles, circular data also applies to measurements such as time of day/year that show periodicity. In general, circular data have no natural ranking, since the origin and sense of rotation is

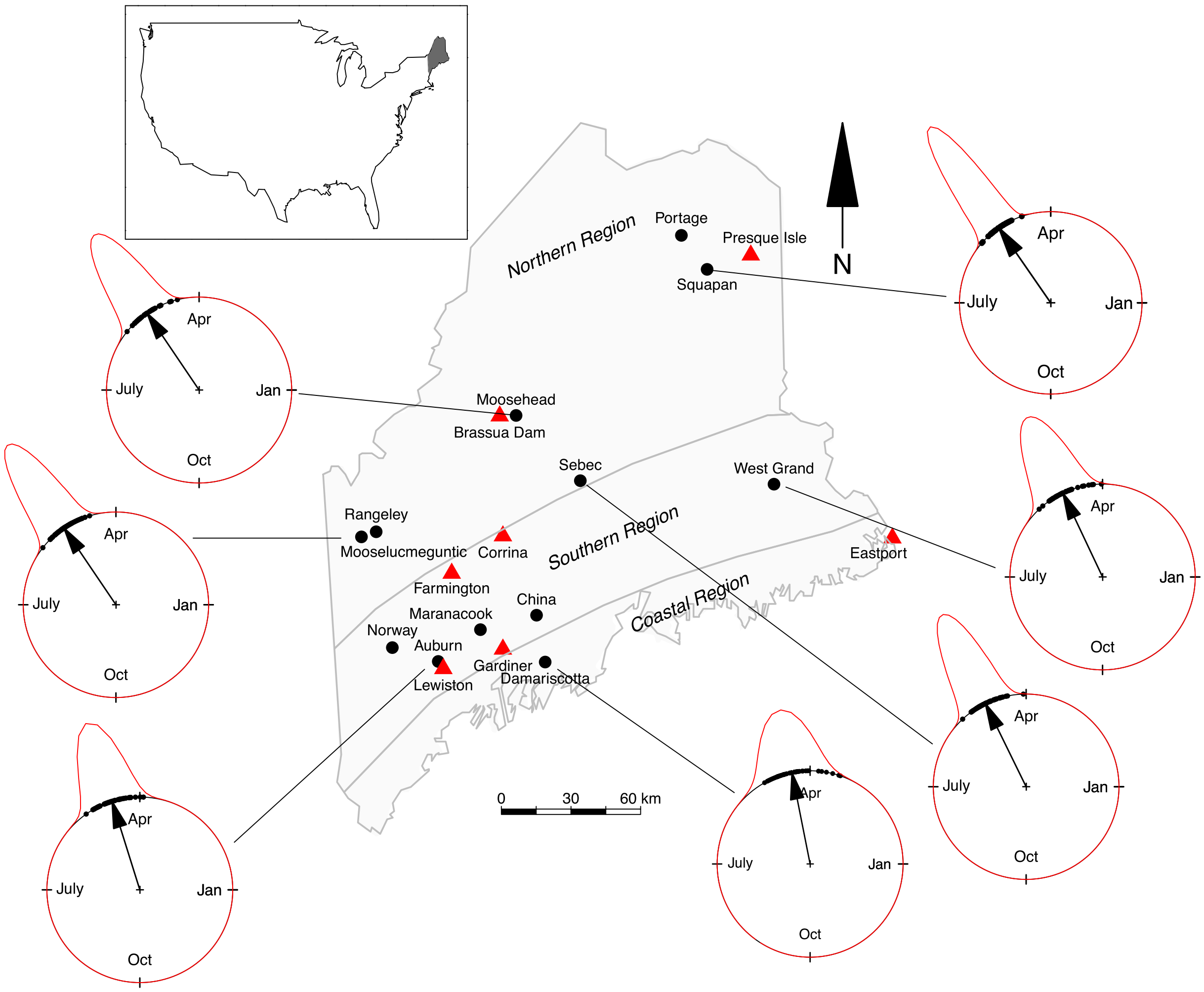

Figure 1. Location of selected Maine lakes (and meteorological stations) and circular plots of their spring ice-out dates. In the map, the filled red triangles and black circles represent the location of U.S. Historical Climatology Network stations and lakes, respectively, while the gray lines indicate the climate divisions in Maine. For each circular plot, the black dots and arrow denote the spring ice-out date of lakes from 1951 to 2010 and the mean ice-out date, respectively, while the red curves represent the fitted von Mises distribution for ice-out dates. Inset map of the United States shows the location of Maine.

arbitrary (e.g., Fisher, 1992). Furthermore, measurements are cyclical just as in a circle, the beginning coincides with the end (i.e., points 0 and 2π coincide). Thus, the use of conventional statistical methods in analyzing circular data often results in misleading or absurd results. For further discussions on the nature of circular data, reader is referred to books by Mardia and Jupp (2009), Jammalamadaka and Sengupta (2001), and Fisher and Lee (1992).

3.2. Summary Statistics for the Ice-Out Dates of Studied Lakes

As illustrated in earlier sections, circular statistical approaches are more appropriate method for describing calendar data such as lake ice-out dates. Thus, the historical ice-out date of studied lakes are transformed to angular data by computing

$$\theta_i = D_i \frac{2\pi}{D_{\text{year}}} \tag{1}$$

where D_{year} represents the number of days in a year. The variable D_i is the spring lake ice-out date in Julian day, and θ_i is its angular value in radians. Since θ_i also corresponds to the polar coordinates ($\cos\theta_i$, $\sin\theta_i$) of a location on a unit circle ($r = 1$), the historical spring ice-out dates of lakes can graphically be depicted as points on a unit circle.

The 1950–2010 climatology of the spring ice-out dates across studied lakes are characterized by estimating the mean and spread. For a sample of n ice-out dates, the sample mean ($\hat{\bar{\theta}}$) and variance ($\hat{\rho}$) are determined

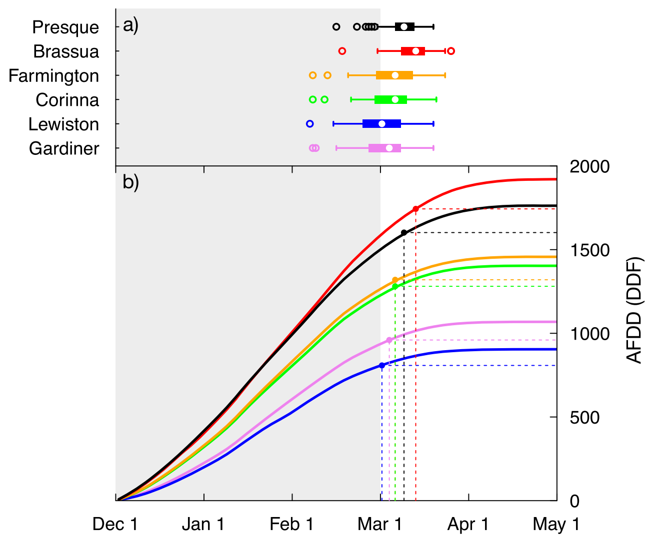

Figure 2. The profile of AFDD growth as a function of days between 1 December and 30 April for six USHCN stations in Maine from 1950 to 2010 . (a) Annual dates when 90% of the winter and spring AFDD is attained for six USHCN stations. Each box plot represents the distribution of the annual dates with the white circle indicating the median, the box denoting the 0.25th and 0.75th quantile range, and the whiskers representing the τ = 0.1 to τ = 0.9 quantile range. (b) The 1950–2010 mean smoothened profile of AFDD as a function of the days between 1 December and 30 April for six USHCN stations. This was generated using nonparametric kernel density estimator (Bowman & Azzalini, 1997). The filled circles on the mean profile denote the mean (median) date at which 90% of the winter and spring AFDD from 1950 to 2010 is attained. The violet, blue, green, orange, red, and black colors signify Gardiner, Lewiston, Corinna, Farmington, Brassua, and Presque Isle stations, respectively. The gray area designates the winter season. AFDD = accumulated freezing degree days; USHCN = U.S. Historical Climatology Network; DDF = Degree Day Fahrenheit.

by computing

$$\hat{\bar{\theta}} = \arctan\left(\frac{S}{C}\right) \tag{2}$$

$$\hat{\rho} = 1 - \sqrt{(S^2 + C^2)} \tag{3}$$

where

$$S = \frac{\sum_{i=1}^{n} \sin\theta_i}{n}$$

and

$$C = \frac{\sum_{i=1}^{n} \cos\theta_i}{n}$$

C and S represent the x and y coordinates of the mean lake ice-out date on the unit circle. The measure of dispersion (ρ) for angular data on a unit circle, ranges from $\rho = 0$ (corresponds to all ice-out dates occurring on the same date of the year) to $\rho = 1$ (indicates maximum variability). Alternatively, the circular standard deviation (σ) can be calculated using the equation

$$\hat{\sigma} = \sqrt{-2\ln(1-\rho)} \tag{4}$$

Figure 1 and Table 1 present the summary statistics for the historical ice-out dates of studied lakes. For the 1950–2010 period, they show that the mean spring ice-out dates of Maine lakes range from mid-April in coastal and southern interior lakes to early May in northern interior lakes. Moreover, contrasting the circular standard deviation of the ice-out dates for studied lakes indicate that coastal lakes have greater variability in their timing of their spring ice-out dates as compared to their inland counterparts.

3.3. Linear-Circular Correlation

Linear-circular (L-C) correlation ($R_{x\theta}$; Mardia, 1976) approach is utilized to measure the linear association between winter/spring climate variables and spring lake ice-out dates. Suppose X and θ denote the variables seasonal temperature/snowfall and spring ice-out date (in radians), respectively. L-C correlation coefficient is defined as the multiple correlations between X and angular components ($\cos\theta$, $\sin\theta$), assuming that the circular variable can be described by a random vector $v = (\cos\theta, \sin\theta)^{\mathrm{T}}$ in a plane (Mardia, 1976). For n pairs of X and θ, this can mathematically be written as

$$R_{x\theta} = \sqrt{\frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2}} \tag{5}$$

where the correlations are $r_{xc} = cor(X, \cos\theta)$, $r_{xs} = cor(X, \sin\theta)$, and $r_{cs} = cor(\cos\theta, \sin\theta)$. If the winter/spring climate variable and spring lake ice-out dates do not exhibit covariability, then $R_{x\theta}$ will approach zero. In contrast, if climate variable and spring lake ice-out dates are strongly associated with each other, then $R_{x\theta}$ will be close to ±1. Under the null hypothesis of no correlation between X and θ, the test statistics follows the distribution

$$\frac{(n-3)R_{x\theta}^2}{1 - R_{x\theta}^2} \sim F_{2,n-3} \tag{6}$$

given that X is normally distributed (page 246, ; Mardia & Jupp, 2009). In this study, correlation coefficients are taken as significant,if the hypothesis that there is no correlation between the two variables is unlikely with a probability of 0.95. For comprehensive expositions on L-C correlation coefficient and other forms of L-C correlation techniques, the reader is referred to books by Mardia and Jupp (2009) and Jammalamadaka and Sengupta (2001).

3.4. Circular-Circular Correlation

The circular-circular (C-C) correlation ($r_{\theta\phi}$) approach is used to assess the rotational association between two time series of lake ice-out dates. For n pairs of θ and ϕ, this can be mathematically expressed as (page 176, ; Jammalamadaka & Sengupta, 2001)

$$r_{\theta\phi} = \frac{\sum_{i=1}^{n} \sin(\theta_i - \bar{\theta}) \sin(\phi_i - \bar{\phi})}{\sqrt{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sin^2(\phi_i - \bar{\phi})}} \tag{7}$$

where $\bar{\theta}$ and $\bar{\phi}$ are the sample mean directions. If $r_{\theta\phi}$ is close to zero, it suggests that the two time series of lake ice-out dates are rotationally independent. On the other hand, when $r_{\theta\phi}$ approaches ± 1, it indicates that there is a strong rotational association between the two time series of lake ice-out dates. Under the null hypothesis of no correlation between ϕ and θ, the test statistics

$$t = \sqrt{f} r_{\theta\phi} \tag{8}$$

follows a standard normal distribution. The term f is given by

$$f = N \frac{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sum_{i=1}^{n} \sin(\phi_i - \bar{\phi})}{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sin^2(\phi_i - \bar{\phi})} \tag{9}$$

Here circular correlation coefficients are taken as significant, if the assumption that there is no correlation between the two time series of ice-out dates is unlikely with a probability of 0.95. For further discussions on C-C correlation coefficient or other forms of L-C correlation techniques, the reader is referred to comprehensive expositions in the published literature (Jammalamadaka & Sengupta, 2001; Mardia & Jupp, 2009).

3.5. L-C Regression

Figure 3 shows the procedural framework applied here for inferring the efficacy of winter and spring climate parameter(s) on the variability of lake ice-out dates. The framework employs preliminary circular diagnostic methods described in earlier sections, as well as assess outputs from candidate ice-out models of varying complexity, developed using the circular regression approach described below. In the latter approach, model building was done in stages beginning with the null model, which presumes that spring ice-out date variability is dependent exclusively on the spring AFDD and AMDD. Subsequently, spring snowfall, winter AFDD and AMDD, and winter snowfall are sequentially added in the following models and fitted. Model parameter significance and outputs across candidate models are then compared to assess the relevance of winter and/or spring climate variables in controlling the variability of spring ice-out dates across Maine lakes.

3.5.1. von Mises Distribution

The von Mises distribution, first proposed by von Mises in 1918, is the most common and best studied of unimodal circular distributions. The reasons for its popularity is (a) its results are easier to interpret, as its inference techniques are well developed; (b) it is flexible with regard to the effect of parameters; (c) it has an in-built measure for scale (dispersion; Jammalamadaka & Sengupta, 2001). Thus, it plays a central role in circular statistics, akin to the Normal distribution for linear data analysis. The von Mises probability density function for random variable ϕ is given by

$$f(\phi; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\phi - \mu)}, \qquad 0 \le \phi, \mu < 2\pi; \kappa > 0 \tag{10}$$

where $I_0()$ is the modified Bessel function of zeroth order and μ is the mean direction and κ is the concentration parameter. When $\kappa > 0$, the density is unimodal and symmetrical about the μ and as κ increases, the distribution increasingly becomes tightly clustered. For $\kappa > 2$, the distribution can be well approximated by a Normal distribution with mean μ and variance $(1/\kappa)$ (Fisher & Lee, 1992).

10.1029/2017WR021731

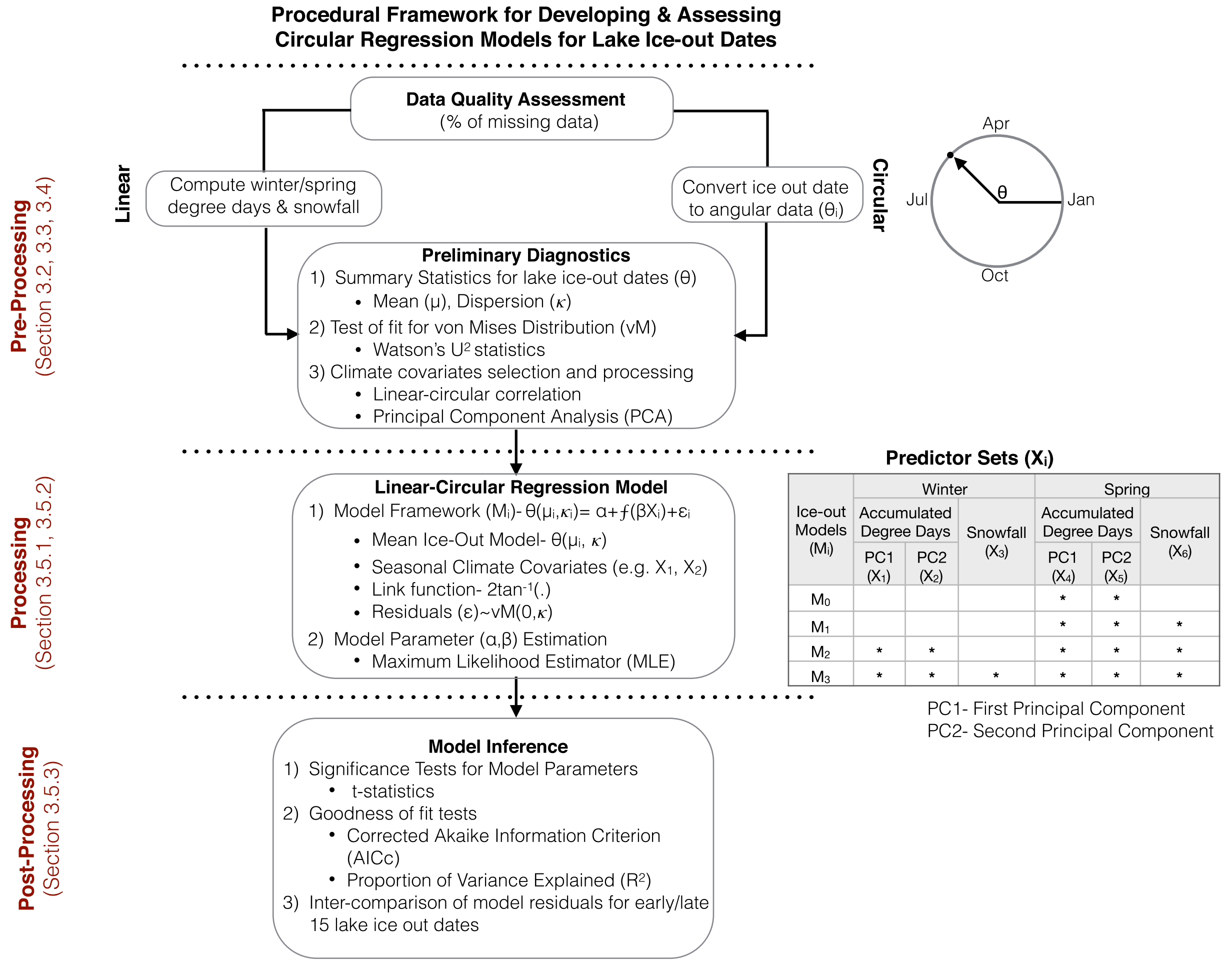

Figure 3. Procedural framework for developing and assessing circular regression models for ice-out dates: three preliminary diagnostic techniques, circular regression model framework, and parameter estimation method and three approaches for assessing model outputs and performance.

The appropriateness of von Mises distribution in representing the probability distribution of ice-out dates for studied lakes are verified using Watson's U^2 test (Lockhart & Stephens, 1985). This test compares the mean squared deviation between the empirical cumulative distribution function and true cumulative distribution function, at all data points, to the deviation against the critical value at alpha level. Results show that the assumption the probability structure for lake ice-out dates has a von Mises distribution cannot be rejected at 90% significance level for all lakes (see Table 1).

3.5.2. Model Framework and Fitting

In the L-C regression models with von Mises distribution, winter and/or spring climate variables are described as linear covariates, and the lake ice-out dates is represented as a circular response variable. In general, the response of lake ice-out date of to seasonal climate variables can be modeled by regressing (a) the mean direction (μ), (b) the dispersion (κ), (c) or both the mean direction (μ) and dispersion of ice-out dates to winter/spring climate variables. In the present study, we focus on the first model and therefore model the mean direction μ of ice-out dates to climate covariates x_i as

$$\mu_i = \mu_0 + g(x_i\beta) + \epsilon_i \tag{11}$$

where β corresponds to the vector of regression parameters to be estimated, μ_0 is the circular mean of the dependent variable, $g()$ is the link function, and ϵ_i is the residual term from the von Mises $vM(0, \kappa)$ distribution. The purpose of the link function $g()$ is to convert linear variables to circular ones. Possible choice of link functions are discussed in Fisher and Lee (1992) and Jammalamadaka and Sengupta (2001); however, in this study we used

$$g(u) = 2\tan^{-1}() \tag{12}$$

The maximum likelihood estimates, for the parameters (μ_0, β, and κ) of a homoscedastic von Mises regression model, are the values that maximize the log likelihood function

$$L(\beta|X) = -N \log I_0(\kappa) + \kappa \sum_{i=1}^{N} \cos(\theta_i - \mu_0 - 2\tan^{-1}(\beta X_i)) \quad (13)$$

Often, the determination of the maximum likelihood estimates requires the use of iterative procedures. The circular package (Agostinelli & Lund, 2013) in the *R* statistical computing environment provided the optimizing algorithm to estimate μ, β, and κ. Furthermore, large sample asymptotic variance is used to estimate standard errors for the parameters, and to test hypotheses (Fisher & Lee, 1992).

3.5.3. Model Diagnostics and Inference

Model inference on the relative importance of winter/spring climate variables on spring ice-out dates is based on (a) parameter significance tests, and (b) comparing model fitness. Significance tests for model parameter estimates indicate whether there is a detectable relationship between the response variable and predictive variable(s) under focus, for a given level of certainty. In the present study, the significance of model parameter estimates for winter/spring degree days and snowfall are determined using *t* statistics. Model parameters are considered significant, if the significance level (p) is less than 0.10. On the other hand, if models are successively fitted in order of increasing complexity, comparing the relative fit for successive pairs of models provides an alternative means of assessing the null hypothesis that the omitted (added) term(s) has no significant contribution on the spring ice-out date variability. In this study, Model 0 (M_0), Model 1 (M_1), and Model 2 (M_2) are special cases of M_1, M_2, and M_3, respectively, and thus, comparison of, say, M_0 and M_1 using goodness of fit tests is a test of the hypothesis that spring snowfall has no influence on the timing of spring lake ice-out dates. The relative fit across models is determined using coefficient of determination (R^2) and bias-corrected Akaike Information Criterion (*AICc*). The coefficient of determination (R^2), which measures the variability explained by the model, was computed for each model by squaring the C-C correlation between observed and model simulated ice-out dates (Lund, 1992). In addition, to balance between model complexity and model fitting of data, the corrected *AICc* was employed. AICc is a likelihood-based criterion that quantifies the relative amount of information lost, if a given ice-out model is used to approximate the underlying process that generates the observed lake ice-out dates. Assuming that the ice-out model residuals have a von Mises distribution with concentration parameter $\hat{\kappa}$, the *AICc* of a given model is given as

$$AICc = 2n \log I_0(\hat{\kappa}) - 2n\hat{\kappa} + \frac{n(n+1)}{n-l-2} \quad (14)$$

where $I_0()$ is the modified Bessel function of zeroth order, n is the sample size, and l is the number of estimated parameters (degrees of freedom) in the model.

4. Results

4.1. Seasonal Meteorological Covariates

Six seasonal meteorological variables were considered in the development of our circular regression models for ice-out dates (see Figure 4). This section provides a rationale, both physical and statistical, for the inclusion of these variables.

4.1.1. Seasonal Winter Degree-Days and Lake Ice-Out

The thickness of winter ice cover over lakes determines the amount of heat energy needed to melt and clear the ice from lake surface. The preceding winter AFDD and AMDD quantities can have strong influence on the timing of lake ice-out dates in spring. In this study, the winter AFDD (AMDD) is computed as the daily degree days below (above) freezing (0 °C or 32 °F) summed over the total number of days during December and February when daily average temperature was below (above) freezing.

L-C correlation tests between winter AFDD and AMDD and spring ice-out dates for studied Maine lakes shows that the preceding winter AFDD has a significant positive correlation (ρ = 0.25–0.53, $p < 0.05$) with the spring ice-out dates for lakes across all three-climate regions in Maine (see Tables S2a–S2f in the supporting information). This indicates that the higher winter AFDDs—that is, the larger the freeze content to grow ice—the longer for the winter ice to clear in spring, and vice versa. Spatial comparison of the correlation coefficient across studied lakes show that the correlation between spring ice-out dates and winter AFDDs is higher in coastal and southern interior regions (ρ = 0.31–0.53) as compared with northern interior regions

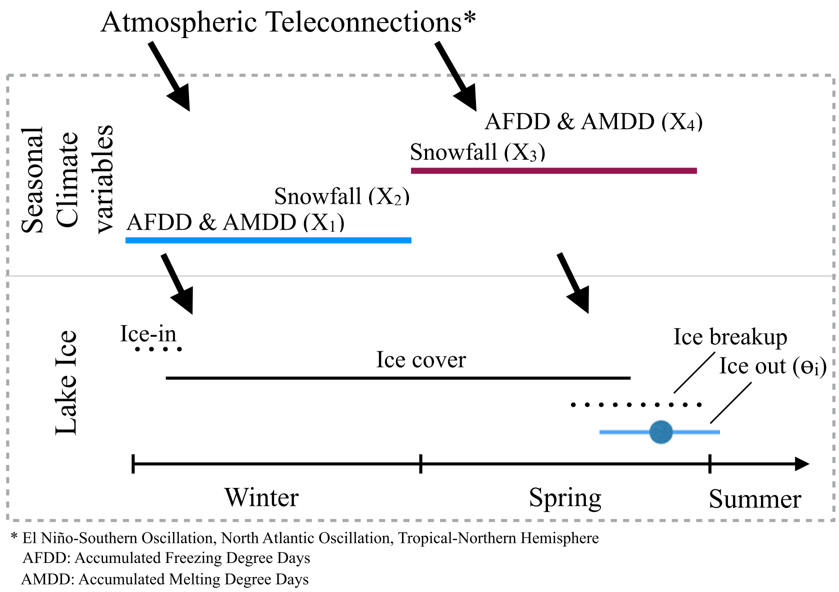

Figure 4. Schema representing the seasonal evolution of lake ice and linked climatic variables. The climate covariates (X) and ice-out date (θ) are shown along the winter-spring seasonal timeline.

(ρ = 0.24–0.45). The underlying physical reasoning for the reduced influence of winter AFDD on the variability of spring ice-out dates of northern interior regions are (a) formation and growth of ice in northern interior Maine lakes begins in fall, which reduces the role of winter AFDD variability on lake ice thickness (b) lake ice-out dates in the northern interior regions often occur later in spring (May–June), which allows spring climate variables to moderate the effect of winter AFDD on the spring ice-out dates.

On the other hand, seasonal winter AMDD shows a significant negative correlation ($|\rho|$ = 0.25-0.55, $p < 0.05$) with spring lake ice-out dates primarily in coastal and southern interior regions (see Table S2). This implies that the higher winter AMDDs—that is, the lower the cold content in the ice cover—the earlier than usual spring ice-out dates and vice versa. Spatially, the relative correlation between winter AMDD and spring lake ice-out dates decreases toward interior Maine regions. The major factor for the reduced influence of winter AMDD on spring ice-out dates of inland Maine lakes is that over the region, daily temperatures during December and February months seldom, if ever, exceed freezing point (0 °C or 32 °F).

4.1.2. Principal Component Analysis for Winter Degree Days

There is a significant ($p < 0.1$) negative correlation between winter AFDD and AMDD for most Maine regions (see Tables S2a–S2f). Consequently, using both winter degree day indices as predictor variables in lake ice-out date regression models generates parameter estimates for winter AFDD and AMDD that are less reliable and physically meaningful, due to collinearity effect. Fekedulegn et al. (2002) showed that transforming the original variables into a new set of orthogonal uncorrelated variables using principal component analysis (PCA) eliminates this effect. Thus, PCA (using the covariance matrix) was performed on the time series of winter AFDD and AMDD at each meteorological station. Across the six meteorological stations, the first principal component (PC1) of winter degree days represents 97–99.5% of the total variability and therefore may be considered as the dominant pattern of winter degree-day variability (see Table S3a). Furthermore given that the magnitude and variance of winter AFDD is much larger than that of winter AMDD for all stations, the loading of winter AFDD in each PC1 pattern is positive and over 0.99, whereas the loading of winter AMDD is negative and less than 0.03 (see Table S3b). The result implies that PC1 patterns primarily reflect the winter AFDD conditions over lakes, and as such positive (negative) PC1 indices represent above (below) average winter AFDDs. Spatially, there is a strong positive correlation ($\rho > 0.91$) between PC1 patterns across stations, which shows strong regional coherence in the temporal pattern of winter AFDD variability in Maine. On the other hand, the second principal component (PC2) reproduces only 0.5–3.0% of the total variability in winter degree days across the six meteorological stations (see Table S3a). Furthermore, for each PC2 pattern, the loading of winter AFDD is negative and less than 0.03, while the loading of winter AMDD is positive and over 0.99 (see Table S3b). This suggests that PC2 scores predominantly represent the winter AMDD conditions over

10.1029/2017WR021731

lakes and positive (negative) PC2 phases imply above (below) normal winter AMDD. There is a strong positive correlation ($\rho > 0.83$) between the PC2 patterns across stations, which implies strong spatial synchroneity in the pattern of winter AMDD variability in Maine.

4.1.3. Winter Snowfall

Winter snowfall can affect the thickness and type of winter ice cover and in turn the timing of spring ice-out dates by (a) reducing rate of ice growth in winter by adding insulation, (b) increasing the winter cold (freeze) content of lake ice, and (c) promoting the development of snow ice, thereby affecting the composition of the winter ice cover over lakes (Adams, 1976). In this study, time series of winter snowfall for the study period was determined by summing the daily total snowfall recorded at each station from the beginning of December to the end of February for each year.

L-C correlation results show that winter snowfall has a significant positive correlation ($\rho = 0.28 - 0.43$, $p < 0.05$) with spring lake ice-out dates chiefly in coastal and southern interior regions (see Tables S2a – S2f). This means that the higher winter snowfall, the longer the duration of winter ice. Furthermore, the coefficient and significance of this linear association decreases toward the interior regions. Again, this is mainly because lake ice-out date in deep interior regions occurs relatively later in spring (May – June), which allows spring climate conditions to have more influence on the timing of ice breakup date. It should be noted that winter snowfall has little or no correlation with the two principal components of winter degree days (see Table S2). This indicates that the relationship between winter snowfall and spring ice-out dates is independent of the prevailing winter temperature conditions.

4.1.4. Seasonal Spring Degree Days and Lake Ice-Out

Spring is the period when the bulk of the ice melting occurs. Williams (1965) showed that spring temperature largely determine the melt rate of winter ice cover. Hodgkins et al. (2002), based on their correlation results, suggested that the prevailing average spring (March – April) temperatures explain 50 – 70% of the variability in the timing of spring ice-out dates in New England lakes. In this study, the spring AFDD (AMDD) is computed as the daily degree days below (above) freezing (0 °C or 32 °F) summed over the total number of days during March and April when daily average temperature was below (above) freezing.

L-C correlation results show that seasonal spring AFDD has significant negative correlation ($|\rho| = 0.35 - 0.65$, $p < 0.05$) with lake ice-out dates across all climate regions in Maine (see Tables S2a – S2f). This implies that the higher the spring AFDDs — that is, the higher the cold content in lake ice — the later the spring ice-out dates and vice versa. Furthermore, the role of spring AFDD on spring lake ice-out dates of Maine lakes across the three climate divisions appears to be uniform, as the correlation coefficients do not show any systematic spatial patterns.

On the other hand, seasonal spring AMDD has significant positive correlation ($\rho = 0.73 - 0.84$, $p < 0.05$) with lake ice-out date in all climate regions (see Tables S2a – S2f). This suggests that the higher the spring AMDD — that is, the higher the melt energy at lake surface — the earlier the timing of ice-out dates. Spatially, the strength of correlation between spring AMDD and lake ice-out dates increases toward the interior regions. This is mainly because lake ice-out date in deep interior regions occurs relatively later in spring (May – June), which allows spring temperatures to have more influence on the timing of ice breakup date.

4.1.5. PCA for Spring Degree Days

There is a significant ($p < 0.1$) negative correlation between spring AFDD and spring AMDD in all stations (see Tables S2a – S2f). To reduce collinearity effect, the spring AFDD and AMDD variables at each lake are orthogonalized into two principal components and these principal components are included as predictor variables in the ice-out date regression models.

Across the six stations, the first principal component (PC1) of spring degree days represents 77 – 87% of the total variability in spring AFDD and AMDD and therefore may be considered as the leading pattern of local spring degree day variability (see Table S4a). Furthermore, in each of the PC1 patterns, the loadings for spring AFDD (AMDD) is of negative (positive) sign, which implies that when PC1 is in the positive phase, spring AFDD (AMDD) is lower (higher) than normal (see Table S4b). However, the loading of spring AFDD and AMDD in PC1 pattern varies across different climate regions in Maine with spring AMDD having relatively higher loading than spring AFDD in stations found in southern interior and coastal regions and vice versa for stations found in northern interior regions. This is because even in spring months, daily temperatures get below 32 °F for a significant period of time in northern interior Maine regions. Spatially, there is a strong positive correlation

10.1029/2017WR021731

($\rho > 0.87$) between spring PC1 patterns across stations, which suggests that there is spatial coherence between spring PC1 patterns across Maine regions.

The second principal component (PC2) for spring degree days reproduces 17–23% of the total variability across the six stations (see Table S4a). Furthermore, in each of the PC2 patterns, the loading for spring AFDD and AMDD are both positive implying positive (negative) PC2 phases are related to higher (lower) than normal spring AFDD and AMDDs over lakes (see Table S4b). AFDD and AMDD values gage the range of the seasonal temperature distribution. As such seasonal conditions where both the AFDD and AMDD are either high or low occur are indicative of a change in the variability of seasonal temperatures. Consequently, PC2 indices show a strong positive correlation ($\rho > 0.68$) with the intraseasonal standard deviation of spring temperatures at all stations where positive (low) spring PC2 phases are related to high (low) intraseasonal spring temperature variability in all stations (see Figure S1). Similar to spring PC1 of spring degree days, the loading of spring AFDD and AMDD in PC1 pattern varies across different climate regions in Maine although here spring AMDD having relatively lower loading than spring AFDD in stations found in southern interior and coastal regions and vice versa for stations found in northern interior regions. There is a strong spatial correlation ($\rho > 0.68$) between the PC2 patterns of spring degree days across the six stations, which implies of a regional synchroneity in the PC2 variability patterns across Maine regions.

Another important result of note is that the two principal components for spring degree days show little or no correlation with that of the winter degree days in all regions (see Table S2). This indicates an absence of climatic persistence between winter and spring degree day variability.

4.1.6. Spring Snowfall

Spring snow accumulation can reduce the melt rate of ice cover by (a) increasing the albedo (thereby lowering radiation absorption) of the ice cover (b) increasing the cold content of the ice cover. In this study, the annual spring snowfall from 1950–2010 was determined by summing the daily total snowfall from the beginning of March to the end of April for each year.

L-C correlation results reveal that spring snowfall has significant positive correlation ($\rho = 0.28$-0.55, $p < 0.05$) with the timing of spring ice-out dates of studied lakes across all climate regions in Maine (see Table S2). This suggests that the more spring snowfall, the longer the duration of ice over lakes. Furthermore, the correlation coefficients across different regions indicate that the correlation coefficient for spring snowfall is lower in northwestern Maine regions (0.28–0.31) as compared with other regions (0.41–0.55).

4.2. Model Output and Inference

The four circular models for each lake describe the variability of spring ice-out dates, as a function of the prevailing winter and/or spring degree days, and snowfall. Key model results are summarized in Tables S5a–S5d for studied lakes. The $2\tan^{-1}(\,)$ used as link function between covariates and ice-out dates (see equation (10) and (11) has both transformative and multiplicative effect on changes in the covariates. For instance, a coefficient of −0.001 associated with spring PC1 implies that an increase by 1 unit in spring PC1 is associated with a multiplicative decrease of $2\tan^{-1}(-0.001)$ radians or 6.6 days in spring ice-out dates.

Model 0 (M_0), a model that explains spring ice-out date variability as a function of the two principal components for spring degree days, captures over 50% of the total variability in ice-out dates of studied lakes (see Table S5a and Figure 5). The prevailing spring temperatures have a strong control over the timing of spring ice-out dates in Maine lakes. However, the efficacy of spring degree days in modulating the timing of spring ice-out dates is not the uniform across Maine lakes, as the performance of M_0 in studied lakes shows variations at regional, and to a lesser extent local scales (see Table S5a). For instance, the explained variance (R^2) by M_0 for studied coastal and southern interior Maine lakes is less than 60%, while for most northern interior lakes, M_0 represents at least 65% of the total variance. Also in Northern interior Maine regions, M_0 captures over 70% of the total variance in high altitude lakes such as Lake Rangeley and Lake Mooselucmeguntic, while this is much lower in relative low altitude lakes such as Lake Portage and Lake Squapan. At local scale, the M_0 for relatively large, deep lakes such as Lake Damariscotta and Moosehead shows higher unexplained variance as compared with that of relatively small, shallow lakes in the same climate division.

In M_0, the coefficient for PC1 of spring degree days is negative, and statistically significant ($p < 0.1$) across all lakes, while the parameter for PC2 is positive and statistically significant for studied lakes, with the exception of Lake Damariscotta and Lake Norway (see Table S5a). Thus, positive PC1 phases (lower than average spring AFDD and higher spring AMDD) are related to earlier than normal spring ice-out dates in lakes, whereas

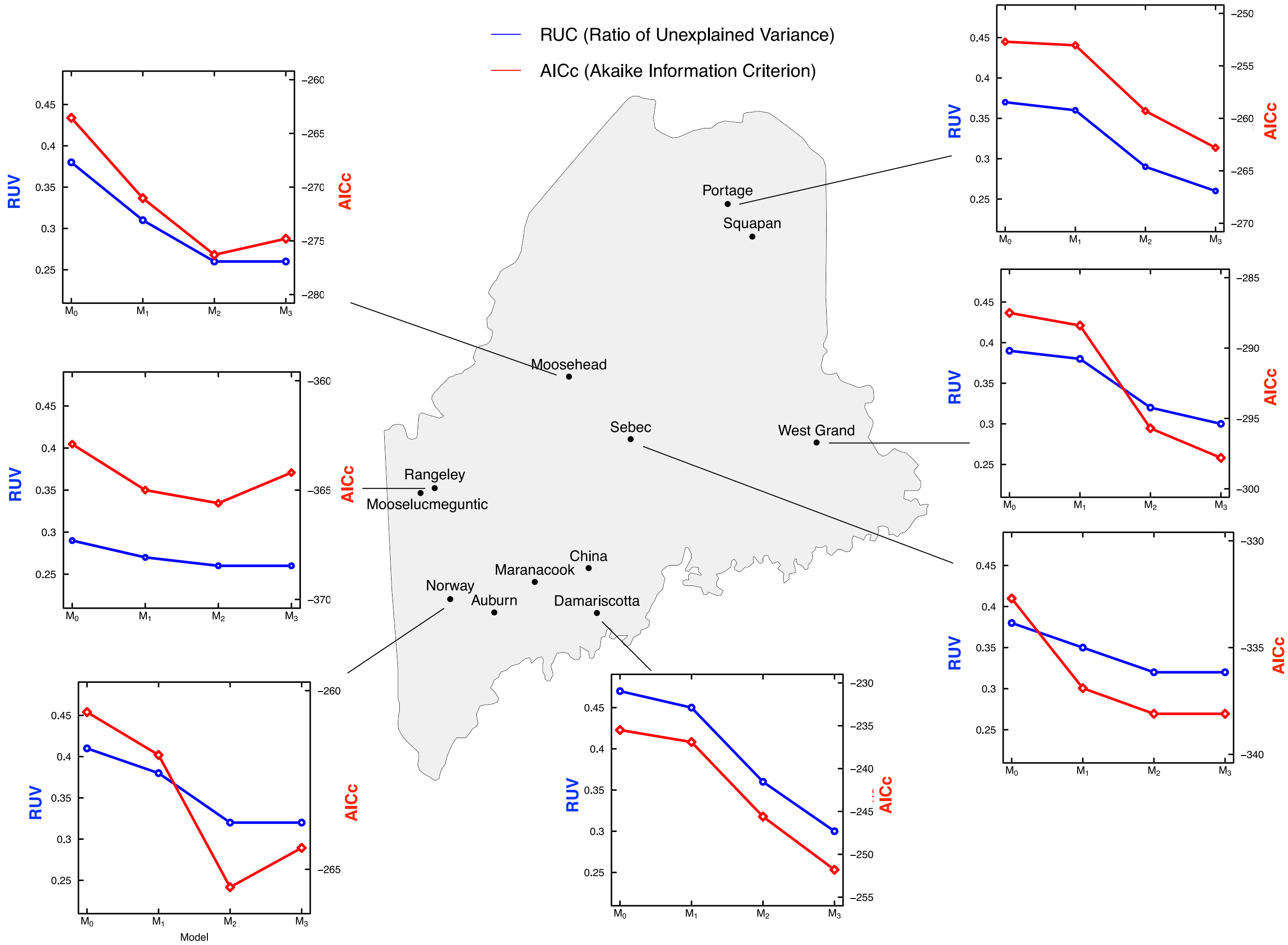

Figure 5. Relative performance of candidate ice-out models for select Maine lakes. M_0, M_1, M_2, and M_3 denote the four candidate ice-out models developed for each lake. The red diamonds and blue circle represent the AICc and RUV values for each lake ice-out date model, respectively.

positive PC2 indices (higher than normal spring AFDD and AMDD) are linked to later than normal spring ice-out dates. Furthermore, PC2 indices represent the intraseasonal variance in spring degree days, and these results suggest that the timing of spring ice-out dates in Maine lakes is sensitive not only to the magnitude of spring degree days but also on the intraseasonal variability of spring temperatures. Comparing the parameter estimates for PC1 and PC2 patterns across the M_0 of studied lakes reveals that in general the coefficient for PC1 (PC2) is of higher (lower) magnitude in southern interior and coastal Maine lakes relative to northern interior lakes (see Table 2a).

Model 1 (M_1), which includes spring snowfall in addition to the two principal components of spring degree days to explain spring ice-out date variability, captures 56–75% of the total variability in the ice-out date of studied lakes (see Table S5b and Figure 5). Assessing the R^2 and AICc of M_1 relative to that of M_0 across studied lakes indicates that the dependence of spring ice-out dates in Maine lakes to spring snowfall shows regional variations based on climatic divisions, and to a lesser extent altitude. For instance, Figure 5 shows that the change in R^2 and *AICc* from M_0 to M_1 is higher in northern interior Maine lakes such as Lake Portage and Lake Sebec, as compared to that of southern and coastal lakes such as Lake Damariscotta and West Grand. Also for northern interior regions, the change in model fitness from M_0 to M_1 is higher in lower altitude lakes such as Lake Moosehead and Lake Portage, relative to that of high-altitude lakes such as Lake Rangeley and Lake Mooselucmeguntic.

In M_1, the coefficient for spring snowfall is positive and statistically significant for all studied lakes, except for Lake Mooselucmeguntic (see Table S5b). This implies that the higher the spring snowfall, the later the spring iceout dates. Comparison of parameter estimates for spring snowfall across the M_1 models studied

Table 2
Comparing Key Statistics of Circular and Linear Regression Models for Lake Squapan (κ = 88)

		Regression coefficients						Model fitness		
		Winter			Spring					
Model type	Model	PC1 (10^{-4})	PC2 (10^{-4})	Snowfall (10^{-4})	PC1 (10^{-4})	PC2 (10^{-4})	Snowfall (10^{-3})	R^2	MAE (days)	RMSE (days)
Circular regression	M_0				−3.7***	3.4***		0.65	2.9	3.6
Linear regression	M_0				−7.4***	6.9***		0.65	2.9	3.6
Circular regression	M_1				−3.3***	3.9***	1.5***	0.73	2.3	3.2
Linear regression	M_1				−6.5***	7.7***	3.0***	0.73	2.3	3.2
Circular regression	M_2	0.5**	−0.1		−3.2***	3.9***	1.3***	0.74	2.2	3.0
Linear regression	M_2	0.9*	−1.8		−6.4***	7.8***	2.5**	0.74	2.2	3.0
Circular regression	M_3	0.4**	−1.0	0.3*	−3.2***	4.0***	1.3***	0.75	2.2	3.0
Linear regression	M_3	0.9*	−1.8	0.5	−6.3***	8.0***	2.6***	0.75	2.2	3.0

Note. The significance of model parameter estimates was determined using *t* statistics method and circular regression parameters. *MAE* = mean absolute error; *RMSE* = root-mean-square error. *Significant at 0.1 significance level. **Significant at 0.05 significance level. ***Significant at 0.01 significance level.

lakes shows that the parameter coefficients for spring snowfall are of lower magnitude in northwestern lakes (Rangeley and Mooselucmeguntic) as compared with lakes in other Maine regions. This result is in consensus with the correlation analysis (section 4.1.6) that the northwestern high altitude lakes have lesser sensitivity to spring snowfall than lakes in other regions.

Model 2 (M_2), with two principal components of winter degree days in addition to the predictor variables in M_1, explains 67–77% of the variability in ice-out dates in studied lakes (see Table S5c and Figure 5). Comparison of model fitness metrics between M_2 and M_1 indicates that the efficacy of winter degree days in modulating the timing of spring ice-out dates in Maine lakes, is higher in large, deep coastal, and southern Maine lakes as compared with small, shallow, and northern interior lakes. For instance, the relative change in R^2 and AICc from M_1 to M_2 is higher for southern and coastal lakes such as Lake Damariscotta and Lake West Grand, as compared to that of northern interior lakes such as Lake Portage and Lake Rangeley. Also in northern interior regions, the improvement in model fitness from M_1 to M_2 is higher in large, deep lakes such as Moosehead, as compared with small, shallow lakes such as Squapan or Portage.

In M_2, the coefficient for PC1 of winter degree days is positive and statistically significant ($p < 0.1$) across all studied lakes, while the parameter for PC2 is negative and statistically significant only for coastal lakes (see Table S5c). In general, this implies that positive PC1 phases (higher than normal winter AFDDs) are associated with later than average spring lake ice-out dates, while positive PC2 phases (higher than normal winter AMDDs) are related to earlier than average spring lake ice-out dates. Comparison of parameter estimates for PC1 and PC2 of winter degree days across the M_2 models of studied lakes reveals that the coefficient for PC2 is of higher magnitude in coastal lakes, relative to lakes in other Maine regions. This indicates that coastal and southern interior lakes have higher sensitivity to PC2 indices (i.e., winter AMDD), as compared with those in northern interior lakes. This conclusion is consistent with the finding in the correlation analysis from earlier section.

Model 3 (M_3), which includes winter snowfall in addition to M_2 predictor variables to model spring ice-out dates, captures 68–77% of the total variability of ice-out dates in studied lakes (see Table S5d and Figure 5). Assessing the change in model fitness metrics between M_2 and M_3 for studied lakes indicates that the modulating influence of winter snowfall on spring ice-out dates is higher for coastal Maine lakes. For instance, Figure 5 shows that the relative change in R^2 and *AICc* from M_3 to M_2 is higher for southern and coastal lakes such as Lake Damariscotta and Lake Auburn as compared to that of northern interior lakes such as Lake Portage and Lake Rangeley.

In M_3, the coefficient for winter snowfall is positive and statistically significant for coastal Maine lakes such as Maranacook, Damariscotta, and China and Auburn (see Table S5d). This implies that the higher the winter snowfall over lakes, the later the spring ice-out dates. Comparison of parameter estimates for winter snowfall in M_3 models across studied lakes shows that the coefficient for winter snowfall in coastal lakes in higher in

10.1029/2017WR021731

coastal lakes as compared to lakes in other Maine regions. This implies that the sensitivity of spring ice-out dates to winter snowfall is higher in coastal lakes as compared with interior Maine regions.

4.2.1. Model Residual Diagnostics

By determining the incremental information added by incorporating winter and/or spring climate variables, the previous section assessed the overall efficacy of winter and spring climate variables, in modulating the timing of spring ice-out dates in Maine lakes. However, this provides limited insight into the role of these variables in producing unusually early/late spring ice-out dates in Maine lakes. Thus, in this section, the import of winter and/or spring variables in producing large departures in the timing of spring ice-out events in Maine lakes was assessed by contrasting the residuals for the earliest and latest 10 ice-out dates, across candidate models (i.e., M_0-M_3). The premise in such assessments is that majority improvement in the model estimation of the earliest/latest ice-out events, after the inclusion of a seasonal climate variable/s, implies the importance of the seasonal climate variable on the occurrence of these events. Figures 6 and S5 depict intermodel residuals for various ice-out dates, and the gray area in these plots represents regions where the estimate made by complex model has lesser error than that of the reduced model(s). In general, the pattern of M_0 residuals for the earliest and latest 10 ice-out dates reveals that the ice-out date models, conditioned on the two principal components of spring degree days, performs poorly when estimating ice-out dates before mid-April for coastal and southern interior lakes and after 10 May for northern interior lakes. Counting the number of earliest/latest 10 lake ice-out events (upper and lower triangles) falling in the gray area of M_1-M_0 residual plot for studied lakes, reveals that for lake Sebec, Portage, Mooselucmeguntic, Squapan, Norway, and Auburn, more than half of the residuals for both the earliest and latest 10 spring ice-out dates are in the gray area of M_1-M_0 subspace. For instance, it can be observed in Figure S5k that the M_1-M_0 residuals for 7 (8) of the 10 earliest (latest) ice-out dates at Lake Portage are in the gray area within M_1-M_0 space. This indicates that the efficacy of spring snowfall in engendering early/late ice-out dates in Maine lakes, is the highest in northern interior regions. Similarly, tallying the number of earliest/latest 10 ice-out events within the gray area of M_2-M_1 residual plots, for studied lakes, shows that for Lake Sebec, Moosehead, West Grand, Norway, Damariscotta, China, and Auburn, more than 60% of both the earliest and latest ice-out events are within the gray area of M_2-M_1 subspace. For instance, Figure 6 shows that the M_2-M_1 residuals for 9 of the 10 earliest (latest) ice-out dates at Lake Norway are in the gray area within M_2-M_1 space. This indicates that the efficacy of daily winter temperatures, in producing early/late ice-out dates in Maine lakes, is the highest in coastal and southern interior lakes, and large northern interior lakes. Finally, counting the number of earliest/latest 10 ice-out events within the gray area of M_3-M_2 residual plots, for studied lakes, shows that for Lake Maranacook, China, Damariscotta and Sebec, 70% or more of the earliest 10 ice-out events are within the gray area of M_3-M_2 subspace. For example, the M_3-M_2 residuals for 8 of the 10 earliest ice-out dates at Lake Damariscotta are in the gray area, within M_3-M_2 space (see Figure S5c). This implies that the antecedent winter snowfall quantity, over coastal regions, has a significant modulating effect on the occurrence/non-occurrence of the earliest ice-out dates of lakes.

High coherence in model residuals for pairs of lakes, indicates how well the ice-out model performs in estimating the spring ice-out dates for these lakes. For instance, if the M_0 residuals for two lakes are highly correlated, in years where M_0 overestimates (underestimates) the ice-out date for one of these lakes, M_0 also tends to overestimate (underestimate) the ice-out date for the other lake as well. Therefore, the pairwise (circular) correlation between model residuals of selected lakes was determined, across the four ice-out models developed. Results show that across the four ice-out models, the strength of correlation between model residuals for two lakes varies depending on the similarity/difference in their respective climate division, and to a lesser extent proximity from each other (see Figure 7). For instance, correlation between model residuals for Lake China and Lake Maranacook across M_0 to M_3 ranges from 0.82 to 0.87, while these correlations between Lake China and Lake Presque ranges from 0.20 to 0.27. This is because ice cover seasons for lakes in the same climate regions have similar sensitivity to winter and spring meteorological variables, given that the prevailing climate conditions over lakes, in the same climate divisions are analogous. On the other hand, the correlation between model residuals for two lakes in general decreases with increasing complexity of ice-out date models (see Figure 7).

In addition to these four seasonal climate variables studied in this paper, the year-to-year variability of spring lake ice-out dates can be modulated by other climatic/nonclimatic variables such as wind, cloudiness, or lake depth. As such, it is not surprising to observe years where all four models for studied lakes underperform by relatively large margins (> 5days). For example, the four ice-out date models developed for all 12 lakes overestimate the timing of spring ice-out dates for the year 2002. To understand the underlying climatic factors, the

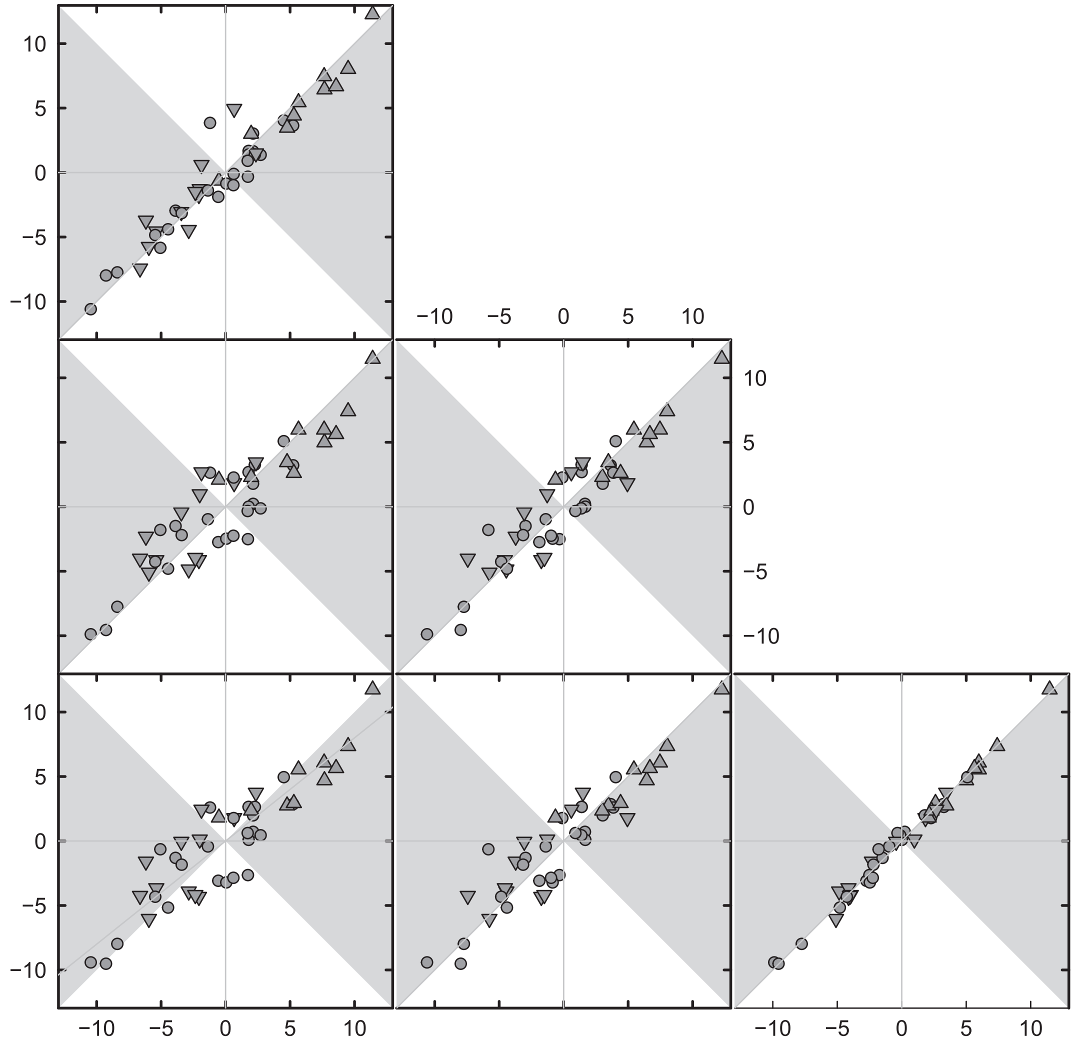

Figure 6. Intermodel residual comparison for ice-out dates at Lake Norway. Model residuals are calculated as the difference between observed and model predicted ice-out dates. The gray region represents the region within bivariate model residual space where estimates from relatively complex models show improvement than that of simpler ones. The filled triangles pointed upward (downward) symbolize the earliest (latest) 15 ice-out dates, while the dots represent other ice-out dates.

prevailing meteorological conditions during the winter and spring of 2002 were scrutinized. It was noted that during the winter and spring of 2002, there was unusually high amount of precipitation in the form of rain. Given that rainfall promotes the melting of lake ice by reducing albedo and freeze content of surface lake ice the four candidate ice-out models are vulnerable to overestimation, when rainfall has a significant influence in modulating the timing of ice-out date of lakes.

4.2.2. Comparison of Circular Regression Models to Traditional Linear Regression Models

As noted in section 1, linear models/methods are not appropriate for analyzing ice-out date(a circular random variable) due to model specification. However, Table 1 shows that the kappa for ice-out dates of studied lakes is greater than 26, and according to Fisher and Lee (1992), von Mises distribution with $\kappa > 2$ can be well approximated using normal distribution. Thus, using the traditional linear regression (TLR) method, ice-out models of varying complexity were developed for studied lakes (see Tables S6a–S6d), and for two of these lakes, the resulting model coefficients and model errors were compared to that of circular models (see Tables 3 and 2). Results reveal that TLR models with only spring degree days explain over 50% of the total variance in ice-out date for Lake Damariscotta and Lake Squapan. They also show that the inclusion of winter meteorological variables in TLR models for ice-out dates reduces model estimate error for both lakes. The consistency in TLR and circular regression model is generally expected for cases with small variance. However, in lakes where

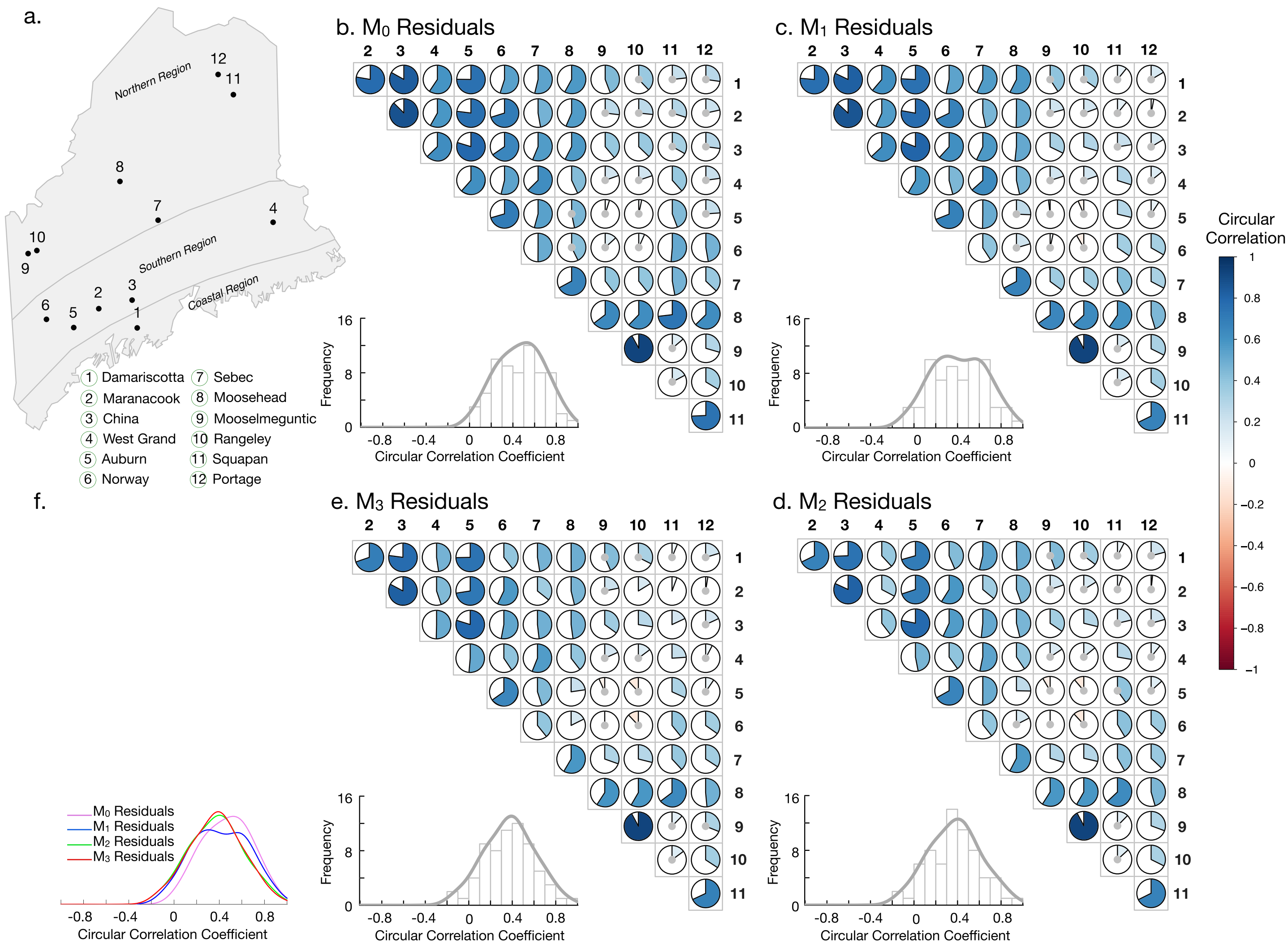

Figure 7. Circular correlations in model residuals for selected Maine lakes across M_0 to M_3. (a). Location of selected lakes in Maine. The dots represent the position of lakes in map while the gray lines denote the climate divisions in Maine. (b–e) Correlation plots of model residuals for selected lakes across M_0 to M_3. In each correlation plot, the filled area in pie charts denotes the degree of correlation between a pair of lakes, while the gray dots inside pie chart signify correlation coefficients that are not significant. (f) Distribution of correlations between model residuals of pair of lakes for the four ice-out date models . The violet, blue, green, and red curves represent distribution of correlations for M_0, M_1, M_2, and M_3.

the timing of ice-out dates shows large variance ($\kappa > 2$), the difference in model results and performance between TLR models and circular regression models is expected to be prominent. As such, the L-C framework developed in this study offers a parsimonious statistical approach to model the effect of linear meteorological variables on lake ice phenology, particularly in a changing climate wherein warmer temperatures are poised to induce increased variability in ice-out dates.

5. Discussion and Conclusions

This study presented a circular regression framework for modeling ice-out dates, conditioned on a suite of winter and spring climate variables (i.e., degree days and snowfall), to determine the import of winter and spring climate conditions on the timing of ice-out dates in Maine lakes. Winter/spring AFDD and AMDD variables were orthogonalized into two principal components to reduce collinearity effect in ice-out date models, and these principal components were included as predictor variables in the circular regression models for ice-out dates. Parameter significance tests and inter-model fitness tests (R^2 and AICc) across candidate ice-out

Table 3
Comparing Key Statistics of Circular and Linear Regression Models for Lake Damariscotta (κ = 25)

		Regression coefficients						Model fitness		
		Winter			Spring					
Model type	Model	PC1 (10^{-4})	PC2 (10^{-4})	Snowfall (10^{-4})	PC1 (10^{-4})	PC2 (10^{-4})	Snowfall (10^{-3})	R^2	MAE (days)	RMSE (days)
Circular regression	M_0				−6.9***	2.2		0.53	5.4	6.9
Linear regression	M_0				−13.3***	4.2		0.53	5.4	6.9
Circular regression	M_1				−6.2***	2.5	1.5***	0.55	5.2	6.7
Linear regression	M_1				−12.3***	4.9	2.9	0.54	5.2	6.7
Circular regression	M_2	1.0***	−9.7***		−5.1***	1.4	1.6**	0.64	4.4	5.8
Linear regression	M_2	2.1**	−19.3***		−10.2***	2.8	3.3*	0.64	4.4	5.8
Circular regression	M_3	1.1***	−7.5***	1.5***	−4.6***	7.9	1.7***	0.70	4.4	5.6
Linear regression	M_3	2.3**	−14.8**	3.0**	−9.2***	1.7	3.3**	0.69	4.4	5.6

Note. The significance of model parameter estimates was determined using *t* statistics method and circular regression parameters. *MAE* = mean absolute error; *RMSE* = root-mean-square error. *Significant at 0.1 significance level. **Significant at 0.05 significance level. ***Significant at 0.01 significance level.

date models revealed that for Maine lakes:

a. The joint effect of seasonal spring degree days (AFDD and AMDD) and the intraseasonal variance in spring temperatures explains more than half of the total variability in spring lake ice-out dates in Maine. Spatially, the modulating influence of spring temperature conditions on lake ice-out dates increases toward northern interior Maine regions.
b. The relative role of spring snowfall in engendering early/late ice-out dates in Maine lakes is the strongest in northern interior region.
c. The efficacy of the antecedent winter degree days (AFDD and AMDD) in modulating variability of lake ice-out dates is significant, across all climate regions in Maine. However, the strongest effects are observed in large, coastal lakes.
d. The relative influence of winter snowfall on lake ice-out date variability is significant, ($p < 0.1$) only in coastal Maine regions.

A diagnostic analysis of years in which all four ice-out models developed most underperformed by more than 5 days, indicated unexplained variance likely stemming from other hydro-climatic processes and lake dynamics. In closing, we put forward the following remarks, and discuss emerging research directions.

1. This study focused on the efficacy of different climatic and nonclimatic variables in modulating the interannual variability of spring ice-out dates of temperate lakes. Results indicate that in addition to spring degree days, the state of the winter degree days and winter and spring snowfall contributes significantly to the overall year-to-year variability of ice-out dates including the occurrence of early/late spring ice-out dates of Maine lakes. Future works on this topic is still needed including determining the role of other climatic/no-climatic variables on the interannual lake ice-out date variability, the use of different link functions in circular ice-out date models and performance of non-parametric circular regression approach for modeling ice-out dates.
2. Large-scale teleconnections patterns produce North American climate anomalies at a regional scale. For Maine, it has been shown that the Tropical/Northern Hemisphere and North Atlantic Oscillation patterns influence interannual winter temperature variability (Beyene & Jain, 2015). Given that the Climate Prediction Centers in North America and Europe routinely provides skillful forecast of these climate patterns, winter meteorological conditions derived from such information can be incorporated in circular ice-out date models conditioned on winter climate variables, to provide season-ahead outlooks on the spring time lake ice season in Maine.
3. Data involving time-of-year variables are prevalent in hydrology and hydrometeorology and to date, conventional approaches that characterize date-of-year variables as linear continuous data are often employed to analyze such data. A number of studies have shown that such approaches produce erroneous results (Jammalamadaka & Sengupta, 2001). This study presents a systematic framework and highlights (a) the

applicability of circular statistical approaches in modeling circular/periodic data such as lake ice phenology (b) the availability of circular counterparts for traditional linear data analysis techniques for environmental systems analysis and modeling.

Acknowledgments
This study is supported by National Science Foundation awards 0904155 and 1055934 and NOAA award NA14OAR4320158. The ice-out date data for the 12 lakes in this study can be accessed at https://pubs.usgs.gov/of/2010/1214/pdf/ofr2010-1214.pdf. The temperature and snowfall data are available at http://cdiac.ornl.gov/ftp/ushcn_daily/. We sincerely thank the Editor, Associate Editor and three anonymous reviewers for their constructive comments toward improving our manuscript.

References

Adams, W. P. (1976). Diversity of lake cover and its implications. *Musk-Ox Journal*, *18*, 86–98.
Agostinelli, C., & Lund, U. (2013). R package circular: Circular statistics (version 0.4-7). CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA. Retrieved from https://r-forge.r-project.org/projects/circular/
Ashton, G. D. (Ed.) (1986). *River and lake ice engineering*. Water Resources Publications.
Beyene, M. T., & Jain, S. (2015). Wintertime weather-climate variability and its links to early spring ice-out in Maine lakes. *Limnology and Oceanography*, *60*(6), 1890–1905.
Bowman, A. W., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations* (Vol. 18). Oxford, UK: Oxford University Press.
Fekedulegn, B. D., Colbert, J. J., Hicks, R. R., & Schuckers, M. (2002). *Coping with multicollinearity: An example on application of principal components regression in dendroecology*. Newton Square, PA: Department of Agriculture, Forest Service, Northeastern Research Station.
Fisher, N. I. (1992). Statistical analysis of circular data: Statistical analysis of circular data.
Fisher, N. I., & Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, *48*(3), 665–677.
Hampton, S. E., Galloway, A. W., Powers, S. M., Ozersky, T., Woo, K. H., Batt, R. D., et al. (2017). Ecology under lake ice. *Ecology Letters*, *20*(1), 98–111.
Hodgkins, G. (2010). Historical ice out dates for 29 lakes in New England, 1807–2008 (Open-File Report 2010-1214). Reston, VA: U. S. Geological Survey.
Hodgkins, G., James, I., & Huntington, T. (2002). Historical changes in lake ice-out dates as indicators of climate change in New England, 1850–2000. *International Journal of Climatology*, *22*(15), 1819–1827.
Jammalamadaka, S. R., & Sengupta, A. (2001). Topics in Circular Statistics. Singapore: World Scientific Publishing.
Kirillin, G., Leppäranta, M., Terzhevik, A., Granin, N., Bernhardt, J., Engelhardt, C., & Efremova, T. (2012). Physics of seasonally ice-covered lakes: A review. *Aquatic Sciences*, *74*(4), 659–682.
Kratz, T. K., Hayden, B. P., Benson, B. J., & Chang, W. Y. B. (2001). Patterns in the interannual variability of lake freeze and thaw dates. *Internationale Vereinigung fur Theoretische und Angewandte Limnologie Verhandlungen*, *27*(5), 2796–2799.
Lee, A. (2010). Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 477–486.
Leppäranta, M. (2010). The impact of climate change on european lakes (Vol. 4). In G. George (Ed.), *The impact of climate change on European lakes, Aquatic Ecology Series* (pp. 63–83). Dordrecht, Netherlands: Springer.
Leppäranta, M. (2015). *Freezing of lakes and the evolution of their ice cover*. Germany: Springer-Verlag Berlin Heidelberg.
Livingstone, D. M. (1997). Break-up dates of alpine lakes as proxy data for local and regional mean surface air temperatures. *Climatic Change*, *37*(2), 407–439.
Lockhart, R. A., & Stephens, M. A. (1985). Tests of fit for the von Mises distribution. *Biometrika*, *72*(3), 647–652.
Lund, U. (1992). Least circular distance regression for directional data. *Journal of Applied Statistics*, *26*(6), 723–733.
Magnuson, J. J, Robertson, D. M, Benson, B. J., Wynne, R. H., Livingstone, D. M., Arai, T., et al. (2000). Historical trends in lake and river ice cover in the northern hemisphere. *Science*, *289*(5485), 1743–1746.
Mallows, C. (1998). The zeroth problem. *The American Statistician*, *52*(1), 1–9.
Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, *63*(2), 403–405.
Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics*. Chichester, UK: John Wiley & Sons.
Mohseni, O., Stefan, H. G., & Erickson, T. R. (1998). A nonlinear regression model for weekly stream temperatures. *Water Resources Research*, *34*(10), 2685–2692.
Prowse, T., Alfredsen, K., Beltaos, S., Bonsal, B. R., Bowden, W. B., Duguay, C. R., et al. (2011). Effects of changes in arctic lake and river ice. *Ambio*, *40*(1), 63–74.
Vavrus, S. J., Wynne, R. H., & Foley, J. A. (1996). Measuring the sensitivity of southern Wisconsin lake ice to climate variations and lake depth using a numerical model. *Limnology and Oceanography*, *41*(5), 822–831.
Weyhenmeyer, G. A., Livingstone, D. M., Meili, M., Jensen, O., Benson, B., & Magnuson, J. J. (2011). Large geographical differences in the sensitivity of ice-covered lakes and rivers in the northern hemisphere to temperature changes. *Global Change Biology*, *17*(1), 822–831.
Williams, G. P. (1965). Correlating freeze-up and break-up with weather conditions. *Canadian Geotechnical Journal*, *2*(4), 313–326.
Williams, C. N., Menne, M. J., Vose, R. S., & Easterling, D. R. (2007). United States Historical Climatology Network monthly temperature and precipitation data (ORNL, CDIAC-187, NDP-019). Oak Ridge, TN: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory.