

<sup>1</sup> Synchrony is more than its top-down and climatic parts: interacting  
<sup>2</sup> Moran effects on phytoplankton in British seas: Supporting information

<sup>3</sup> L. W. Sheppard, E. J. Defriez, P. C. Reid, D. C. Reuman

<sup>4</sup> **Appendix S1 Details of the numerical examples**

<sup>5</sup> For the first example given in the Introduction, suppose  $\alpha(n, t)$  and  $\beta(n, t)$  are biotic or abiotic envi-  
<sup>6</sup> ronmental random variables in locations  $n = 1, 2$  at times  $t$ , and assume these are independent through  
<sup>7</sup> time and standard-normally distributed for all  $n$  and  $t$ . If a population index  $\gamma(n, t)$  follows a stationary  
<sup>8</sup> autoregressive (AR) process such as  $\gamma(n, t+1) = c\gamma(n, t) + \alpha(n, t)$  for  $n = 1, 2$ , then by the Moran theorem  
<sup>9</sup> [Moran, 1953], the population correlation through time  $\text{cor}(\gamma(1), \gamma(2))$  equals  $\text{cor}(\alpha(1), \alpha(2))$ . If instead  
<sup>10</sup>  $\gamma(n, t+1) = c\gamma(n, t) + a\alpha(n, t) + b\beta(n, t)$ , then the Moran theorem implies

$$\text{cor}(\gamma(1), \gamma(2)) = \text{cor}(a\alpha(1) + b\beta(1), a\alpha(2) + b\beta(2)) \quad (1)$$

$$= \frac{a^2 \text{cov}(\alpha(1), \alpha(2)) + ab\text{cov}(\alpha(1), \beta(2)) + ab\text{cov}(\beta(1), \alpha(2)) + b^2 \text{cov}(\beta(1), \beta(2))}{\sqrt{a^2 + 2ab\text{cov}(\alpha(1), \beta(1)) + b^2} \sqrt{a^2 + 2ab\text{cov}(\alpha(2), \beta(2)) + b^2}}. \quad (2)$$

<sup>11</sup> Thus the population correlation depends not only on covariances  $\text{cov}(\alpha(1), \alpha(2))$  or  $\text{cov}(\beta(1), \beta(2))$  of  $\alpha$  or  
<sup>12</sup>  $\beta$  between locations (synchrony of the environmental variables), but also on covariances  $\text{cov}(\alpha(n), \beta(n))$ ,  
<sup>13</sup> and hence includes interaction effects between these Moran drivers. An AR( $N$ ) process could be used  
<sup>14</sup> instead of the AR(1) process used above, with similar result.

<sup>15</sup> We work out some specific examples. Suppose first that  $\beta$  is independent of  $\alpha$ ,  $a = 1$ ,  $b = 1$ ,  
<sup>16</sup>  $\text{cov}(\alpha(1), \alpha(2)) = 0.9$ , and  $\text{cov}(\beta(1), \beta(2)) = 0.5$ . Then it is straightforward to show  $\text{cor}(\gamma(1), \gamma(2)) = 0.7$ .  
<sup>17</sup> If again  $a = 1$ ,  $b = 1$ ,  $\text{cov}(\alpha(1), \alpha(2)) = 0.9$ , and  $\text{cov}(\beta(1), \beta(2)) = 0.5$ , but now  $\text{cov}(\alpha(1), \beta(1)) = 0.5$ ,  
<sup>18</sup>  $\text{cov}(\alpha(2), \beta(2)) = 0.5$ ,  $\text{cov}(\alpha(1), \beta(2)) = 0.5$ ,  $\text{cov}(\alpha(2), \beta(1)) = 0.5$ , then  $\text{cor}(\gamma(1), \gamma(2)) = 0.8$ . Thus  
<sup>19</sup> interactions between Moran drivers in the second example are responsible for a 0.1 increase in population  
<sup>20</sup> synchrony.

<sup>21</sup> We now present supplementary details of our second example, from Results, the more sophisticated  
<sup>22</sup> and detailed example making use of phase-shifted relationships between variables and timescale-specific  
<sup>23</sup> synchrony, and showing explicitly how interaction effects can increase or decrease synchrony.

<sup>24</sup> In the main text we defined  $\gamma_S^{(3)}(n, t) = -(\alpha(n, t) + \beta(n, t - 5))/f$ , where  $f = \sqrt{2 + \frac{1}{2} \cos(\varphi\pi)}$  was  
<sup>25</sup> chosen so that the variance through time of  $\gamma_S^{(3)}(n, t)$  is 1, the same as the variance through time of  $\gamma_S^{(1)}$   
<sup>26</sup> and  $\gamma_S^{(2)}$ . We know  $\text{var}_t(\gamma_S^{(1)}(n, t)) \approx \text{var}_t(\gamma_S^{(2)}(n, t)) \approx 1$  because  $\alpha(n, t)$  and  $\beta(n, t)$  were constructed to

27 have variance 1. We know

$$\text{var}_t[-\alpha(n, t) - \beta(n, t - 5)] = \text{var}_t[-\sqrt{\frac{1}{4}}(\alpha_S(n, t) + \beta_S(n, t - 5)) - \sqrt{\frac{3}{4}}(\alpha_L(n, t) + \beta_L(n, t - 5))] \quad (3)$$

$$\approx \frac{1}{4} \text{var}_t[\alpha_S(n, t) + \beta_S(n, t - 5)] + \frac{3}{2} \quad (4)$$

$$\approx \frac{1}{2} \text{var}_t[\sin(wt) + \cos(wt - \phi\pi - \frac{\pi}{2})] + \frac{3}{2} \quad (5)$$

$$= \frac{1}{2} [\text{var}_t(\sin(wt)) + \text{var}_t(\cos(wt - \phi\pi - \frac{\pi}{2}))] \quad (6)$$

$$+ 2\text{cov}_t(\sin(wt), \cos(wt - \phi\pi - \frac{\pi}{2})) + \frac{3}{2} \quad (7)$$

$$\approx 2 + \text{cov}_t[\sin(wt), \cos(wt - \phi\pi - \frac{\pi}{2})] \quad (8)$$

$$= 2 + \text{cov}_t[\sin(wt), \sin(wt - \phi\pi)] \quad (9)$$

$$= 2 + \frac{1}{2} \cos(\phi\pi), \quad (10)$$

28 where the last equality follows from lemma 3. So  $\text{var}_t(\gamma_S^{(3)}(n, t)) \approx 1$ .

29 **Appendix S2 Thoughts and uncertainties on future patterns of PCI  
30 synchrony**

31 *C. finmarchicus* is in decline in UK seas because it is a cold-water species and is shifting northward with  
32 climatic warming [Beaugrand and Reid, 2003]; it is being replaced by *C. helgolandicus*. Our long-timescale  
33 wavelet models can be used to highlight uncertainties in what the consequences of this change may be  
34 for long-timescale PCI synchrony. *C. finmarchicus* is a synchronising variable for PCI in the sense that  
35 if it were completely asynchronous our model of PCI would have lower synchrony (Fig. 3, magenta line).  
36 However, if the contribution of *C. finmarchicus* to PCI variability were reduced to zero with no other  
37 changes to the plankton system, PCI should become more synchronous because the main driver of PCI  
38 variability would then be temperature alone, and temperature is highly synchronous (Fig. S3). This  
39 projection is based on setting the *C. finmarchicus* coefficient  $\beta_k(\sigma)$  (see Eq. 1 in the main text) to zero in  
40 our best long-timescale model. This result may have value as a projection given the assumptions involved,  
41 but it likely has little value as a prediction of future events since reductions in *C. finmarchicus* have  
42 already been associated with large-scale changes in the plankton community [Beaugrand and Reid, 2003,  
43 Beaugrand et al., 2003]. Relative to *C. finmarchicus*, *C. helgolandicus* had different phase relationships  
44 with both PCI and growing season temperature (details not shown), so even if *C. helgolandicus* replaced  
45 *C. finmarchicus* with no other changes to the community (still an inaccurate idealization) one could  
46 expect major changes in PCI synchrony stemming from both changes in direct top-down Moran effects  
47 and changes in interactions between Moran effects.

48 **Appendix S3 Data**

49 All measurements of biological and physical variables, which were made at irregularly spaced but numerous  
50 times and locations, were compiled into time series representing the changing values found inside 2 by 2  
51 degree areas of sea (26 different areas spread around the UK; Fig. S4). Where measurements of a variable  
52 were available for a given month in a given year within a given 2 by 2 degree area, they were averaged  
53 to obtain a single value for that location and month. Where no readings were available, an average value  
54 across other years for that calendar month and location was used to fill the gap. The values in this filled  
55 monthly time series were then averaged within each year to obtain annualized time series for each of the

variables in each of the 26 areas. Physical variables were sampled independently of biological variables but assembled into time series by the same procedures, except that for biological variables all 12 months were included in a yearly average, whereas for physical variables we explored using averages over different sets of months corresponding to different seasons (table 2) which might coincide with the growing season of certain plankton groups.

As stated in the main text, PCI samples are categorized as No Green (NG), Very Pale Green (VPG), Pale Green (PG) and Green (G) by visual comparison. Colebrook and Robinson [1965] gives the dilution factors needed to render green samples with different colors the same color (VPG 1, PG 2, G 6.5), but alternatives are available. Raitsos et al. [2013] obtains NG 0, VPG 1, PG 2.4 and G 4.6 for ratios of chlorophyll  $\alpha$  obtained from samples of different color categories. They note that using these values instead of the dilutions of Colebrook and Robinson [1965] produces little change in time series. Raitsos et al. [2013] also quotes the satellite data comparison of Raitsos et al. [2005], giving the following SeaWiFS chlorophyll  $\alpha$  levels for each CPR category: NG 1.15, VPG 1.34, PG 2.06, G 3.92 mg/m<sup>3</sup>, with some caveats. We used the original dilution factors of Colebrook and Robinson [1965] to construct monthly time series by taking the mean of available samples in each area in each month in each year.

The biological variables to be compared with the PCI index were zooplankton groups sampled simultaneously with PCI. Individual samples were averaged into monthly time series in the same way as for PCI. Of the very many plankton groups counted in the CPR survey, the species we analyzed were chosen prior to any analyses for their commonness and likely importance for plankton ecology and human concerns.

Two degree summaries of monthly sea surface temperature, cloud and wind data were drawn from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS), release 2.4, in Monthly Summary Groups (MSG1) format. Cloud cover was recorded as proportion of the sky area that was observed to be obscured by cloud, discretized into oktas, or eighths, and was included because of its likely influences on sunlight levels, temperature and salinity variation (related to freshwater input/output). Temperature measurements from the top ten meters of the water column were included, as was wind speed at the sea surface. Salinity sample data was downloaded from <http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html> on 16 January 2015. Individual samples from the top ten meters of the water column were averaged into monthly time series.

Areas in which PCI samples were unavailable for two or more consecutive years at a time were excluded. The choice of 2 by 2 degree areas was a compromise between spatial resolution and the number of samples available within a typical area to produce useable plankton time series. In the 26 useable areas, CPR data were missing for an average of 2.9 months from each year. For each zooplankton group and for PCI, values for these months were filled with the median value of the time series in that month in that location in other years. The median value was used because plankton data were non-normally distributed (see Appendix S4). Physical data were compiled from monthly time series in the same way, except that when missing months were filled in, the mean value in that month in that location in other years was used instead of the median that was used for the plankton time series. The mean value was used because the physical variables were close to normally distributed (see Appendix S4). For sea surface temperature, over all 26 useable areas an average of only 0.032 months were missing from each year. For cloud cover data, over all 26 useable areas an average of only 0.041 months were missing from each year. For wind speed data, over all 26 useable areas an average of only 0.023 months were missing from each year. For salinity data, over all 26 useable areas an average of 5.3 months were missing for each year. Thus our salinity variables were the least reliable time series used. Other physical variables (e.g., nutrient concentrations) were excluded from our analysis due to even fewer monthly values.

## Appendix S4 Box-Cox transformations

For each variable the 26 annualized time series were subjected to Box-Cox normalization. This avoids variability in each data set being dominated by spikes corresponding to outbreaks. Population data

were roughly log-normally distributed. We used a version of the standard optimal Box-Cox transformation technique [Sakia, 1992] to render the marginals of our time series as close to normal as possible. Fourier surrogate techniques (Appendix S8) produce data with normal marginals [Schreiber and Schmitz, 2000], so significance testing based on statistics derived from these surrogates is only fair if the data themselves are normally distributed.

For each variable at each site we found the optimal Box-Cox transformation using the following procedure. As the transformation is only possible for strictly positive values, standard Box-Cox transformations of population data involve the addition of a constant to each variable to give a minimum value of one. We call this artificial minimum the ‘offset’. Since some of our variables are not population densities but physical variables we allowed for a range of possible best offsets when selecting the best Box-Cox coefficient. We began with a starting offset equal to the minimum value of the time series (i.e. no constant was added), or if the minimum was  $\leq 0$ , equal to the minimum increment found between any two successive ordered values. We transformed the time series using a range of 400 candidate Box-Cox coefficients from -2 to +2, and found the maximum likelihood variance, mean, and slope parameters for a model of the transformed variable with a Gaussian distribution about a linear trend. We then repeated this exercise with a range of 100 offset values decreasing in five percent increments from the starting offset, plus 100 more values increasing in five percent increments from the starting offset. We recorded the log maximum likelihood value found using each candidate offset and Box-Cox coefficient. The offset and coefficient giving the greatest mean log likelihood were selected as optimal and the corresponding Box-Cox transformation was applied to the time series. The linear trend was subtracted from each time series, and the variance re-scaled to one. Thus for each time series we obtained zero-mean time series with no trend and distributions as close to Gaussian as possible for a transformation in the Box-Cox family.

The mean (over locations) of the Box-Cox coefficients found for time series of each variable is given in table S2. Box Cox coefficients for biological variables tended to be around 0, for physical variables around 1, indicating that the raw physical values were approximately normally distributed and biological counts were approximately log normal prior to Box-Cox transformation.

## Appendix S5 The wavelet transform and normalizations

If  $x(t)$  ( $t = 1, \dots, T$ ) is a time series, we denote its *wavelet transform* at timescale (period)  $\sigma$  and time  $t$  by  $W_\sigma(t)$ . Wavelets are increasingly common in ecology [Grenfell et al., 2001, Viboud et al., 2006, Keitt and Fischer, 2006, Cazelles et al., 2008, Rouyer et al., 2008, Keitt, 2008, Bell et al., 2012, Cazelles et al., 2014, Sheppard et al., 2016] and general introductions are available [Torrence and Compo, 1998, Addison, 2002]. We used a continuous complex Morlet wavelet transform [Addison, 2002]. The mother wavelet is  $\Psi(t) = (e^{i2\pi f_0 t} - e^{-(2\pi f_0)^2/2}) \exp(-t^2/2)$ , with  $f_0 = 0.5$ . Wavelets associated with a range of timescales were produced using a rescaling technique:  $\Psi_\sigma(t) = s^{-1/2} (e^{i2\pi f_0 t/s} - e^{-(2\pi f_0)^2/2}) \exp(-t^2/2s^2)$ ; following earlier convention [Cazelles et al., 2014] we identify each wavelet with a timescale,  $\sigma = s/f_0$ , and frequency,  $f = f_0/s$ . Convolving the time series  $x(t)$  with wavelets having different periods produces a set of complex components  $W_{n,\sigma}(t) = \sum_{t'} x_n(t+t') \Psi_\sigma(t')$  which comprise the wavelet transform of  $x(t)$ . Wavelets with periods from two years to over 26 years were generated, starting with  $\sigma = 2$  and multiplying each period by 1.05 to get the next. The longest timescale used for the data of this study was 26.5499 years. The transform is scalloped to remove poorly estimated values [Addison, 2002], specifically where the wavelet envelope at the time series edge in the convolution has 50% or more of its maximum amplitude. The scalloping procedure determines the shape of the useable wavelet transform values by removing the unreliable edge values, with the high-frequency end having the most useable values and the low-frequency end having fewer useable values due to the wider wavelets overhanging the edge of the time series in the convolutions that define the wavelet transform. Thus the plots are ‘scalloped’. The center frequency  $f_0$  of the mother wavelet, which has width  $s = 1$  and  $\sigma = 2$ , was taken to be 0.5 to give a high degree of temporal resolution, but making necessary the subtraction of a constant to keep the mean of the wavelet equal to zero [Addison, 2002]. The mother wavelet was scaled so that one oscillation

equaled two years, i.e.,  $\sigma = 2$ , because a two-year period is the highest-frequency fluctuation that can be identified in an annual time series. The actual peak in the Morlet wavelet power spectrum of a sinusoidal signal with frequency  $f'$  and period  $\sigma' = 1/f'$  is at  $s = ((2\pi f_0 + (2 + (2\pi f_0)^2)^{1/2})/4\pi)(\sigma')$ , so  $f \approx f'$  only [Meyers et al., 1993].

If  $x_n(t)$  ( $t = 1, \dots, T$ ) is the  $n$ th of  $N$  time series from different locations, and  $W_{n,\sigma}(t)$  is its wavelet transform, then we refer to

$$w_{n,\sigma}(t) = W_{n,\sigma}(t)/|W_{n,\sigma}(t)|$$

as the *phasor-normalized transform*. We also define

$$w_{n,\sigma}(t) = \frac{W_{n,\sigma}(t)}{\sqrt{\frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T W_{n,\sigma}(t) \overline{W_{n,\sigma}(t)}}},$$

where overbar denotes complex conjugation, as the *power-normalized transform*, using the same notation,  $w_{n,\sigma}(t)$ , as was used for the phasor-normalized transforms. The normalization used in any given context will be specified textually if it is not immediately clear. The terminology *power-normalized transform* was chosen because the denominator in the expression for the power-normalized transforms  $w_{n,\sigma}(t)$  is the square root of the average wavelet power of time series over locations. That denominator is only a single positive rescaling factor, so the  $w_{n,\sigma}(t)$  contain essentially the same information as the  $W_{n,\sigma}(t)$ . The phases of  $w_{n,\sigma}(t)$  and  $W_{n,\sigma}(t)$  are the same, and are equal to the phase of oscillation in  $x_n(t)$  at time  $t$  and timescale  $\sigma$ .

## Appendix S6 Spatial coherence

The *spatial coherence* of two variables  $x_n^{(0)}(t)$  and  $x_n^{(1)}(t)$  is the magnitude of the quantity

$$\Pi_\sigma^{(01)} = \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T w_{n,\sigma}^{(0)}(t) \overline{w_{n,\sigma}^{(1)}(t)}, \quad (11)$$

where the  $w$  are power-normalized transforms (Appendix S5). The spatial coherence takes values between 0 and 1 (lemma 4). Because  $w_{n,\sigma}^{(0)}(t) \overline{w_{n,\sigma}^{(1)}(t)}$  is a complex number with phase equal to the phase difference between the two wavelet components, the sum over  $n$  and  $t$  is large if the phase difference is consistent over time and locations, and small otherwise. Wavelet components can also have varying magnitudes, and spatial coherence is larger still if there are correlations in amplitudes of fluctuations. For  $N = 1$ , the spatial coherence is the (standard) wavelet coherence [Torrence and Compo, 1998, Addison, 2002]. The measure also relates to the phase coherence which is defined for a single pair of transforms ( $N = 1$ ) [Grenfell et al., 2001, Bandrivskyy et al., 2004, Viboud et al., 2006, Stefanovska, 2007], using the same formula but with phasor-normalized transforms (Appendix S5). The phase of (11) is the average phase difference between the variables, as a function of  $\sigma$ .

## Appendix S7 Linear models for wavelet transforms

To understand potential causes of synchrony in the  $x_n^{(0)}(t)$  we use complex linear models expressing the transforms,  $w_{n,\sigma}^{(0)}(t)$ , of the  $x_n^{(0)}(t)$  as nearly as possible in terms of the transforms,  $w_{n,\sigma}^{(k)}(t)$ , of the  $x_n^{(k)}(t)$ :

$$w_{n,\sigma}^{(0)}(t) \approx \beta_1(\sigma) w_{n,\sigma}^{(1)}(t) + \cdots + \beta_K(\sigma) w_{n,\sigma}^{(K)}(t). \quad (12)$$

We use power-normalized transforms in this section. The coefficient  $\beta_k(\sigma)$  is a complex number with phase equal to the phase difference between the fluctuations in  $x_n^{(k)}(t)$  at timescale  $\sigma$  and the contribution of this potential driver of dynamics to the fluctuations in  $x_n^{(0)}(t)$  at the same timescale. The magnitude

182 of  $\beta_k(\sigma)$  is the corresponding ‘gain’, i.e., the strength of the contribution of the driver at timescale  $\sigma$  to  
 183  $x_n^{(0)}(t)$ , relative to the strength of the oscillation at that timescale in the driver itself.

184 For a frequency component  $\sigma$ , we consider the space  $\mathbb{C}^D$  where  $D$  is the product of  $N$  and the number of  
 185 time steps,  $t$ , for which the wavelet components  $w_{n,\sigma}^{(0)}(t)$  and  $w_{n,\sigma}^{(k)}(t)$  are defined after scalloping to remove  
 186 poorly estimated values (Appendix S5). The components  $w_{n,\sigma}^{(0)}(t)$  for  $n = 1, \dots, N$  combine naturally to  
 187 form a point in this space, as do the components  $w_{n,\sigma}^{(k)}(t)$ ,  $n = 1, \dots, N$ , for any  $k$ . The space  $\mathbb{C}^D$  is a  
 188 Hilbert space under the inner product  $\langle w_{n,\sigma}^{(0)}(t), w_{n,\sigma}^{(k)}(t) \rangle = \frac{1}{NT} \sum_n \sum_t w_{n,\sigma}^{(0)}(t) \overline{w_{n,\sigma}^{(k)}(t)}$ , and we denote the  
 189 associated norm by  $\|\cdot\|$ .

190 Applying lemma 5 with  $V = \text{Span}(w_{n,\sigma}^{(1)}(t), \dots, w_{n,\sigma}^{(K)}(t))$ , there is a unique  $v \in V$  such that  $w_{n,\sigma}^{(0)}(t) =$   
 191  $v + d$ , where  $d \perp V$ . The point  $v$  is an approximation of  $w_{n,\sigma}^{(0)}(t)$ , in the sense that the residuals,  $d$ , have  
 192 minimal magnitude,  $\|d\|$  (lemma 5). The point  $v$  also has maximal spatial coherence with  $w_{n,\sigma}^{(0)}(t)$ , in the  
 193 sense that, for any other  $w \in V$ , the spatial coherence,  $\frac{|\langle w_{n,\sigma}^{(0)}(t), w \rangle|}{\|w\|}$ , of  $w_{n,\sigma}^{(0)}(t)$  with the power-normalization  
 194 of  $w$  is less than or equal to the spatial coherence,  $\frac{|\langle w_{n,\sigma}^{(0)}(t), v \rangle|}{\|v\|}$ , of  $w_{n,\sigma}^{(0)}(t)$  with the power-normalization of  
 195  $v$ . This follows from lemma 5.

196 The point  $v$  can be written explicitly as a complex linear model expressing the transforms  $w_{n,\sigma}^{(0)}(t)$  as  
 197 nearly as possible in terms of the transforms  $w_{n,\sigma}^{(k)}(t)$  by writing  $v = \beta_1 w_{n,\sigma}^{(1)}(t) + \dots + \beta_K w_{n,\sigma}^{(K)}(t)$ , where  
 198  $\beta = (\beta_1, \dots, \beta_K)^t$  is a solution of the equation  $M\beta = p$ , with the  $ij^{\text{th}}$  component of the matrix  $M$  equal  
 199 to  $\langle w_{n,\sigma}^{(j)}(t), w_{n,\sigma}^{(i)}(t) \rangle$  (here  $1 \leq i, j, \leq K$ ) and the  $j^{\text{th}}$  component of  $p$  equal to  $\langle w_{n,\sigma}^{(0)}(t), w_{n,\sigma}^{(j)}(t) \rangle$  (lemma 5).  
 200 We know a solution of this equation exists because  $v \in V = \text{Span}(w_{n,\sigma}^{(1)}(t), \dots, w_{n,\sigma}^{(K)}(t))$ . The solution will  
 201 be unique if and only if the  $w_{n,\sigma}^{(k)}(t)$  for  $k = 1, \dots, K$  are linearly independent. Solutions to the equation  
 202  $M\beta = p$  (ie. complex wavelet coefficients  $\beta$  for each frequency) were calculated by means of the left  
 203 matrix divide operation in Matlab for numerical stability.

## 204 Appendix S8 Surrogates

205 Starting from a spatiotemporal dataset  $x_n(t)$  for  $n = 1, \dots, N$  and  $t = 1, \dots, T$ , we used two types  
 206 of artificial ‘surrogate’ (i.e., resampled) data, *spatially synchronous surrogates* and *asynchronous surro-*  
 207 *gates*. Surrogate approaches to significance testing for wavelet analyses are standard [Theiler et al., 1992,  
 208 Prichard and Theiler, 1994, Schreiber and Schmitz, 2000, Rouyer et al., 2008, Cazelles et al., 2014]. Both  
 209 procedures provide synthetic spatiotemporal data for the same number of locations over the same times as  
 210 the original data, with certain features of the original data randomized away and other features retained.  
 211 Both procedures first Fourier transform the  $x_n(t)$ , then randomize the phases of the Fourier components  
 212 by adding a random uniformly distributed phase at each frequency, and finally inverse Fourier transform  
 213 to obtain the surrogate data. Thus both procedures are based on the standard Fourier surrogate approach  
 214 [Schreiber and Schmitz, 2000]. These steps result in time series that have the same power spectrum as  
 215 the original time series, and therefore the same autocorrelation properties, but have different temporal  
 216 structure [Schreiber and Schmitz, 2000]. Box-Cox transformations were performed (Appendix S4) to nor-  
 217 malize data during the data preparation stage, in part because Fourier surrogates tend to normalize time  
 218 series marginal distributions [Schreiber and Schmitz, 2000]. Comparisons are therefore fairer between  
 219 statistics calculated for surrogate and real data when the real data has been pre-normalized. For spatially  
 220 synchronous surrogates, for each timescale  $\sigma$  the same phase is added to the Fourier components at  $\sigma$  for  
 221 each of the  $x_n(t)$ . Thus surrogates also preserve cross-spectrum and cross-correlation properties of the  
 222 time series. This is the approach developed by Prichard and Theiler [1994]. For asynchronous surrogates,  
 223 independent phases are added for each  $x_n(t)$ , so that synchrony between the  $x_n(t)$  is eliminated at all  
 224 timescales.

225 **Appendix S9 Model construction**

226 We developed a significance test for adding variables to the linear model constructed in Appendix S7.  
 227 Consider adding another variable to an explanatory model consisting of  $K$  variables. If  $x_n^{(k)}(t)$  for  $k =$   
 228  $0, \dots, K + 1$  are variables measured at locations  $n = 1, \dots, N$  and times  $t = 1, \dots, T$ , then we can use  
 229 spatially synchronous surrogates to determine whether the linear model

$$w_{n,\sigma}^{(0)}(t) \approx \beta_1(\sigma)w_{n,\sigma}^{(1)}(t) + \dots + \beta_{K+1}(\sigma)w_{n,\sigma}^{(K+1)}(t) \quad (13)$$

230 is significantly better than the linear model

$$w_{n,\sigma}^{(0)}(t) \approx \alpha_1(\sigma)w_{n,\sigma}^{(1)}(t) + \dots + \alpha_K(\sigma)w_{n,\sigma}^{(K)}(t). \quad (14)$$

231 The resulting test, now described in detail, is the wavelet linear model analogue of standard  $F$ -tests used  
 232 to compare nested models in a classic general linear modeling framework.

233 Using the notation of Appendix S7 and  $V = \text{Span}(w_{n,\sigma}^{(1)}(t), \dots, w_{n,\sigma}^{(K+1)}(t))$ , first  $w_{n,\sigma}^{(0)}(t)$  is written  
 234 as  $v_{\text{data}} + d_{\text{data}}$  for each timescale,  $\sigma$ . Then define  $w_{j,n,\sigma}^{(K+1)}(t)$  as the power-normalized transforms of  
 235 a  $j^{\text{th}}$  synchrony-preserving surrogate of  $x_n^{(K+1)}(t)$  for  $j = 1, \dots, 10000$ , and write  $w_{n,\sigma}^{(0)}(t)$  as  $v_j + d_j$   
 236 for  $V = \text{Span}(w_{n,\sigma}^{(1)}(t), \dots, w_{n,\sigma}^{(K)}(t), w_{j,n,\sigma}^{(K+1)}(t))$  for each timescale,  $\sigma$ . Thus we have  $d_{\text{data},\sigma}$  and  $d_{j,\sigma}$  for  
 237  $j = 1, \dots, 10000$ , making the  $\sigma$  dependence explicit. We consider that the model (13) is a significant  
 238 improvement over the model (14), for timescale  $\sigma$  at significance level  $\alpha$ , if  $\|d_{\text{data},\sigma}\|^2$  is smaller than a  
 239 fraction  $1 - \alpha$  of the  $\|d_{j,\sigma}\|^2$ . In words, the more complex model is considered better at  $\sigma$  if the additional  
 240 predictor allows a reliably greater improvement in the representation of  $w_{n,\sigma}^{(0)}(t)$  than would a randomized  
 241 surrogate data set with the same spatial and temporal autocorrelation properties.

242 The requirement that  $\|d_{\text{data},\sigma}\|^2$  is smaller than a fraction  $1 - \alpha$  of the  $\|d_{j,\sigma}\|^2$  at timescale  $\sigma$  is  
 243 equivalent to the requirement that the spatial coherence at  $\sigma$  between the right side of (13) and  $w_{n,\sigma}^{(0)}(t)$   
 244 is greater when the right side of (13) is fitted using the actual  $w_{n,\sigma}^{(K+1)}(t)$  than it is when fitted using any  
 245 of a fraction  $1 - \alpha$  of the surrogate transforms  $w_{j,n,\sigma}^{(K+1)}(t)$  in place of  $w_{n,\sigma}^{(K+1)}(t)$ . Hence the more complex  
 246 model is considered better at  $\sigma$  if the additional predictor improves the spatial coherence of the model  
 247 with  $w_{n,\sigma}^{(0)}(t)$  more than would a randomized, surrogate predictor. To see this, we write the squares of the  
 248 spatial coherences to be compared as

$$\frac{|\langle w_{n,\sigma}^{(0)}(t), v_{\text{data}} \rangle|^2}{\|v_{\text{data}}\|^2} = \frac{|\langle v_{\text{data}} + d_{\text{data}}, v_{\text{data}} \rangle|^2}{\|v_{\text{data}}\|^2} \quad (15)$$

$$= \frac{|\|v_{\text{data}}\|^2|^2}{\|v_{\text{data}}\|^2} \quad (16)$$

$$= \|v_{\text{data}}\|^2 \quad (17)$$

$$= \|w_{n,\sigma}^{(0)}(t)\|^2 - \|d_{\text{data}}\|^2 \quad (18)$$

249 and

$$\frac{|\langle w_{n,\sigma}^{(0)}(t), v_j \rangle|^2}{\|v_j\|^2} = \frac{|\langle v_j + d_j, v_j \rangle|^2}{\|v_j\|^2} \quad (19)$$

$$= \frac{|\|v_j\|^2|^2}{\|v_j\|^2} \quad (20)$$

$$= \|v_j\|^2 \quad (21)$$

$$= \|w_{n,\sigma}^{(0)}(t)\|^2 - \|d_j\|^2, \quad (22)$$

where (18) and (22) follow because  $v_{\text{data}} \perp d_{\text{data}}$  and  $v_j \perp d_j$ . Thus the spatial coherence (15) is larger than the spatial coherence (19) for exactly the same  $j$  for which  $\|d_{\text{data}}\|^2$  is smaller than  $\|d_j\|^2$ .

The procedures described above give one statistical test for each timescale  $\sigma$ . But the results of the tests are not independent across values of  $\sigma$ , particularly for similar values. To determine whether (13) is a significant improvement over (14) for a whole band of timescales  $\sigma$  in aggregate, the rank of  $\|d_{\text{data},\sigma}\|^2$  relative to the surrogate values  $\|d_{j,\sigma}\|^2$  was found for each timescale in the band (rank 1 was the lowest) and the mean rank was computed inside the band. The same procedure applied to each surrogate separately (i.e., ranking each  $\|d_{j,\sigma}\|^2$  against the  $\|d_{j',\sigma}\|^2$  for all  $j' \neq j$ ) produced surrogate mean rank values for the timescale band, and the proportion of these mean ranks less than the actual mean rank provided a  $p$ -value for the band.

Two timescale bands, as opposed to three or more bands, were used for simplicity and because wavelet methods have finite capacity for resolving timescales of fluctuation in finite time series. A frequency of 1 cycle every 4 years was selected as the boundary between bands both because this is half the Nyquist frequency for annual sampling (1 cycle every 2 years), and because it is the boundary between persistent and anti-persistent behavior in Fourier components, as measured with lag-1 autocorrelation [Sheppard et al., 2016].

Statistical comparisons between (14) and models adding more than one predictor variable are straightforwardly possible via the same approach. Independent spatially synchronous surrogates are generated for each of the added variables.

The use of spatially synchronous surrogates is appropriate for the tests of this section because spatial and temporal autocorrelation properties of  $x_n^{(K+1)}(t)$  can inflate spatial coherences between  $x_n^{(K+1)}(t)$  and  $x_n^{(0)}(t)$ , producing small values of  $\|d_{\text{data},\sigma}\|^2$ . Using spatially synchronous surrogates ensures that the surrogates of  $x_n^{(K+1)}(t)$  have the same spatial and temporal autocorrelation properties as  $x_n^{(K+1)}(t)$  itself, and the values  $\|d_{j,\sigma}\|^2$  are subject to the same biases. Spatially synchronous surrogates represent the null hypothesis of no relationship between  $x_n^{(0)}(t)$  and  $x_n^{(K+1)}(t)$  while retaining the spectral properties of  $x_n^{(K+1)}(t)$ .

We used the above significance testing methods as part of a model selection and validation procedure for constructing optimal models of PCI transforms for long ( $> 4$  years) and short ( $< 4$  years) timescales, separately. None of the wavelet components we used had a scale of exactly 4 years. We examined all possible combinations of four variables or fewer for each timescale band separately. This choice to limit models to four variables or fewer followed a similar choice of Defriez and Reuman [2017a] and Defriez and Reuman [2017b]. The choice served to keep computation times reasonable, and was consistent with our desire to focus on only the most important determinants of plankton dynamics and synchrony. Models including more than one temperature variable, more than one salinity variable, more than one wind variable or more than one cloud cover index were discarded. Of the remaining models, and separately for long and short timescale, we retained only those models for which each variable included in the model significantly improved the fit over not including it (using the significance methods described above). This step ensured the selected model was not merely the best model among poor alternatives, but also explains a significant amount of variation in the response transform  $w_{n,\sigma}^{(0)}(t)$  (this is what the word ‘testing’ in ‘model selection and testing’ refers to). Models for which not all variables could individually pass this test were discarded. This prevents the inclusion of redundant explanatory variables. Remaining models were ranked (again, separately for long and short timescales) via a leave-one-out cross validation criterion.

The leave-one-out cross validation was performed as follows. For each model remaining after the above steps, and for each of the 26 regions of British seas for which PCI time series and other time series and their transforms were available (figure S4), we re-fitted to obtain model parameters using the data in the other 25 locations, and then evaluated goodness of fit for the data in the omitted location. In other words, when leaving out location  $n'$  we wrote (using the notation of section Appendix S7)  $w_{n \neq n',\sigma}^{(0)}(t)$  as  $v_{\text{data}} + d_{\text{data}}$  and found the optimal (minimal  $\|d_{\text{data},\sigma}\|^2$ ) coefficients  $\beta_k$  for  $n \neq n'$ . We then found  $\|d_{\text{data},\sigma}\|^2$  at location  $n'$  using these coefficients. The average of  $\|d_{\text{data},\sigma}\|^2$  over  $\sigma$  in the desired frequency

band gave the leave-one-out goodness of fit for this location  $n'$  for the model. The lower the value the better the fit. We averaged over the 26 locations, left out in turn, to obtain a single index of leave-one-out goodness of fit for each model for the frequency band. The resulting measure was intended to be robust against overfitting by the addition of more variables. The remaining model with the best leave-one-out goodness of fit was selected at high frequencies, and separately at low frequencies.

## Appendix S10 Checking phase relationships

The phases of the model coefficients  $\beta_k(\sigma)$  are supposed to represent the phase lag of effects (or associations) at timescale  $\sigma$  of  $x_n^{(k)}(t)$  on  $x_n^{(0)}(t)$  (see Eq.2 from Methods for notation). Strong associations which are consistent over time and space should therefore produce a correspondence between the phase of  $\beta_k(\sigma)$  and the average phase, over space and/or time, of  $\overline{w_{n,\sigma}^{(0)}(t)w_{n,\sigma}^{(k)}(t)}$ , where the overbar denotes complex conjugation; the phase of this quantity is the phase difference between  $x_n^{(0)}(t)$  and  $x_n^{(k)}(t)$  at timescale  $\sigma$ , time  $t$ , and location  $n$ . We tested for this phase correspondence and found it to be good (Fig. S1). Especially at long timescales, a particular phase relationship between predictor and PCI transforms is maintained consistently through time (Fig. S1a, b). PCI and temperature had a consistent phase relationship on long timescales, corresponding to a phase shift of one quarter of a cycle (Fig. S1a,e). The average phase difference between PCI and temperature transforms (average of the green line in Fig. S1e over long timescales) was  $0.42\pi$  radians at long timescales ( $\pi/2$  is exactly a quarter cycle). PCI and *C. finmarchicus* were found to be in a consistent anti-phase relationship (Fig. S1b,f). The average phase difference between PCI and *C. finmarchicus* transforms (average of the green line in Fig. S1f) was  $0.78\pi$  radians for long timescales and  $0.99\pi$  radians for short timescales, both of which were reasonably close to anti-phase ( $\pi$  radians phase difference). The average phase differences between PCI and echinoderm and decapod larvae, respectively, on short timescales were  $0.09\pi$  and  $0.01\pi$  radians, both of which are approximately in-phase relationships.

## Appendix S11 Location permutations

For each predictor variable, the spatial coherence with PCI was computed, as was the spatial coherence between PCI and 1000 surrogate versions of the predictor in which time series locations were randomly permuted but time series were otherwise unmanipulated. For each timescale in the relevant band (long timescales for growing season temperature and *C. finmarchicus* abundance, short timescales for *C. finmarchicus* abundance and for echinoderm larvae and decapod larvae) the spatial coherence for the unmanipulated data was ranked in the distribution of surrogate spatial coherences and ranks were averaged across the timescale band. The same procedure was then applied to each surrogate spatial coherence separately, i.e., surrogate spatial coherences were then ranked in the other surrogate spatial coherences and these values were averaged for each surrogate across the timescale band. The proportion of the surrogate mean ranks less than the actual mean rank was then computed. On long timescales, mean spatial coherence rank was higher than that of 989 of 1000 such surrogates values for growing season temperature, and 1000 of 1000 for *C. finmarchicus*. At short timescales, mean spatial coherence rank was higher than that of 997 of 1000 such surrogates for *C. finmarchicus*, 1000 of 1000 for echinoderm larvae, and 988 of 1000 for decapod larvae.

These results are consistent with spatial inhomogeneity observed in the raw values of the PCI time series. Our constructed PCI monthly time series mean values (before the Box-Cox and other standardization), when averaged over all months and years, produced values that varied by location between a minimum of 0.6207 and a maximum of 2.1956 (mean of 1.1157 and standard deviation of 0.4187). These values do not have units because PCI is an index.

## 342 Appendix S12 Link to species-specific results

343 Spatial coherences between nine individual phytoplankton species abundances from the CPR data set and  
 344 *C. finmarchicus* were tested in Sheppard et al. [2017]. Results were consistent with our findings that PCI  
 345 and *C. finmarchicus* were significantly spatially coherent and in anti-phase. At short-timescales, three of  
 346 the four *Ceratium* species tested were significantly spatially coherent ( $p < 0.05$ ) with *C. finmarchicus*, and  
 347 were in an approximate anti-phase relationship with it. Average phase shifts and  $p$ -values were: *Ceratium*  
 348 *fusus*,  $0.9581\pi$ ,  $p = 0.0088$ ; *Ceratium furca*,  $-0.9451\pi$ ,  $p = 0.0143$ ; *Ceratium macroceros*,  $-0.8356\pi$ ,  
 349  $p = 0.0372$ . No other significant ( $p < 0.05$ ) spatial coherences were found between phytoplankton species  
 350 and *C. finmarchicus* at short or long timescales. Sheppard et al. [2017] used the same definitions of ‘short’  
 351 and ‘long’ timescales used here, and the same spatial grid.

## 352 Appendix S13 Mean fields

353 Given a spatiotemporal data set  $x_n(t)$  for  $n = 1, \dots, N$  and  $t = 1, \dots, T$ , and letting  $w_{n,\sigma}(t)$  represent  
 354 the phasor-normalized wavelet transforms (Appendix S5), we define the *wavelet phasor mean field* as

$$r_\sigma(t) = \frac{1}{N} \sum_{n=1}^N w_{n,\sigma}(t). \quad (23)$$

355 The magnitude of this quantity is between 0 and 1. The wavelet phasor mean field is one natural choice  
 356 for a time- and timescale-specific measure of synchrony: if the phasors,  $w_{n,\sigma}(t)$ , are asynchronous their  
 357 average tends to have small magnitude, as the phasors conduct a random walk in the complex plane when  
 358 summed; but if phases are synchronized, the phasors produce a directed walk when summed, giving a  
 359 large value for  $|r_\sigma(t)|$ . The value  $|r_\sigma(t)| = 1$  indicates the time series have identical phases of oscillation  
 360 at timescale  $\sigma$  and time  $t$ .

361 The significance of a wavelet phasor mean field magnitude  $|r_\sigma(t)|$  at timescale  $\sigma$  and time  $t$  is obtained  
 362 by comparing to a distribution of the mean field magnitude of  $N$  random phasors,

$$\left| \frac{1}{N} \sum_{n=1}^N \exp(2\pi u_n i) \right|, \quad (24)$$

363 where the  $u_n$  are independent uniformly distributed random variables on the unit interval. For large  $N$ ,  
 364 this is a Rayleigh distribution [Strutt, 1902]. For any  $N$  a distribution of possible magnitudes for the sum  
 365 can quickly numerically generated by summing a large number of sets of  $N$  random unit phasors.

366 If phase synchrony occurs for  $\sigma$  and  $t$ , so that the unit phasors  $w_{n,\sigma}(t)$  have similar phase for all or  
 367 most  $n = 1, \dots, N$ , then  $|r_\sigma(t)|$  will tend to be larger than the vast majority of the random quantities  
 368 of (24). The test applies for one pair of  $\sigma$  and  $t$  values at a time, so one must be cognizant of multiple  
 369 testing errors when examining ranges of  $\sigma$  and  $t$ . The approach used here is the same as that used in  
 370 Sheppard et al. [2013]. The significance of the wavelet phasor mean field of PCI time series is used to  
 371 plot the significance contours of the phase synchrony depicted in figure 3a

372 We also define a *wavelet mean field*, another measure of synchrony which includes information about  
 373 the magnitudes of oscillation in the  $x_n(t)$  in addition to the phase-synchrony information on which the  
 374 wavelet phasor mean field relies exclusively. The same equation (23) is used to define the wavelet mean  
 375 field, but now with the  $w_{n,\sigma}(t)$  representing power-normalized transforms (Appendix S5).

376 In addition to the reasoning that the magnitude of the the wavelet mean field will tend to be big-  
 377 ger when fluctuations at time  $t$  and timescale  $\sigma$  have similar phase at all sampling locations, and are  
 378 therefore synchronized, the wavelet mean field is also a natural choice for a time- and timescale-specific  
 379 measure of strength of synchrony because of its mathematical properties. The mean squared magnitude  
 380  $\frac{1}{T} \sum_{t=1}^T |r_\sigma(t)|^2$  is between 0 and 1. We refer to this quantity as the *mean squared synchrony*. It equals 1

for all  $\sigma$  if and only if the time series  $x_n(t)$  ( $t = 1, \dots, T$ ) are identical (lemma 6). Also, this quantity is the power of the average time series divided by the average of the powers of all the time series (lemma 6). If time series are unsynchronized, power in the average time series will be reduced, as unsynchronized fluctuations will tend to cancel. In contrast, synchronized fluctuations will reinforce each other and contribute power to the average. In this way  $\frac{1}{T} \sum_{t=1}^T |r_\sigma(t)|^2$  represents synchrony for each timescale,  $\sigma$ . It is a generalization of

$$\frac{\text{var}\left(\frac{1}{N} \sum_{n=1}^N x_n(t)\right)}{\frac{1}{N} \sum_{n=1}^N \text{var}(x_n(t))},$$

<sup>376</sup> a familiar quantity interpretable as synchrony because it represents the extent to which fluctuations in  
<sup>377</sup> local time series reinforce each other or cancel in the average (lemma 6). The wavelet mean field [Sheppard  
<sup>378</sup> et al., 2016] and related methods [Grenfell et al., 2001, Viboud et al., 2006, Keitt and Fischer, 2006, Keitt,  
<sup>379</sup> 2008] have been used previously to study synchrony in ecology.

## <sup>380</sup> Appendix S14 A Moran theorem for wavelet linear models

<sup>381</sup> This section parallels the development of the wavelet Moran theorem in Sheppard et al. [2016], but  
<sup>382</sup> this time incorporating multiple drivers of synchrony. If, as in previous sections,  $x_n^{(0)}(t)$  and  $x_n^{(k)}(t)$ ,  
<sup>383</sup>  $k = 1, \dots, K$  are variables measured at locations  $n = 1, \dots, N$  and times  $t = 1, \dots, T$ ,  $w_{n,\sigma}^{(0)}(t)$  and  $w_{n,\sigma}^{(k)}(t)$   
<sup>384</sup> are the corresponding power-normalized transforms, and we consider the Hilbert space  $\mathbb{C}^D = \mathbb{C}^{D_\sigma}$  of  
<sup>385</sup> Appendix S7 and the subspace  $V = V_\sigma = \text{Span}(w_{n,\sigma}^{(1)}(t), \dots, w_{n,\sigma}^{(K)}(t))$  then we make use of the model  
<sup>386</sup>  $v = v_{n,\sigma}(t)$  described in Appendix S7, i.e.,  $w_{n,\sigma}^{(0)}(t) = v_{n,\sigma}(t) + d_{n,\sigma}(t)$ . Here, dependencies on  $\sigma$ ,  $n$  and  $t$   
<sup>387</sup> are initially made explicit for clarity. By the reasoning of Appendix S7, we can write

$$v_{n,\sigma}(t) = \beta_1(\sigma)w_{n,\sigma}^{(1)}(t) + \dots + \beta_K(\sigma)w_{n,\sigma}^{(K)}(t). \quad (25)$$

<sup>388</sup> This is the right side of the best-fitting model (Eq.2) from the main text.

<sup>389</sup> We let  $h_{n,\sigma}(t) = \frac{v_{n,\sigma}(t)}{\|v_{n,\sigma}(t)\|}$  and  $\Pi_\sigma^{(0h)} = \|v_{n,\sigma}(t)\|$ , so that  $v_{n,\sigma}(t) = \Pi_\sigma^{(0h)}h_{n,\sigma}(t)$  and  $w_{n,\sigma}^{(0)}(t) =$   
<sup>390</sup>  $\Pi_\sigma^{(0h)}h_{n,\sigma}(t) + d_{n,\sigma}(t)$ . Because  $w_{n,\sigma}^{(0)}(t)$  is power normalized,

$$1 = \|w_{n,\sigma}^{(0)}(t)\| \quad (26)$$

$$= \|v_{n,\sigma}(t) + d_{n,\sigma}(t)\| \quad (27)$$

$$= \sqrt{\langle v_{n,\sigma}(t) + d_{n,\sigma}(t), v_{n,\sigma}(t) + d_{n,\sigma}(t) \rangle} \quad (28)$$

$$= \sqrt{\langle v_{n,\sigma}(t), v_{n,\sigma}(t) \rangle + \langle d_{n,\sigma}(t), d_{n,\sigma}(t) \rangle} \quad (29)$$

$$= \sqrt{\|v_{n,\sigma}(t)\|^2 + \|d_{n,\sigma}(t)\|^2}, \quad (30)$$

<sup>391</sup> so  $\Pi_\sigma^{(0h)}$ , which is a nonnegative real number, is less than or equal to 1. The larger  $\Pi_\sigma^{(0h)}$  is, the closer  
<sup>392</sup>  $h_{n,\sigma}(t)$  comes to  $v_{n,\sigma}(t)$ , and the closer  $v_{n,\sigma}(t)$  comes to  $w_{n,\sigma}^{(0)}(t)$ . Because

$$\langle w_{n,\sigma}^{(0)}(t), h_{n,\sigma}(t) \rangle = \langle v_{n,\sigma}(t) + d_{n,\sigma}(t), \frac{v_{n,\sigma}(t)}{\|v_{n,\sigma}(t)\|} \rangle \quad (31)$$

$$= \langle v_{n,\sigma}(t), \frac{v_{n,\sigma}(t)}{\|v_{n,\sigma}(t)\|} \rangle \quad (32)$$

$$= \|v_{n,\sigma}(t)\| \quad (33)$$

$$= \Pi_\sigma^{(0h)}, \quad (34)$$

393  $\Pi_\sigma^{(0h)}$  equals the spatial coherence of  $w_{n,\sigma}^{(0)}(t)$  and  $h_{n,\sigma}(t)$ , which explains the choice of notation  $\Pi_\sigma^{(0h)}$  (see  
394 Appendix S6). By orthogonality of  $v_{n,\sigma}(t)$  and  $d_{n,\sigma}(t)$  (i.e.,  $\langle v_{n,\sigma}(t), d_{n,\sigma}(t) \rangle = 0$ ),  $\langle h_{n,\sigma}(t), d_{n,\sigma}(t) \rangle = 0$ ,  
395 as well.

396 **Theorem 1. A Moran theorem for wavelet linear models** *Using the notation defined above, and if*  
397  $r_\sigma^{(0)}(t) = \frac{1}{N} \sum_{n=1}^N w_{n,\sigma}^{(0)}(t)$  *is the wavelet mean field of the*  $x_n^{(0)}(t)$ ,  $r_\sigma^{(h)}(t) = \frac{1}{N} \sum_{n=1}^N h_{n,\sigma}(t)$  *and*  $r_\sigma^{(d)}(t) =$   
398  $\frac{1}{N} \sum_{n=1}^N d_{n,\sigma}(t)$ , *then the time-averaged squared magnitude of*  $r_\sigma^{(0)}(t)$  *is*

$$\frac{1}{T} \sum_{t=1}^T |r_\sigma^{(0)}(t)|^2 = |\Pi_\sigma^{(0h)}|^2 \frac{1}{T} \sum_{t=1}^T |r_\sigma^{(h)}(t)|^2 \quad (35)$$

$$+ \frac{1}{TN^2} \sum_{t=1}^T \sum_{n=1}^N \sum_{m \neq n} 2 \operatorname{Re}(\Pi_\sigma^{(0h)} h_{n,\sigma}(t) \overline{d_{m,\sigma}(t)}) \quad (36)$$

$$+ \frac{1}{T} \sum_{t=1}^T |r_\sigma^{(d)}(t)|^2. \quad (37)$$

399 If populations  $x_n^{(0)}(t)$  do not interact with populations  $x_m^{(0)}(t)$  (for  $m \neq n$ ) at neighboring sites, nor are  
400 they directly affected by other conditions (environmental or biotic) in neighboring sites, interpreted as the  
401 term in (36) being negligible (see justification in the proof below), then

$$\frac{1}{T} \sum_{t=1}^T |r_\sigma^{(0)}(t)|^2 \approx |\Pi_\sigma^{(0h)}|^2 \frac{1}{T} \sum_{t=1}^T |r_\sigma^{(h)}(t)|^2 + \frac{1}{T} \sum_{t=1}^T |r_\sigma^{(d)}(t)|^2. \quad (38)$$

402 In other words, synchrony of  $x_n^{(0)}(t)$  can be divided into a component attributable to the variables  $x_n^{(k)}(t)$ ,  
403  $k = 1, \dots, K$ , and residual synchrony. It follows that

$$\{|r_\sigma^{(0)}(t)|\}_t \geq |\Pi_\sigma^{(0h)}|^2 \{|r_\sigma^{(h)}(t)|\}_t, \quad (39)$$

404 where  $\{x\}_t$  represents the time average of the square of  $x$ . If, in addition, the only synchronizing influence  
405 on  $x_n^{(0)}(t)$  is the variables  $x_n^{(k)}(t)$ ,  $k = 1, \dots, K$ , interpreted as the second term on the right of (38) being  
406 negligible, then

$$\{|r_\sigma^{(0)}(t)|\}_t \approx |\Pi_\sigma^{(0h)}|^2 \{|r_\sigma^{(h)}(t)|\}_t. \quad (40)$$

407 Thus  $|\Pi_\sigma^{(0h)}|^2 \{|r_\sigma^{(h)}(t)|\}_t$  gives the amount of synchrony explained solely by the synchronizing influence of  
408 the variables  $x_n^{(k)}(t)$ ,  $k = 1, \dots, K$ .

409 *Proof.* We omit subscripts  $\sigma$  and time arguments  $t$ , sums over  $n$  and  $m$  are understood to be from 1 to  
410  $N$ , and sums over  $t$  are understood to be from 1 to  $T$ .

$$|r^{(0)}|^2 = \left( \frac{1}{N} \sum_n w_n^{(0)} \right) \left( \frac{1}{N} \sum_m \overline{w_m^{(0)}} \right) \quad (41)$$

$$= \frac{1}{N^2} \left( \sum_n (\Pi^{(0h)} h_n + d_n) \right) \left( \sum_m (\overline{\Pi^{(0h)} h_m} + \overline{d_m}) \right) \quad (42)$$

$$= \frac{1}{N^2} \sum_n \sum_m (\Pi^{(0h)} h_n \overline{\Pi^{(0h)} h_m} + d_n \overline{\Pi^{(0h)} h_m} + \Pi^{(0h)} h_n \overline{d_m} + d_n \overline{d_m}) \quad (43)$$

$$= |\Pi^{(0h)}|^2 \frac{1}{N^2} \sum_n \sum_m h_n \overline{h_m} + \frac{1}{N^2} \sum_n \sum_m (d_n \overline{\Pi^{(0h)} h_m} + \Pi^{(0h)} h_n \overline{d_m}) + \frac{1}{N^2} \sum_n \sum_m d_n \overline{d_m} \quad (44)$$

$$= |\Pi^{(0h)}|^2 \left| \frac{1}{N} \sum_n h_n \right|^2 + \frac{1}{N^2} \sum_n \sum_m 2 \operatorname{Re}(\Pi^{(0h)} h_n \overline{d_m}) + \left| \frac{1}{N} \sum_n d_n \right|^2. \quad (45)$$

411 Taking the time average and using the fact that  $\langle h_n, d_n \rangle = 0$  to exclude summands with  $n = m$  in the  
 412 middle term of (45), we get

$$\frac{1}{T} \sum_t |r^{(0)}|^2 = |\Pi^{(0h)}|^2 \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n h_n \right|^2 + \frac{1}{TN^2} \sum_t \sum_n \sum_{m \neq n} 2 \operatorname{Re}(\Pi^{(0h)} h_n \overline{d_m}) + \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n d_n \right|^2, \quad (46)$$

413 which proves the first statement of the theorem, i.e., (35)-(37). The remaining claims follow straightfor-  
 414 wardly.

415 The assumption of non-interacting populations not affected by the environments of neighboring sites  
 416 was interpreted to mean (36) is negligible. This is reasonable because for such populations, the magnitude  
 417 of  $\frac{1}{TN^2} \sum_t \sum_n \sum_{m \neq n} h_n \overline{d_m}$  should be negligible, as it quantifies relationships between environmental  
 418 fluctuations at site  $n$ , built into the model  $h_n$ , and residuals at other sites,  $d_m$ . We tested the assumption  
 419 that the cross terms of (36) were negligible for our best short- and long-timescale models, finding that  
 420 they were (Appendix S15). ■

## 421 Appendix S15 Predicted synchrony, fractions of synchrony explained, 422 and cross terms

423 Using the reasoning behind theorem 1 of Appendix S14, and continuing the notation from that section,  
 424 we can consider  $|\Pi_\sigma^{(0h)}| |r_\sigma^{(h)}(t)|$  (see (45)) as it varies with  $\sigma$  and  $t$  as a ‘predicted synchrony’ function,  
 425 comparable with the observed value of  $|r_\sigma^{(0)}(t)|$ . See, e.g., Fig.2b, which shows synchrony predicted by  
 426 our best long-timescale model on long timescales, and which shows synchrony predicted by our best  
 427 short-timescale model on short timescales.

428 Using the wavelet Moran theorem, we can further quantify the proportion of synchrony explained by  
 429 the model as

$$q(\sigma) = \frac{|\Pi_\sigma^{(0h)}|^2 \frac{1}{T} \sum_{t=1}^T |r_\sigma^{(h)}(t)|^2}{\frac{1}{T} \sum_{t=1}^T |r_\sigma^{(0)}(t)|^2}. \quad (47)$$

430 This quantity was averaged over long ( $> 4$  years) or short ( $< 4$  years) timescales, respectively, to provide  
 431 the fractions of synchrony explained, and quoted in the main text, by our best long- and short-timescale  
 432 models of PCI, respectively: 61.1% for the long-timescale model, which had predictors growing season  
 433 temperature and *C. finmarchicus* abundance; and only 3.1% for the short-timescale model, which had  
 434 predictors *C. finmarchicus*, decapod larvae abundance, and echinoderm larvae abundance (see Results in  
 435 main text). Because the short-timescale model explained such a small fraction of observed synchrony in  
 436 PCI, we did not base any conclusions about the nature of PCI synchrony on the model.

437 We tested the assumption that the cross terms of (36) were negligible for our best short- and long-  
 438 timescale models of PCI as follows. The quantity

$$\frac{\frac{1}{TN^2} \sum_{t=1}^T \sum_{n=1}^N \sum_{m \neq n} 2 \operatorname{Re}(\Pi_\sigma^{(0h)} h_{n,\sigma}(t) \overline{d_{m,\sigma}(t)})}{\frac{1}{T} \sum_{t=1}^T |r_\sigma^{(0)}(t)|^2} \quad (48)$$

439 was averaged over long timescales for our best long-timescale model of PCI. The result was 11%. The  
 440 quantity was averaged over short time scales for our best short-timescale model of PCI. The result was  
 441 1.4%. The long-timescale value was relatively small compared to the the fraction of synchrony explained  
 442 (see main text and above). As stated above, the fraction of synchrony explained by the short-timescale  
 443 model was very small so the model was not considered informative about synchrony. The 11% contribution  
 444 of cross terms in the long-timescale model indicates a tendency for the model at a location to be slightly  
 445 positively associated with model residuals at other locations; these associations may be due to mixing  
 446 effects, which render PCI values in each area of the sea slightly dependent on the driving effects of  
 447 temperature and *C. finmarchicus* fluctuations in nearby parts of the sea.

448 The quantity  $q$  is also, as we now show, the fraction of the total power in the mean signal  $\frac{1}{N} \sum_n x_n^{(0)}(t)$   
449 explained by the synchrony of the  $x_n^{(k)}(t)$ . Consider the power  $P$  of the mean of the un-normalized wavelet  
450 transforms  $W_{n,\sigma}(t)$ , which is also the power of the wavelet transform of the mean  $\frac{1}{N} \sum_n x_n^{(0)}(t)$ , by linearity  
451 of the transform. This represents the amount of variability in the mean signal at scale  $\sigma$ :

$$P = \frac{1}{T} \sum_{t=1}^T \left( \frac{1}{N} \sum_{n=1}^N W_{n,\sigma}(t) \right) \overline{\left( \frac{1}{N} \sum_{n=1}^N W_{n,\sigma}(t) \right)} \quad (49)$$

$$= \left[ \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T W_{n,\sigma}(t) \overline{W_{n,\sigma}(t)} \right] \left[ \frac{1}{T} \sum_{t=1}^T \left( \frac{1}{N} \sum_{n=1}^N w_{n,\sigma}(t) \right) \overline{\left( \frac{1}{N} \sum_{n=1}^N w_{n,\sigma}(t) \right)} \right] \quad (50)$$

$$= \left[ \frac{1}{N} \sum_{n=1}^N \frac{1}{T} \sum_{t=1}^T W_{n,\sigma}(t) \overline{W_{n,\sigma}(t)} \right] \left[ \{|r_\sigma^{(0)}(t)|\}_t \right]. \quad (51)$$

452 Here we use the definitions of the power normalized transform and the mean field. Thus the power of  
453 the mean signal depends on the mean of wavelet powers and on synchrony  $\{|r_\sigma^{(0)}(t)|\}_t$ . The component  
454 of synchrony attributable to the model  $v$  is  $|\Pi_\sigma^{(0h)}|^2 \{|r_\sigma^{(h)}(t)|\}_t$ , so the component of  $P$  explained by  $v$  is

$$P_v = \left[ \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T W_{n,\sigma}(t) \overline{W_{n,\sigma}(t)} \right] \left[ |\Pi_\sigma^{(0h)}|^2 \{|r_\sigma^{(h)}(t)|\}_t \right] \quad (52)$$

455 and the fraction of the total power of the mean explained by synchrony in the  $x_n^{(k)}(t)$  is

$$\frac{P_v}{P} = \frac{|\Pi_\sigma^{(0h)}|^2 \frac{1}{T} \sum_{t=1}^T |r_\sigma^{(h)}(t)|^2}{\frac{1}{T} \sum_{t=1}^T |r_\sigma^{(0)}(t)|^2} = q(\sigma). \quad (53)$$

456 If the number of sampling locations is large, a case for which local effects are negligible, the remaining  
457 power in the fluctuations of the total population may be explained by the synchronizing influence of one  
458 or more other variables unrelated to the  $x_n^{(k)}(t)$  for  $k = 1, \dots, K$ .

## 459 Appendix S16 Testing for significance of interaction effects

460 We tested whether the synchrony attributable to our model included a substantial effect due to interac-  
461 tions between the influences of the predictor variables occasioned by their non-independence. We here  
462 describe the case in which the best model had two predictors,  $K = 2$  in Eq.2 of the main text, for simplic-  
463 ity and because the best long-timescale model on which we focus in Results had  $K = 2$ . Having fitted the  
464 model with data to determine  $\beta_k(\sigma)$  for  $k = 1, 2$ , we replaced  $x_n^{(2)}(t)$  by *spatially synchronous surrogates*  
465 (Appendix S8),  $\tilde{x}_n^{(2)}(t)$ , which had the same patterns of autocorrelation and synchrony as the unmanipu-  
466 lated  $x_n^{(2)}(t)$ , but were unrelated to  $x_n^{(1)}(t)$ . Mean squared model synchrony  $\{|\tilde{r}_\sigma^{(h)}(t)|\}_t$  was recomputed  
467 using the surrogates and their transforms  $\tilde{w}_{n,\sigma}^{(2)}(t)$  in Eq.2 of the main text in place of  $w_{n,\sigma}^{(2)}(t)$  (but still  
468 using the original  $\beta_k(\sigma)$ ). This was done for 1000 independent sets of spatially synchronous surrogates  
469 and the average surrogate mean squared model synchrony was computed and compared to  $\{|r_\sigma^{(h)}(t)|\}_t$ .  
470 For comparison, the same procedure was carried out with *asynchronous surrogates* (Appendix S8) of  
471  $x_n^{(2)}(t)$ , which randomize away both relationships with  $x_n^{(1)}(t)$  and synchrony of  $x_n^{(2)}(t)$ , while preserving  
472 the autocorrelation structure of  $x_n^{(2)}(t)$ .

473 **Appendix S17 Attributing synchrony to individual predictors and in-**  
 474 **teractions among predictors**

475 The interaction terms in the following theorem indicate synchrony is not only the ‘sum of its parts’.

476 **Theorem 2. Synchrony attribution theorem** *Using the notation of the previous two sections, the*  
 477 *portion of synchrony explained by a wavelet model can be partitioned as*

$$|\Pi_{\sigma}^{(0h)}|^2 \frac{1}{T} \sum_t |r_{\sigma}^{(h)}(t)|^2 = \sum_k |\beta_k(\sigma)|^2 \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n w_{n,\sigma}^{(k)}(t) \right|^2 + \sum_j \sum_{k \neq j} \frac{1}{T} \sum_t \left( \beta_j(\sigma) \frac{1}{N} \sum_m w_{m,\sigma}^{(j)}(t) \right) \overline{\left( \beta_k(\sigma) \frac{1}{N} \sum_n w_{n,\sigma}^{(k)}(t) \right)}. \quad (54)$$

478 The  $k$ th summand in the first term on the right is the timescale-specific portion of synchrony attributable to  
 479 the  $k$ th predictor. The second term on the right is the timescale-specific portion of synchrony attributable  
 480 to interaction effects. The second term is real-valued and can be positive or negative.

481 *Proof.* The following equations omit explicit dependencies on  $\sigma$  and  $t$  for brevity.

$$|\Pi^{(0h)}|^2 \frac{1}{T} \sum_t |r^{(h)}|^2 = \frac{1}{T} \sum_t \left| \Pi^{(0h)} \frac{1}{N} \sum_n h_n \right|^2 \quad (55)$$

$$= \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n v_n \right|^2 \quad (56)$$

$$= \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n \sum_k \beta_k w_n^{(k)} \right|^2 \quad (57)$$

$$= \frac{1}{TN^2} \sum_t \left[ \left( \sum_j \sum_m \beta_j w_m^{(j)} \right) \overline{\left( \sum_k \sum_n \beta_k w_n^{(k)} \right)} \right] \quad (58)$$

$$= \frac{1}{TN^2} \sum_t \left[ \sum_k \left| \sum_n \beta_k w_n^{(k)} \right|^2 + \sum_j \sum_{k \neq j} \left( \sum_m \beta_j w_m^{(j)} \right) \overline{\left( \sum_n \beta_k w_n^{(k)} \right)} \right] \quad (59)$$

$$= \frac{1}{T} \sum_t \sum_k \left| \frac{1}{N} \sum_n \beta_k w_n^{(k)} \right|^2 + \frac{1}{T} \sum_t \sum_j \sum_{k \neq j} \left( \frac{1}{N} \sum_m \beta_j w_m^{(j)} \right) \overline{\left( \frac{1}{N} \sum_n \beta_k w_n^{(k)} \right)} \quad (60)$$

$$= \sum_k |\beta_k|^2 \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n w_n^{(k)} \right|^2 + \sum_j \sum_{k \neq j} \frac{1}{T} \sum_t \left( \beta_j \frac{1}{N} \sum_m w_m^{(j)} \right) \overline{\left( \beta_k \frac{1}{N} \sum_n w_n^{(k)} \right)}. \quad (61)$$

482 The remaining statements of the theorem follow straightforwardly. ■

483 Applying the synchrony attribution theorem, we can partition the fraction of synchrony explained  
 484 (47) as  $q(\sigma) = \sum_k q_k(\sigma) + q_{\text{int}}(\sigma)$ , where

$$q_k(\sigma) = \frac{|\beta_k|^2 \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_n w_n^{(k)} \right|^2}{\frac{1}{T} \sum_{t=1}^T |r_{\sigma}^{(0)}(t)|^2} \quad (62)$$

$$q_{\text{int}}(\sigma) = \frac{\sum_j \sum_{k \neq j} \frac{1}{T} \sum_t \left( \beta_j \frac{1}{N} \sum_m w_m^{(j)} \right) \overline{\left( \beta_k \frac{1}{N} \sum_n w_n^{(k)} \right)}}{\frac{1}{T} \sum_{t=1}^T |r_{\sigma}^{(0)}(t)|^2}. \quad (63)$$

485 These quantities were averaged over long timescales to provide the fractions quoted in the main text  
486 of long-timescale synchrony in PCI explained by growing season temperature (41.4%), *C. finmarchicus*  
487 abundance (5.4%), and interactions between these two (14.3%). The sum of the quantities, 61.1%, is  
488 the previously-quoted fraction of long-timescale synchrony explained by our best long-timescale model,  
489 as expected.

490 Attribution of synchrony to particular drivers and interaction effects as well as the construction of  
491 best models on which this attribution is based are both complicated by the regime shift that occurred  
492 in the 1980s in the seas around the UK (Beaugrand and Reid, 2003; PCI increased, *C. finmarchicus*  
493 abundance decreased, and there were other effects). The regime shift may mean there were changes  
494 in biological dependencies from before to after the shift, so that our results represent an average of  
495 distinct phenomena before and after the regime shift instead of being a reflection of one consistent set of  
496 relationships between variables. This possibility is difficult to assess because time series are too short to  
497 reliably perform wavelet analyses separately for data before and after the shift.

## 498 Appendix S18 Tests of methods using the numeric example

499 Treating the simulated time series of the numeric example ( $\alpha(n, t)$  and  $\beta(n, t)$  from Fig.1a,d and  $\gamma^{(3)}(n, t)$ )  
500 in the same way as real ecological data, we evaluated our wavelet methods by checking they gave results  
501 consistent with the construction of those data.

502 We constructed and tested a wavelet model of  $\gamma^{(3)}(n, t)$  as described in Appendix S9. Working back  
503 from the time series only, we constructed the best fit model of the wavelet transform of  $\gamma^{(3)}(n, t)$  in terms  
504 of the wavelet transforms of its two drivers  $\alpha(n, t)$  and  $\beta(n, t)$ , as in equation 2. We applied the method  
505 of spatially synchronous surrogates (Appendix S8) to determine that adding the second variable to a  
506 one-variable model incorporating only either  $\alpha(n, t)$  or  $\beta(n, t)$  would produce a statistically significant  
507 improvement ( $p = 0.000$  for 1000 surrogates in both cases) in goodness of fit.

508 We verified that the synchrony explicable by the best fit model (according to the wavelet Moran  
509 theorem, see Appendix S14 and Appendix S15) is consistent with the actual synchrony of  $\gamma^{(3)}(n, t)$ , as  
510 expected since the local noise added to  $\gamma^{(3)}$  was asynchronous. At the 20 year timescale, the expected  
511 synchrony according to the best fit model was 1.022 times the synchrony observed in the wavelet mean  
512 field of  $\gamma(n, t)$ , i.e., all of the synchrony in  $\gamma^{(3)}$  was attributable to  $\alpha(n, t)$  and  $\beta(n, t)$ , as expected. In the  
513 analysis of real PCI data, the synchrony associated with the statistical model we constructed was only  
514 some fraction of the observed synchrony, due to the other, unattributed components of variability having  
515 non-zero spatial synchrony of their own.

516 We evaluated the size of the interaction effects between  $\alpha(n, t)$  and  $\beta(n, t)$  as in Appendix S16: 1)  
517 we evaluated the mean squared synchrony of the model (evaluated at the 20-year timescale); and we  
518 compared it to 2) the same quantity calculated after replacing  $\alpha(n, t)$  by synchrony-preserving surrogates  
519 (we used 1000 surrogates and took the average). The former quantity was 1.24 times the latter, showing  
520 substantial interaction effects were detected.

## 521 Appendix S19 Lemmas and extended computations

522 **Lemma 3.**  $\text{cov}_t(\sin(wt), \sin(wt - \varphi\pi)) \approx \frac{1}{2} \cos(\varphi\pi)$ , where  $w = 2\pi/20$ ,  $\text{cov}_t$  is the sample covariance  
523 computed over  $t = 1, \dots, T - 1$ , and  $T$  is divisible by 20.

*Proof.*

$$\text{cov}_t(\sin(wt), \sin(wt - \varphi\pi)) = \frac{1}{T-1} \sum_{t=0}^{T-1} \sin(wt) \sin(wt - \varphi\pi) \quad (64)$$

$$= \frac{1}{T-1} \sum_{t=0}^{T-1} \sin(wt)[\sin(wt) \cos(\varphi\pi) - \cos(wt) \sin(\varphi\pi)] \quad (65)$$

$$= \cos(\varphi\pi) \frac{1}{T-1} \sum_{t=0}^{T-1} \sin^2(wt) - \sin(\varphi\pi) \frac{1}{T-1} \sum_{t=0}^{T-1} \sin(wt) \cos(wt) \quad (66)$$

$$\approx \frac{1}{2} \cos(\varphi\pi) - \sin(\varphi\pi) \frac{1}{T-1} \sum_{t=0}^{T-1} \frac{\sin(2wt)}{2} \quad (67)$$

$$= \frac{1}{2} \cos(\varphi\pi), \quad (68)$$

where (65) and (67) follow from the trigonometric identities  $\sin(x \pm y) = \sin(x)\cos(y) \pm \cos(x)\sin(y)$  and  $\sin(2x) = 2\sin(x)\cos(x)$ , respectively.  $\blacksquare$

The following lemma was taken, verbatim, from Sheppard et al. [2016].

**Lemma 4.** *The spatial coherence,  $|\Pi_\sigma^{(01)}|$ , is between 0 and 1.*

*Proof.* See lemma 2 of Sheppard et al. [2016].  $\blacksquare$

**Lemma 5.** *Let  $X$  be a complex Hilbert space, let  $\langle \cdot, \cdot \rangle$  be its inner product, let  $V$  be a finite-dimensional subspace of  $X$ , and take  $x \in X$ . Then there exists a unique  $v \in V$  such that  $x = v + d$  and  $d \perp V$ . Furthermore,  $\|d\| \leq \|x - w\|$  for all  $w \in V$ , with equality if and only if  $w = v$ . And  $\frac{|\langle x, w \rangle|}{\|x\| \|w\|} \leq \frac{|\langle x, v \rangle|}{\|x\| \|v\|}$  for all  $w \in V$ , with equality if and only if  $w = \alpha v$  for some complex number  $\alpha$ . Additionally, if  $v_1, \dots, v_K \in X$ ,  $V = \text{Span}(v_1, \dots, v_K)$ , and  $v = \beta_1 v_1 + \dots + \beta_K v_K$ , then  $\beta = (\beta_1, \dots, \beta_K)^t$  is a solution of the equation  $M\beta = p$  where  $M = \left[ \frac{\langle v_j, v_i \rangle}{\|v_i\|^2} \right]_{ij}$  and  $p = (p_1, \dots, p_K)^t$  with  $p_j = \frac{\langle x, v_j \rangle}{\|v_j\|^2}$ .*

*Proof.* By theorem 2.1 on page 253 in Loomis and Sternberg [1968],  $X = V \oplus V^\perp$ , so that gives a unique  $v \in V$  and  $d \in V^\perp$  such that  $x = v + d$ . By lemma 2.1 on page 252 of Loomis and Sternberg [1968],  $\|d\| \leq \|x - w\|$  for all  $w \in V$ , with equality if and only if  $w = v$ .

Now let  $w_1, \dots, w_m$  be an orthonormal basis for  $V$  such that  $w_1 = \frac{v}{\|v\|}$ . Then, given  $w \in V$ , write  $w = \alpha_1 w_1 + \dots + \alpha_m w_m$ . We have

$$\frac{|\langle x, w \rangle|}{\|x\| \|w\|} = \frac{|\langle \|v\| w_1 + d, \alpha_1 w_1 + \dots + \alpha_m w_m \rangle|}{\|x\| \sqrt{|\alpha_1|^2 + \dots + |\alpha_m|^2}} = \frac{\|v\| |\alpha_1|}{\|x\| \sqrt{|\alpha_1|^2 + \dots + |\alpha_m|^2}} \leq \frac{\|v\|}{\|x\|} = \frac{|\langle x, v \rangle|}{\|x\| \|v\|}. \quad (69)$$

Equality holds if and only if  $\alpha_2 = \dots = \alpha_m = 0$ , in other words, if and only if  $w = \alpha v$ .

The last statement of the lemma is equivalent to  $\sum_{j=1}^K \frac{\langle v_j, v_i \rangle}{\|v_i\|^2} \beta_j = \frac{\langle x, v_i \rangle}{\|v_i\|^2}$  for all  $i = 1, \dots, K$ , but

$$\frac{\langle x, v_i \rangle}{\|v_i\|^2} = \frac{\langle v, v_i \rangle}{\|v_i\|^2} = \frac{\langle \sum_{j=1}^K \beta_j v_j, v_i \rangle}{\|v_i\|^2} = \sum_{j=1}^K \beta_j \frac{\langle v_j, v_i \rangle}{\|v_i\|^2}. \quad (70)$$

$\blacksquare$

The following lemma was taken, verbatim, from Sheppard et al. [2016].

<sup>544</sup> **Lemma 6.** *The mean squared magnitude of the wavelet mean field,  $\frac{1}{T} \sum_{t=1}^T |r_\sigma(t)|^2$ , satisfies*

$$\frac{1}{T} \sum_{t=1}^T |r_\sigma(t)|^2 = \frac{\frac{1}{T} \sum_{t=1}^T \left[ \left( \frac{1}{N} \sum_{n=1}^N W_{n,\sigma}(t) \right) \left( \frac{1}{N} \sum_{n=1}^N \overline{W_{n,\sigma}(t)} \right) \right]}{\frac{1}{N} \sum_{n=1}^N \frac{1}{T} \sum_{t=1}^T W_{n,\sigma}(t) \overline{W_{n,\sigma}(t)}} \quad (71)$$

<sup>545</sup> and

$$0 \leq \frac{1}{T} \sum_{t=1}^T |r_\sigma(t)|^2 \leq 1, \quad (72)$$

<sup>546</sup> with

$$\frac{1}{T} \sum_{t=1}^T |r_\sigma(t)|^2 = 1 \quad (73)$$

<sup>547</sup> if and only if all the time series are identical. The denominator of the right side of (71) is the average of  
<sup>548</sup> the powers of all the time series,  $x_n(t)$ , and the numerator is the power of the average time series, and  
<sup>549</sup> hence equation (71) is a timescale-specific generalization of

$$\frac{\text{var} \left( \frac{1}{N} \sum_{n=1}^N x_n(t) \right)}{\left( \frac{1}{N} \sum_{n=1}^N \text{var}(x_n(t)) \right)}. \quad (74)$$

<sup>550</sup> *Proof.* See lemma 1 of Sheppard et al. [2016]. ■

551 Appendix S20 Figures and tables

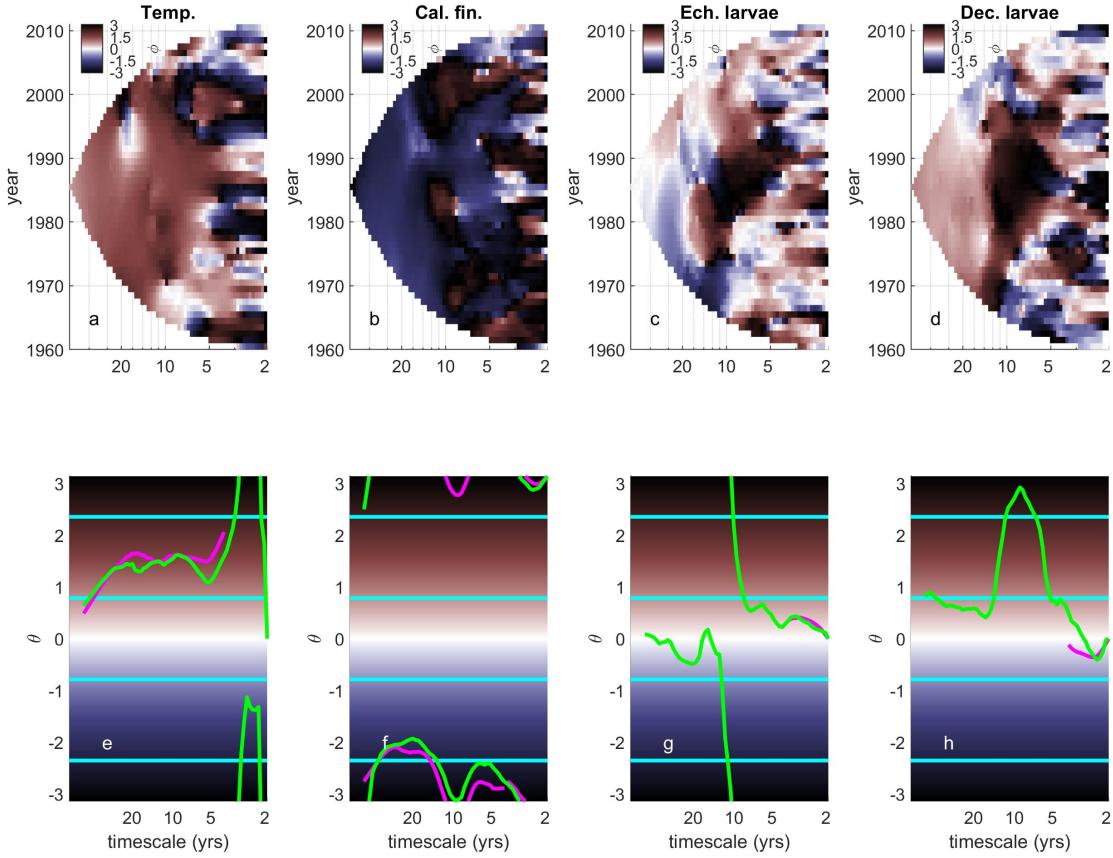


Figure S1: Phytoplankton wavelet phase shifts relative to four explanatory variables, compared to phases of coefficients of top-ranked models, for the purposes of testing and interpreting top-ranked models. Top panels: The typical phase difference,  $\phi$ , i.e. the phase of  $\frac{1}{N} \sum_n e^{i\eta(n,t)}$ , where  $\eta(n,t)$  is the phase of  $w_{n,\sigma}^{(0)}(t)w_{n,\sigma}^{(k)}(t)$ , for the four predictors  $k$  appearing in the top-ranked models at long and short timescales: growing season temperature (a); *C. finmarchicus* abundance (b); echinoderm larvae abundance (c); and decapod larvae abundance (d). Bottom panels: The typical phase difference,  $\theta$ , i.e. the phase of  $\frac{1}{NT} \sum_{n,t} e^{i\eta(n,t)}$  (green), compared to phases of corresponding coefficients  $\beta_k(\sigma)$  from top-ranked models (magenta). Magenta lines extend across long timescales for predictors included in the top-ranked long-timescale model, across short timescales for predictors included in the top-ranked short-timescale model, and across long and short timescales for *C. finmarchicus*, since that variable was in the top-ranked models for both long and short timescales. Temp.=growing season temperature; *C. fin.*=*C. finmarchicus*; Ech.=echinoderm; Dec.=decapod.

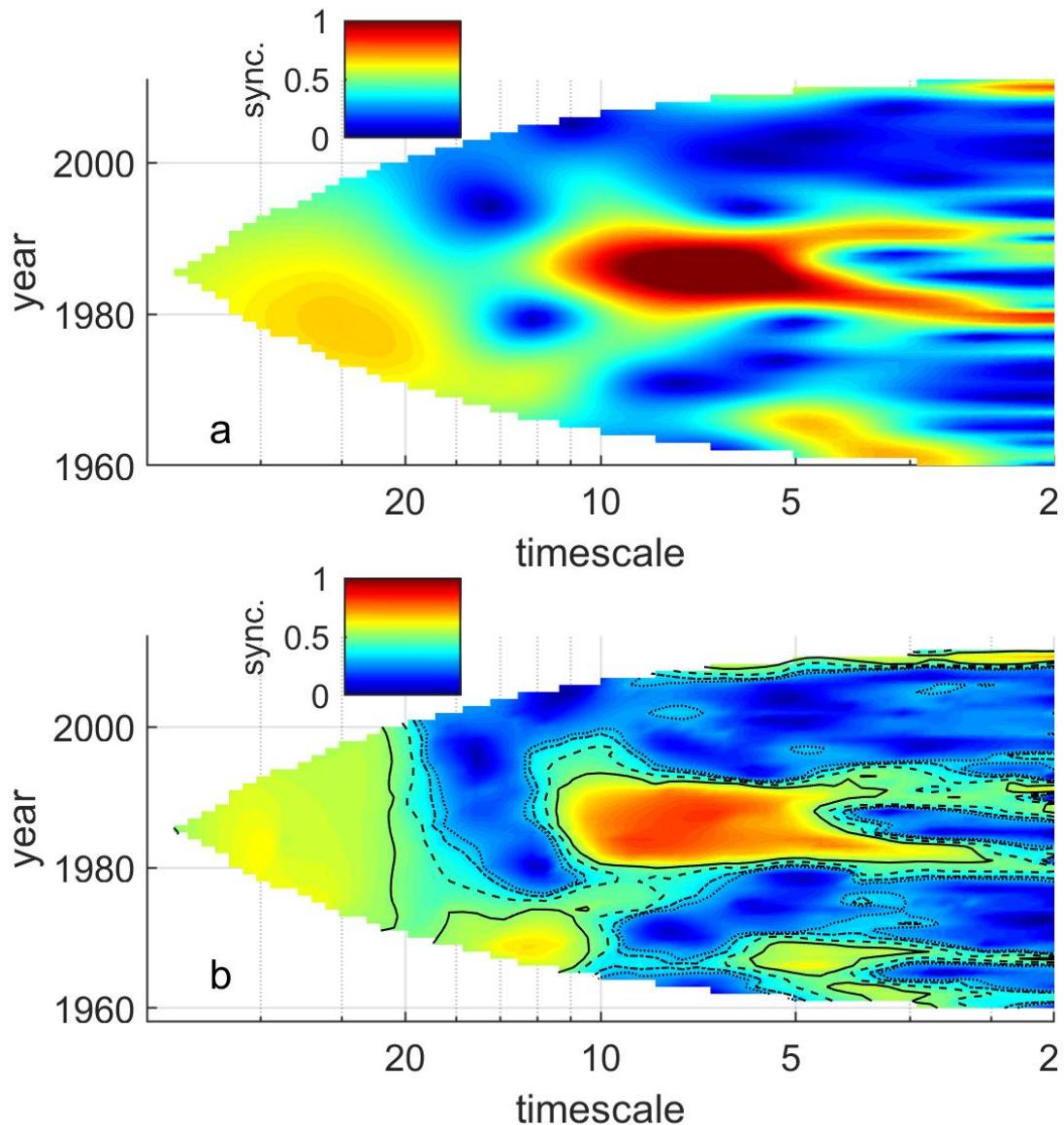


Figure S2: The synchrony of PCI expressed as wavelet mean field magnitude (a) and wavelet phasor mean field magnitude (b) plots. Statistical significance thresholds on the WPMFM are plotted as contours showing actual phase agreement between locations greater than the 90th, 95th, 99th and 99.9th percentile of a distribution of unsynchronized unit phasors (dotted, dash-dotted, dashed and line contours respectively).

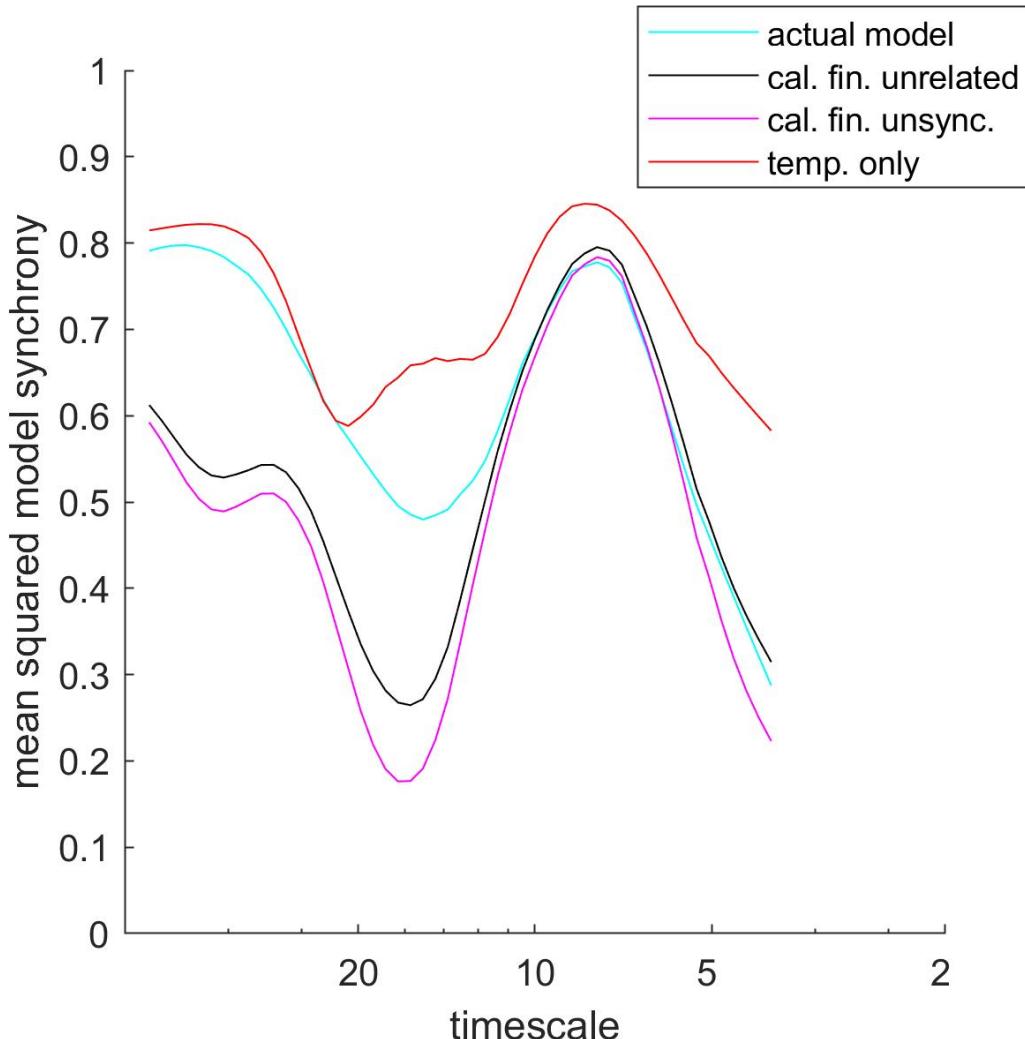


Figure S3: Like Fig.3 from the main text, but with one additional line (the red line) representing mean squared model synchrony after setting the coefficient  $\beta_k(\sigma)$  of the *C. finmarchicus* term in the best long-timescale model (see main text) to 0.

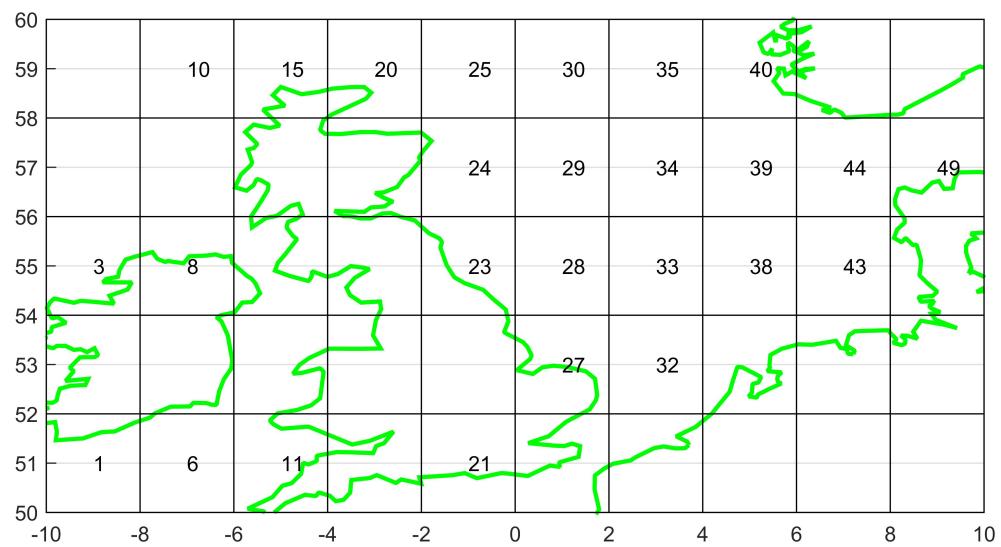


Figure S4: The 26 areas of the North Sea and British seas which were used.

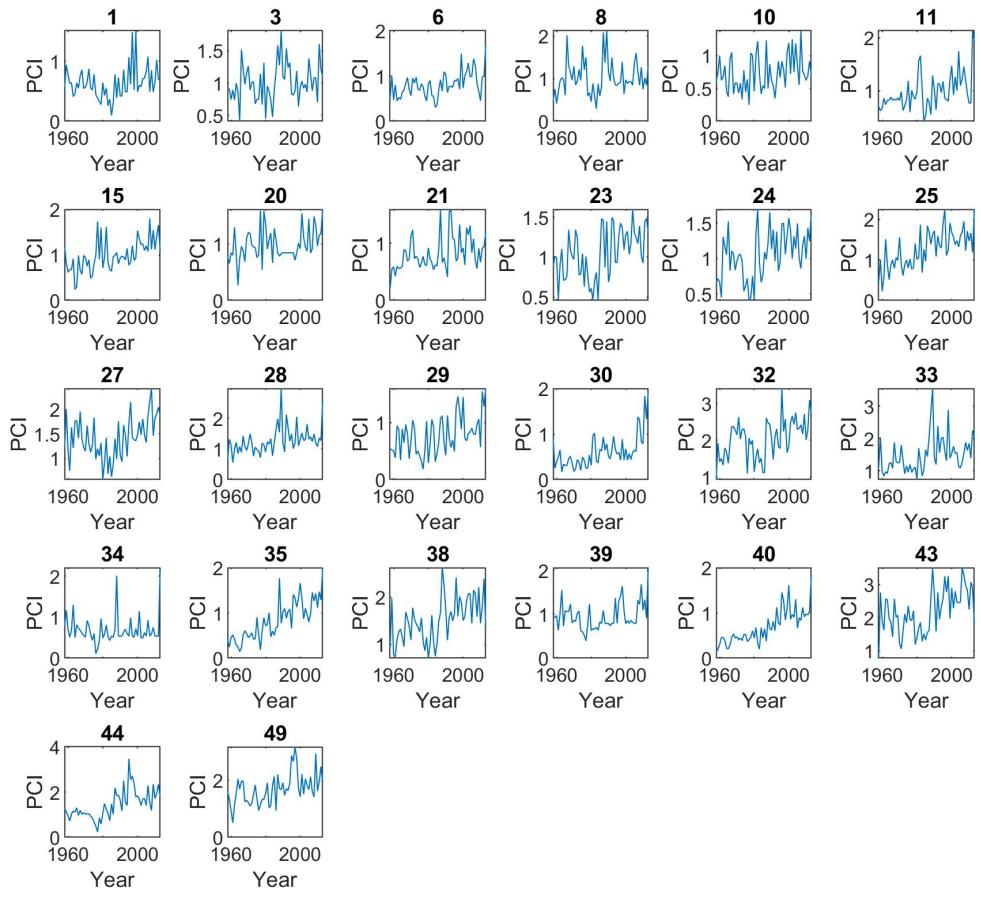


Figure S5: The annualized time series for PCI (mean of 12 monthly values).

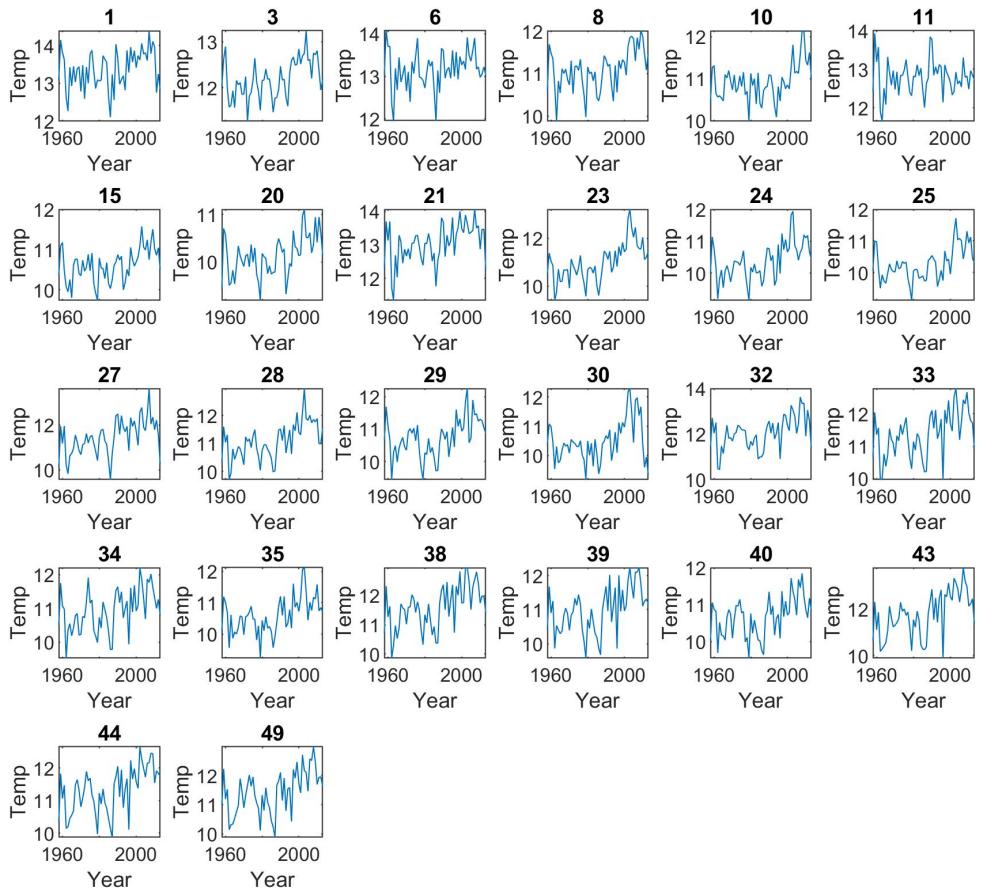


Figure S6: The annualized time series for growing season temperature (mean of monthly values for March to September).

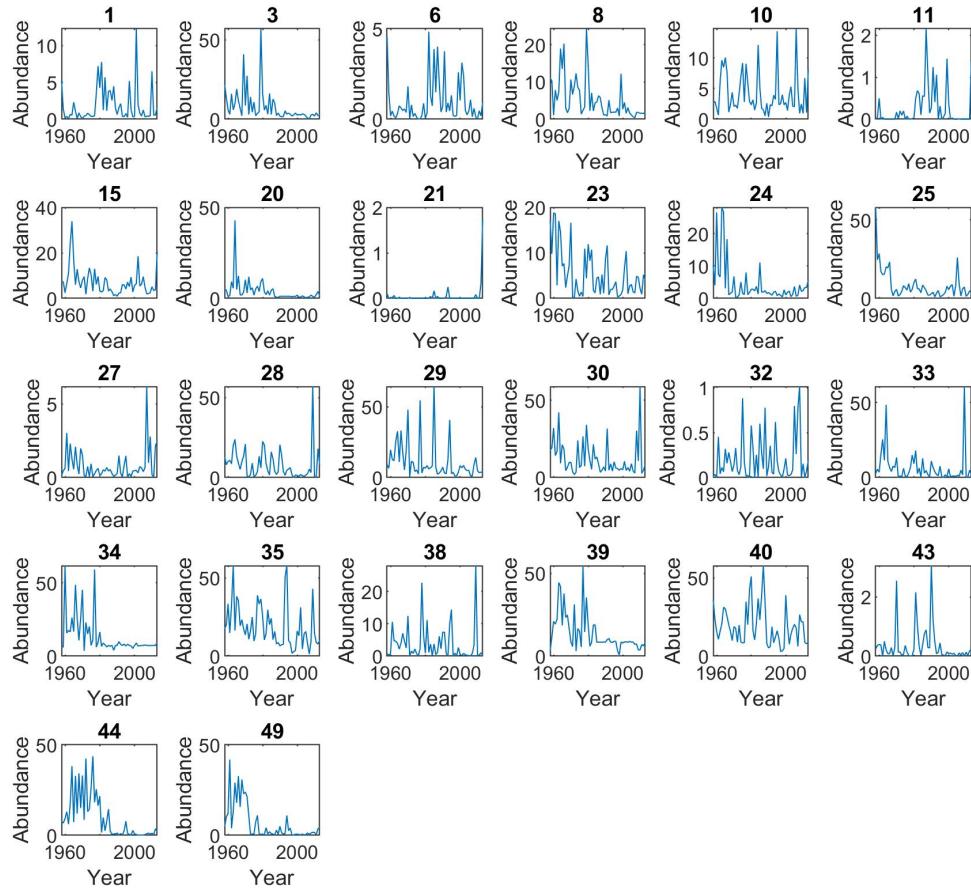


Figure S7: The annualized time series for *C. finmarchicus* abundance (mean of 12 monthly values).

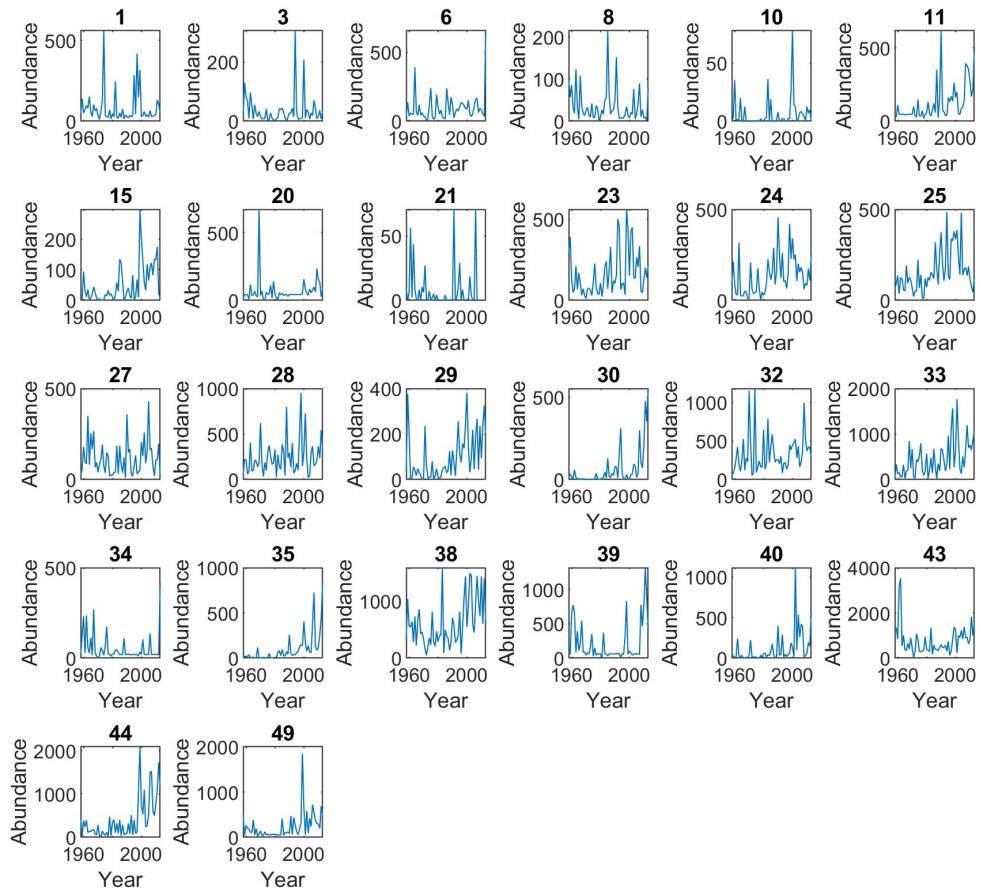


Figure S8: The annualized time series for echinoderm larvae abundance (mean of 12 monthly values).

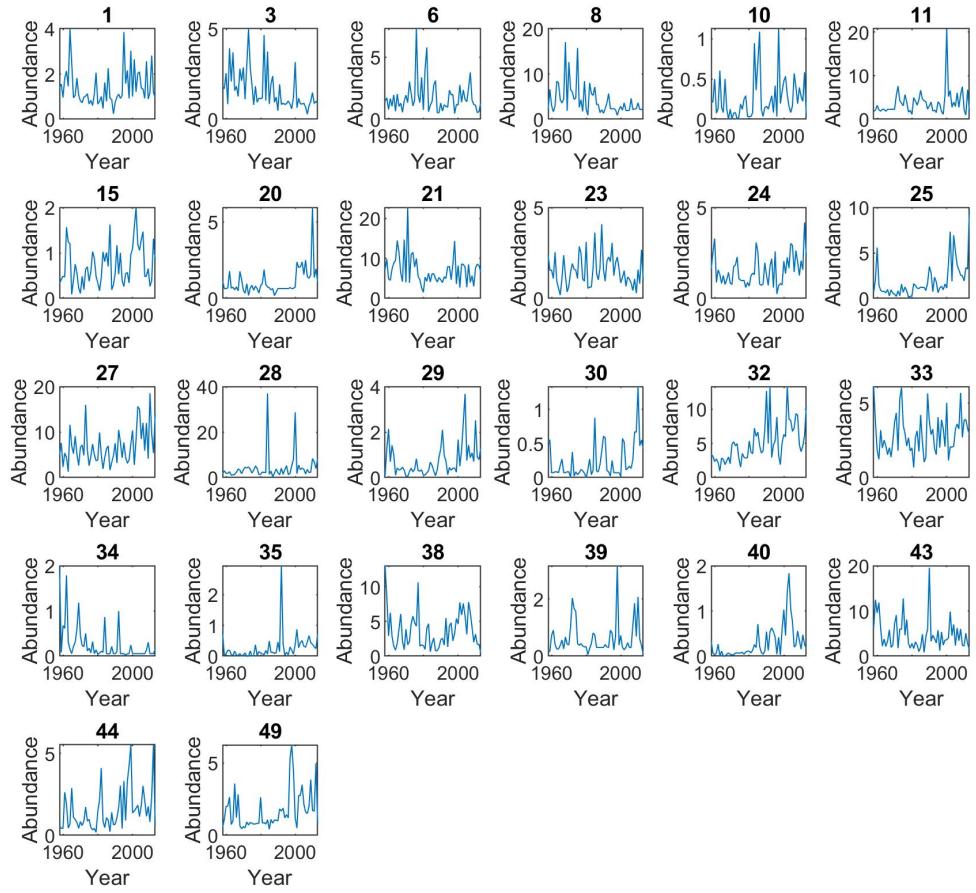


Figure S9: The annualized time series for decapod larvae abundance (mean of monthly values).

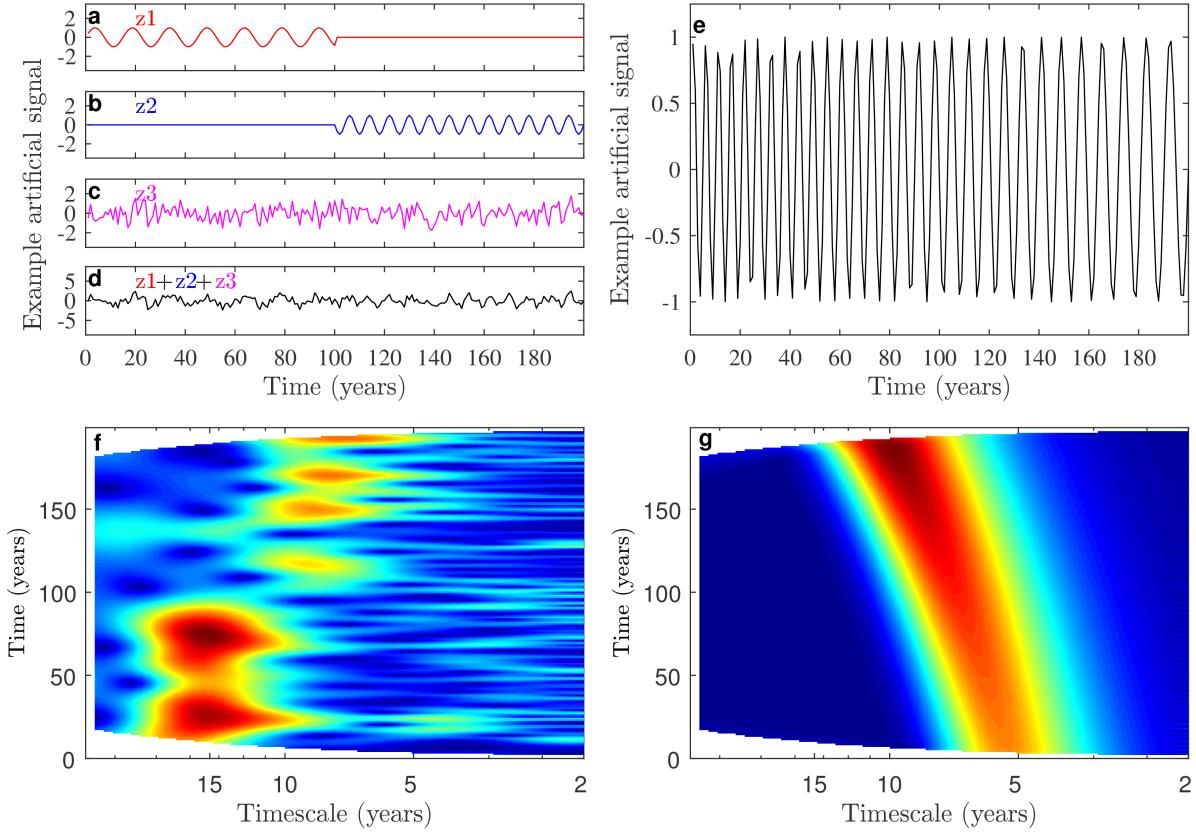


Figure S10: Demonstrations of the detection of oscillations on different timescales for different times using the magnitude of the wavelet transform. The time series of panel d was the sum of: 1) a sine wave of amplitude 1 and period 15 that operated for the first half of the time series (a); 2) a sine wave of amplitude 1 and period 8 that operated for the second half of the time series (b); and 3) normally distributed white noise of standard deviation 0.5 (c). Although periodicities in (d) and changes therein are difficult to detect by eye with any certainty, the magnitude of the wavelet transform (f; Appendix S5) reveals them clearly. The time series of (e) changes period gradually from 5 to 10. The magnitude of the wavelet transform (g) shows the change.

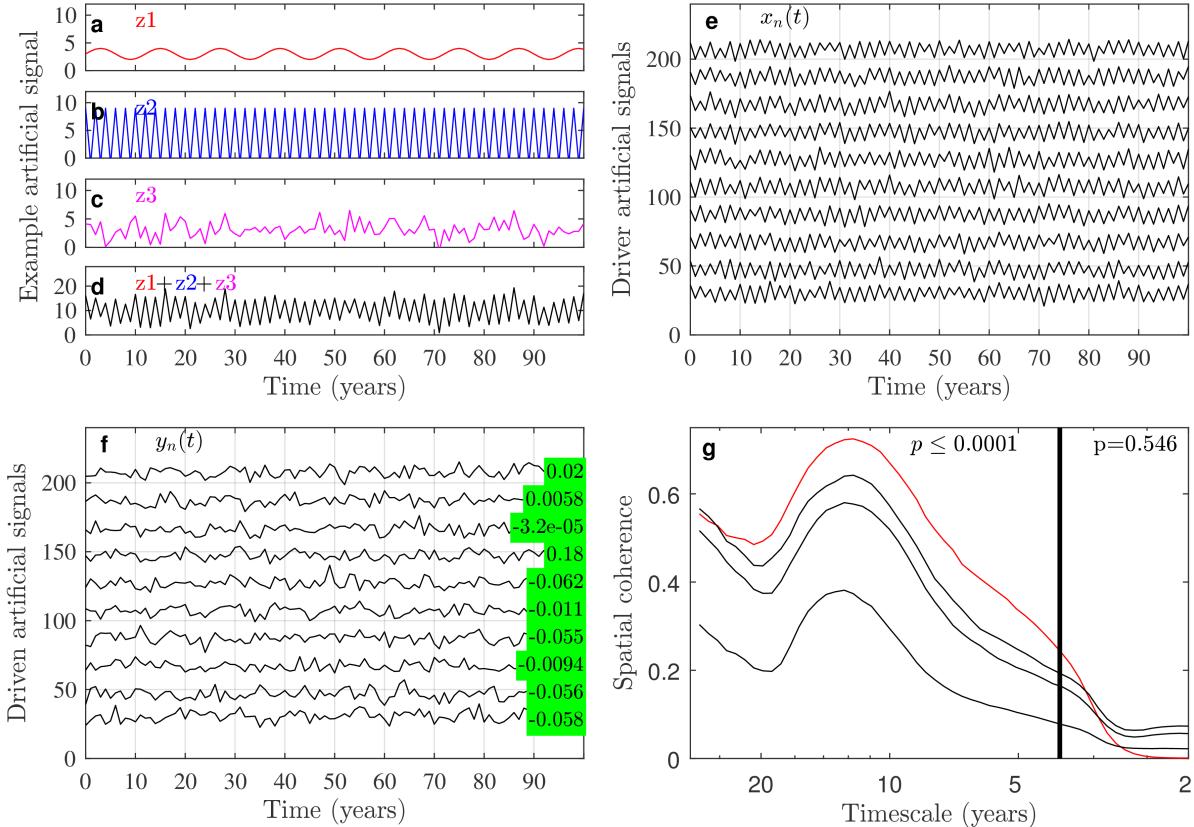


Figure S11: Measuring relationships between variables using spatial coherence. This figure was adapted with only minor changes from supplementary figure 5 of Sheppard et al. [2016]. The time series of panel e were used as drivers in producing the time series of panel f and the figure shows how this relationship can be detected with the spatial coherence technique. The time series of panel e were constructed as the sum of: 1) a single common signal of amplitude 1 and period 12 years (a); 2) a single common signal of amplitude 5 and period 2 years (b); and 3) normally distributed white noise of standard deviation 1.5, independently generated for each of the 10 time series (c). The time series of panel f were produced via the relationship  $y_n(t) = (x_n(t) + x_n(t-1))/2 + \epsilon_n(t)$  where the  $\epsilon_n(t)$  were independent normal random numbers of mean 0 and standard deviation 3. This transmits the period-12 component of the  $x$  signals to the  $y$  but not the period-2 component because averaging covers a whole period for that component. Correlations (f, green numbers) between  $x_n(t)$  and  $y_n(t)$  did not indicate any particular relationship. Correlations cannot detect the relationship between the  $x$  and  $y$  because the technique confounds phenomena occurring on different timescales. Spatial coherences revealed a highly significant relationship at periods around 12 years (g) and on average over long timescales (left  $p$ -value on panel g, long timescales defined as  $> 4$  years) but no relationship (right  $p$ -value) for short timescales ( $< 4$  years). The red line on g is the spatial coherence and black lines are 50th, 95th, and 99th percentiles of spatial coherences of synchrony-preserving surrogate data sets (Appendix S8) appropriately representing the null hypothesis of no relationship between the  $x$  and  $y$ . See also Appendix S9 for a description of how the aggregate long- and short-timescale  $p$ -values were computed.

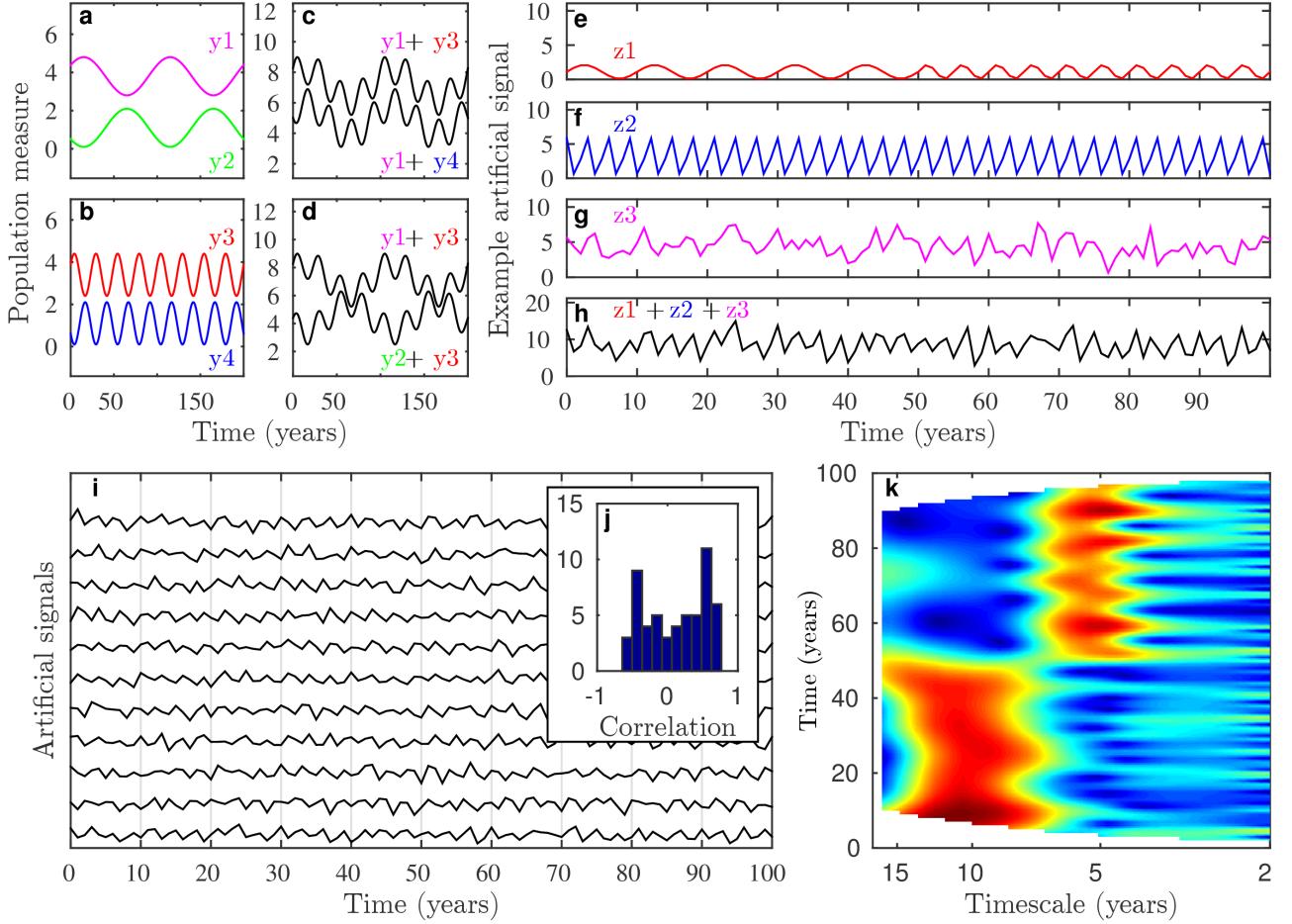


Figure S12: Detecting time- and timescale-specific synchrony with the wavelet mean field. This figure was taken without change from Sheppard et al. [2016]. Panels a-d show the principle of how synchrony can differ for different timescales of dynamics. Time series  $y_1$  and  $y_2$  (a) are exactly anti-correlated (out of phase), as are  $y_3$  and  $y_4$  (b). Combining  $y_1$  with  $y_3$  and  $y_1$  with  $y_4$  gives two time series (c) which are synchronized on long timescales, but anti-synchronized on short timescales. The reverse is also possible (d). This timescale-specific structure of synchrony cannot be detected with correlation coefficients, which are 0 for both c and d, because contributions from different timescales cancel. In practice, real population and environmental signals are broadband, and exact cancellation is unlikely, but asynchrony at some frequencies can strongly conceal important synchrony at other frequencies. Panels e-k demonstrate this concealment, using artificial data, and also show how the wavelet mean field detects time- and timescale-specific synchrony. Each of 11 artificial time series were constructed as the sum of: 1) a single common signal of amplitude 1 that changes its oscillatory period at  $t = 50$  from 10 years to 5 years (e); 2) oscillations of amplitude 3 that have the same oscillatory period (3 years), but random and independent phases in each of the 11 constructed time series (f); and 3) white noise of standard deviation 1.5, again independently generated for each of the 11 time series (g). Synchrony in the resulting time series cannot be visually detected (i), nor is it readily apparent by examining the 55 pairwise correlation coefficients between time series (j), which spanned a wide range of values including 0. But the wavelet mean field magnitude (k) showed clear color bands at 10-year period for  $t < 50$  and 5-year period for  $t > 50$ . The wavelet mean field magnitude displays strength of synchrony as a function of timescale of dynamics and time, here with red indicating synchrony and blue asynchrony. Wavelet phasor mean fields provide plots similar to (k) but with values between 0 and 1 that indicate the strength of phase synchrony and that can be straightforwardly significance tested as described in Appendix S13.

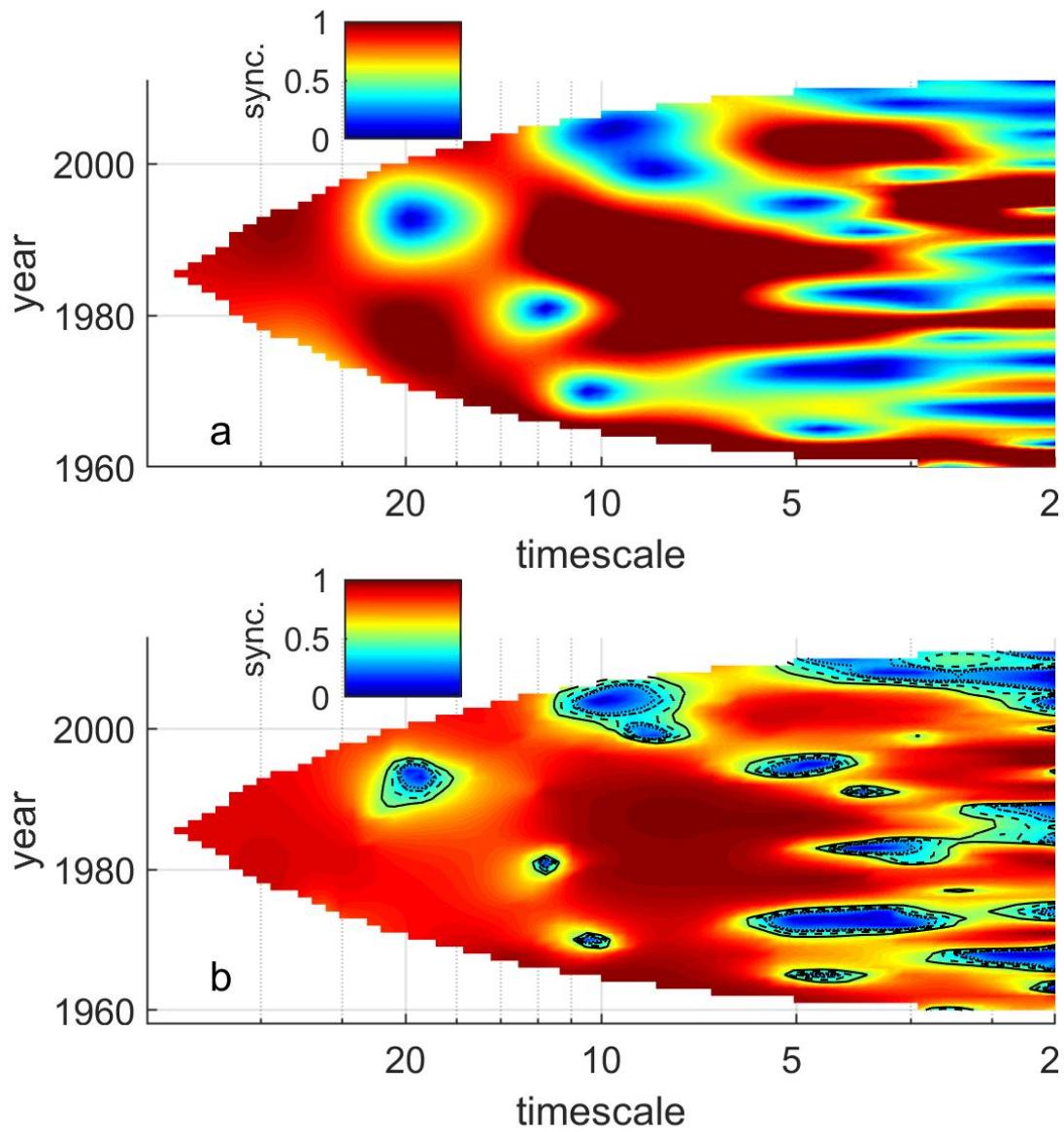


Figure S13: The synchrony of temperature expressed as wavelet mean field magnitude (a) and wavelet phasor mean field magnitude (b) plots. Statistical significance thresholds on the WPMFM are plotted as contours showing actual phase agreement between locations greater than the 90th, 95th, 99th and 99.9th percentile of a distribution of unsynchronized unit phasors (dotted, dash-dotted, dashed and line contours respectively).

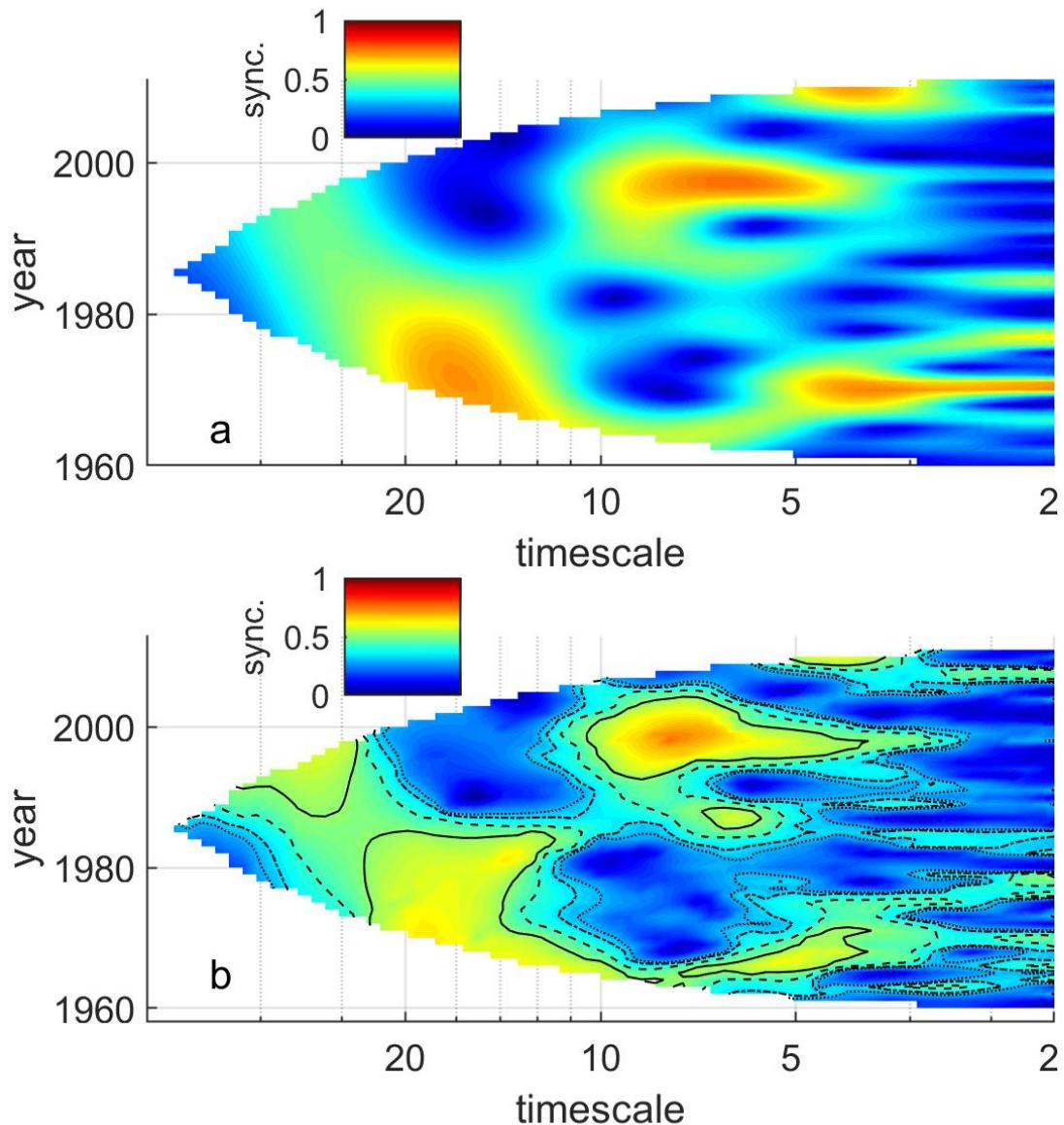


Figure S14: The synchrony of *C. finmarchicus* abundance fluctuations expressed as wavelet mean field magnitude (a) and wavelet phasor mean field magnitude (b) plots. Statistical significance thresholds on the WPMFM are plotted as contours showing actual phase agreement between locations greater than the 90th, 95th, 99th and 99.9th percentile of a distribution of unsynchronized unit phasors (dotted, dash-dotted, dashed and line contours respectively).

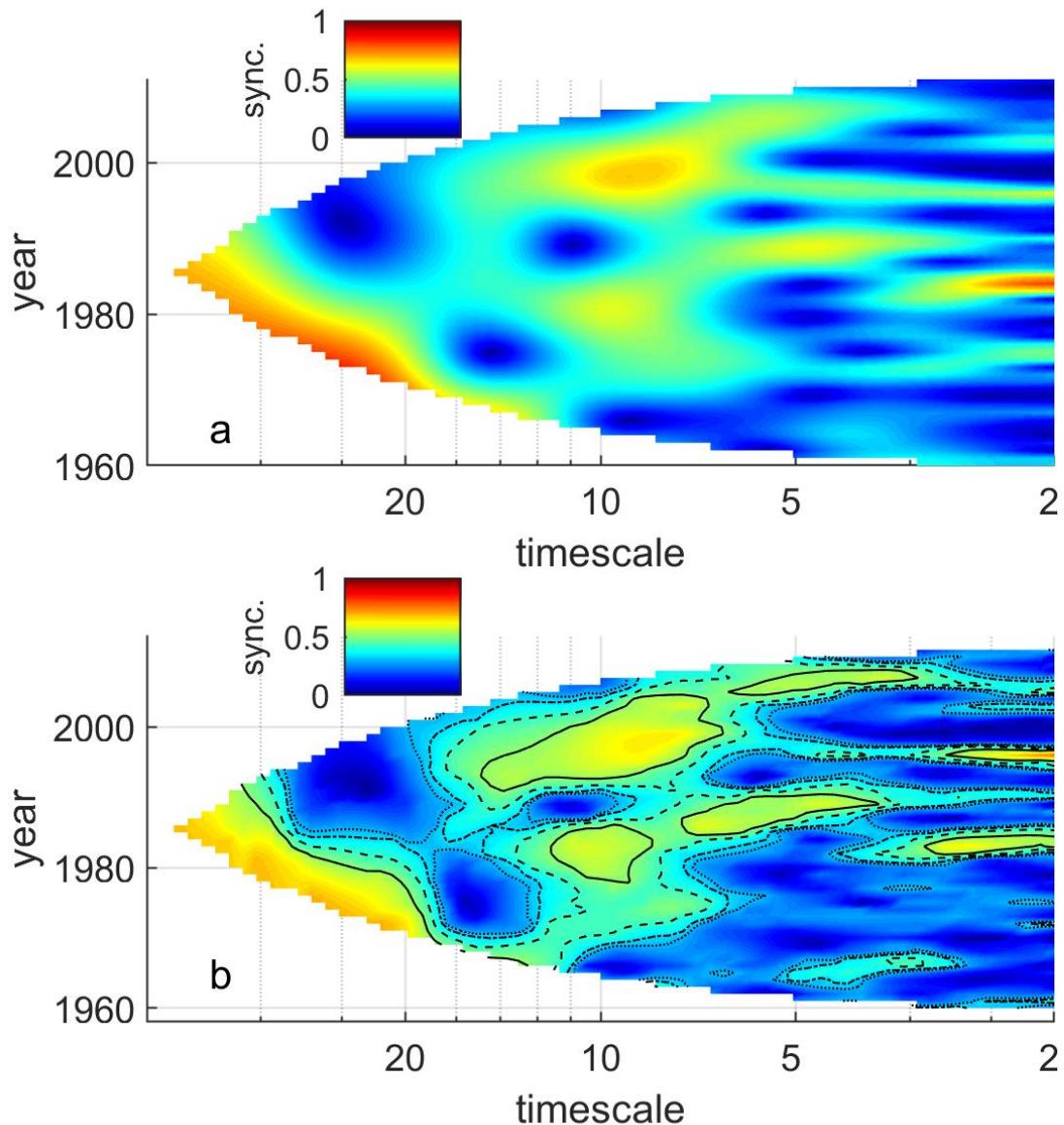


Figure S15: The synchrony of echinoderm larvae abundance fluctuations expressed as wavelet mean field magnitude (a) and wavelet phasor mean field magnitude (b) plots. Statistical significance thresholds on the WPMFM are plotted as contours showing actual phase agreement between locations greater than the 90th, 95th, 99th and 99.9th percentile of a distribution of unsynchronized unit phasors (dotted, dash-dotted, dashed and line contours respectively).

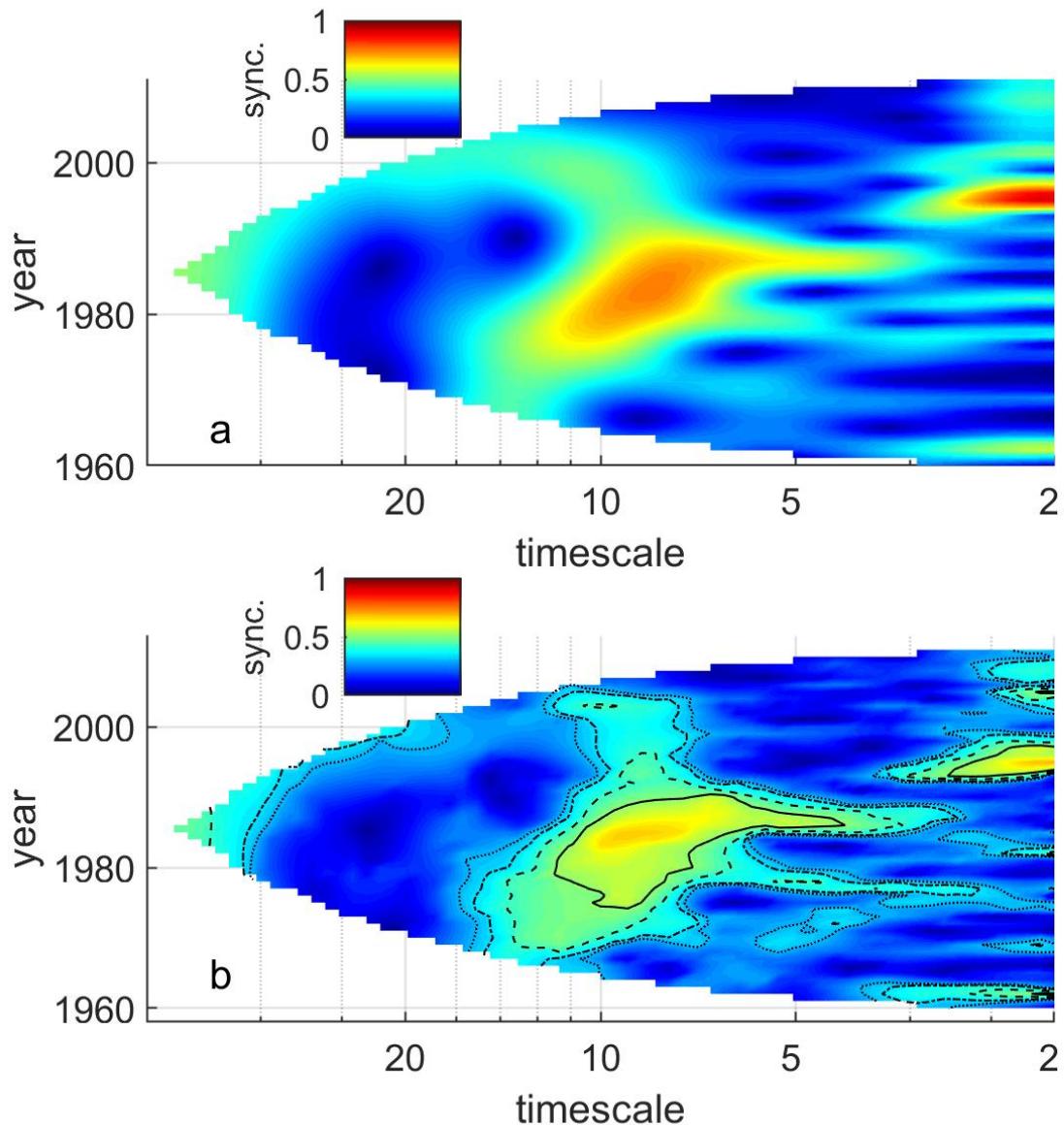


Figure S16: The synchrony of decapod larvae abundance fluctuations expressed as wavelet mean field magnitude (a) and wavelet phasor mean field magnitude (b) plots. Statistical significance thresholds on the WPMFM are plotted as contours showing actual phase agreement between locations greater than the 90th, 95th, 99th and 99.9th percentile of a distribution of unsynchronized unit phasors (dotted, dash-dotted, dashed and line contours respectively).

<b>Low frequency model</b>	<b>Leave-one-out score</b>
Growing season temperature, <i>C. finmarchicus</i>	0.2045
Yearly temperature, Summer salinity, Autumn cloud cover, <i>C. finmarchicus</i>	0.1991
Yearly temperature, <i>C. finmarchicus</i>	0.1982
Growing season temperature, Summer salinity, <i>C. finmarchicus</i>	0.1918
Yearly temperature, Summer salinity, <i>C. finmarchicus</i>	0.1878

Table S1: Table of best long-timescale models. Of all models we considered (Methods), the model with the highest leave-one-out goodness of fit score for which no variables could be dropped without significantly reducing model fit was the model listed in the top row and analyzed in the main text. The table includes all models considered for which the leave-one-out goodness of fit score was at least 90% that of the top model and for which no variables could be dropped without significantly reducing model fit.

Variable	Box-Cox coefficient
<i>Calanus</i> I-IV	-0.116
<i>Para-Pseudocalanus</i> spp.	-0.036
<i>Acartia</i> spp. (unidentified)	0.102
<i>Oithona</i> spp.	0.193
<i>Pseudocalanus elongatus</i> adult	0.135
<i>Temora longicornis</i>	0.137
<i>Centropages typicus</i>	0.197
<i>Calanus finmarchicus</i>	0.088
<i>Calanus helgolandicus</i>	0.110
<i>Metridia lucens</i>	0.125
Echinoderm larvae	0.042
Decapoda larvae (total)	0.061
Euphausiacea (total)	0.099
Yearly temperature	0.942
Spring temperature	1.235
Summer temperature	0.583
Autumn temperature	0.241
Growing season temperature	0.972
Yearly wind speed	1.201
Spring wind speed	0.811
Summer wind speed	0.602
Autumn wind speed	0.951
Growing season wind speed	0.972
Yearly salinity	1.443
Spring salinity	1.616
Summer salinity	0.713
Autumn salinity	1.904
Growing season salinity	1.257
Yearly cloud cover	1.633
Spring cloud cover	1.482
Summer cloud cover	1.755
Autumn cloud cover	1.379
Growing season cloud cover	1.257

Table S2: The mean of the Box-Cox coefficients found at each location, for every variable.

552 **References**

- 553 P. S. Addison. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in*  
554 *Science, Engineering, Medicine and Finance*. Taylor and Francis, New York, 2002.
- 555 A. Bandrivskyy, A. Bernjak, P. V. E. McClintock, and A. Stefanovska. Wavelet phase coherence analysis:  
556 Application to skin temperature and blood flow. *Cardiovasc. Eng.*, 4(1):89–93, 2004.
- 557 G. Beaugrand and P.C. Reid. Long-term changes in phytoplankton, zooplankton and salmon related to  
558 climate. *Global Change Biology*, 9:801–817, 2003.
- 559 G. Beaugrand, K.M. Brander, J.A. Lindley, S. Souissi, and P.C. Reid. Plankton effect on cod recruitment  
560 in the North Sea. *Nature*, 426:661–664, 2003.
- 561 J. Bell, E.C. Burkness, A.E. Milne, D.W. Onstad, M. Abrahamson, K.L. Hamilton, and W.D. Hutchinson.  
562 Putting the brakes on a cycle: bottom-up affects damp cycle amplitude. *Ecology Letters*, 15:310–318,  
563 2012.
- 564 B. Cazelles, M. Chavez, D. Berteaux, F. Ménard, J.O. Vik, S. Jenouvrier, and N.C. Stenseth. Wavelet  
565 analysis of ecological timeseries. *Oecologia*, 156:287–304., 2008.
- 566 B. Cazelles, K. Cazelles, and M. Chavez. Wavelet analysis in ecology and epidemiology: imapct of  
567 statistical tests. *Journal of the Royal Society Interface*, 11:20130585, 2014.
- 568 J.M. Colebrook and G. A. Robinson. Continuous plankton records: Seasonal cycles of phytoplankton and  
569 copepods in the north-eastern Atlantic and the North Sea. *Bulletins of Marine Ecology*, 6:123–139,  
570 1965.
- 571 Emma J. Defriez and Daniel C. Reuman. A global geography of synchrony for marine phytoplankton.  
572 *Global Ecology and Biogeography*, 26(8):867–877, 2017a. ISSN 1466-8238. doi: 10.1111/geb.12594. URL  
573 <http://dx.doi.org/10.1111/geb.12594>.
- 574 Emma J. Defriez and Daniel C. Reuman. A global geography of synchrony for terrestrial vegetation.  
575 *Global Ecology and Biogeography*, 26(8):878–888, 2017b. ISSN 1466-8238. doi: 10.1111/geb.12595.  
576 URL <http://dx.doi.org/10.1111/geb.12595>.
- 577 B.T. Grenfell, O.N. Bjørnstad, and J. Kappey. Traveling waves and spatial hierarchies in measles epi-  
578 demics. *Nature*, 414:716–723, 2001.
- 579 T.H. Keitt. Coherent ecological dynamics induced by large-scale disturbance. *Nature*, 454:331–335, 2008.
- 580 T.H. Keitt and J. Fischer. Detection of scale-specific community dynamics using wavelets. *Ecology*, 87:  
581 2895–2904, 2006.
- 582 L.H. Loomis and S. Sternberg. *Advanced Calculus*. Addison-Wesley Series in Mathematics, New York,  
583 1968.
- 584 S.D. Meyers, B.G. Kelly, and J.J. O'Brien. An introduction to wavelet analysis in oceanography and  
585 meteorology: with application to the dispersion of yanai waves. *Mon. Weather Rev.*, 121:2858–2866,  
586 1993.
- 587 P.A.P. Moran. The statistical analysis of the Canadian Lynx cycle. *Australian Journal of Zoology*, 1:  
588 291–298, 1953.
- 589 D. Prichard and J. Theiler. Generating surrogate data for time series with several simultaneously measured  
590 variables. *Phys. Rev. Lett.*, 73(7):951–954, 1994.

- 591 D. E. Raitsos, P. C. Reid, S. J. Lavender, M. Edwards, and A.J. Richardson. Extending the SeaWiFS  
592 chlorophyll data set back 50 years in the northeast Atlantic. *Geophysical Research Letters*, 32:L06603,  
593 2005.
- 594 D. E. Raitsos, A. Walne, S. J. Lavendar, P. Licandro, P. C. Reid, and M. Edwards. A 60-year ocean colour  
595 data set from the continuous plankton recorder. *Journal of Plankton Research*, 35:158–164, 2013.
- 596 T. Rouyer, J.M. Fromentin, N.C. Stenseth, and B. Cazelles. Analysing multiple time series and extending  
597 significance testing in wavelet analysis. *Marine Ecology Progress Series*, 359:11–23, 2008.
- 598 R.M. Sakia. The Box-Cox transformation technique: a review. *The Statistician*, 41:169–178, 1992.
- 599 T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142(3-4):346–382, AUG 15 2000. ISSN  
600 0167-2789.
- 601 L. W. Sheppard, A. C. Hale, S. Petkoski, P. V. E. McClintock, and A. Stefanovska. Characterizing an  
602 ensemble of interacting oscillators: The mean-field variability index. *Phys.Rev.E*, 87:012905, 2013.
- 603 L.W. Sheppard, J.R. Bell, R. Harrington, and D.C. Reuman. Changes in large-scale climate alter spatial  
604 synchrony of aphid pests. *Nature Climate Change*, 6:610–613, 2016. doi: 10.1038/nclimate2881.
- 605 L.W. Sheppard, P.C. Reid, and D.C. Reuman. Rapid surrogate testing of wavelet coherences. *European  
606 Physical Journal Nonlinear Biomedical Physics*, 5:1–9, 2017.
- 607 A. Stefanovska. Coupled oscillators - Complex but not complicated cardiovascular and brain interactions.  
608 *IEEE Engineering in Biology and Medicine Magazine*, 26(6):25–29, 2007. ISSN 0739-5175. doi: {10.  
609 1109/MEMB.2007.907088}.
- 610 J. W. Strutt. *Scientific Papers by John William Strutt, Baron Rayleigh*, Vol. 3, pp. 47-187 (see in  
611 particular p. 52). Cambridge University Press, Cambridge, 1902.
- 612 J. Theiler, S. Eubank, A. Longtin, B Galdríkian, and J.D. Farmer. Testing for nonlinearity in time series:  
613 the method of surrogate data. *Physica D*, 58(1-4):77–94, 1992.
- 614 C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.*, 79(1):  
615 61–78, 1998.
- 616 C. Viboud, O.N. Bjørnstad, D.L. Smith, L. Simonsen, M.A. Miller, and B.T. Grenfell. Synchrony, waves,  
617 and spatial hierarchies in the spread of influenza. *Science*, 312:447–451, 2006.