**Supplement: Detailed explanation of breakpoint analysis**

*Continuous Segmented Regression*

We used segmented regression to test for abrupt changes in the trend of ice dates in Lake Suwa and the river Tornio. Specifically, we wanted to test when a shift in the temporal trend of ice date may have occurred. To estimate the timing and magnitude of a change in the slope of ice dates for Tornio and Lake Suwa, we used continuous segmented regression (CSR) models. In CSR, trend lines on either side of the estimated breakpoint intersect (hence making them “continuous”), but are allowed to have different slopes. In general, a CSR takes the form

[1]



where yi are observations of ice date, is a latent variable representing potentially unobserved ice dates ( and yi only differ in Tobit model, described below), x­i are the years of the time series, β0 is the intercept of the regression (ice date on year 0), β1 is the temporal trend in ice date (change in ice date per change in year), the ak are the breakpoints (k was either 1 or 2), the βk+1 are the changes in the temporal trend at each of the k breakpoints, and the εi are the errors. Note that the βk+1 parameters indicate the effect of years elapsed since the breakpoint once the breakpoint has passed on ice date.



*Fitting CSR in Tornio (OLS)*

The Tornio time series began in 1693 and ended in 2013, thus x = 1, 2, … 321. Ice breakup dates for Tornio ranged from day 117 to day 160, and the ice melted each year of the time series. For Tornio, we fit CSR parameters using ordinary least squares using the lm() function in the statistical programming language R.

*Fitting CSR in Lake Suwa (Tobit)*

The Lake Suwa time series began in 1443 and ended in 2004 (x = 1, 2, … 562), and ice observations were made for 427 of the 562 years (See Fig X in Main Text). The day that Lake Suwa froze ranged from day -54 to day 41 (negative values indicate freezing before January 1st of the designated “year”); however, there were 37 years when the lake did not freeze. Treating no-freeze years as missing data or as a constant date would result in biased results if we employed the regression techniques used for Tornio. Thus, calculating trends and breakpoints for Lake Suwa ice dates required statistical approach distinct from that used in Tornio. If the lake is considered as an instrument that measures a value, which we call ice date, that indicates the favorability of conditions for ice formation, and if we understand the lake instrument to censor these measurements at 41, then the no-freeze years can be encoded as ice dates of 41. We consider Lake Suwa as an instrument with output of ice date that is censored at an upper limit, L = 41. As such, the observed yi are related to L and the latent variable in the following manner:



[2]



To address this censoring of Lake Suwa ice dates while fitting the parameters in Eq. 1, we used a Tobit regression model. For a Tobit regression model with an upper limit (right censoring) of the response variable, the log likelihood of observing data given the parameters β (as in Eq. 1) and σ2 (the variance of ε in Eq. 1), can be calculated as:

[3]



where φ(.) and Φ(.) are the probability and cumulative density functions of the normal distribution, respectively. The first term is the standard normal likelihood, and applies to observations for which an ice date was observed. The second term reflects the probability of the observation being censored, and applies to no-freeze years. Given parameter values, Eq. 3 reflects the probabilities of observing the ice dates (yi) during freeze years, as well as the probabilities that ice date was censored (unobserved) during no-freeze years. Thus, the β in the Tobit regression model indicate the effect of unit change in X on the latent variable, . We used Tobit regression models as implemented by the vglm() function in the R package VGAM to fit parameters in Eq. 1 to Lake Suwa data.



*Finding Breakpoint Locations*

When fitting models with one breakpoint, breakpoints were searched exhaustively, and the breakpoint location whose model had the lowest AIC was selected. The same procedure applied to fitting the two-breakpoint model for Tornio, which was fit with OLS (Figure S1). Because The Tobit model requires substantially more computational power, and because the Lake Suwa time series is longer (thus more possible breakpoint combinations), breakpoint combinations were not searched exhaustively for Lake Suwa. Instead, we used the genetic optimization algorithm in the rgenoud package (Walter R. Mebane, Jr., Jasjeet S. Sekhon. 2011. Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software* **42**:11, 1-26). This algorithm permits integer optimization, and is robust to rough likelihood surfaces and local minima.

*Model Selection*

We compared AIC values from CSR models containing one or two breakpoints to multiple regression models containing only year or only year and year2 as predictor variables. We fit models either by OLS (Tornio) or by maximum likelihood of the corresponding Tobit regression (Lake Suwa) (Table S1). For Tornio and Lake Suwa, the biggest decrease in AIC between models of consecutive complexity was the comparison in AIC for the 2nd order polynomial and the one breakpoint model (ΔAIC for Tornio and Suwa was -1.8 and -2.2, respectively; Figure S1 for Tornio). Although modest, these changes in AIC suggest that a model with a single breakpoint was a reasonable descriptor of the trend in ice date in both systems.

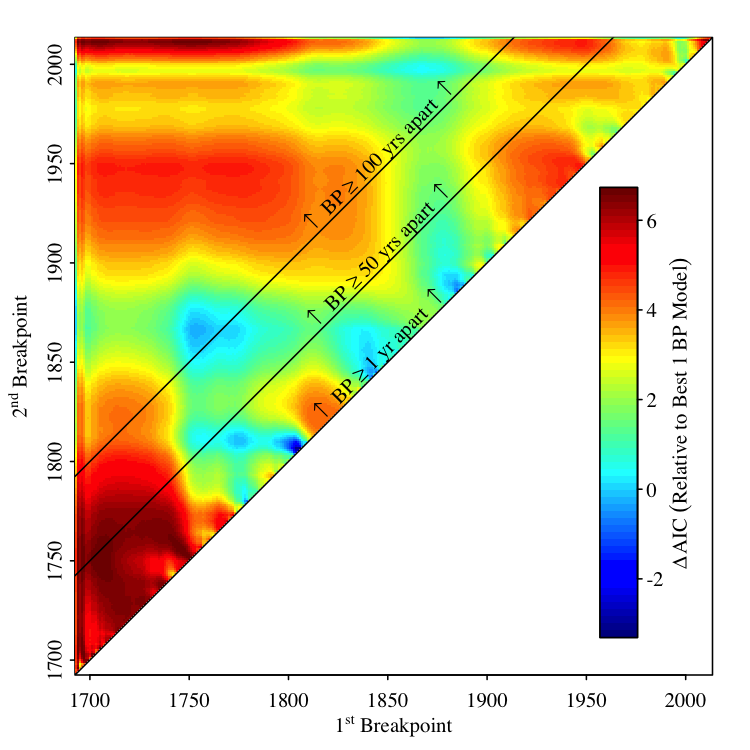
**Table S1.** AIC values of fitted regression models relating ice date () to years elapsed (xi).



|  |  |  |
| --- | --- | --- |
| Model | Tornio AIC | Suwa AIC |
|  | 2155.825 | 3536.38 |
|  | 2154.881 | 3515.407 |
|  | 2153.072 | 3513.241 |
|  | 2152.774\* | 3511.682\*\* |

\*Breakpoints restricted to being at least 10 years apart; See Figure S1.

\*\*Breakpoints restricted to being at least 50 years apart; if restricted to 25 years, AIC = 3510.898.



**Figure S1.** Relative probabilities of two versus one breakpoint in the Tornio time series. Colors indicate change in AIC for the two breakpoint model relative to the one breakpoint model (Table S1) for all combinations of first and second breakpoint years in the two breakpoint model. Sloped lines2 indicate boundaries where the first and second breakpoints are separated by the indicated period of time. Note that when at least 25 years separates breakpoints, the one breakpoint model is always more parsimonious than the two breakpoint model (Table S1).

**Supplement: Detailed explanation of driver analysis**

Results from our breakpoint analysis suggest that both systems experienced shifts in the temporal trend of ice date. Despite the geographical distance separating these time series of river ice breakup and lake ice formation, breakpoints were identified at relatively similar points in time (1807 in Lake Suwa, 1867 in Tornio). Thus, the abrupt changes in ice date trends are unlikely to be driven solely by system-specific forcings. If the abrupt nature of the shift in ice trend is driven by climate, then climatic drivers of ice date should either exhibit the same abrupt transition, or the relationship between ice date and its climatic drivers must change. To test these hypotheses, we explored linear relationships between ice date and the following climate drivers (unless otherwise specified, drivers apply to both systems): air temperature (AirT), aerosol optic depth (AOD), atmospheric CO2, El Niño Southern Oscillation index (ENSO, Suwa only), North Atlantic Oscillation index (NAO, Tornio only), Sunspots (Tornio only), and years elapsed since the beginning of the time series (Year). In each system, the relationships between ice date and each of the climate drivers were explored using data before and after the breakpoint. In Lake Suwa, the duration of the period on both sides of the breakpoint was 101 years (1581–1681 and 1897–1997), and the time periods were chosen to maximize duration on either side of the breakpoint while minimizing the inclusion of years for which ice data are missing. In Tornio, we used the full time series on either side, giving 174 years in the first portion (1693 – 1866) and 146 years in the second portion (1867 – 2013). For Tornio, air temperature data were not available prior to 1803, thus the early time period for analyses involving Tornio AirT was only 64 years (1803 – 1866).

For each ice date – driver pair in each system, we performed separate linear regressions for the period before and the period after the breakpoint (periods as described above). This procedure resulted in twenty-two separate regressions, and each before-after pair of regressions is equivalent to fitting a single model of the form

y = β0 + β1\*x1 + β2\*x2 + β3\*x1\*x2 + ε [4]

where y is ice date, β0 is the intercept, x1 is the driver variable and β1 its effect, x2 is a dummy variable that is 0 before the breakpoint and 1 after, ergo β2 is the post-breakpoint intercept, and β3 is the change in the relationship between the driver and ice date after the breakpoint (i.e., β3 is the adjustment made to β1 after the breakpoint). As described for the breakpoint analysis, for Tornio Eq. 4 was fit with ordinary least squares, and with the Tobit regression for Lake Suwa. For each regression, we assumed there was a possibility that the residuals of the regression would be autocorrelated; to estimate coefficient standard errors in the presence of autocorrelation, we used a bootstrapping procedure where the randomized residuals retained the autocorrelation structure of the regression residuals.

To characterize the autocorrelation structure of the residuals in Eq. 4, we fit an autoregressive moving average (ARMA(p,q)) model with *p* AR parameters and *q* MA parameters. An ARMA(p,q) model has the general form

 [5]

The yt are the residuals ε from Eq. 4, and μ is the mean of the residuals, which is 0. The βi are the AR parameters, the αj are the MA parameters, and the εt-j are the residuals. We applied Eq. 5 to the ε of Eq. 4. We selected among ARMA models using AIC, and allowed model complexity to vary from ARMA(1,0) to ARMA(5,5) (including all orders in between). These models were fit and selected using the stepwise procedure implemented in the auto.arima function in the forecast R package (Rob J. Hyndman and Yeasmin Khandakar. 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* **27**:3). We then simulated an ARMA process using the fitted ARMA parameters; the variance of the innovations in the simulated ARMA process was the maximum likelihood estimate acquired in fitting Eq. 5 to the residuals.

The ARMA-simulated residuals formed the basis of our bootstrapping procedure. These simulated residuals were then added to the fitted regression values, and the regression was re-fit. This procedure was repeated 1,000 times. The standard deviation of these 1,000 parameter estimates was then used as the standard error of the parameters in Eq. 4. In summary, our bootstrapping procedure was as follows:

1. Fit Equation 4
2. Divide residuals from #1 into before and after period
3. Fit time series model to each set of residuals (Eq. 5)
4. Simulate new sets of residuals from fitted time series model
5. Add simulated residuals from #4 to fitted values () from Eq. 4
6. Re-fit Eq. 4
7. Repeat steps 2-6 1,000 times
8. The standard error of parameters in #1 is the standard deviation of all estimates in #6

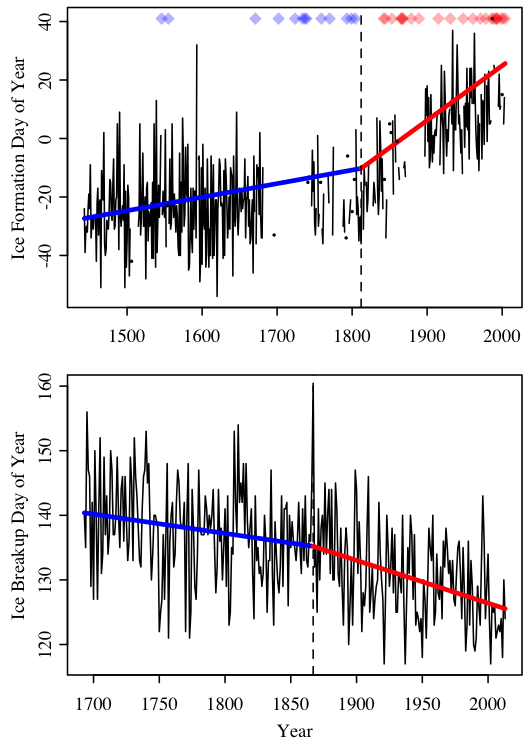
We calculated p-values corresponding to the probability that regression coefficients between drivers and ice dates differed before and after the breakpoints in the following manner:

Z = (β2 – β1)/(s.e.2 + s.e.1)-1/2

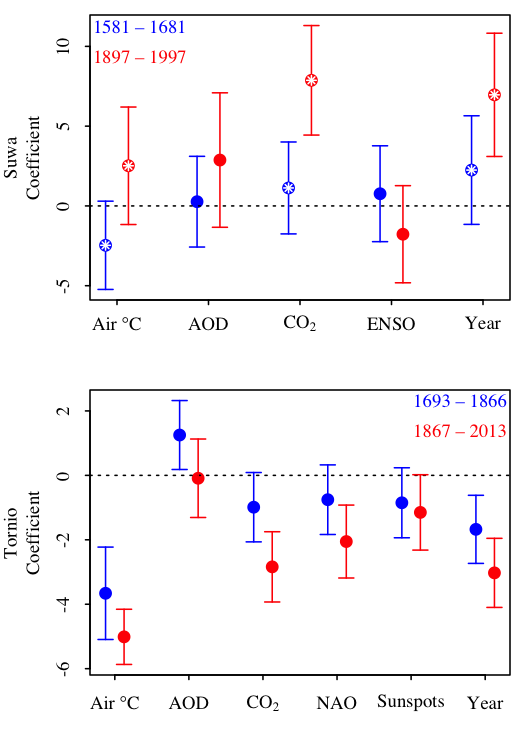
where Z is the z-score to be compared under the standard normal curve, the β are the regression coefficients, and s.e. are the standard errors of those coefficients. We also corrected these p-values to control for multiple tests and to maintain constant family wise error rates. We performed the Holm-Bonferroni correction, and no conclusions about significance among pairs changed.

For the Tornio data set (for which the response is ice breakup), negative coefficients indicate that a driver is unfavorable to ice, whereas in the Lake Suwa data set (for which the response is ice formation) positive coefficients indicate that a driver is unfavorable to ice. Of the 11 pairs of coefficients, 10 of them (all except ENSO in Lake Suwa) indicated that a unit change in driver led to less favorable ice conditions in the post-breakpoint part of the time series relative to the first part of the time series. I.e., the per-unit impact of these climate drivers was heightened in a manor that was unfavorable to ice. However, examined individually, no pairs of coefficients were significantly different in Tornio, and only 3 pairs were significantly different in Lake Suwa.

**Rough Figure Captions for Sapna & John**

****

**Figure 1 (time series w/ breakpoints).** Long-term observational records of ice dates for Lake Suwa in Japan (top panel, ice formation) and for the river Tornio in Finland (bottom panel, ice breakup). Solid black lines indicate ice dates. Solid black lines and small black dots indicate ice dates (black dots are used when no observation were made during adjacent years). For Lake Suwa, colored diamonds indicate years when observations were made, but during which the lake never froze. Vertical dashed lines are placed at the year of the breakpoint in the trend of ice date. Thick blue lines indicate the temporal trend before the breakpoint, and thick red lines indicate the temporal trend after the breakpoint.



**Figure 2 (deltaDrivers).** Regression coefficients between drivers and ice dates for Lake Suwa and the river Tornio. [Sapna, fill in months and units? Note that I standardized predictors, so the units don’t matter much b/c I divided by them; I’ve come to worry that that might reduce the statistical significance of our results (too conservative), but I’m not sure] Colored dots indicate estimated coefficients, with blue being the coefficient estimated from data before the breakpoint, and red the period after. Error bars are the 95% CI’s. Pairs of colored dots inlaid with white asterisks indicate significant (α = 0.05; family wise error rate maintained within each system) difference between coefficients prior to and following the breakpoint.