Oikos

OIK-02262

Hammond, M. P. and Kolasa, J. 2015. Predicting the occurrence of persistent hotspots in ecosystem variables. – Oikos doi: 10.1111/oik.02262

Appendix 1

Analytical decomposition of persistent spatial variation (PSV) into its component patterns

1. Decomposition of F-value into its spatial and temporal components

Y_{ik} is the value of an ecosystem variable at site *i* and time *k.* Persistent spatial variation (PSV) occurs when value differences between sites *i* and *j* persist over times *k* and *l* (see Fig. 1c in main text). These spatial and temporal sources of variation are contrasted by the *F*-value of an ANOVA comparing site means. This makes *F* an effective index of PSV, formulated as:

$$F = \frac{MS_{among\ sites}}{MS_{within\ sites}} \quad \text{(A1)}$$

$MS_{among\ sites}$ is the mean squares of $\bar{Y}_i$, the temporal mean of site *i*. And $MS_{within\ sites}$ is the mean squares within site *i* (i.e. variation over times *k* and *l*). The numerator and denominator can be expressed as coefficients of variation (*CV*'s), but must first be converted into variances. $MS_{among\ sites}$ is equal to the variance of site means (i.e. $var(\bar{Y}_i, \bar{Y}_j \ldots \bar{Y}_n)$), which we term $Var(\bar{Y}_i)$, times the number of temporal observations per site n_k (Sokal and Rohlf 1981). Thus:

$$F = \frac{n_k Var(\bar{Y}_i)}{MS_{within\ sites}} \quad \text{(A2)}$$

Meanwhile, $MS_{within\ sites}$ is equal to the summed temporal variances of sites *i* and *j* divided by the number of sites n_i when groups are of equal size. We therefore have:

$$F = \frac{n_k Var(\bar{Y}_i)}{n_i^{-1} \sum_{i=1}^{n} var_i} \quad \text{(A3)}$$

We now express these variances in relative terms by multiplying by $\bar{\bar{Y}}_i^2 / \bar{\bar{Y}}_i^2$ where $\bar{\bar{Y}}_i$ is the grand mean of site means. Rearranging, we get:

$$F = \frac{Var(\bar{Y}_i)}{\bar{\bar{Y}}_i^2} \cdot \frac{\bar{\bar{Y}}_i^2}{\sum var_i} \cdot n_i n_k \quad \text{(A4)}$$

The first term is equivalent to the squared coefficient of variation of the site means (i.e. $CV(\bar{Y}_i, \bar{Y}_j \ldots \bar{Y}_n)^2$) that we term CV_s^2 with subscript *s* to emphasize that it is a relative measure of spatial variation. *F* thus becomes:

$$F = CV_s^2 \cdot \frac{\bar{\bar{Y}}_i^2}{\sum var_i} \cdot n_i n_k \quad \text{(A5)}$$

To turn the second term into a squared temporal Coefficient of variation, we convert $\bar{\bar{Y}}_i^2$ – a spatial grand mean – into the regional temporal mean $\bar{Y}_K^2$. The regional temporal mean is the mean obtained by summing values across sites and averaging over time (i.e. $\bar{Y}_K$ where $K = \sum_{i=1}^{n} Y_{ik}$). $\bar{Y}_K^2$ is substituted into Eq. A5 using the equality $\bar{\bar{Y}}_i^2 = (1/n_i^2)\bar{Y}_K^2$. Simplifying, we obtain:

$$F = CV_s^2 \cdot \frac{\bar{Y}_K^2}{\sum var_i} \cdot \frac{n_k}{n_i} \quad \text{(A6)}$$

The new, second term of Eq. A6 is the quotient of the squared regional temporal mean and the summed temporal variances of each site. This term is the reciprocal of Proulx et al.'s (2010) *Variance CV*. It is thus an index of temporal stability or constancy of sites which we call CV_t^{-1}. Applying this term, we get:

$$F = CV_s^2 \cdot CV_t^{-1} \cdot \frac{n_k}{n_i} \quad \text{(A7)}$$

The third term n_k/n_i changes value with the number of observations over time and across space. It is possible that this term's influence on *F* may complicate comparisons between data sets of different sizes. As a precaution against this, we only compared datasets of similar sizes and values of n_k/n_i.

2. Decomposing spatial variation (CV_s^2) into its two spatial components

CV_s^2 is a dimensionless index of spatial variation among site means defined, as in Eq. A4, as:

$$CV_s^2 = \frac{Var(\bar{Y}_i)}{\bar{\bar{Y}}_i^2} \quad \text{(A8)}$$

The numerator $Var(\bar{Y}_i)$ is the (sample) variance of site means $\bar{Y}_i$ and integrates two sources of spatial variation that are common in landscapes: first, variation arising from having unoccupied sites with densities of zero and, second, variation in population density among occupied sites (Fig. A1). These sources can be partitioned from $Var(\bar{Y}_i)$ using a model II (random effects), one-way ANOVA that compares two groups: occupied sites (*o*) and unoccupied sites (*u*).

In this scheme, $Var(\bar{Y}_i)$ is expressed as sums of squares (*SS*) and decomposed into among-group and within-group sources of variation, as follows:

$$SS_I = SS_{o,u} + SS_{within} \quad \text{(A9)}$$

where SS_I is the total sums of squares, $SS_{O,u}$ is the sums of squares between the occupied and unoccupied group, and SS_{within} is the sums of squares within the groups. Among- and within-variation are converted into their corresponding variance components, as follows using formulae that can accommodate groups of unequal size (Sokal and Rohlf 1981 p. 216):

$$s^2_{o,u} = \frac{MS_{o,u} - MS_{within}}{n_o} \quad \text{(A10)}$$

$$\mathrm{s}^2 = \mathrm{MS}_{\mathrm{within}} \quad \text{(A11)}$$

where MS are the mean squares corresponding to the above *SS* and n_o is a modified sample size for groups that differ in their respective n (see Sokal and Rohlf 1981 p. 214 for details). $s^2_{o,u}$ is the variance added by value differences between the occupied and unoccupied group, and so is the variance attributable to having zeros in unoccupied sites (Fig. A1). Meanwhile, s^2 is an estimate of the variance within groups. The definition of s^2 (Eq. A11) can be rewritten to isolate the contributions of *SS* from the occupied ($SS_{within.o}$) and unoccupied group ($SS_{within.u}$):

$$s^2 = \frac{SS_{within}}{n_i - 2} = \frac{SS_{within.o}}{n_i - 2} + \frac{SS_{within.u}}{n_i - 2} \quad \text{(A12)}$$

We note that all values within the unoccupied group are always zero, rendering $SS_{within.u}$ zero also. Thus:

$$\mathrm{s}^2 = \frac{\mathrm{SS}_{\mathrm{within.o}}}{\mathrm{n_i} - 2} + 0 \quad \text{(A13)}$$

Eq. A13 shows that only variation within the occupied group contributes to s^2. Thus, in our special case, s^2 provides an estimate of variation among populations in occupied sites.

Variance components are typically expressed in relative terms as the ratio of a single component to the total of all components. When $s^2_{o,u}$ is standardized to the component total, it is called the interclass correlation coefficient (ICC) and measures the proportion of total variance in the sample that arises between the occupied and unoccupied group. If we then multiply this proportion by the original spatial variance $Var(\bar{Y}_i)$, we obtain a new, absolute variance $Var_{s.occ}$ that expresses variance arising from occupancy patterns:

$$Var_{s.occ} = \left(\frac{s^2_{o,u}}{s^2 + s^2_{o,u}}\right) Var(\bar{Y}_i) \quad \text{(A14)}$$

Similarly, if we multiply $Var(\bar{Y}_i)$ by the proportion of variance from within-groups, we obtain $Var_{s.pop}$ that expresses variance arising among populations in occupied sites:

$$\mathrm{Var}_{\mathrm{s.pop}} = \left(\frac{\mathrm{s}^2}{\mathrm{s}^2 + \mathrm{s}^2_{\mathrm{o,u}}}\right) \mathrm{Var}(\bar{\mathrm{Y}}_{\mathrm{i}}) \quad \text{(A15)}$$

$Var_{s.occ}$ and $Var_{s.pop}$ sum to $Var(\bar{Y}_i)$ and so may be substituted into Eq. A8 to yield:

$$CV_s^2 = \frac{Var_{s.occ}}{\bar{\bar{Y}}_i^2} + \frac{Var_{s.pop}}{\bar{\bar{Y}}_i^2} \quad \text{(A16)}$$

Eq. A16 shows that CV_s^2, an index of spatial variation among site means, decomposes into two dimensionless indices that we call $CV^2_{s.occ}$ and $CV^2_{s.pop}$.

We note that estimates of variance components can be biased by several factors which would, in turn, affect estimates of $Var_{s.occ}$ and $Var_{s.pop}$ and their dimensionless forms in Eq. A16. First, biased estimates may occur when there are few observations (sites) per group (Gray 2012), which is the case in the dataset containing only seven lakes. Second, ANOVA is not robust to unequal group variances, which is certain to occur because the unoccupied group always has zero variance. However, these biases should not affect our results because all variables within a dataset would be subjected to the same degree of bias. Thus, while the absolute magnitude of $Var_{s.occ}$ and $Var_{s.pop}$ may reflect minor bias, the relationships sought among variables (e.g. mean-variance scaling) should not be affected.

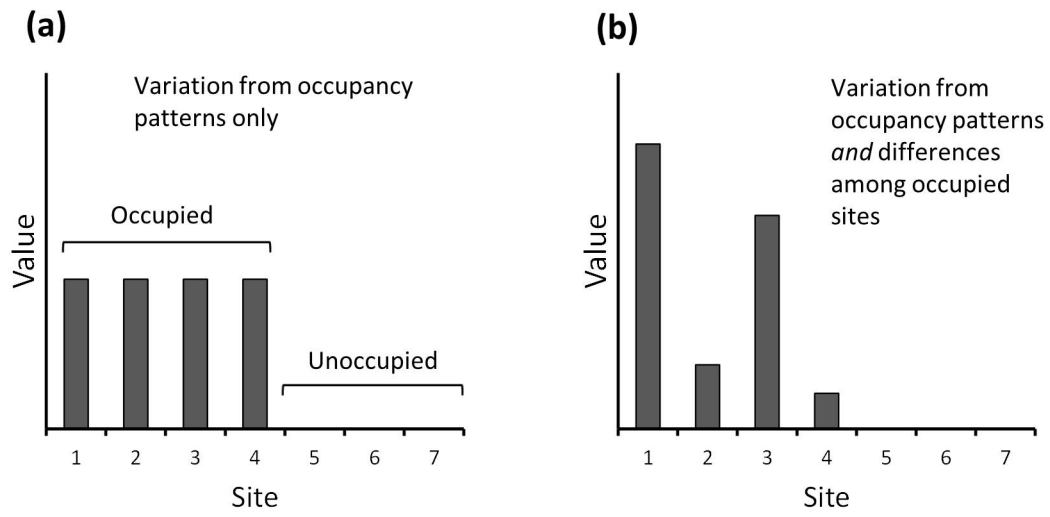

Figure A1. Plots showing two sources of variation among sites. (a) Spatial variation arising solely from differences between occupied and unoccupied sites i.e. no variation among populations in occupied sites (Interclass correlation coefficient = 1.0). (b) Spatial variation arising from both differences between occupied and unoccupied sites and variation among populations in occupied sites (*ICC* = 0.5).

References

Gray, B. R. 2012. Variance components estimation for continuous and discrete data, with emphasis on cross-classified sampling designs. – In: Gitzen, R. A. et al. (eds), Design and analysis of long-term ecological monitoring studies. Cambridge Univ. Press, pp. 200–227.

Proulx, R. et al. 2010. Diversity promotes temporal stability across levels of ecosystem organization in experimental grasslands. – PLoS ONE 5: e13382.

Sokal, R. R. and Rohlf, F. J. 1981. Biometry, 2nd edn. – W. H. Freeman and Co, p. 859.

Appendix 2

Using mean-variance scaling to quantify the effects of abundance on PSV and its components

In this appendix, we use a hypothetical example to illustrate how mean-variance scaling can quantify the effect of abundance on the spatial and temporal components of persistent spatial variation (PSV). From the equation for *F* and its components (Eq. B1), we note that the spatial component CV_s^2 (Eq. B2) and the temporal component CV_t^{-1} (Eq. B3) are squared *CV*'s and thus ratios of variances and squared means.

$$F = \frac{MS_{among\ sites}}{MS_{within\ sites}} = CV_s^2 \cdot CV_t^{-1} \cdot \frac{n_k}{n_i} \quad \text{(B1)}$$

$$CV_s^2 = \frac{Var(\bar{Y}_i)}{\bar{\bar{Y}}_i^2} \quad \text{(B2)}$$

$$CV_t^{-1} = \frac{\bar{Y}_K^2}{\sum_{i=1}^{n} var_i} \quad \text{(B3)}$$

Scaling of means and variances on a log-log plot is a common means of determining whether variation changes with mean abundance. The null expectation is that the scaling exponent *b* =2 when variability neither increases nor decreases with mean abundance (Taylor 1961, Kilpatrick and Ives 2003). Different values of b also have specific meanings for *CV*'s, as follows (McArdle et al. 1990, Ballantyne and Kerkhoff 2007):

1. when *b* =2, the corresponding CV does not change with mean abundance
2. when *b* >2, CV is an increasing function of mean abundance
3. when *b* < 2, CV is a decreasing function of mean abundance

Thus, scaling the variances in Eqs. B2–B3 and their corresponding means measures the effect of mean density on the spatial and temporal *CV*'s; CV_s^2 and CV_t^{-1}. Furthermore, these *CV* terms combine in a known way to create PSV (Eq. B1) and so the effects of abundance on these terms can be traced up to PSV. Note that $Var(\bar{Y}_i)$ in Eq. B2 decomposes into two further variances which are scaled with theirs means in the main text but that these are not included in our example here.

Figure B1 shows two possible ways in which abundance can affect CV_s^2 and CV_t^{-1} and, in turn, PSV measured by the F-value. In Fig. B1a, spatial variability increases with mean (*b* > 2) while temporal variability does not (*b* = 2). The net result is that *F* grows with abundance because abundance increases the value of CV_s^2 but does not affect CV_t^{-1}. In contrast, Fig. B1b shows the case where temporal variability increases with mean (*b* > 2) while spatial variability is unaffected (*b* = 2). In net, *F* decreases with abundance because abundance increases temporal variability which decreases its inverse CV_t^{-1} and it does not change the value of CV_s^2 . Because scaling slopes may

reinforce each other (e.g. both < 2) or may act in opposition (e.g. one > 2, one <2), the net effect of abundance on *F* will result from their particular combination.

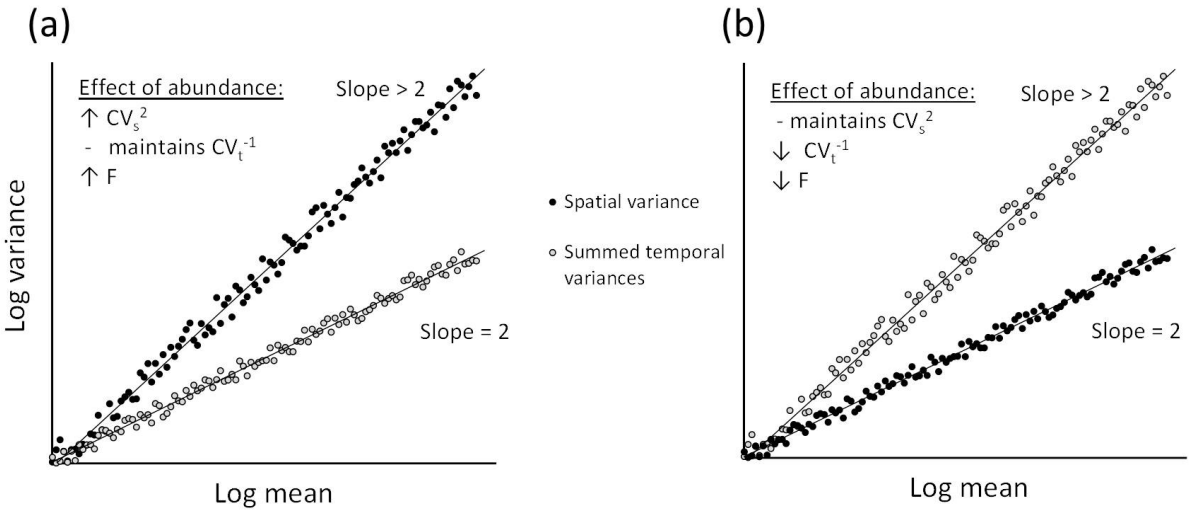

Figure B1. Hypothetical mean-variance scaling plots illustrating possible effects of abundance on variance terms of Eq. B1 (Eq. 1 in main text). Through its effects on spatial variance and summed temporal variances (corresponding to CV_s^2 and CV_t^{-1}, respectively), abundance could: (a) Accentuate spatial variability (scaling slope $b > 2$) and not affect temporal variability ($b = 2$), thereby increasing *F,* or (b) not affect spatial variability ($b = 2$) and increase temporal variability ($b \geq 2$), thereby decreasing *F*.

References

Ballantyne, F. and Kerkhoff, A. J. 2007. The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. – Oikos 116: 174–180.

Kilpatrick, A. M. and Ives, A. R. 2003. Species interactions can explain Taylor's power law for ecological time series. – Nature 422: 65–68.

McArdle, B. H. et al. 1990. Variation in the size of animal populations: patterns, problems and artefacts. – J. Anim. Ecol. 59: 439–454.

Taylor, L. R. 1961. Aggregation, variance and the mean. – Nature 189: 732–735.

Appendix 3

Chemical hardness (η) of lake ions

Table C1. Sources of chemical hardness (η) estimates used in analyses. Also indicated is the related chemical species used as an alternative if an estimate could not be found.

Variable	Substitute used	Source
Mg^{2+}		Parr and Pearson 1983
Na^{+}		Parr and Pearson 1983
SO_4^{2-}		Parr and Pearson 1983
Ca^{2+}		Parr and Pearson 1983
Cl^{-}		Parr and Pearson 1983
K^{+}		Pearson 1988
Fe^{3+}		Parr and Pearson 1983
Mn^{2+}		Parr and Pearson 1983
Dissolved inorganic carbon	CO_2	Parr and Pearson 1983
Total organic carbon	Mean of six carbohydrates	Torrent-Sucarrat et al. 2010
NO_3^{-}/NO_2^{-}	NO_2^{-}	Pearson 1988
NH_4^{+}	NH_3	Parr and Pearson 1983
Dissolved oxygen	O_2	Pearson 1988
Total P	P	Parr and Pearson 1983

References

Parr, R. G. and Pearson, R. G. 1983. Absolute hardness: companion parameter to absolute electronegativity. – J. Am. Chem. Soc. 105: 7512–7516.

Pearson, R. G. 1988. Absolute electronegativity and hardness: application to inorganic chemistry. – Inorg. Chem. 27: 734–740.

Torrent-Sucarrat, M. et al. 2010. On the applicability of local softness and hardness. – Phys. Chem. Chem. Phy. PCCP 12: 1072–1080.

Appendix 4

Best predictors of PSV: Full and reduced multiple regression model results

We used multiple regression to identify the best predictors of persistent spatial variation (PSV) from those hypothesized (see Introduction and Methods). We entered all factors for which data were available into standard multiple regression models and sought terms that significantly predicted *F* (Table D1). Because overfitting may occur in cases with too many factors, we repeated the analyses using the backward stepwise routine in Statistica ver. 8.0 to see whether the important factors were retained in reduced models (Table D2). The two methods consistently indicated the importance of the same predictors (mainly temporal CV^{-1} and autocorrelation time) within datasets.

Table D1. Results from standard multiple regression models predicting log(*F*). Predictors are ranked by their semi-partial correlation coefficient. Significant predictors appear in bold. Superscripts: * = log-transformed; †= square-root transformed.

Data	Predictor	t	p	Semi-partial correlation	R^2	DF	F	p
Rock pool species	**Temporal CV^{-1}***	3.07	0.005	0.31	0.71	6,29	11.59	<0.0001
	OMI†	2.40	0.023	0.24				
	Mean density*	1.37	0.180	0.14				
	Autocorrelation time	1.26	0.217	0.13				
	Skewness	0.66	0.513	0.07				
	Tolerance	–0.25	0.805	–0.03				
Rock pool species and environmental variables	**Temporal CV^{-1}***	3.25	0.003	0.35	0.60	3,34	16.67	<0.0001
	Autocorrelation time*	0.98	0.336	0.11				
	Skewness	0.84	0.407	0.092				
Lake ions	**Temporal CV^{-1}***	3.95	0.003	0.48	0.86	4,9	14.04	0.0006
	Chemical hardness†	1.86	0.096	0.23				
	Autocorrelation time*	0.47	0.652	0.06				
	Skewness	0.25	0.809	0.03				
Lake environmental variables	**Temporal CV^{-1}***	3.11	0.005	0.44	0.58	3,21	9.60	0.0003
	Autocorrelation time*	3.11	0.005	0.44				
	Skewness	–0.16	0.876	–0.02				
Lake species	Temporal CV^{-1}*	1.25	0.429	0.40	0.90	12,1	0.73	0.734
	Tolerance	0.63	0.640	0.20				
	Doubling time	0.55	0.678	0.18				
	Autocorrelation time*	0.23	0.858	0.08				
	OMI	0.18	0.887	0.06				
	Northern range limit	–0.09	0.945	–0.03				
	Trophic level	–0.14	0.911	–0.05				
	Skewness	–0.26	0.840	–0.08				
	Mean density*	–0.34	0.791	–0.11				
	Mass	–0.36	0.779	–0.12				
	Southern range limit	–0.37	0.773	–0.12				
	Fecundity*	–0.48	0.71	–0.16				
Lake species and environmental variables	**Temporal CV^{-1}***	7.22	0.000	0.44	0.83	3,48	75.69	<0.0001
	Autocorrelation time*	2.36	0.022	0.14				
	Skewness	–1.58	0.120	–0.10				

Table D2. Results from backward stepwise regression models predicting log(*F*). Predictors are ranked by their semi-partial correlation coefficient. Significant predictors appear in bold. Superscripts: * = log-transformed; †= square-root transformed.

Data	Predictor	t	p	Semi-partial correlation	R^2	DF	F	p
Rock pool species	**Temporal CV^{-1}***	6.94	0.000	0.77	0.59	1,34	48.12	<0.0001
Rock pool species and environmental variables	**Temporal CV^{-1}***	6.88	0.000	0.56	0.57	1,36	47.28	<0.0001
Lake ions	**Temporal CV^{-1}***	5.52	0.000	0.62	0.86	2,11	34.19	<0.0001
	Chemical hardness†	3.25	0.008	0.36				
Lake environmental variables	**Temporal CV^{-1}***	3.55	0.002	0.49	0.58	2,22	15.06	<0.0001
	Autocorrelation time*	3.19	0.004	0.44				
Lake species	**Temporal CV^{-1}***	4.41	0.001	0.40	0.59	1,12	19.44	0.0009
Lake species and environmental variables	**Temporal CV^{-1}***	13.84	0.000	0.89	0.79	1,50	191.6	<0.0001

Appendix 5

Names and abbreviations of study variables as plotted in Figure 2

Table E1. Abbreviation and names of study variables plotted in Fig. 2 for Wisconsin lakes and Jamaican rock pools.

Abbreviation	Variable	System
Mg	magnesium (Mg^{2+})	Wisconsin temperate lakes
Na	sodium (Na^{+})	
SO_4	sulfate (SO_4^{2-})	
Ca	calcium (Ca^{2+})	
Cl	chloride (Cl^{-})	
K	potassium (K^{+})	
Fe	iron (Fe^{3+})	
Mn	manganese(Mn^{2+})	
SiO_2	silica	
DIC	dissolved inorganic carbon	
DOC	dissolved organic carbon	
TOC	total organic carbon	
NO_3	nitrate (NO_3^{-})	
NH_4	ammonium (NH_4^{+})	
DO	dissolved oxygen	
P	total phosphorus	
pH	pH	
Cond.	conductivity	
Alk.	alkalinity	
Temp.	water temperature	
Light input	light intensity at surface	
Light atten.	light attenuation	
Species richness	total species count	
Total density	total density of organisms	
Rock bass	*Ambloplites rupestris*	
Mysis	*Mysis relicta*	
Bluegill	*Lepomis macrochirus*	
Temp.	water temperature	Jamaican rock pools
pH	pH	
DO	dissolved oxygen	
Chl a	chlorophyll a	
Turbidity	water turbidity	
Species richness	total species count	
Total abundance	total density of organisms	

Appendix 6

Supplementary analysis of PSV predictors

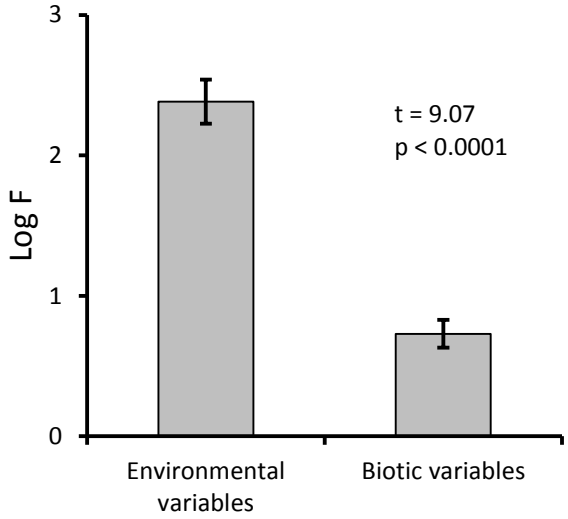

Figure F1. Persistent spatial variation was significantly higher in lake environmental variables than biotic variables.

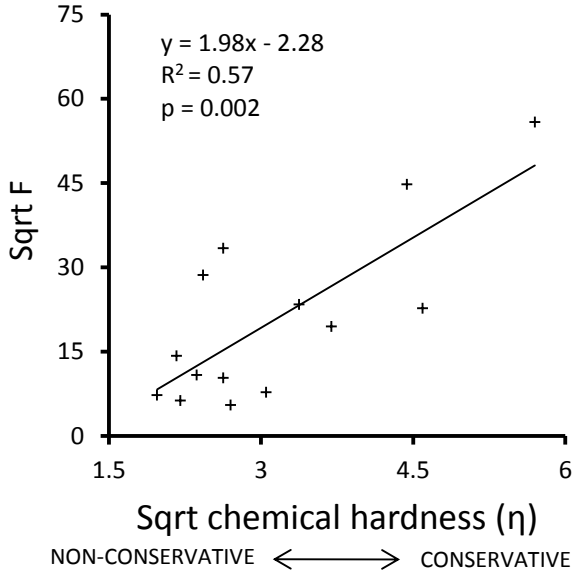

Figure F2. In lakes, published values of hardness for a chemical species were significantly related to persistent spatial variation.

Appendix 7

Representative patterns of PSV for different types of ecosystem variables in temperate lakes

PSV differed among types of ecosystem variables (Fig. 2; Supplementary material Appendix 6). To help visualize these differences, we plot data for representative variables in the seven LTER North Temperate Lakes in Wisconsin over time. Figure G1 plots concentrations of conservative ions, which typically had clear and persistent hotspots associated with high PSV. PSV was slightly lower for physical variables (Fig. G2), followed by nutrients (Fig. G3) and by population variables, which had the lowest PSV values (Fig. G4).

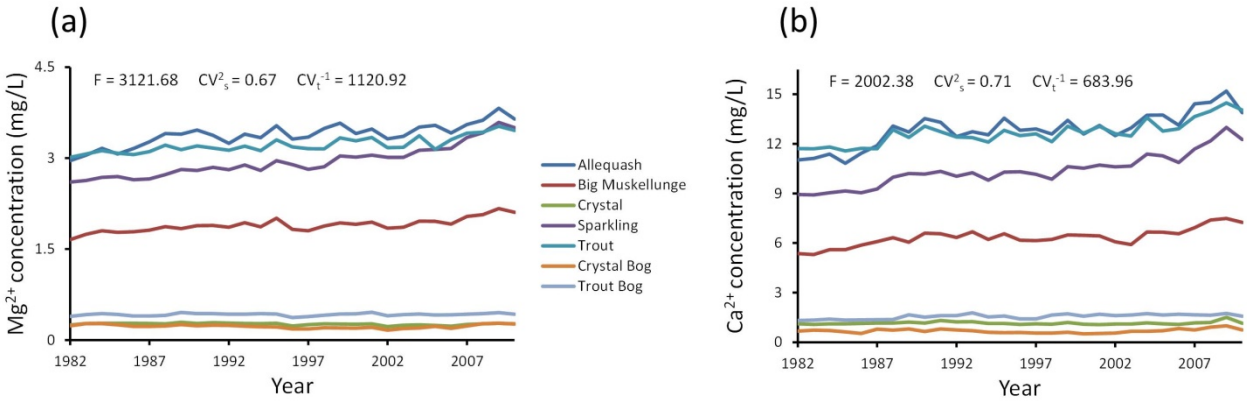

Figure G1. Time series of representative conservative ions in seven Wisconsin lakes, 1982–2010: (a) Magnesium and (b) calcium. *F*-values are reported as well as its components, spatial variation (CV_s^2) and temporal stability (CV_t^{-1}).

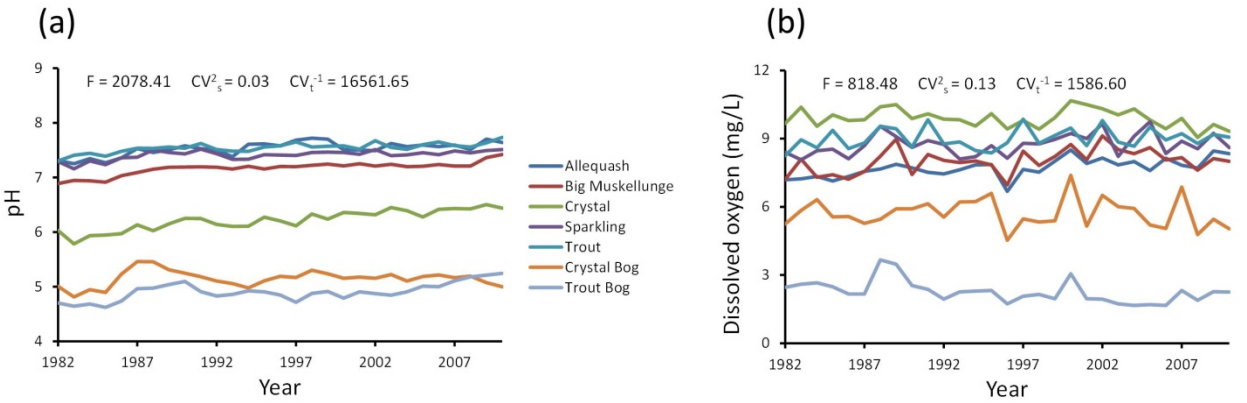

Figure G2. Time series of representative physical variables in seven Wisconsin lakes, 1982–2010: (a) pH and (b) dissolved oxygen. *F*-values are reported as well as its components, spatial variation (CV_s^2) and temporal stability (CV_t^{-1}).

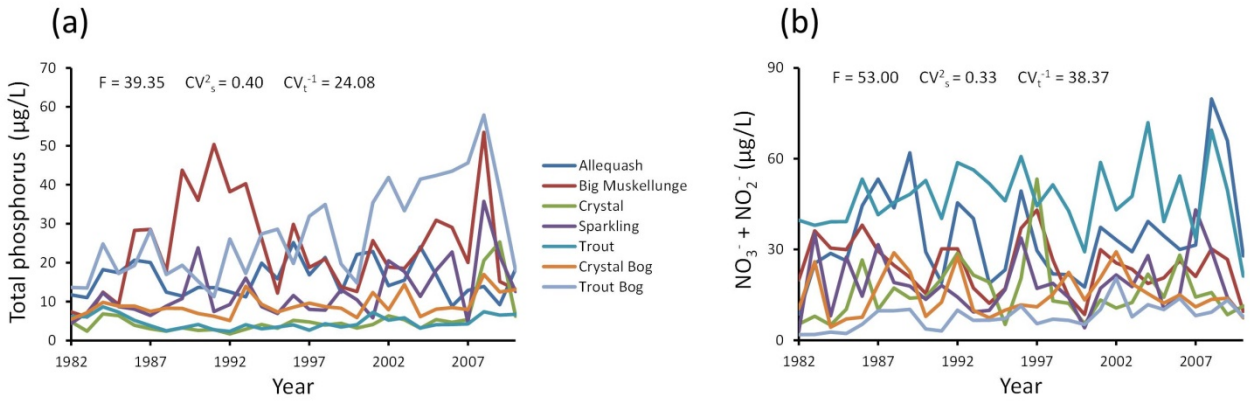

Figure G3. Time series of representative biologically-important nutrients in seven Wisconsin lakes, 1982-2010: (a) Total phosphorus and (b) nitrate/nitrite. *F*-values are reported as well as its components, spatial variation (CV_s^2) and temporal stability (CV_t^{-1}).

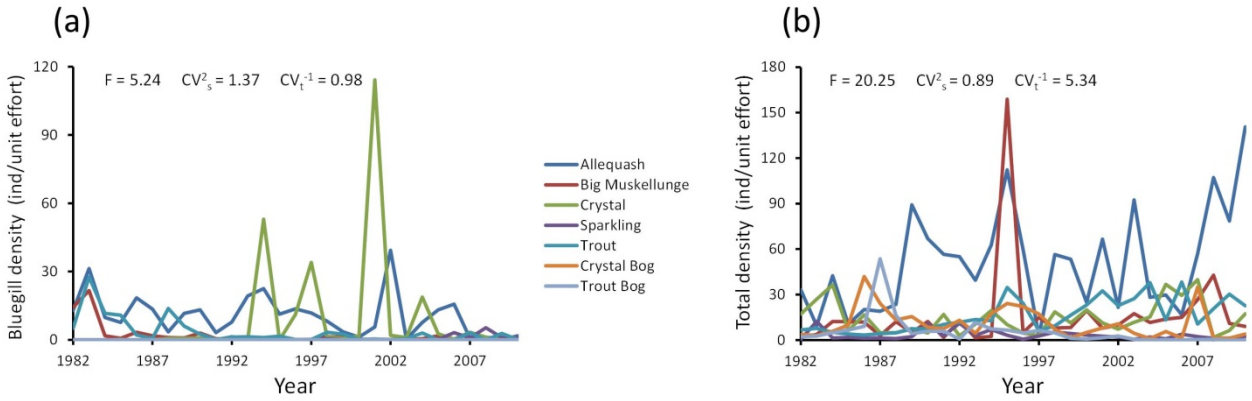

Figure G4. Time series of representative population varibles in seven Wisconsin lakes, 1982–2010: (a) Density of Bluegill, *Lepomis macrochirus* and (b) mean density of all populations. *F*-values are reported as well as its components, spatial variation (CV_s^2) and temporal stability (CV_t^{-1}).

Appendix 8

Effects of spatial and temporal scale on the detection of PSV and its components

Patterns of variation are scale-dependent (Levin 1992) unless they are from the special class of scale-invariant patterns (e.g. fractals). Thus, when ecologists take measurements at discrete locations and times, they may detect persistent spatial variation (PSV) occurring at those scales but fail to detect it if it occurs at other scales. For example, pH values may persist within a narrow range from year to year, but fluctuate considerably over a day. Similarly, pH may be similar among lakes but differ among basins within a single lake. We ask, therefore, at what scales of variation for an ecosystem process will sampling (at fixed locations and times) detect PSV? We first derive a theoretical expectation, then validate it by manipulating spatial and temporal scales of simulated data, and finally provide a general rule for the conditions under which PSV will be detectable.

Theoretical expectations

Fourier spectral analysis breaks down series with complex fluctuations, whether temporal or spatial, into cyclic signals at different frequencies or scales (Platt and Denman 1975). Because theory for this method is well-developed, we can use it to derive expectations for which spatial and temporal scales of variation will be detected as PSV for a given sampling design.

The upper limit for detecting fluctuations is known as the Nyquist frequency f_s, which corresponds to the finest-grain detail that can be discerned given the frequency of sampling (for a helpful discussion of spectral analysis, see Supplement of Denny et al. 2004). Cycles of a measured ecological process y at frequencies higher than f_s are simply too rapid to be accurately captured by the sampled series. The lower limit for detecting fluctuations is called the fundamental frequency f_0, which corresponds to the largest cycle that can be detected given the extent of sampling. Cycles at frequencies lower than f_0 are too gradual to be detected because they do not complete a full cycle within the sampling period. Thus, the measurable variance of a time or space series is bounded by f_s and f_0 such that:

$$var(y) = \int_{f_0}^{f_s} S(f)df \qquad \text{(H1)}$$

where $S(f)$ is the autospectral density function, which is the contribution to overall variance of y for a given frequency f (Denny et al. 2004).

Each frequency in Eq. H1 can be replaced by its inverse, the period of the cycle. Such a formulation has the advantage of corresponding better to the common definition of scale as an extent of either time or space. The variance of a time series y_i is thus:

$$var(y_i) = \int_{T_0}^{T_s} S(T) d\,T \tag{H2}$$

T_s is the smallest detectable period of variation, which we call the Nyquist period because it is the period analog of the Nyquist frequency. Its value is twice the interval between samples (i.e. $2\Delta k$ where k is a sampled timepoint) such that fluctuations with shorter periods are not detected. T_0 is the fundamental period, the largest detectable period and corresponds to the extent over which sampling occurs (i.e. $(n\text{-}1)\,\Delta k$).

Similarly, the variance of a spatial series y_k is the variance contributed by periods between the spatial Nyquist period S_s and fundamental period S_0:

$$var(y_k) = \int_{S_0}^{S_s} S(S) d\,S \tag{H3}$$

Our measure of PSV, F, is a ratio of spatial (among-site) and temporal (within-site) variances which, as seen above, can be expressed in terms of variance spectra. From Appendix 1, we can define F as:

$$F = \frac{n_k Var(\bar{Y}_i)}{n_i^{-1} \sum_{i=1}^{n} var_i} \tag{H4}$$

where $Var(\bar{Y}_i)$ is the spatial variance of site means, var_i is the temporal variance of site i, and n_k and n_i are the number of time points and locations sampled, respectively.

Substituting the spectral definitions of spatial and temporal variance in Eq. H2 and H3 into Eq. H4, we rewrite F in spectral terms:

$$F = \frac{n_k \int_{S_0}^{S_s} S(S) d\,S}{n_i^{-1} \sum_{i=1}^{n} \int_{T_0}^{T_s} S(T_i) d\,T_i} \tag{H5}$$

Equation H5 suggests the spatial and temporal scales that will be most detectable and contribute most to F of a given ecological process. Specifically, it indicates that F will be greater when:

1) the numerator is large when spatial variation occurs at detectable scales, that is between twice the spatial sampling interval (S_s) and the sampling extent (S_0), and

2) the denominator is small when temporal variation of patches occurs outside detectable scales, that is at scales less than twice the temporal sampling interval (T_s) or more than the sampling extent (T_0).

While theory provides expectations for which spatial and temporal scales of variation will be detected as PSV, practice may differ. For instance, variation that occurs outside of the limits in Eq. H5 can influence the measured variance through aliasing. Aliasing occurs when variation at periods smaller than the minimum detectable period spuriously appears as longer-period fluctuations (Denny et al. 2004; Fig. H3). For such empirical reasons, we validated theoretical expectations with numerical simulations.

Detectability of PSV for differently-scaled processes – a numerical simulation

Methods

We tested the detectability of PSV for different scales of variation by simulating pH values while manipulating the dominant spatial and temporal periods of fluctuation. To match theoretical expectations from spectral analysis, pH values at time k were modelled as a cosine function, modified from Arrigoni et al. (2008), as follows:

$$y_k = (0.5R_t)cos\big((k - p_k)c_t\big) + M_i + \varepsilon \quad \text{(H6)}$$

where R_t is the amplitude of temporal fluctuations, k is the day, p_k is the day of peak amplitude, c_t is the factor that converts radians to days by dividing 2π by the period of interest and was used to manipulate the period of the cosine wave, M_i is the mean value of site i and ε is stochastic Gaussian noise.

To introduce cyclic variation across sites, the value of M_i in Eq. H6 was also made a cosine function:

$$M_i = (0.5R_s)cos\big((i - p_i)c_s\big) + \overline{M_\iota} \quad \text{(H7)}$$

where R_s is the amplitude of spatial fluctuations, i is the site, p_i is the site of peak amplitude, c_s is the factor that converts radians to sites by dividing 2π by the period of interest and was used to manipulate the period of the cosine wave, and $\overline{M_\iota}$ is the grand mean of all sites. Parameter values were estimated from pH data for the seven North Temperate LTER lakes.

To mimic a discrete sampling regime, pH values were simulated in Excel ver. 14.0 (Microsoft Corp. 2010) for a linear array (i.e. a strip) of 1000 sites and sampled at 100-site intervals. In time, simulated daily values were sampled at 365-day intervals over ten years. The resulting 10 × 10 site by time data matrix was used to calculate F and its components, CV_t^{-1} and CV_s^2. This process was repeated for a wide range of spatial and temporal periods of pH fluctuation.

Simulation results and discussion

Simulation results largely agreed with theoretical expectations and refined the scaling conditions under which PSV can be detected. Because F as a measure of PSV depends on both spatial variation (CV_s^2) and temporal stability (CV_t^{-1}), we present the effects of scale on these components before considering its effects on F.

Effect of spatial and temporal scale on stability

Similar to the expectations from Eq. H2, temporal variation was least when pH varied at very small or very large temporal periods. Figure H1a shows that high stability values, measured as CV_t^{-1}, occurred: 1) at periods larger than twice the fundamental temporal period T_0, and 2) at periods smaller than the temporal Nyquist period T_s, where temporal variation is poorly detected.

Variation with a period more than twice the fundamental period registers as more stable because change occurs so slowly that not all variation is captured within the sampling period. Fig. H2a shows variation occurring at the fundamental period such that a full cycle of change occurs over the sampling period and an estimate of variation would capture the full amplitude of the cycle. In contrast, Fig. H2b shows a process with a much longer period so that sampling only captures a portion of the full amplitude of the cycle. The result of this sampling phenomenon is that, beyond twice the fundamental period, CV_t^{-1} gradually increases as the temporal period of the sampled process grows to decades and centuries (Fig. H1b). In essence, change is so gradual that it barely registers.

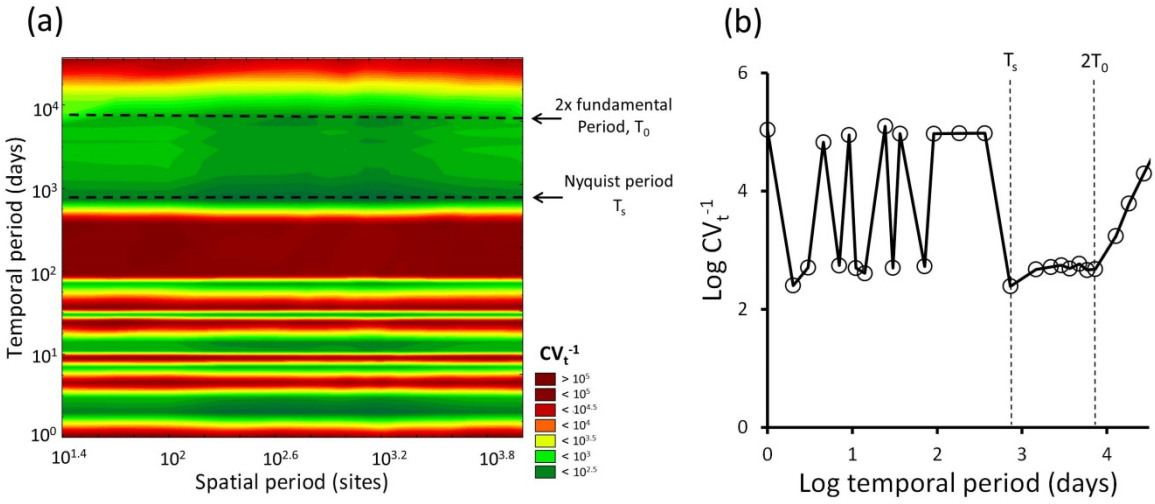

Figure H1. Effect of spatial and temporal scale of variation on temporal stability of simulated pH. (a) Stability of pH, measured as CV_t^{-1}, peaked when it fluctuated at temporal periods greater than twice the fundamental period and less than the Nyquist period. (b) Example of change in stability with increasing temporal period of variation.

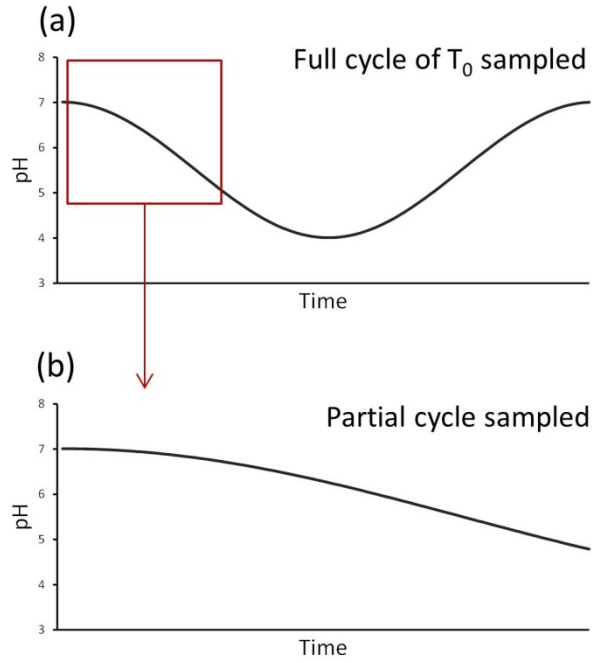

Figure H2. Effect of long temporal period on the variation of pH detected by sampling. (a) Variation occurring at the fundamental period T_0 completes a full cycle within the sampling period and the full amplitude is sampled. (b) Variation occurring at more than twice T_0 does not complete a full cycle in the sampling period and so only a portion of its amplitude and variation is sampled.

Variation at periods shorter than the Nyquist period often registered as high stability, but only for periods that were a harmonic of the Nyquist period. This gave rise to the bands of high stability in Fig. H1a and appears to be result of aliasing where short, periodic variation registers as longer-period variation when sampled (Denny et al. 2004). High stability occurs at harmonics of the Nyquist period because it is at these that pH cycles line up with sampling intervals so that only e.g., peaks or troughs are sampled (Fig. H3a). As a result, little to no variation is detected. In contrast, if the period is not a harmonic, peaks, troughs and intermediate points are all captured by sampling and variation is detected (Fig. H3b). Thus, when variation occurs at short periods below the Nyquist period, CV_t^{-1} may be high but these values are sensitive to the sampling interval chosen relative to period of variation.

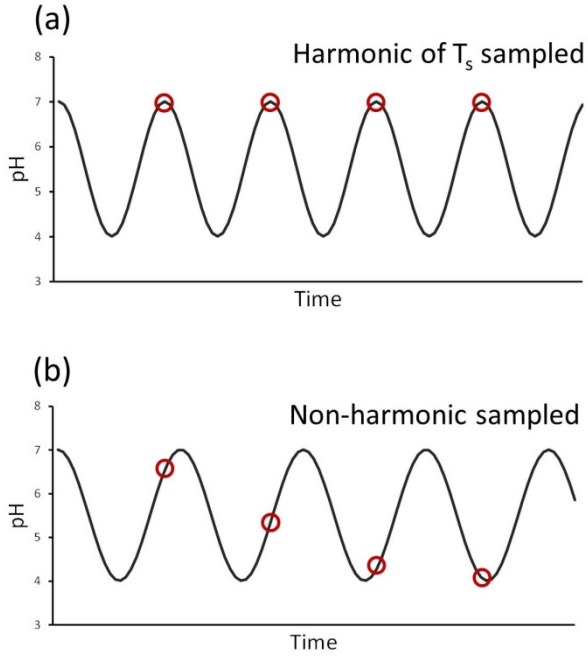

Figure H3. Effect of short temporal period on the variation of pH detected by sampling. (a) Variation occurring at periods below the Nyquist period T_s are incompletely sampled. But if they are a harmonic of T_s, sampling occurs at the same parts of the phase (e.g., peak), resulting in the detection of low variation. (b) Variation that is not a harmonic of the fundamental period is sampled at different parts of the phase, leading to the sampling of a wider range of values and thus greater variation.

Effect of spatial and temporal scale on spatial variation

Equation H3 indicated that spatial variation would be most detectable when it occurred at scales greater than twice the sampling interval and less than the sampled extent. Simulations largely agreed with this prediction, with Fig. H4a showing peak spatial variation, measured as CV_s^2, above the Nyquist period S_s and below twice the fundamental period S_0. Consistent with theory, fluctuations in this scale interval register clearly as high spatial variation but taper off outside of the interval (Fig. H4a).

Spatial variation is low at periods larger than twice the fundamental period for the same reason as the temporal case (Fig. H2): sampling captures only a portion the entire cycle of variation, leading to the detection of less amplitude and spatial variation. Spatial variation is also low when its dominant period falls below the Nyquist period. This is because the pattern is inadequately sampled and variation is not detected in its periodic form. Rather, sampling only captures a small subset of the full range of values which registers as weak spatial variation.

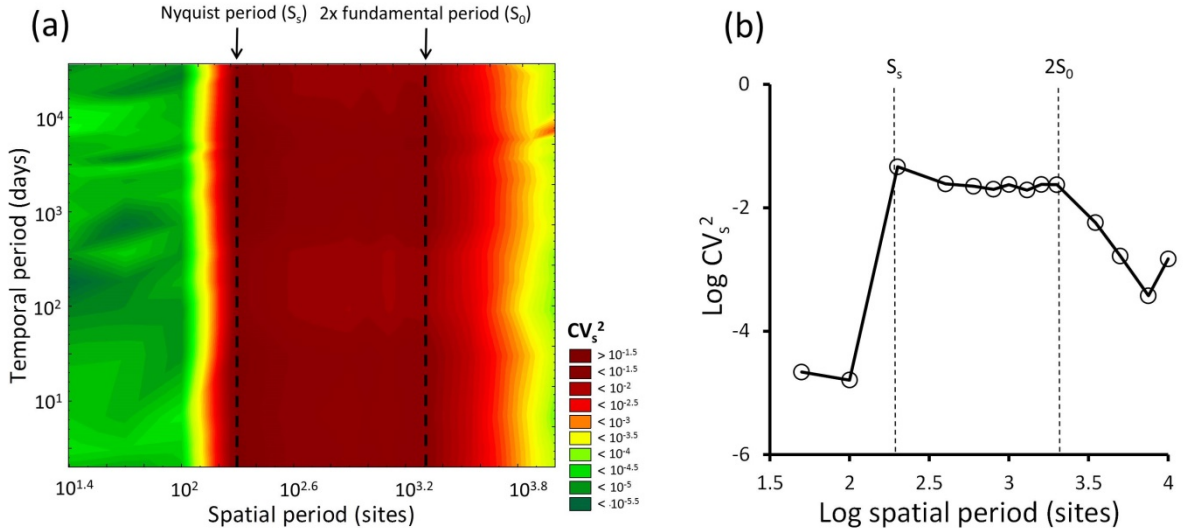

Figure H4. Effect of spatial and temporal scale of variation on spatial variation of simulated pH. (a) Spatial variation of pH, measured as CV_s^2, peaked when it fluctuated at spatial periods less than twice the fundamental period and greater than the Nyquist period. (b) Example of change in spatial variation with increasing spatial period of variation.

Effect of spatial and temporal scale on persistent spatial variation

Because F is a composite of spatial variation and temporal stability, its behavior at different spatial and temporal scales should be a cross between the behaviors of spatial variation and stability. This assertion proved true, with the value of F varying with scale as a blend of the patterns seen in Fig. H1a and H4a. The response of F also fit well with the theoretical expectations of Eq. H5. In particular, Fig. H5 shows that F:

1) was high when spatial variation occurred in the detectable range above the spatial Nyquist period S_s and below the spatial fundamental period S_0, though the empirical upper limit was $2S_0$ and not S_0 as predicted, and

2) was high when temporal variation occurred outside the detectable range, that is below the temporal Nyquist period T_s and above the temporal fundamental period T_0. However, the empirical upper limit was found to be $2T_0$ and not T_0 as predicted.

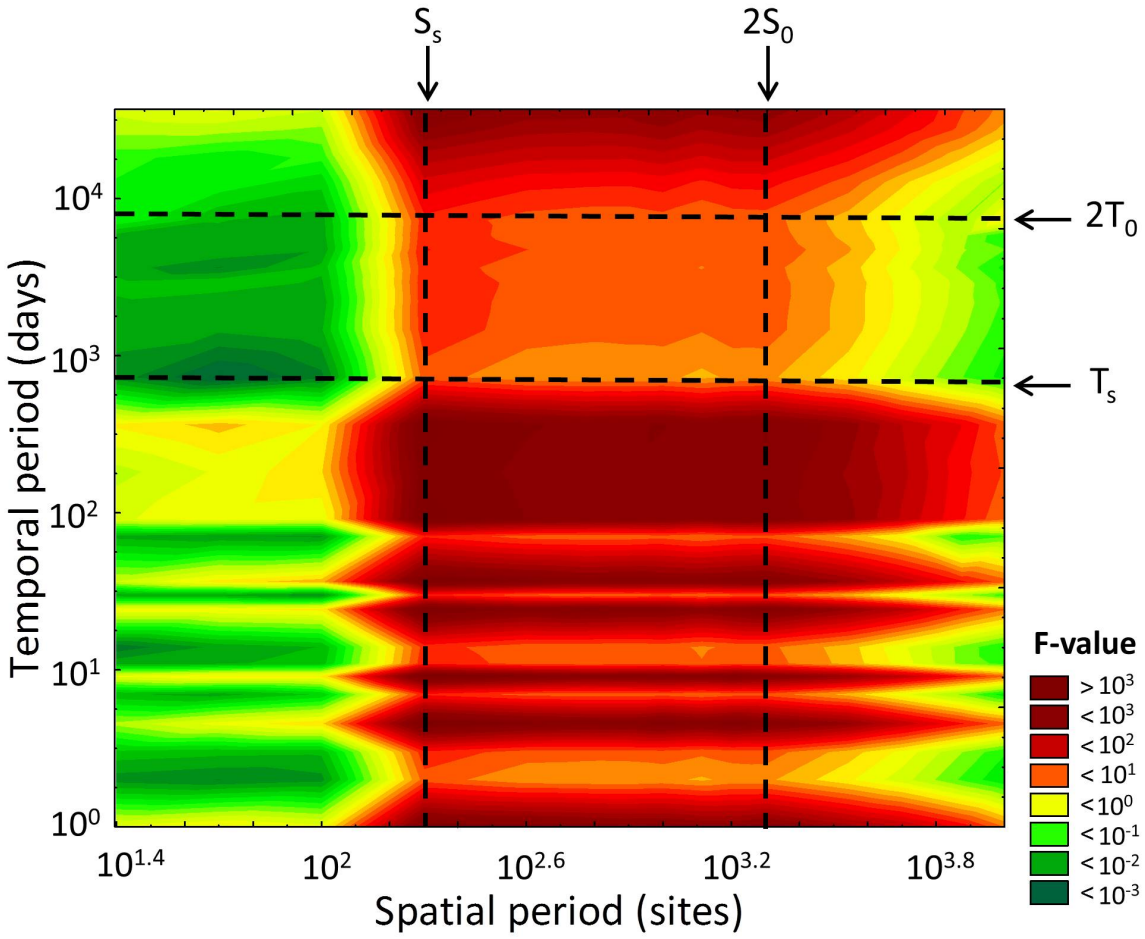

Figure H5. Effect of spatial and temporal scale of variation on PSV of simulated pH. PSV of pH, measured as *F*, peaked at the spatial and temporal periods corresponding to the highest values of its components, stability and spatial variation. See text for details.

General rule for detecting PSV

Theory and simulation results suggest some simple rules for when PSV will be detected by a given sampling regime. Specifically, *F* will be high when a process:

1) has a long temporal period (e.g. buffered physicochemical processes or long-lived organisms) which reduces the amount of temporal variation captured in the sampling period, thereby increasing CV_t^{-1},

2) has a spatial period within the range of detection that is bounded by twice the sampling interval and extent (e.g. intermediate dispersal or flow distance), which enhances the amount of spatial variation measured as CV_s^2, and/or

3) has a temporal period so short (e.g. rapid physicochemical reactions or short-lived organisms) that fluctuations are not detected as meaningful temporal variation and register as

high stability.

We found that the latter, a very short temporal period, can produce high CV_t^{-1} and F at certain harmonics of the Nyquist period (Fig. H1 and H5) likely due to aliasing. Such a signal may or may not be a false-positive for the existence of PSV, depending on the research question asked. For instance, if one were interested in the persistence of summer pH values, then it would be perfectly reasonable to sample only the seasonal peaks or troughs, as in Fig. H3, to assess persistence. But it would not be reasonable to assert that pH values were persistent over any finer a timescale (e.g. days). Similarly, averaging of finer-scale data into annual values, as we did for lakes, provides insight into the inter-annual persistence of pH but these mean values say nothing of persistence at sub-annual timescales. Thus, the term persistent in PSV only has meaning at the scales measured (or averaged), for instance in the recurrence of similar values year upon year, and cannot be easily extrapolated to scales that have not been adequately sampled.

The flipside of this discussion is that only processes with particular spatial and temporal scales will be detected as PSV. From theory and from simulation we found these rough criteria to be:

$$F \xrightarrow{\text{increases}} \begin{cases} S_s > \text{spatial period} < 2S_0 \\ T_s < \text{temporal period} > 2T_0 \end{cases} \quad \text{(H8)}$$

These criteria may be useful for anticipating how much PSV an ecological variable will display given its spatial and temporal variance spectra and details of the sampling protocol. In the absence of measured variance spectra, proxies for scale (e.g. organism lifespan and dispersal distance) might effectively estimate a process's dominant periods of variation.

References

Arrigoni, A. S. et al. 2008. Buffered, lagged, or cooled? Disentangling hyporheic influences on temperature cycles in stream channels. – Water Resour. Res. 44: 1–13.

Denny, M. et al. 2004. Quantifying scale in ecology: lessons from a wave-swept shore. – Ecol. Monogr. 74: 513–532.

Levin, S. A. 1992. The problem of pattern and scale in ecology: The Robert H. MacArthur Award lecture. – Ecology 73: 1943–1967.

Platt, T. and Denman, K. L. 1975. Spectral analysis in ecology. – Annu. Rev. Ecol. Evol. Syst. 6: 189–210.