APPLICATION

Synchrony: quantifying variability in space and time

Tarik C. Gouhier[1]* and Frederic Guichard[2]

[1]*Marine Science Center, Northeastern University, 430 Nahant Road, Nahant, MA 01908, USA; and* [2]*Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montreal, QC H3A 1B1, Canada*

Summary

1. There is growing recognition that linking patterns to their underlying processes in interconnected and dynamic ecological systems requires data sampled at multiple spatial and temporal scales.

2. However, spatially explicit and temporally resolved data sets can be difficult to analyze using classical statistical methods because the data are typically autocorrelated and thus violate the assumption of independence.

3. Here, we describe the synchrony package for the R programming environment, which provides modern parametric and nonparametric methods for (i) quantifying temporal and spatial patterns of auto- and cross-correlated variability in univariate, bivariate, and multivariate data sets, and (ii) assessing their statistical significance via Monte Carlo randomizations.

4. We illustrate how the methods included in the package can be used to investigate the causes of spatial and temporal variability in ecological systems through a series of examples, and discuss the assumptions and caveats of each statistical procedure in order to provide a practical guide for their application in the real world.

Key-words: synchrony, phase-locking, time-series analysis, spatial statistics, variogram

Introduction

Empirical and theoretical research is increasingly focusing on processes operating at multiple spatial and temporal scales to understand the dynamics of complex and interconnected ecological systems (Loreau, Mouquet & Holt 2003; Menge *et al.* 2003; Borcard *et al.* 2004; Leibold *et al.* 2004; Gouhier, Guichard & Gonzalez 2010a; Gouhier, Guichard & Menge 2010b). Quantifying patterns and processes across scales is likely to yield novel insights into classical ecological questions such as the relative influence of local and regional processes on the spatiotemporal distribution of species across a range of ecosystems (Ricklefs 2008). However, the use of spatially and temporally replicated data sets in ecological studies can present practical challenges because the data typically violate the assumption of independence common to many classical statistical tests (e.g. Legendre 1993; Fortin & Dale 2005). Such non-independence (or autocorrelation) in the data must be accounted for when fitting statistical models to avoid making spurious conclusions (Hurlbert 1984). Yet, non-independence in the form of auto- and cross-correlated variability in space and time need not be the bane of our statistical existence. By embracing and quantifying correlated variability via proper statistical procedures, we can turn the bane of non-independence into a veritable boon and reveal previously hidden relationships between spatiotemporal ecological patterns and processes (Legendre 1993; Gouhier, Guichard & Gonzalez 2010a; Dray *et al.* 2012). Below, we present the

synchrony package for the R programming environment (R Development Core Team 2013) and use three examples to demonstrate how it can be used to identify novel relationships between variables in spatially- and/or temporally resolved data sets.

Description

The synchrony package includes functions to (i) generate random matrices with specified levels of auto- and cross-correlation that are useful for developing ecological theory (Vasseur & Fox 2007; Vasseur 2007; Gouhier, Guichard & Menge 2010b), (ii) identify temporally correlated variability between multiple time series via parametric and nonparametric methods (Buonaccorsi *et al.* 2001; Cazelles & Stone 2003; Gouhier & Guichard 2007), and (iii) estimate spatial, temporal, and spatiotemporal patterns of auto- and cross-correlated variability in univariate, bivariate, and multivariate data sets (Bjornstad, Ims & Lambin 1999; Bjornstad & Falck 2001; Fortin & Dale 2005; Gouhier, Guichard & Gonzalez 2010a, see Table 1). The methods included in the package have either not been implemented in R before or have been augmented with Monte Carlo randomization procedures that account for temporal autocorrelation and thus generate appropriate type I errors. Hence, synchrony extends and complements existing packages such as ncf (Bjornstad & Falck 2001), geoR (Ribeiro & Diggle 2001), and vegan (Oksanen *et al.* 2013). We now describe the functionality of the package using three examples and provide all the code used to generate the analyses in Appendix S1.

*Correspondence author. E-mail: tarik.gouhier@gmail.com

Table 1. Description of the main functions included in the `synchrony` package

Function name	Description
`community.sync`	Compute the correlation (and its statistical significance) between multiple time series within a community (Loreau & de Mazancourt 2008)
`meancorr`	Compute the mean correlation (and its statistical significance) between all pairs of time series using the Pearson product–moment correlation
`kendall.w`	Compute the concordance (and its statistical significance) between multiple variables (Zar 1999; Legendre 2005)
`phase.sync`	Determine the strength of phase-locking (and its statistical significance) between pairs of quasiperiodic time series (Cazelles & Stone 2003)
`peaks`	Determine the proportion of concurrent local extrema (and its statistical significance) between pairs of time series (Buonaccorsi *et al.* 2001)
`vario`	Compute variograms and correlograms of univariate (one observation per location) or multivariate (multiple observations per location) data sets using the Pearson product–moment correlation, the Spearman's ranked correlation, Kendall's W, Geary's C, Moran's I, the covariance, or the semivariance (Bjornstad, Ims & Lambin 1999; Fortin & Dale 2005)
`vario.fit`	Fit spherical, Gaussian, nugget, linear, exponential, sill, periodic, or hole theoretical models to empirical variograms obtained using function vario (Fortin & Dale 2005; Gouhier, Guichard & Gonzalez 2010a)
`correlated.matrix`	Generate a matrix of values with a specific mean, standard deviation, and column-wise cross-correlation (Legendre 2000)
`phase.partnered`	Generate two vectors of values with a specific mean, standard deviation, autocorrelation, and cross-correlation (Vasseur 2007)
`plot;print`	Default methods to plot and print `synchrony`, `vario` and `vario.fit` objects

Example 1: Community synchrony

Synchrony in the local abundance of species can serve as an important indicator of stability and persistence (Gouhier, Guichard & Menge 2010b). Although patterns of community synchrony alone cannot be used to identify their causal drivers (Loreau & de Mazancourt 2008; Gouhier, Guichard & Menge 2010b), they can certainly promote our understanding of the phenomenon. There are several metrics that have been proposed for measuring community synchrony: the mean correlation coefficient, Kendall's W and Loreau and de Mazancourt's ϕ. Their performance can be compared by measuring synchrony in randomly assembled communities with a specified number of species (columns), time steps (rows) and level of synchrony (Fig. 1, Appendix S1). For example, function `correlated.matrix` can be used to generate a community time series with `nspecies=10`, `ntimes=100` and the desired level of synchrony among species `rho=0.7`:

```
library(synchrony)
## synchrony 0.2.1 loaded.
comm<-correlated.matrix(rho=0.7,nspecies=10,
  ntimes=100)$community
```

The level of synchrony in the randomly assembled community `comm` can then be assessed using each metric. The most intuitive measure of community synchrony is the mean correlation coefficient, which is provided by the `meancorr` function (Houlahan *et al.* 2007):

$$\bar{\rho} = \frac{\sigma^2\left[\sum_{i=1}^{N} x_i(t)\right] - \sum_{i=1}^{N} \sigma^2[x_i(t)]}{2\sum_{i<j}^{N} \sigma[x_i(t)]\sigma[x_j(t)]} \qquad \text{eqn 1}$$

where σ^2 is the variance operator. However, the mean correlation coefficient is somewhat flawed because its range is mathematically constrained by the number of species N. Indeed, because there are only N terms contributing to the numerator of eqn 1 but $N(N-1)$ terms contributing to its denominator, the minimum correlation coefficient assuming all species have the same population variance is (Loreau & de Mazancourt 2008): $\rho_{min} = -1/(N-1)$. Increasing species richness N thus increases the minimum average correlation (Fig. 1). To compute synchrony via the mean correlation coefficient and its significance via `nrands=999` randizations, one can use function `meancorr`:

```
meancorr(data=comm,nrands=999,alternative=
  "two.tailed",type=1,quiet=TRUE)
```

```
## Mean Pearson correlation: 0.601.
## Mean correlation p-value (two-tailed test): 0.001
```

By default, the *P*-value is based on a two-tailed test and generated by a naive randomization procedure specified via argument `type=1` that destroys both the temporal autocorrelation within each species and the cross-correlation among species (Legendre 2005). Alternatively, one can specify a one-tailed test (e.g. `alternative=greater` or `alternative=less`). Additionally, one can specify argument type = 2 to select the 'caterpillar' randomization procedure, which preserves the temporal autocorrelation within species but destroys the cross-correlation among species by displacing the time series by a random amount for each randomization (Purves & Law 2002). By preserving the temporal autocorrelation within each species, the 'caterpillar' procedure generates the correct type I error regardless of the level of autocorrelation within each time series.
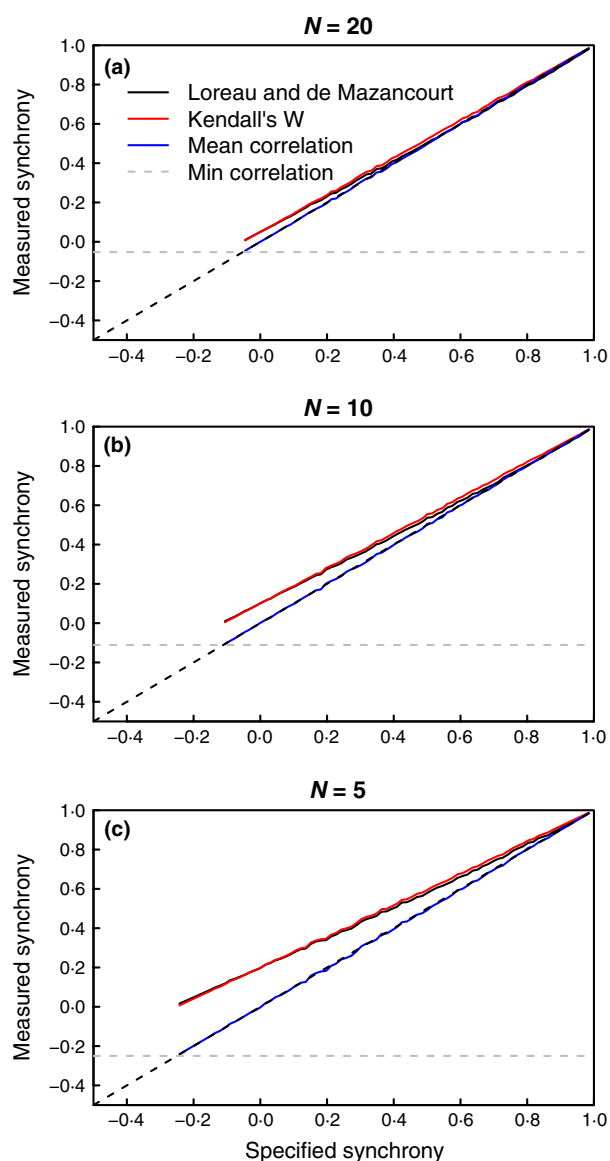
Fig. 1. Comparing three measures of community synchrony. For each desired level of synchrony (i.e., specified synchrony between −0·5 and 1), the function `correlated.matrix` was used to generate 100 random replicate N × T community matrices consisting of T = 100 time steps for (a) N = 20, (b) N = 5, and (c) N = 3 species. The three measures of synchrony were then computed and averaged across all replicates. The dashed black line represents the 1:1 line, and the horizontal dashed grey line represents the minimum average correlation for each community size (i.e., −1/(N−1)).

Two alternative metrics that overcome the limitations of the mean correlation coefficient have been proposed. The first is Kendall's W, a nonparametric statistic bound between 0 and 1 that measures the level of agreement between several ranked variables (i.e., species; Zar 1999; Buonaccorsi *et al.* 2001; Legendre 2005):

$$W = 12 \frac{\sum R_i^2 - \frac{(\sum R_i)^2}{T}}{N^2(T^3 - T) - N\sum \tau} \qquad \text{eqn 2}$$

where R_i is the ranked time series of species i, N is the number of species, T is the number of time steps, and $\sum \tau$ is a correction for ties such that (Zar 1999): $\sum_{i=1}^{j} (t_i^3 - t_i)$. Here, t_i is the number of tied ranks in each group i of j groups of ties. Kendall's W has several desirable characteristics. First, its range does not contract with increasing species richness (Fig. 1). Secondly, its statistical significance can be determined using a standard χ^2_{T-1} test (Zar 1999). Because this test has been shown to be too conservative, a Monte Carlo randomization procedure that shuffles the columns of the matrix independently and produces the correct rates of type I and type II errors in the absence of autocorrelation in the data (Legendre 2005) has also been included in the `synchrony` package. Thirdly, Kendall's W is related to the mean Spearman's ranked correlation $\bar{\rho}_s$ between all pairs of species:

$$W = \frac{1 + (N-1)\bar{\rho}_s}{N} \qquad \text{eqn 3}$$

Hence, despite the fact that Kendall's W cannot distinguish asynchrony (negatively correlated fluctuations) from the lack of synchrony (independent fluctuations) because its range falls between 0 and 1 regardless of the sign of the mean correlation (Fig. 1), the mean Spearman's ranked correlation can help distinguish those two scenarios. The last equation (eqn 3) also shows that Kendall's W converges to the mean correlation coefficient with increasing species richness (Fig. 1). However, because Kendall's W depends on species richness, one cannot directly compare levels of synchrony across communities with different numbers of species. For instance, species-poor communities undergoing independent fluctuations (i.e., specified synchrony of 0) are characterized by a higher Kendall's W than species-rich communities undergoing the same level of independent fluctuations (Fig. 1a vs. c). To compute synchrony via Kendall's W and its significance via `nrands=999` randomizations, one can use function `kendall.w`:

```
kendall.w(data=comm, nrands=999, type=1,
  quiet=TRUE)
## Kendall's W (uncorrected for ties): 0.595
## Kendall's W (corrected for ties): 0.595
## Spearman's ranked correlation: 0.55
## Kendall's W p-value (one-tailed test [greater]):
  0.001
```

By default, the *P*-value is based on a one-tailed test and generated via the same naive randomization procedure described above. Alternatively, one can specify argument `type=2` to employ the caterpillar randomization procedure. The other main metric used to measure community synchrony was introduced by Loreau & de Mazancourt (2008): $\phi = \sigma^2\left[\sum_{i=1}^{N} x_i(t)\right] / \left(\sum_{i=1}^{N} \sigma[x_i(t)]\right)^2$, where the numerator represents the community variance and the denominator represents the sum of the population variances squared. This metric also varies between 0 (lack of synchrony) and 1 (perfect synchrony). Like Kendall's W, the range of this metric does not depend on species richness, but its value does. Specifically, if all species have the same population variance (Loreau & de Mazancourt 2008):

$$\phi = \frac{1 + (N-1)\bar{\rho}_r}{N} \qquad\qquad \text{eqn 4}$$

Hence, the only difference between ϕ (eqn 4) and Kendall's W (eqn 3) is that the former depends on the mean Pearson product–moment correlation $\bar{\rho}_r$, whereas the latter depends on the mean Spearman ranked correlation $\bar{\rho}_s$. Given their strong structural similarities, it is not surprising that these two metrics behave very similarly, with Kendall's W typically converging to the specified mean correlation coefficient more readily (Fig. 1). To compute synchrony via Loreau and de Mazancourt's ϕ and its significance via `nrands=999` randomizations, one can use function `community.sync`:

```
community.sync(data=comm,nrands=999,
  alternative="greater",type=1,quiet=TRUE)
```

```
## Community synchrony: 0.641
## Mean pairwise correlation: 0.601
## Community synchrony p-value (one-tailed test
  [greater]): 0.001
```

By default, the *P*-value is based on a one-tailed test and generated via the same naive randomization procedure described above. Alternatively, one can specify argument `type=2` to employ the caterpillar randomization procedure.

Example 2: Noise and synchrony in the real world

In a noisy and interconnected world, quantifying synchrony can be challenging because multiple processes can generate complex dynamics that may either prevent the detection of synchronized fluctuations when they occur (i.e., type II error) or lead to their false detection (i.e., type I error). To illustrate this issue, one can generate two independent realizations of a second-order auto-regressive (AR(2)) process:

```
# Set random seed
set.seed(65)
t1<-arima.sim(n=500,list(ar=c(1.61,-0.77)),
  sd=0.1)+1.055
t2<-arima.sim(n=500,list(ar=c(1.61,-0.77)),
  sd=0.1)+1.055
```

We can then generate perfectly cross-correlated (`rho=1`) white noise (`gamma=0`) by using function `phase partnered`:

```
(corr<-phase.partnered(n=length(t1),rho=1,
  gamma=0))
```

```
## Cross-correlation: 1
## Autocorrelation: 0
## Standard deviation: 0.1
## Mean: 0
```

Here, argument `rho` controls the cross-correlation between the time series (varies between -1 and 1), and argument `gamma` controls the autocorrelation of each time series. Setting gamma to values between -2 and 0 will generate time series dominated by high frequencies (i.e., blue noise) whereas setting gamma to values between 0 and 2 will generate time series dominated by low frequencies (i.e., red noise). We can then add the correlated white noise to each independent AR(2) time series to determine whether synchrony metrics are able to correctly conclude that they are unrelated (i.e., have correct type I error; Fig. 2):

```
t1.corr<-t1+corr$timeseries[,1]
t2.corr<-t2+corr$timeseries[,2]
```

Conversely, one can generate two perfectly correlated time series (sinusoidal models):

```
t1<-10*sin(seq(from=0,to=20*pi,
  length.out=500))+50
t2<-10*sin(seq(from=0,to=20*pi,
  length.out=500))+50+2
```

We can then generate uncorrelated and negatively correlated noise:

```
# Set random seed
set.seed(1)
uncorr<-phase.partnered(n=500,rho=0,gamma=0,
  sigma=8,mu=0)
negcorr<-phase.partnered(n=500,rho=-1,gamma=0,
  sigma=8,mu=0)
```

Finally, we can add the uncorrelated and negatively correlated noise to the original sinusoidal models to determine whether synchrony metrics are able to accurately conclude that the time series are synchronized (i.e., have correct type II error; Fig. 3):

```
t1.uncorr<-t1+uncorr$timeseries[,1]
t2.uncorr<-t2+uncorr$timeseries[,2]
t1.negcorr<-t1+negcorr$timeseries[,1]
t2.negcorr<-t2+negcorr$timeseries[,2]
```

We then use the mean correlation coefficient, Kendall's W, and Loreau & de Mazancourt (2008)'s ϕ to measure spatial synchrony between these two time series. These metrics erroneously suggest that two independent (unrelated) realizations of an AR(2) process are synchronized in the presence or absence of correlated noise (Table 2). Furthermore, these metrics also fail to detect synchrony between the sinusoidal models in the presence of negatively correlated noise (Table 2). Noise can thus significantly hamper our ability to detect synchrony in the real world by both masking truly synchronized dynamics (false negatives) and making unrelated dynamics appear synchronized (false positives).

To contend with the disruptive effect of noise, it can be useful to turn to nonlinear measures of synchrony. These metrics quantify the relationship between the phases of pairs of data sets (Blasius, Huppert & Stone 1999; Cazelles & Stone 2003). Hence, they are ideally suited for linking noisy time series whose amplitudes are uncorrelated or imperfectly correlated, but whose phase relationship remains relatively constant or locked over time (Cazelles & Stone 2003). The simplest such metric is what we have termed concurrency
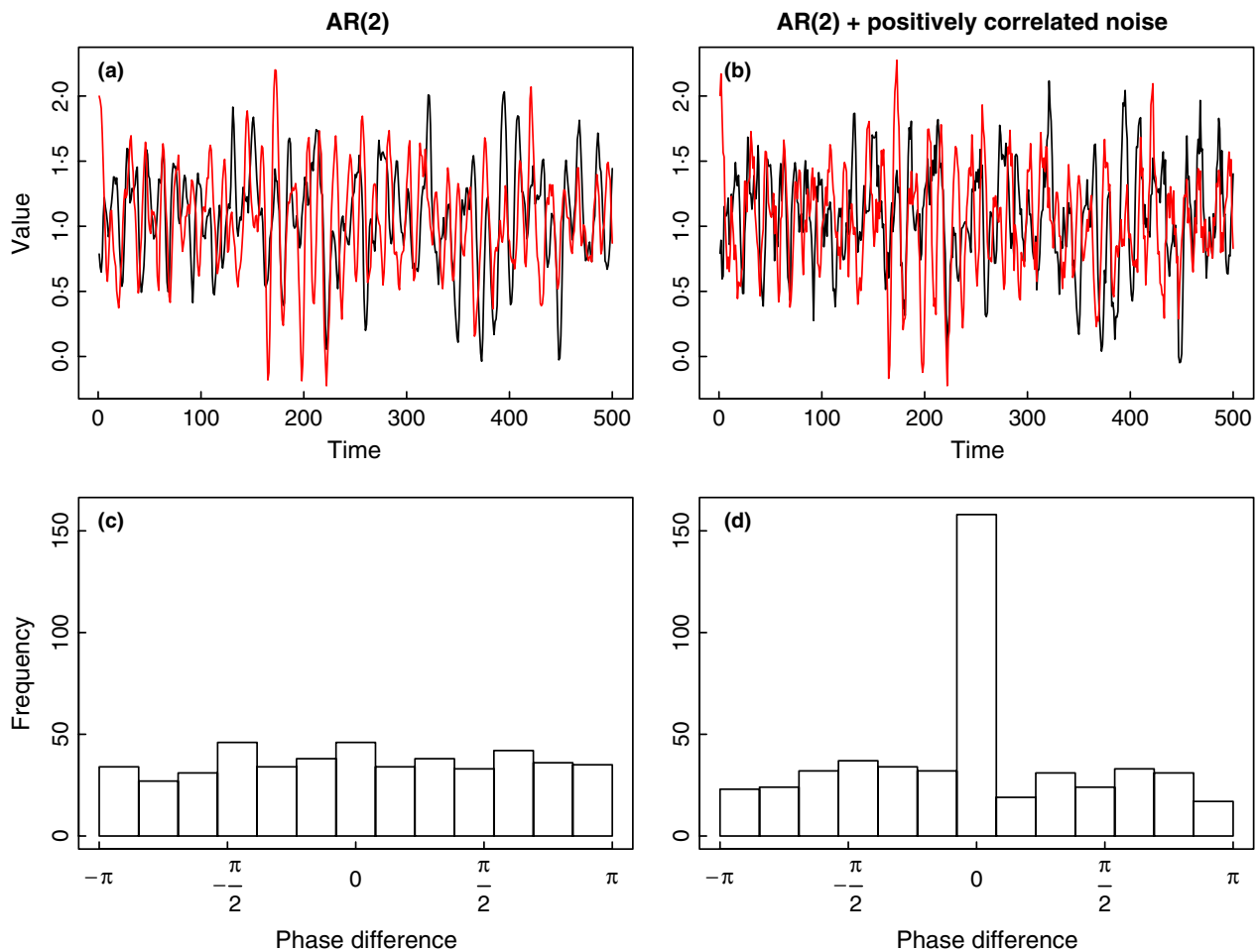
Fig. 2. The effect of stochastic noise on detecting phase synchrony between two unrelated time series. The two time series depicted in (a) represent independent realizations of an AR(2) process with first- and second-order coefficients 1·61 and −0·77, mean 1·055, and standard deviation 0·1. The function `phase.partnered` was used to add (b) positively correlated white noise (cross-correlation of 1) to the AR(2) time series. (c, d) Phase analysis of the AR(2) time series without (c) vs. with (d) correlated noise.

(implemented in function `peaks`), which simply measures the proportion of concurrent peaks (local maxima) and troughs (local minima) between pairs of time series (Buonaccorsi *et al.* 2001). This metric varies between 0 when the time series never peak and trough together, and 1 when the time series always peak and trough simultaneously. The statistical significance of concurrency can be determined via a Monte Carlo randomization procedure that either (i) shuffles the time series independently (`type=1`) or (ii) shuffles the order of the time series and thus maintains the level of autocorrelation in each time series while destroying their cross-correlation (`type=2`).

The second main nonlinear metric, implemented in function `phase.sync`, measures phase synchrony between quasiperiodic times series (Cazelles & Stone 2003). This is done by (i) finding the local extrema (minima or maxima) of each time series, (ii) assigning them successive phase values (0 for first extremum, 2π for second extremum, 4π for the third, etc.), and (iii) assigning points that fall between the extrema phase values via linear interpolation (Cazelles & Stone 2003). This procedure thus converts a time series of amplitudes into a time series of phases that is more robust to noise. The relationship between

the time series of phases can be determined by computing and then plotting the frequency distribution of their instantaneous difference at each time step (Fig. 2). Time series that are phase synchronized or locked exhibit a modal distribution with a prominent peak at a given phase difference (Fig. 2d), whereas unrelated times series are characterized by a uniform or diffuse distribution (Fig. 2c). The strength of phase synchrony can be quantified by a Q index that falls between 0 (no phase synchrony) and 1 (full phase synchrony) and is computed as follows (Cazelles & Stone 2003):

$$Q = \frac{S_{\max} - S}{S_{\max}} \qquad \text{eqn 5}$$

where $S = -\sum_{k=1}^{N_h} p_k \ln(p_k)$ is the Shannon entropy of the frequency histogram of phase differences, with N_h being the number of phases in the frequency histogram and p_k being the proportion of points in bin k of the frequency histogram. Here, $S_{\max} = \ln(N_h)$ is the maximum entropy possible (i.e., uniform frequency distribution). The statistical significance of the Q values can be determined via Monte Carlo randomizations that shuffle the time series while maintaining their temporal (autocorrelation) structure (Cazelles & Stone
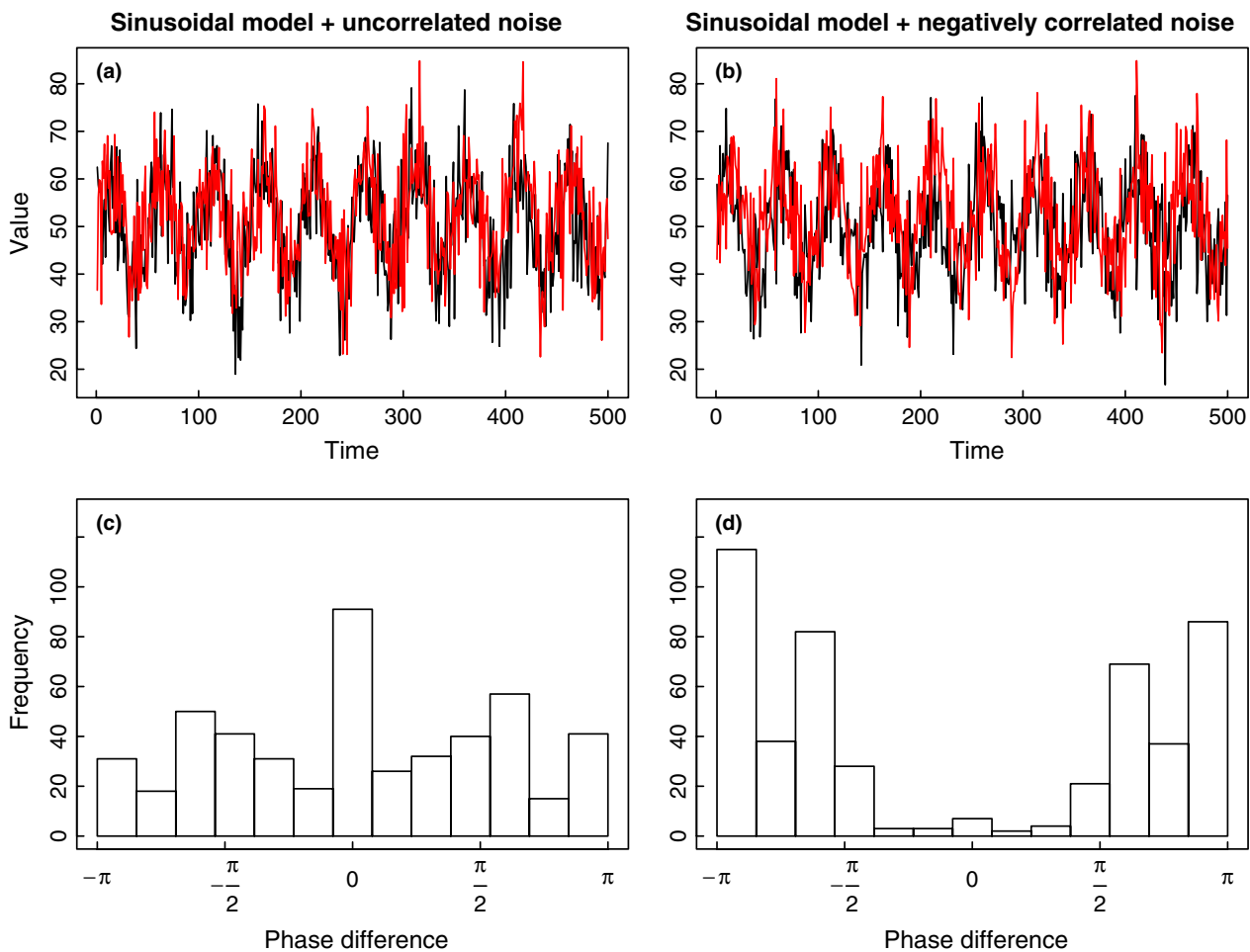
Sinusoidal model + uncorrelated noise **Sinusoidal model + negatively correlated noise**

Fig. 3. The effect of stochastic noise on detecting phase synchrony between two time series with perfectly correlated underlying skeletons. The skeleton time series were generated by two in-phase sinusoidal waves with the same frequency (10) and slightly different means (50 and 52, respectively). The `phase.partnered` function was then used to add (a) uncorrelated noise (cross-correlation of 0) or (b) negatively correlated noise (cross-correlation of −1). (c, d) Phase analysis of the sinusoidal model time series with (c) uncorrelated or (d) negatively correlated noise.

2003). We can put these nonlinear synchrony metrics to the test by analyzing the previously-generated time series:

```
# Compute Phase synchrony/locking
phase.t1t2.negcorr<-phase.sync(t1.negcorr,
  t2.negcorr,nrands=999,quiet=TRUE)
# Compute Concurrency
peaks.t1t2.negcorr<-peaks(t1.negcorr,t2.negcorr,
  nrands=999,type=1,quiet=TRUE)
```

Here, both nonlinear metrics are able to correctly identify the independent second-order autoregressive (AR(2)) realizations as uncorrelated (Table 2, Fig. 2). However, when the AR(2) processes are overlain with correlated noise, both metrics suggest that the time series are synchronized (Table 2, Fig. 2). Furthermore, concurrency is unable to detect the correlated sinusoidal models in the presence of either uncorrelated or negatively correlated noise (Table 2). Phase synchrony detects that the time series are locked in phase in the presence of the uncorrelated noise, but incorrectly suggests that the time series are locked in anti-phase (phase difference of π) in the presence of negatively correlated noise even though their underlying sinusoidal models are positively correlated (Table

2, Fig. 3). Overall, this shows that there is no 'magic bullet' method for quantifying synchrony in noisy observational data, especially if multiple factors are operating simultaneously.

Example 3: Spatial synchrony

Correlated fluctuations in species abundance can also occur between spatially isolated populations. Such spatial synchrony can be caused by endogenous factors such as dispersal between populations and trophic interactions with species whose dynamics are spatially synchronized, or exogenous factors such as spatially correlated environmental noise (Bjornstad, Ims & Lambin 1999; Liebhold, Koenig & Bjornstad 2004). Unfortunately, the multiplicity of causal factors makes identifying the drivers of spatial synchrony difficult unless some can be excluded a priori because of 'natural barriers'. For instance, in systems where distinct populations do not 'interact' via dispersal (e.g. sheep on isolated islands or fish in different lakes), spatially synchronized dynamics can be attributed to correlated environmental noise (Grenfell *et al*. 1998; Tedesco *et al*. 2004). However, can spatially synchronized dynamics and

Table 2. The effect of stochastic noise on detecting synchronized fluctuations between pairs of time series

	Synchrony metric									
	Mean correlation (ρ)		Kendall's concordance (W)		LdM (ϕ)		Concurrency (C)		Phase synchrony (Q)	
	Value	*P*-value	Value	*P*-value	Value	*P*-value	Value	*P*-value	Value	*P*-value
AR(2)	0·201	0·028	0·584	0·006	0·601	0·038	0·132	1·000	0·005	0·996
AR(2) + positively correlated noise	0·212	0·024	0·591	0·003	0·606	0·009	0·504	0·001	0·052	0·001
Sinusoidal model + uncorrelated noise	0·450	0·085	0·735	0·000	0·725	0·043	0·338	0·388	0·041	0·049
Sinusoidal model + negatively correlated noise	−0·123	0·782	0·474	0·793	0·439	0·638	0·000	1·000	0·253	0·001

The sinusoidal model represents two 500-step time series simulated by two in-phase sinusoidal waves (cross-correlation of 1) with the same frequency (10), and slightly different means (50 and 52, respectively). The two AR(2) time series are independent realizations of a second-order autoregressive process with first- and second-order coefficients 1·61 and −0·77, mean 1·055, and standard deviation 0·1. The uncorrelated, negatively correlated, and positively correlated noise was generated with the `phase.partnered` function by adding two independent, positively correlated, or negatively correlated white noise signals with the same mean (0) and standard deviation (8) to the sinusoidal model and the AR(2) time series. The statistical significance of the mean correlation (ρ), Kendall's W, Loreau and de Mazancourt's (LdM) index of synchrony ϕ, concurrency (C), and phase synchrony (Q) was determined via 999 Monte Carlo randomizations using functions `meancorr`, `kendall.w`, `comm.sync`, `peaks`, and `phase.sync`.

their drivers be identified in ecological systems that lack such 'natural barriers'?

One way of limiting this issue is to quantify the spatial scale of variation of potential causal processes. For instance, synchronized fluctuations between populations that lie beyond the spatial range of autocorrelated variation of a potential causal process are unlikely to be driven by that process. Hence, we can erect 'statistical barriers' that would allow us to largely exclude certain processes and thus make it easier to identify the drivers of spatially synchronized population fluctuations by (i) reducing the pool of candidate factors and (ii) limiting false positives. These 'statistical barriers' are analogous to the 'natural barriers' that have long been exploited to ascribe patterns of synchrony to their underlying cause in nature (Grenfell *et al.* 1998; Post & Forchhammer 2002; Tedesco *et al.* 2004).

Spatial patterns of variability are typically quantified by computing a specific statistic (e.g. correlation, covariance, semivariance) between pairs of locations and plotting that value as a function of the lag distance that separates them (Figs 4 and 5). These types of analyses are commonly referred to as variograms and have a rich history in geostatistics (Rossi *et al.* 1992; Fortin & Dale 2005). The most widely used metric is the semivariance (Fortin & Dale 2005):

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N} [z(x_i) - z(x_i + h)]^2 \qquad \text{eqn 6}$$

where h is the lag distance, $n(h)$ is the number of points in lag distance h, and z is the value of the variable at location x. The semivariance function thus measures how the (semi) variance of a given variable between pairs of locations changes with the (lag) distance that separates them. These so-called empirical variograms can be computed by using function `vario`. Function `vario.fit` can then be used to fit statistical models to these empirical variograms to extract important characteristics

such as (i) the nugget, which represents the semivariance at the smallest lag distance and indicates observational error; (ii) the spatial range which corresponds to the maximum distance at which the variable is spatially autocorrelated; (iii) and the sill, the semivariance at the spatial range (Fortin & Dale 2005). Variogram fitting is performed via (optionally weighted) non-linear least squares with the standard function `nls`. Because `nls` can be finicky about initial parameter estimates, `vario.fit` seamlessly shifts to fallback functions when `nls` fails. Multiple statistical models with varying levels of complexity can be fit with `vario.fit`, and their relative performance can be determined via AIC (Burnham & Anderson 2002):

$$\text{AIC} = 2p + n\log\sqrt{\frac{1}{n}\sum_{i}^{n}(\hat{x}_i - x_i)^2} \qquad \text{eqn 7}$$

where n is the number of data points in the variogram, x_i are the variogram values, $\hat{x}_i$ are the variogram values predicted by the model, $\sqrt{\frac{1}{n}\sum_{i}^{n}(\hat{x}_i - x_i)^2}$ is the root-mean-square error (RMSE), and p is the number of parameters in the model. Function `vario` can also be used to compute (cross) correlograms using the covariance, correlation, Geary's C, and Moran's I (Fortin & Dale 2005).

The Mantel correlogram can be used to quantify spatial synchrony in multivariate data sets (i.e., multiple observations per location) by computing the correlation between the time series of pairs of locations as a function of the lag distance that separates them (Bjornstad, Ims & Lambin 1999; Bjornstad & Falck 2001). Statistical significance is assessed via Monte Carlo randomizations, whereby the data are randomly assigned to each lag distance, and the correlation values are then calculated. This procedure is repeated multiple times, and the p-value for each lag distance is then computed as the proportion of randomizations that produce correlation values that are equal to or more extreme than those observed in the
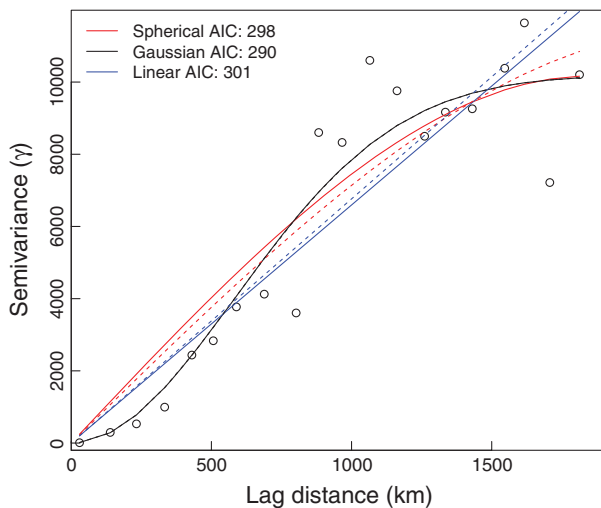
Fig. 4. Fitting statistical models (lines) to the empirical (semi)variogram (open circles) of mean annual upwelling along the West Coast of the United States for the year 2000 computed using function `vario`. The dashed lines represent weighted versions of the statistical models described in the legend using the proportion of points within each lag distance as weights when calling function `vario.fit`.

original data set. By default, the p-values obtained are for a two-tailed test where the null hypothesis is that the correlation values within each lag distance are equal to the regional mean. Alternatively, the `vario` function can also compute a one-tailed test and determine its direction automatically for each lag distance based on the observed correlation value.

As a practical example, three separate statistical models were fit to the empirical variogram of mean annual upwelling currents along the West Coast of the United States (Fig. 4, Table 2), where these currents have been shown to affect population growth (Menge, *et al*. 2003, 2004). We begin by loading the PISCO data from the `synchrony` package and selecting the variables of interest for year 2000 via the `subset` function:

```
data(pisco.data)
d<-subset(pisco.data,subset=year==2000,
  select=c("latitude","longitude","upwelling"))
```

We then compute the empirical (semi)variogram over the full spatial extent of the data set by specifying argument `extent=1`

```
semiv<-vario(data=d,extent=1)
```

Finally, we can fit three different theoretical models to the empirical (semi)variogram:

```
var.gaussian<-vario.fit(semiv$vario,
  semiv$mean.bin.dist,type="gaussian")
var.spherical<-vario.fit(semiv$vario,
  semiv$mean.bin.dist,type="spherical")
var.linear<-vario.fit(semiv$vario,
  semiv$mean.bin.dist,type="linear")
```

The best fit was obtained using the Gaussian model, which predicts a range of autocorrelated variability of approximately
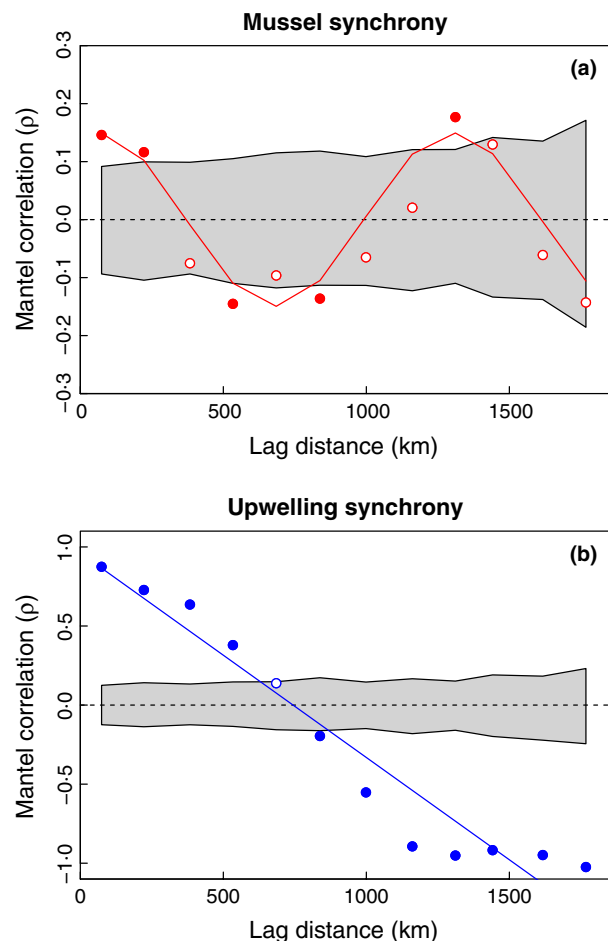

Fig. 5. Fitting statistical models to (centered Mantel) correlograms of (a) mean annual mussel cover and (b) mean annual upwelling along the West coast of the United States from 2000 to 2003. Filled circles represent statistically significant levels of synchrony ($\alpha < 0.05$; two-tailed test) based on 999 Monte Carlo randomizations computed via function `vario`. The gray region represents the 95% confidence intervals of the Monte Carlo randomizations. The horizontal dashed line represents the mean correlation across the entire region.

Table 3. Summary statistics obtained by fitting multiple statistical models to the empirical (semi)variogram of mean annual upwelling along the West Coast of the United States for the year 2000

Model	Parameter estimates			Model fit	
	Nugget (c_0)	Sill (c_1)	Range (a)	RMSE	AIC
Spherical	0·00	10157·07	1845·23	1473·00	297·80
Gaussian	0·00	10188·00	1427·09	1217·74	290·19
Linear	7·23	–	–	1665·07	300·70

The nugget (c_0) and the sill (c_1), respectively, represent the semivariance at the smallest lag distance and as it begins to plateau. The range (a) corresponds to the lag distance of the sill.

1426 km (Table 3). These results suggest that upwelling may exert an effect on biological patterns of abundance at lag distances of up to 1426 km. However, a closer look at the variogram shows a sharp change in semivariance at a lag

distance of about 800 km, with the autocorrelation (semivariance) in upwelling declining (increasing) markedly beyond that distance (Fig. 4).

Mussel populations have been shown to depend on upwelling currents for larval and food supply at local scales (e.g. Connolly, Menge & Roughgarden 2001; Menge *et al*. 2004). We can compute the spatial synchrony of upwelling and mussel cover to determine whether this relationship holds at regional to continental scales. We begin by extracting the relevant variables from the PISCO data set:

```
upw<-subset(pisco.data,select=
  c("latitude","longitude","year","upwelling"))
mus<-subset(pisco.data,select=c
  ("latitude","longitude","year","mussel_abund"))
```

Then, we `reshape` the data from 'long' to 'wide' format:

```
upw.wide<-reshape(data=upw,idvar=c("latitude",
  "longitude"),timevar=c("year"),
  direction="wide")
mus.wide<-reshape(data=mus,idvar=c("latitude",
  "longitude"),timevar=c("year"),
  direction="wide")
```

Finally, we compute spatial synchrony for each variable by calculating the average correlation within `n.bins=12` equidistant lag distances:

```
sync.upw<-vario(n.bins=12,data=upw.wide,
  type="pearson",extent=1,nrands=999,
  is.centered=TRUE,alternative="two",quiet=TRUE)
sync.mus<-vario(n.bins=12,data=mus.wide,
  type="pearson",extent=1,nrands=999,
  is.centered=TRUE,alternative="two",quiet=TRUE)
```

Despite the established link at local scales, the Mantel correlograms of mean annual mussel abundance and upwelling along the West Coast of the United States show strikingly different patterns of spatial synchrony (Fig. 5). Upwelling exhibits a statistically significant linear decay with lag distance, whereas mussel abundance exhibits a statistically significant nonlinear (periodic) pattern with lag distance, going from synchrony (lag distance <200 km), to asynchrony (200 km < lag distance <1000 km), and back to synchrony (lag distance ~1300 km). Because upwelling becomes asynchronous at intermediate lag distances (~800 km), we can safely rule it out as the main driver of synchrony in mussel abundance at lag distances greater than 800 km (Gouhier, Guichard & Gonzalez 2010a). Hence, this example shows that in interconnected ecological systems where multiple plausible drivers of spatial synchrony operate and cannot be ruled out *a priori* because of 'natural barriers', 'statistical barriers' may be erected so that processes whose synchrony patterns do not match those of the response variable of interest can be excluded *a posteriori*.

Conclusion

The examples above demonstrate how the `synchrony` can be used to help understand the relationship between ecological patterns and processes across scales. Future versions of `synchrony` will both (i) extend existing functionality by providing methods to analyze anisotropic (or directional) spatial synchrony patterns (Hagen *et al*. 2008) and (ii) provide additional approaches such as symbolic methods to identify associations between multiple time series based on their intrinsic rhythms (Cazelles 2004).

Acknowledgements

Data accessibility

The R script used to produce all the analyses and figures has been uploaded as online supporting information.

References

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

Bjornstad, O.N., Ims, R.A. & Lambin, X. (1999) Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends in Ecology & Evolution*, **14**, 427–432.

Blasius, B., Huppert, A. & Stone, L. (1999) Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature*, **399**, 354–359.

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.

Buonaccorsi, J.P., Elkinton, J.S., Evans, S.R. & Liebhold, A.M. (2001) Measuring and testing for spatial synchrony. *Ecology*, **82**, 1668–1679.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach.* Springer, New York.

Cazelles, B. (2004) Symbolic dynamics for identifying similarity between rhythms of ecological time series. *Ecology Letters*, **7**, 755–763.

Cazelles, B. & Stone, L. (2003) Detection of imperfect population synchrony in an uncertain world. *Journal of Animal Ecology*, **72**, 953–968.

Connolly, S.R., Menge, B.A. & Roughgarden, J. (2001) A latitudinal gradient in recruitment of intertidal invertebrates in the northeast pacific ocean. *Ecology*, **82**, 1799–1813.

Dray, S., Plissier, R., Couteron, P., Fortin, M.J., Legendre, P., Peres-Neto, P.R. *et al*. (2012) Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, **82**, 257–275.

Fortin, M.J. & Dale, M.R.T. (2005) *Spatial Analysis: A Guide for Ecologists.* Cambridge University Press, Cambridge, UK.

Gouhier, T.C. & Guichard, F. (2007) Local disturbance cycles and the maintenance of spatial heterogeneity across scales in marine metapopulations. *Ecology*, **88**, 647–657.

Gouhier, T.C., Guichard, F. & Gonzalez, A. (2010a) Synchrony and stability of food webs in metacommunities. *The American Naturalist*, **175**, E16–E34.

Gouhier, T.C., Guichard, F. & Menge, B.A. (2010b) Ecological processes can synchronize marine population dynamics over continental scales. *Proceedings of the National Academy of Sciences*, **107**, 8281–8286.

Grenfell, B.T., Wilson, K., Finkenstadt, B.F., Coulson, T.N., Murray, S., Albon, S.D., Pemberton, J.M., Clutton-Brock, T.H. & Crawley, M.J. (1998) Noise and determinism in synchronized sheep dynamics. *Nature*, **394**, 674–677.

Hagen, S.B., Jepsen, J.U., Yoccoz, N.G. & Ims, R.A. (2008) Anisotropic patterned population synchrony in climatic gradients indicates nonlinear climatic forcing. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 1509–1515.

Houlahan, J.E., Currie, D.J., Cottenie, K., Cumming, G.S., Ernest, S.K.M., Findlay, C.S. *et al*. (2007) Compensatory dynamics are rare in natural ecological communities. *Proceedings of the National Academy of Sciences*, **104**, 3273–3277.

Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, **54**, 187–211.

Legendre, P. (1993) Spatial autocorrelation: Trouble or new paradigm? *Ecology*, **74**, 1659.

Legendre, P. (2000) Comparison of permutation methods for the partial correlation and partial mantel tests. *Journal of Statistical Computation and Simulation*, **67**, 37–73.

Legendre, P. (2005) Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 226–245.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.

Liebhold, A., Koenig, W.D. & Bjornstad, O.N. (2004) Spatial synchrony in population dynamics. *Annual Review of Ecology Evolution and Systematics*, **35**, 467–490.

Loreau, M. & de Mazancourt, C. (2008) Species synchrony and its drivers: Neutral and non-neutral community dynamics in fluctuating environments. *American Naturalist*, **172**, E48–E66.

Loreau, M., Mouquet, N. & Holt, R.D. (2003) Meta-ecosystems: a theoretical framework for a spatial ecosystem ecology. *Ecology Letters*, **6**, 673–679.

Menge, B.A., Lubchenco, J., Bracken, M.E.S., Chan, F., Foley, M.M., Freidenburg, T.L. *et al.* (2003) Coastal oceanography sets the pace of rocky intertidal community dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12229–12234.

Menge, B.A., Blanchette, C., Raimondi, P., Freidenburg, T., Gaines, S., Lubchenco, J. *et al.* (2004) Species interaction strength: Testing model predictions along an upwelling gradient. *Ecological Monographs*, **74**, 663–684.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* (2013) *vegan: Community Ecology Package*. http://CRAN.R-project.org/package = vegan.

Post, E. & Forchhammer, M.C. (2002) Synchronization of animal population dynamics by large-scale climate. *Nature*, **420**, 168–171.

Purves, D.W. & Law, R. (2002) Fine-scale spatial structure in a grassland community: quantifying the plant's-eye view. *Journal of Ecology*, **90**, 121–129.

R Development Core Team (2013) *R: A language and environment for statistical computing*. R Development Core Team, Vienna, Austria.

Ribeiro, P.J. & Diggle, P.J. (2001) geoR: a package for geostatistical analysis. *R-NEWS*, **1**, 1418.

Ricklefs, R.E. (2008) Disintegration of the ecological community. *American Naturalist*, **172**, 741–750.

Rossi, R.E., Mulla, D.J., Journel, A.G. & Franz, E.H. (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, **62**, 277–314.

Tedesco, P.A., Hugueny, B., Paugy, D. & Fermon, Y. (2004) Spatial synchrony in population dynamics of west African fishes: a demonstration of an intraspecific and interspecific Moran effect. *Journal of Animal Ecology*, **73**, 693–705.

Vasseur, D.A. (2007) Environmental color intensifies the Moran effect when population dynamics are spatially heterogeneous. *Oikos*, **116**, 1726–1736.

Vasseur, D.A. & Fox, J.W. (2007) Environmental fluctuations can stabilize food web dynamics by increasing synchrony. *Ecology Letters*, **10**, 1066–1074.

Zar, J.H. (1999) *Biostatistical Analysis*, 4th edn. Prentice-Hall Inc., Upper Saddle River, NJ.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. R code used to conduct the analyses and produce the figures.