The geography of spatial synchrony: Supporting information appendices

Walter, Sheppard, Anderson, Kastens, Bjørnstad, Liebhold, Reuman

Computing was done in R version 3.2.5 and Matlab version 2014b.

S1 The VARMA model as a linearization of a general population model

Let $x_i(t)$ be the population density in patch i at time t and suppose dynamics in the patch in the absence of dispersal are given by

$$x_i(t) = g_i(x_i(t-1), \ldots, x_i(t-a), \epsilon_i^{(1)}(t), \ldots, \epsilon_i^{(1)}(t-c), \epsilon_i^{(2)}(t), \ldots, \epsilon_i^{(2)}(t-c), \ldots, \epsilon_i^{(b)}(t), \ldots, \epsilon_i^{(b)}(t-c)), \quad (1)$$

where, for $j = 1, 2, \ldots, b$, $\epsilon^{(j)} = (\epsilon_1^{(j)}, \ldots, \epsilon_P^{(j)})$ is a second-order-stationary stochastic process with absolutely summable autocovariance function (Reinsel, 1997, section 1.1) and with $\mathrm{E}(\epsilon^{(j)}) = 0_P$. The $\epsilon^{(j)}$ represent environmental variables. Suppose x_i^* is a stable equilibrium of the "deterministic skeleton" of g_i, so that

$$x_i^* = g_i(x_i^*, \ldots, x_i^*, 0, \ldots, 0). \quad (2)$$

Here x_i^* represents the carrying capacity of patch i.

To include dispersal, we assume

$$D \begin{pmatrix} x_1^* \\ \vdots \\ x_P^* \end{pmatrix} = \begin{pmatrix} x_1^* \\ \vdots \\ x_P^* \end{pmatrix} \quad (3)$$

and we consider the model

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_P(t) \end{pmatrix} = D \begin{pmatrix} g_1(x_1(t-1), \ldots, x_1(t-a), \epsilon_1^{(1)}(t), \ldots, \epsilon_1^{(1)}(t-c), \ldots, \epsilon_1^{(b)}(t), \ldots, \epsilon_1^{(b)}(t-c)) \\ \vdots \\ g_P(x_P(t-1), \ldots, x_P(t-a), \epsilon_P^{(1)}(t), \ldots, \epsilon_P^{(1)}(t-c), \ldots, \epsilon_P^{(b)}(t), \ldots, \epsilon_P^{(b)}(t-c)) \end{pmatrix}. \quad (4)$$

The deterministic skeleton of this model has equilibrium $(x_1^*, \ldots, x_P^*)^t$. Linearizing at this equilibrium gives

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_P(t) \end{pmatrix} \approx D \left[\begin{pmatrix} \sum_{l=1}^{a} m_{1l} w_1(t-l) + \sum_{j=1}^{b} \sum_{l=0}^{c} q_{1l}^{(j)} \epsilon_1^{(j)}(t-l) \\ \vdots \\ \sum_{l=1}^{a} m_{Pl} w_P(t-l) + \sum_{j=1}^{b} \sum_{l=0}^{c} q_{Pl}^{(j)} \epsilon_P^{(j)}(t-l) \end{pmatrix} + \begin{pmatrix} x_1^* \\ \vdots \\ x_P^* \end{pmatrix} \right], \quad (5)$$

where $w_i(t) = x_i(t) - x_i^*$, $m_{il} = \frac{\partial g_i}{\partial x_i(t-l)}(x_i^*, \ldots, x_i^*, 0, \ldots, 0)$, and $q_{il}^{(j)} = \frac{\partial g_i}{\partial \epsilon_i^{(j)}(t-l)}(x_i^*, \ldots, x_i^*, 0, \ldots, 0)$. By assumption (3), this is equivalent to the presentation of the model in the main text. Assumption (3) decreases generality, but not unduly for the demonstration purposes to which we put the model in the study. For simplicity, for the demonstrations in the main text, we assume the $\epsilon^{(j)}$ are independent through time and of each other, and that $\epsilon^{(j)}(t)$ is Gaussian with covariance matrix Ω_j.

S2 Model simulations

For each simulation, the model was run for 150 time steps, with a 100-year burn-in period (only data from the last 50 time steps were analyzed). Model parameters for the runs analyzed in the main text that implemented mechanisms A-D operating in isolation are given in Table 1. The first two environmental variables ($j = 1, 2$) were considered "regional" or synchronizing environmental variables, in the sense that Ω_j for these variables had nonzero entries off the diagonal. The third environmental variable ($j = 3$) represented local noise, i.e., its off-diagonal entries were zero. We incorporated local variability in density dependence, sensitivities to environmental drivers, and dispersal by adding a small amount of normally distributed random noise (mean 0, standard deviation 0.01) to autoregressive parameters (m_{il}), driver sensitivity parameters ($q_{il}^{(j)}$), and entries of the dispersal matrix (D). Environmental sensitivity parameters were further constrained to be non-negative, and in the mechanism C simulations were constrained to sum, in each location, to 1 across drivers 1 and 2.

To develop a strong test of matrix regression methods for discriminating mechanisms of geography of synchrony, we simulated a fifth scenario involving: the combination of two Moran driver variables, $\epsilon^{(1)}(t)$ and $\epsilon^{(2)}(t)$, having different spatial structures, with populations equally sensitive to each driver; as well as spatially structured dispersal. We let $S_1^{(1)} = \{1, ..., 8\}$, $S_2^{(1)} = \{9, ..., 16\}$, $S_1^{(2)} = \{1, ...4, 9, ..., 12\}$, $S_2^{(2)} = \{5, ..., 8, 13, ...16\}$, $S_1^{(d)} = \{1, 2, 5, 6, 9, 10, 13, 14\}$, and $S_2^{(d)} = \{3, 4, 7, 8, 11, 12, 15, 16\}$. We let $\epsilon^{(j)}(t)$ have Ω_j a block matrix with entries 0.6 within $S_1^{(j)}$ or $S_2^{(j)}$ and 0.3 between these groups, for $j = 1, 2$. Individuals dispersed such that $\approx 40\%$ of each population dispersed approximately evenly to other populations within $S_1^{(d)}$ or $S_2^{(d)}$, with no dispersal between these groups. The autoregressive parameters $m_{i1} \approx 0.375$ and $m_{i2} \approx -0.368$ were used for all locations, i. The moving average lag $c = 1$ and the sensitivity parameters $q_{i1}^{(1)} \approx 1$, $q_{i1}^{(2)} \approx 1$, and $q_{i1}^{(3)} \approx 0.1$ were used for all i. Approximate equalities ($\approx$) or use of the word "approximately" here indicate parameters for which a small amount of normally distributed random noise (mean 0, standard deviation 0.01) was introduced, as in the mechanism-specific simulations.

S3 Similarity of density dependence matrix

Consider a two-habitat-patch model,

$$w_i(t) = m_{i1}w_i(t-1) + \cdots + m_{ia}w_i(t-a) + \epsilon_i(t) \tag{6}$$

for $i, j = 1, 2$. Here, $w_i(t)$ is the deviation of a population index, $x_i(t)$, in habitat patch i at time t, from an equilibrium value x_i^*. The parameters $m_{i1}, \ldots, m_{ia}$ define intrinsic population dynamics in patch i, and $\epsilon_i(t)$ represents an environmental variable in patch i at time t. We assume (ϵ_1, ϵ_2) is Gaussian white noise with the ϵ_i standard normal. A necessary and sufficient condition [Brillinger, 2001, Reinsel, 1997, Shumway and Stoffer, 2000] for equation (6) to define a stationary stochastic process is that the complex roots of $1 - m_{i1}z - \cdots m_{ia}z^a$ have $|z| > 1$ for $i = 1, 2$, which we assume. We here derive $\mathrm{cor}(w_1, w_2)/\mathrm{cor}(\epsilon_1, \epsilon_2)$ as a function of the $m_{i1}, \ldots, m_{ia}$.

We will first derive the spectral matrix for the model, equation (6). The spectral matrix, denoted f_{ww}, is a 2×2 matrix

$$f_{ww} = \begin{bmatrix} f_{w_1w_1} & f_{w_1w_2} \\ f_{w_2w_1} & f_{w_2w_2} \end{bmatrix}, \tag{7}$$

where $f_{w_iw_i}$ is the power spectrum of population fluctuations in habitat patch i (for $i = 1, 2$), and $f_{w_iw_j}$ is the cross spectrum of population fluctuations in habitat patches i and j ($i = 1, 2$, $i \neq j$). $\mathrm{Re}(f_{w_iw_j})$ is the cospectrum. We use similar notation, $f_{\epsilon\epsilon}$, for the spectral matrix of the noise process.

Defining $w(t) = (w_1(t), w_2(t))^t$ and $\epsilon(t) = (\epsilon_1(t), \epsilon_2(t))^t$, the model (equation (6)) is a vector autoregressive moving average (VARMA) model,

$$\mathcal{M}w = \epsilon \tag{8}$$

for the linear filter

$$\mathcal{M} = I - M_1B - M_2B^2 - \cdots - M_aB^a, \tag{9}$$

where

$$M_l = \begin{bmatrix} m_{1l} & 0 \\ 0 & m_{2l} \end{bmatrix}, \tag{10}$$

and B is the back shift operator, $B^j\delta_t = \delta_{t-j}$, for a stochastic process δ_t. See Brillinger [2001], Reinsel [1997], and Shumway and Stoffer [2000] for background information on VARMA models and linear filters.

Defining $\mu = \exp(-2\pi i\nu)$, where ν is frequency in units of cycles per time step, and computing the spectral matrices of both sides of equation (8), we get $T(\mathcal{M})f_{ww}T(\mathcal{M})^* = f_{\epsilon\epsilon}$, where

$$T(\mathcal{M}) = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{11}$$

for $\lambda_i = 1 - m_{i1}\mu - \cdots - m_{ia}\mu^a$ the transfer function matrix of the linear filter $\mathcal{M}$. Therefore

$$f_{ww} = T(\mathcal{M})^{-1} f_{\epsilon\epsilon} (T(\mathcal{M})^*)^{-1} \tag{12}$$

$$= \begin{bmatrix} \frac{f_{\epsilon_1\epsilon_1}}{|\lambda_1|^2} & \frac{f_{\epsilon_1\epsilon_2}}{\lambda_1\overline{\lambda_2}} \\ \frac{f_{\epsilon_2\epsilon_1}}{\lambda_2\overline{\lambda_1}} & \frac{f_{\epsilon_2\epsilon_2}}{|\lambda_2|^2} \end{bmatrix}. \tag{13}$$

The desired correlation is

$$\mathrm{cor}(w_1, w_2) = \frac{\mathrm{Re}\int_{-1/2}^{1/2} f_{w_1w_2} d\nu}{\sqrt{\int_{-1/2}^{1/2} f_{w_1w_1} d\nu \int_{-1/2}^{1/2} f_{w_2w_2} d\nu}}, \tag{14}$$

$$= \frac{\mathrm{Re}\int_{-1/2}^{1/2} \frac{f_{\epsilon_1\epsilon_2}}{\lambda_1\overline{\lambda_2}} d\nu}{\sqrt{\int_{-1/2}^{1/2} \frac{f_{\epsilon_1\epsilon_1}}{|\lambda_1|^2} d\nu \int_{-1/2}^{1/2} \frac{f_{\epsilon_2\epsilon_2}}{|\lambda_2|^2} d\nu}}. \tag{15}$$

$$\tag{16}$$

But since ϵ is white noise, $f_{\epsilon\epsilon}$ is a constant, real function of ν, so this is

$$\mathrm{cor}(w_1, w_2) = \frac{f_{\epsilon_1\epsilon_2}}{\sqrt{f_{\epsilon_1\epsilon_1} f_{\epsilon_2\epsilon_2}}} \frac{\mathrm{Re}\int_{-1/2}^{1/2} \frac{1}{\lambda_1\overline{\lambda_2}} d\nu}{\sqrt{\int_{-1/2}^{1/2} \frac{1}{|\lambda_1|^2} d\nu \int_{-1/2}^{1/2} \frac{1}{|\lambda_2|^2} d\nu}} \tag{17}$$

$$= \mathrm{cor}(\epsilon_1, \epsilon_2) \frac{\mathrm{Re}\int_{-1/2}^{1/2} \frac{1}{\lambda_1\overline{\lambda_2}} d\nu}{\sqrt{\int_{-1/2}^{1/2} \frac{1}{|\lambda_1|^2} d\nu \int_{-1/2}^{1/2} \frac{1}{|\lambda_2|^2} d\nu}}. \tag{18}$$

Therefore the desired ratio, $\mathrm{cor}(w_1, w_2)/\mathrm{cor}(\epsilon_1, \epsilon_2)$, is the second factor on the right in equation (18), and this can be computed quickly and easily for any model coefficients $m_{i1}, \ldots, m_{ia}$. Various generalizations of this result are straightforward, but the development in the main text only uses what is derived here.

S4 A means of model selection for matrix regression

We begin with notational setup. Let R be a response-variable matrix which is a synchrony (e.g., correlation) matrix for populations in different locations. Let $Q_1, \ldots, Q_d$ be predictor matrices which are similarity or dissimilarity matrices for the same locations, and which therefore have the same dimensions as R. This is the standard set up for matrix regression with a synchrony matrix response variable. Let $M_1, \ldots, M_s$ be several alternative models, each of which constitutes a choice of which of the predictors, $Q_1, \ldots, Q_n$ are used. Each model represents a hypothesis about which mechanisms are important for driving synchrony. We want to determine the relative support the data provide to the alternative models/hypotheses. Given a subset $\mathcal{S} = \{i_1, \ldots, i_n\}$ of the index set of locations $\{1, \ldots, P\}$, we denote

by $c(\mathcal{S})$ the complement of $\mathcal{S}$ in $\{1,\ldots,P\}$, and we denote by $R[\mathcal{S}]$ the matrix obtained by deleting the rows and columns of R with an index in $c(\mathcal{S})$.

Models are ranked according to their out-of-sample predictive power, a standard approach. A leave-n-out approach is applied to locations in the set $\{1,\ldots,P\}$:

1. For a random choice, $\mathcal{S} = \{i_1,\ldots,i_n\}$ of n locations to leave out:
   (a) Estimate the coefficients of each of the models $\mathcal{M}_1,\ldots,\mathcal{M}_s$ using the reduced matrices $R[c(\mathcal{S})]$ and $Q_j[c(\mathcal{S})]$.
   (b) Use these estimated coefficients to generate predicted values for $R[\mathcal{S}]$ from the matrices $Q_j[\mathcal{S}]$; call the predicted matrix $\hat{R}_{\mathcal{S}}$.
   (c) Compute sum squared errors between entries of $R[\mathcal{S}]$ and entries of $\hat{R}_{\mathcal{S}}$.
2. Repeat 1 above for all possible index sets $\mathcal{S}$ of size n, aggregating results from all index sets to compute a mean squared error for out-of-sample predictions.

In the event that, for a particular model and a particular index set $\mathcal{S}$, coefficients of the fitted model were not determined, prediction errors could not be generated and that index set was not used for that model.

Ranks obtained from the above procedure are considered ordinal, not cardinal: they indicate which models are best supported, but they do not indicate, for instance, whether the top model was overwhelmingly supported over the others, or was only marginally better. Model weights are needed for the latter purpose. Models are weighted through the following procedure based on resampling locations with replacement (bootstrapping). The procedure is essentially what is recommended by Burnham and Anderson [2002] (their section 2.13.2):

1. For a random resampling (with replacement) of P locations from the location index set $1,\ldots,P$:
   (a) Construct matrices R_r and $Q_{j,r}$ for the resampled location set from the matrices R and Q_j. These are the same dimensions as R and Q_j, but contain measures of synchrony (R) or other similarity/dissimilarity measures (Q_j) between the resampled locations.
   (b) Eliminate "self comparisons" from R_r and $Q_{j,r}$, replacing them with NA. "Self-comparisons" occur whenever the same location is selected twice in the resampling. For instance, in a synchrony matrix using correlations, self-comparisons produce off-diagonal entries of 1. In a geographic distance matrix, self-comparisons produce off-diagonal entries of 0.
   (c) Perform model ranking as described above using the matrices R_r and $Q_{j,r}$, and tabulate which model was ranked on top.
2. Repeat step 1, 100 or more times. The weight for a particular model was taken to be the fraction of resamplings for which it was ranked on top.

This resampling procedure is essentially the same as that proposed by Bjørnstad, O.N. and Falck, W. [2001] in the context of their now widely used spline correlogram methods. Because the statistical procedure (model ranking) applied to each resampling above is somewhat complex, it should be studied how the resampling procedure performs numerically. This is a large task, but we have begun the task and results are promising so far (Table 1 results). Code, in the R programming language, for these leave-n-out and resampling procedures is included in Appendix S10.

Lillegård et al. [2005a] analyzed the coverage probability of the confidence envelopes produced by the resampling method of Bjørnstad, O.N. and Falck, W. [2001] in the context of estimating correlation-versus-distance relationships for spatiotemporal data, and on that basis criticized the method; however, they had made an implementation error, and when they later corrected it [Lillegård et al., 2005b], coverage rates were more reasonable. Lillegård et al. [2005a] compared algorithms for estimating correlation-versus-distance relationships for spatiotemporal population (or other) data such as we have considered, i.e., time series $x_i(t)$ for times $t = 1,\ldots,T$ and locations $i = 1,\ldots,P$. In this context, coverage is the probability, as a function of distance, that the confidence envelopes provided by an algorithm contain the true correlation-versus-distance relationship. For a reasonable but particular generating model, Lillegård

et al. [2005a] originally found that the coverage of the algorithm of Bjørnstad, O.N. and Falck, W. [2001] was poor, as low as 59%. However, upon fixing their implementation, coverage was revealed to always be at least 70% [Lillegård et al., 2005b].

Lillegård et al. [2005a] provided alternative resampling schemes which had better coverage rates when estimating correlation-versus-distance relationships when data were generated from the particular model they used; but these schemes are not appropriate for our applications. The nonparametric bootstrapping approaches advocated by Lillegård et al. [2005a] are based on resampling of time points. As Lillegård et al. [2005a] recognized, straightforward resampling of time points (their Method III) is inappropriate when time series have non-negligible temporal autocorrelation, as population time series often do. For cases with substantial temporal autocorrelation, Lillegård et al. [2005a] also suggest (their Method IV) a method based on resampling of residuals after a procedure sometimes called "prewhitening," but the method as they describe it relies on estimating parameters of a dynamical model, for which Lillegård et al. [2005a] use their generating model. We are interested primarily in wholly nonparametric methods that can be applied in cases for which dynamical models are unavailable. Furthermore, for some of our applications, the response-variable matrix is not a correlation matrix, but is instead a matrix of synchrony values that explicitly consider the timescale-specific structure of synchrony (based on wavelet or other methods). Resampling time points in straightforward ways tends to destroy or alter the timescale-specific structure in which we are interested.

S5 Details on mapping synchrony for AVHRR data

Geographical patterns of synchrony in annual peak terrestrial vegetation greenness were examined for the conterminous United States using Normalized Difference Vegetation Index (NDVI) data from the Advanced Very High Resolution Radiometer (AVHRR) [Eidenshink, 1992, 2006]. These data have 1x1 km pixel resolution and are released at weekly intervals as biweekly maximum value composites, running from 1989-present. Prior to analysis, water pixels were removed using the mask described by Carroll et al. [2009]. For each pixel, we created an annual time series of growing season maximum NDVI, a vegetation greenness metric related to primary productivity [Running et al., 2000]. Spatial synchrony of each grid cell with its neighbors was computed as the mean Pearson correlation coefficient relating the time series of maximum NDVI at the focal grid cell with all other non-water grid cells within a 10 km radius. This procedure was applied to two non-overlapping 10-year time intervals, 1989-1998 and 2006-2015. Change in synchrony and its geography were highlighted by differencing results from the two periods.

S6 Details on producing synchrony networks for gypsy moth data

Gypsy moth outbreak synchrony networks were constructed from 40 years (1975-2014) of annual defoliation records. Defoliation is mapped by aerial survey. We used as a proxy for population size the annual area defoliated by gypsy moth in 64x64 km grid cells spanning the gypsy moth's invaded range in the United States, circa 1975. This area spans New England, New York, New Jersey, and eastern Pennsylvania. These data are also analyzed in Haynes et al. [2013, 2009]. Prior to analysis, grid cells experiencing detectable defoliation in fewer than 5 years were removed; out of 69 grid cells, 63 were retained for analysis. Data were 5th root transformed to normalize distributions, and were then centered and scaled to unit variance.

Synchrony between location pairs was assessed using the real part of the cross-wavelet transform, power-normalized to give values -1 to 1. This quantity takes the form

$$S(\sigma) = \frac{\frac{1}{T}\sum_{t=1}^{T} \mathrm{Re}(W_1(\sigma,t)W_2^*(\sigma,t))}{\sqrt{(\frac{1}{T}\sum_{t=1}^{T} W_1(\sigma,t)W_1^*(\sigma,t))(\frac{1}{T}\sum_{t=1}^{T} W_2(\sigma,t)W_2^*(\sigma,t))}}. \quad (19)$$

Here, t is time, σ is timescale, W_1 and W_2 are wavelet transforms, * indicates complex conjugation, and Re denotes taking the real part. This makes a sensible synchrony metric because it takes large values when oscillations have correlated magnitudes through time and are in-phase. It has the further advantage of being timescale specific. We used the complex Morlet wavelet transform.

Statistical significance of synchrony between location pairs was assessed by comparison with synchrony of a distribution of surrogate datasets produced using amplitude adjusted Fourier transform (AAFT) surrogate methods [Schreiber and Schmitz, 2000]. These surrogate datasets provide a strong null hypothesis against which to test synchrony because the surrogates have the same power spectra but randomized phase information. AAFT methods, rather than Fourier surrogate methods [Sheppard et al., 2015], were used because AAFT produced surrogates more similar in their marginal distributions to our data. To test for statistical significance within timescale ranges of interest, we ranked the empirical synchrony (which was a function of timescale - equation (19)) against the synchrony of surrogates at each timescale and averaged ranks within each specified range of timescales, following Sheppard et al. [2015]. Statistical significance was based on how this value compared to the distribution of analogous values for the surrogates [Sheppard et al., 2015]. Excepting the use of AAFT in place of Fourier surrogate methods, our significance testing followed the procedure of Sheppard et al. [2015].

Synchrony networks were constructed from the adjacency matrix consisting of pairwise statistical significances of synchrony. Hence, we produced and analyzed unweighted, undirected networks. This choice was for simplicity. The 'biwavelet' R package was used.

S7 A network null model

Empirical degree distributions and statistical significance of graph metrics were evaluated by comparison with a null model for spatial synchrony networks. Systematic differences from this null model indicate deviation from isotropic distance decay in synchrony. Our null model consisted of random graphs which preserved the number of nodes and randomly created edges between nodes with distance-dependent probabilities conditioned on the empirical data. The dependence of edge probabilities on distance was modeled using logistic regression, but other choices for the functional relationship between distance and edge probability may be more appropriate in other applications. We also tested a negative exponential function with the gypsy moth synchrony network and found overwhelmingly similar results as with logistic regression; the pattern of distance-decay was similar between logistic regression and the negative exponential function, and results of statistical tests were the same.

An important advantage of a spatially explicit null model such as this is that it produces a statistical test that minimizes one type of bias associated with network structure. As a simple consequence of the geography of sampling locations, locations near the geographical center of a network tend to have more close neighbors, and thus a larger number of probable network links, than locations on the periphery. Null models which preserve node locations and which are based on spatial relationships such as distance will also share this property and therefore yield significance tests accounting for this bias. This is especially important for testing single-node metrics, such as our application of degree centrality to the gypsy moth synchrony network (main text). The 'igraph' R package was used.

S8 Details on matrix regression analyses for gypsy moth data

Matrix regression with leave-n-out cross-validation and resampling (Supplementary Material section 4) was used to examine mechanisms underlying geography of gypsy moth defoliation synchrony at long timescales (4-12 year periods). The response matrix was pairwise synchrony (the power-normalized real part of the cross-wavelet transform, averaged over 4-12 year periods; Appendix S6); note that we used values of this statistic, not statistical significance thereof. Our analysis of causes of geography of synchrony was restricted to grid cells containing at least one weather station. Of the 63 grid cells used to construct the synchrony network, 47 were retained for this analysis.

Predictor matrices included: spatial proximity; similarity in density dependence (Appendix S3); similarity in forest type; and two different matrices of weather synchrony. Spatial proximity was the Euclidean distance between locations, transformed according to $1 - (x/\max(x))$. Similarity of density dependence was determined by fitting an AR(2) model to the defoliation time series of each location and using the fitted AR(2) coefficients to determine the expected correlation between locations as described in Appendix S3. Similarity in forest type was computed following Haynes et al. [2013]. Forest type maps by Ruefenacht et al. [2008] were used to compute the proportion of each grid cell comprised of the 7 most common forest types. Using these proportions, differences between grid cell forest

composition were computed using Mahalanobis distances. Differences were transformed to similarities according to $1 - (x/\max(x))$.

Matrices of weather synchrony were developed from a principal components decomposition of the following weather variables on monthly timescales: mean daily minimum temperature, mean daily maximum temperature, and total precipitation. Matrices of weather synchrony were developed from a principal components decomposition of the following weather variables: January through December minimum temperature (12 variables), January through December maximum temperature (12 variables), and January through December total precipitation (12 variables). These variables were generated from daily weather station data by averaging. Raw weather data were extracted from weather stations in the US Historical Climate Network [Easterling et al., 1996]. The first principal component (PC1) primarily represented variation in temperature and the second principal component (PC2) primarily represented variation in precipitation [Haynes et al., 2013]. Synchrony in time series of PC1 and PC2 was then calculated using the power-normalized real part of the cross wavelet transform.

In our model selection procedure we considered all possible combinations of predictor variables. In leave-n-out model selection, n was taken to be 4 sites, and model weights were based on 250 resamplings of the data (Appendix S4). Variable importance weights were assigned by summing the weights of all models containing each predictor. See also the R code used, Appendix S10.

S9 Network module structure in climate PC2

We examined the degree to which module structure in the network of climate PC2 synchrony corresponds with module structure in the long-timescale gypsy moth synchrony network. Although synchrony in climate PC2 was a highly weighted and statistically significant predictor of gypsy moth synchrony, the modules of these networks did not obviously correspond. Despite some visual similarities (Figure 1), a χ^2 test showed the two classifications to be independent of one another (df = 1, $p = 0.975$). The lack of correspondence is consistent with the fact that matrix models of gypsy moth synchrony matrices, though highly significant, did not explain a large fraction of the geographic variation in synchrony.

S10 Code for matrix model selection

See separate file AppendixS10.R.

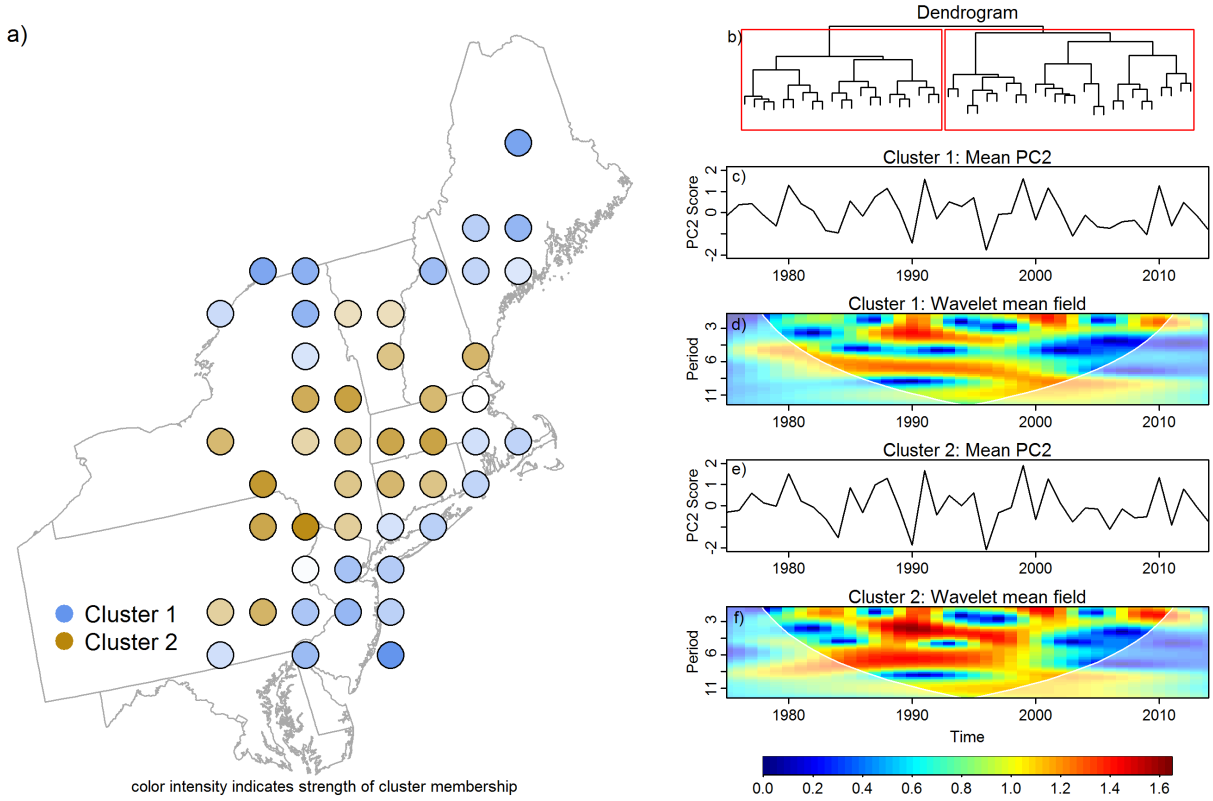

Figure 1: Module structure in the climate PC2 synchrony network. Panel a) cluster identity (color) and membership strength (shading); b) wavelet clustering dendrogram; c) mean time series for cluster 1; d) wavelet mean field magnitude for cluster 1; e) mean time series for cluster 2; f) wavelet mean field magnitude for cluster 2

Mechanism	Groups	AR parameters	Driver sensitivity	Driver covariance	Dispersal
A) Driver pattern	$S_1 = \{1, \ldots, 8\}$ $S_2 = \{9, \ldots, 16\}$	$a = 2$ $m_{i1} \approx 0.375$ $m_{i2} \approx -0.368$	$c = 1$ $q_{i1}^{(1)} \approx 1$ $q_{i1}^{(2)} \approx 0$ $q_{i1}^{(3)} \approx 0.1$	Ω_1 and Ω_2 are block matrices, entries 0.6 within S_1 or S_2, 0.3 between these groups	$\approx 5\%$ of each pop. disperses approx. evenly among other populations
B) Different density dependence	$S_1 =$ $\{1, \ldots, 4, 9, \ldots, 12\}$ $S_2 =$ $\{5, \ldots, 8, 13, \ldots, 16\}$	$a = 2$ S_1: $m_{i1} \approx 0.375$ $m_{i2} \approx -0.368$ S_2: $m_{i1} \approx 1.141$ $m_{i2} \approx -0.368$	$c = 1$ $q_{i1}^{(1)} \approx 1$ $q_{i1}^{(2)} \approx 0$ $q_{i1}^{(3)} \approx 0.1$	Ω_1 and Ω_2 have all off-diagonal entries equal to 0.6	$\approx 5\%$ of each pop. disperses approx. evenly among other populations
C) Sensitivity to different drivers	$S_1 =$ $\{1, \ldots, 4, 13, \ldots, 16\}$ $S_2 = \{5, \ldots, 12\}$	$a = 2$ $m_{i1} \approx 0.375$ $m_{i2} \approx -0.368$	$c = 1$ S_1: $q_{i1}^{(1)} \approx 1$ $q_{i1}^{(2)} \approx 0$ S_2: $q_{i1}^{(1)} \approx 0$ $q_{i1}^{(2)} \approx 1$ $q_{i1}^{(3)} \approx 0.1$	Ω_1 and Ω_2 have all off-diagonal entries equal to 0.6	$\approx 5\%$ of each pop. disperses approx. evenly among other populations
D) Unequal dispersal	$S_1 =$ $\{1, 2, 5, 6, 9, 10, 13, 14\}$ $S_2 =$ $\{3, 4, 7, 8, 11, 12, 15, 16\}$	$a = 2$ $m_{i1} \approx 0.375$ $m_{i2} \approx -0.368$	$c = 1$ $q_{i1}^{(1)} \approx 1$ $q_{i1}^{(2)} \approx 0$ $q_{i1}^{(3)} \approx 0.1$	Ω_1 and Ω_2 have all off-diagonal entries equal to 0.6	$\approx 40\%$ of each pop. disperses approx. evenly to other pops. within S_1 or S_2; no dispersal between these groups

Table 1: Model parameters for simulations. Ω_3 was always the identity matrix. Ω_1 and Ω_2 always had 1s along the diagonal. Approximate equalities ($\approx$) or use of the word "approximately" indicate parameters for which a small amount of normally distributed random noise (mean 0, standard deviation 0.01) was introduced.

References

Bjørnstad, O.N. and Falck, W. Nonparametric spatial covariance functions: Estimation and testing. *Environmental and Ecological Statistics*, 8:53–70, 2001.

D. Brillinger. *Time Series: Data Analysis and Theory*. SIAM, Philadelphia, expanded edition edition, 2001.

K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, 2nd edition, 2002.

M.L. Carroll, J.R. Townshend, C.M. DiMiceli, P. Noojipady, and R.A. Sohlberg. A new global raster water mask at 250 meter resolution. *Internation Journal of Digital Earth*, 2:291–308, 2009.

D.R. Easterling, T.R. Karl, E.H. Mason, P.Y. Hughes, and D.P. Bowman. United States Historical Climatology Network (U.S. HCN) monthly temperature and precipitation data. Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 1996.

J.C. Eidenshink. The 1990 conterminous u.s. avhrr data set. *Photogrammetric Engineering and Remote Sensing*, 58: 809–813, 1992.

J.C Eidenshink. A 16-year time series of 1 km avhrr satellite data of the conterminous united states and alaska. *Photogrammetric Engineering and Remote Sensing*, 72:1027–1035, 2006.

K.J Haynes, A.M. Liebhold, and D.M. Johnson. Spatial analysis of harmonic oscillation of gypsy moth outbreak intensity. *Oecologia*, 159:249–256, 2009.

K.J. Haynes, O.N. Bjornstad, J.A. Allstadt, and A.M. Liebhold. Geographical variation in the spatial synchrony of a forest-defoliating insect: isolation of environmental and spatial drivers. *Proceedings of the Royal Society B*, 280: 20122373, 2013.

M. Lillegård, S. Engen, and B.-E. Sæther. Bootstrap methods for estimating spatial synchrony of fluctuating populations. *Oikos*, 109:342–350, 2005a.

M. Lillegård, S. Engen, and B.-E. Sæther. Bootstrap methods for estimating spatial synchrony of fluctuating populations: an addendum. *Oikos*, 110:629, 2005b.

G.C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, 2nd edition, 1997.

B. Ruefenacht, M.V. Finco, M.D. Nelson, R. Czaplewski, E.H. Helmer, J.A. Blackard, G.R. Holden, A.J. Lister, D. Salajanu, D. Weyermann, and K. Winterberger. Conterminous u.s. and alaska forest type mapping using forest inventory and analysis data. *Photogrammetric Engineering and Remote Sensing*, 11:1379–1388, 2008.

Steven W Running, Peter E Thornton, Ramakrishna Nemani, and Joseph M Glassy. Global terrestrial gross and net primary productivity from the earth observing system. In *Methods in ecosystem science*, pages 44–57. Springer, 2000.

T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.

L.W Sheppard, J.R. Bell, R. Harrington, and D.C. Reuman. Changes in large-scale climate alter spatial synchrony of aphid pests. *Nature Climate Change*, 2015. doi: 10.1038/nclimate2881.

R.H. Shumway and D.S. Stoffer. *Time Series Analysis and its Applications*. Springer-Verlag, New York, 2000.