UC Merced
Frontiers of Biogeography

Title

How perilous are broad-scale correlations with environmental variables?

Permalink

https://escholarship.org/uc/item/9mh596tr

Journal

Frontiers of Biogeography, 12(2)

Authors

Currie, David J.
Pétrin, Cloé
Boucher-Lalonde, Véronique

Publication Date

2020

DOI

10.21425/F5FBG44842

Supplemental Material

https://escholarship.org/uc/item/9mh596tr#supplemental

Copyright Information

Peer reviewed

RESEARCH ARTICLE

How perilous are broad-scale correlations with environmental variables?

David J. Currie[1]* (iD), Cloé Pétrin[1] and Véronique Boucher-Lalonde[1]

[1] Biology Department, University of Ottawa, Ottawa Ont., K1N 6N5, Canada. *Email: dcurrie@uottawa.ca

Abstract

Many studies correlate geographic variation of biotic variables (e.g., species ranges, species richness, etc.) with variation in environmental variables (climate, topography, history). Often, the resulting correlations are interpreted as evidence of causal links. However, both the dependent and independent variables in these analyses are strongly spatially structured. Several studies have suggested that spatially structured variables may be significantly correlated with one another despite the absence of a causal link between them. In this study we ask: if two variables are spatially structured, but causally unrelated, how strong is the expected correlation between them? As a specific example, we consider the correlations between broad-scale variation in gamma species richness and climatic variables. Are these correlations likely to be statistical artefacts? To answer these questions, we randomly generated pseudo-climatic variables that have the same range and spatial autocorrelation as temperature and precipitation in the Americas. We related mammal and bird species richness both to the real and the pseudo-climatic variables. We also observed the correlations among pseudo-climate simulations. Correlations among randomly generated, spatially unstructured, variables are very small. In contrast, the median correlations between spatially structured , but causally unrelated, variables are near $r^2 = 0.1 - 0.3$, and the 95% confidence limits extend to $r^2 = 0.6 - 0.7$. Viewing this as a null expectation, published richness–climate correlations are typically marginally stronger than these values. However, many other published richness–environment correlations would fail this test. Tests of the "predictive ability" of a correlation cannot reliably distinguish correlations due to spatial structure from causal relationships. Our results suggest a three-part update of Tobler's "First Law of Geography": #1) Everything in geography that is spatially structured will be collinear. #2) Near things are more related than distant things. #3) The more strongly spatially structured two variables are, the stronger the collinearity between them will be.

Highlights

- Most variables in biogeographic and macroecological studies are spatially structured. Spatially structured variables are almost always significantly correlated with one another.

- These correlations will often be in the range of $r^2 = 0.1$ to 0.3, but they can be very much stronger.

- If a biotic variable Y is driven by an environmental variable X_1 that is spatially structured, then Y will be significant correlated with nearly any other spatially structured environmental variable $X_2, X_3, \ldots$. Moreover, $X_1, X_2, X_3 \ldots$ are likely to be collinear, purely by chance.

- Spurious correlations resulting from spatially structured variables are likely to be rampant in the biogeographic and macroecological literatures. Other evidence must be adduced before suggesting that a correlation reflects a causal link.

Introduction

The most basic questions in ecology and biogeography are: why do some places on Earth have more species, or higher productivity, or greater densities of individuals, or more herbivory, etc., than other places (Wallace 1876; Rosenzweig 1995; Adams 2009)? A first step toward answering these questions is to ask which characteristics of the environment are most strongly correlated with variation in the biotic variable of interest (Currie 2019). Consequently, the goal of many biogeographic or macroecological studies is to characterize the relationship between some biotic variable Y and environmental variables X_1, X_2, ...:

$$Y = f\left(X_1, X_2, \ldots\right) + \varepsilon \qquad (1)$$

where ε is an error term. Of course, correlations such as these do not imply causation, but correlations are "a recognition of the possible" (Rigler 1982). Correlations invite the researcher to propose and test hypotheses about underlying causal processes. But before suggesting that a correlation may reflect a causal link, it makes sense first to ask what correlations would arise in the absence of causal links (Gotelli and Ulrich 2012).

In the present study, we focus on correlations between broad-scale variation in species richness and environmental variables. The strongest of these correlations ($r^2 > 0.8$) are typically with climatic variables: heat, water, and/or an interaction between the two (Currie 1991; Wright et al. 1993; Field et al. 2009). Other richness–environment correlations are comparatively weak ($0.0 < r^2 < 0.1$), but still statistically significant when sample sizes are large (e.g., Fig. 6.1 in Wright et al. 1993).

Richness–environment correlations can arise in several different ways. Variation in environmental variable X_1 may *cause* variation in richness. Alternatively, a correlation between richness and X_1 may be indirect: some other variable X_2 causes the variation in richness, and X_1 happens to be collinear with X_2. For example, continental-scale variation in species richness is strongly correlated with contemporary climate (Field et al. 2009), as well as with climate during the Last Glacial Maximum (Hortal et al. 2011), and with time since the retreat of the last glaciers (Hawkins and Porter 2003). These three environmental variables are strongly collinear (Hawkins and Porter 2003). Not surprisingly, richness correlates with them all. It is unlikely that they all represent causal links.

Further, variables that are spatially structured have an elevated probability of being correlated with one another, even in the absence of any causal link (Lennon 2000). For example, geographic ranges of individual species are often correlated with environmental variables (Kearney and Porter 2009; Araújo et al. 2013). Yet, randomization studies have shown that species' ranges sometimes correlate equally strongly with randomly generated, spatially structured independent variables (Chapman 2010; Bahn and McGill 2013; Fourcade et al. 2018). This led Beale et al. (2008) to conclude dramatically that

"opening the climate envelope reveals no macroscale association with climate in European birds". This may be a very general problem, considering that much of biogeography and macroecology is built upon correlations between spatially structured biological and environmental variables (Lennon 2000).

Ecological studies most commonly deal with spatially structured dependent variables by partitioning the error term ε in eq. 1 into a spatially structured component and a random component (Section 5.1.3 in Fortin and Dale 2005). This makes sense if endogenous population processes such as dispersal, territoriality, etc., cause neighbouring samples to be more similar (or sometimes more dissimilar) than would be expected, given the environmental conditions (Legendre 1993). Fortin and Dale (2005) call this non-independence in the residuals of statistical models "inherent autocorrelation". Inherent autocorrelation causes hypothesis tests to be too liberal because the individual observations do not each contribute an independent degree of freedom (Clifford et al. 1989). An extensive literature discusses statistical models that include, and thereby control for, spatially autocorrelated residual error (Dutilleul et al. 1993, Legendre 1993, Dale and Fortin 2002, Fortin and Dale 2005, Dormann 2007, Dormann et al. 2007, Kühn 2007, Bini et al. 2009, Diniz-Filho et al. 2009, Beale et al. 2010, Peres-Neto and Legendre 2010).

However, spatial structure in biotic variable Y can also arise if the driving (independent) variables X_1, X_2, ..., are spatially structured. Fortin and Dale (2005) call this "induced spatial dependence". Often, the goal of a biogeographic or macroecological study is to characterize the shape and the strength of the relationship the relationship between Y and X_1, X_2,... . If the relationship is modelled with an autocorrelated error term, that term will likely be collinear with X_1, X_2, ... It will therefore capture, and control for, some part of the induced spatial dependence, which was the focus of the study in the first place. Using an autocorrelated error term affects parameter estimates and the apparent strength of the biology–environment relationship in idiosyncratic, method-dependent ways (Kühn 2007; Bini et al. 2009), depending on the strength of the collinearity between the autocorrelated error term and X_1, X_2, ... When data are spatially structured, significance tests are more conservative in models that include an autocorrelated error term than in models without that term. However, when statistical power is high (which is often the case in biogeographic/macroecological studies), significance tests are not particularly informative. In sum, models that include an autocorrelated error term do not appear to be appropriate when the independent variables are spatially structured.

Rather than controlling for spatial autocorrelation, the present study asks: how strong are correlations between spatially structured variables expected to be when there is no causal link between the two? How does this compare to observed correlations between broad-scale variation in species richness and environmental variables? Since the effect of inherent autocorrelation becomes small when the number of sites is large (Dutilleul et al. 1993), we shall ignore inherent autocorrelation in species richness here.

To answer whether broad-scale variation in species richness and environmental variables are stronger than predicted from induced spatial dependence alone, we used a randomization study, as advocated by Lennon (2000). We focused on temperature and precipitation because the strongest richness–environment correlations reported in the literature often involve those two variables (Wright et al. 1993). We were interested to know how strong correlations are likely to be in practice; we therefore used a real geographic domain: the Americas (whereas Lennon 2000 used a 32 x 32 quadrat matrix). We used the algorithm of Chapman (2010) to generate simulated climatic variables that have the same spatial autocorrelation structure as temperature and precipitation in the Americas. We shall refer to these as pseudo-temperature and pseudo-precipitation. We then estimated bird species richness using the species' ranges reported by Birdlife International and mammal richness from ranges reported on NatureServe. We related richness to observed temperature and precipitation, and to pseudo-temperature and pseudo-precipitation. We compared the strength of these relationships. Finally, we tested how well models using real climate data in North America predict the spatial variation of richness in an independent area: South America. We compare this to richness predicted using pseudo-climate data.

Methods

Study area

The study area is North and South America, divided at the Darien Gap of Panama. We superimposed a grid of 100 km x 100 km quadrats over this area. Because richness is a strongly non-linear function of area, we excluded quadrats that were >50% water. Off-shore islands were also excluded. This yielded 1978 quadrats in North America and 1764 in South America.

Data

Mean annual temperature, total annual precipitation, and elevation were taken from WorldClim (Hijmans et al. 2005) Version 1.4 at a resolution of 30 arc-seconds. The data were resampled to the grid system, and the mean value for each quadrat was extracted. The frequency distribution of precipitation (mm/year) is strongly positively skewed. We therefore cube-root transformed precipitation to yield an approximately normally distributed independent variable in order to reduce the weight of extreme values. Temperature (K) was untransformed.

Ranges of the native bird species of the Americas were obtained from Birdlife International[1]. Ranges of the native mammals of the Americas were taken from NatureServe (Patterson et al. 2007). Richness was determined by superimposing ranges on the grid system. We counted a species as present in any quadrat into which its range extended, partially or entirely. Richness represents a tally of all the presences in a given quadrat. The frequency distribution of richness is also strongly positively skewed; a logarithmic transformation yielded an approximately normal distribution and improved residuals in regression models.

Pseudo-climate

We created pseudo-temperature and pseudo-precipitation datasets as follows. We used the algorithm presented by Chapman (2010). Broadly, the algorithm first models the spatial correlation structure in a real data set. It then generates a stationary Gaussian field with the same spatial structure. Chapman's algorithm assumes that autocorrelation is homogeneous across the study area, except near the coasts, where there is a buffering effect of the ocean. We rescaled the resulting pseudo-temperature surface to have the same range of values as the real temperature data. This yields temperature surfaces that resemble a real temperature map, except that the extremes of temperature were not generally in the expected places and the gradients were not necessarily in the expected directions (e.g., Fig. 1). We carried out the same process for precipitation, and we repeated the simulations 1000 times.
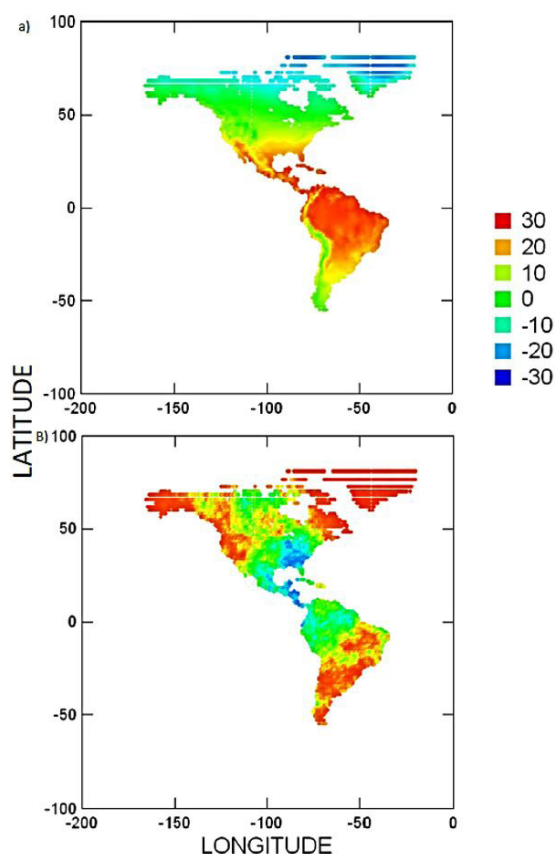

Figure 1. Mean annual temperature (°C) observed in North America (top panel) and an example of a simulation of pseudo-temperature produced using Chapman's (2010) algorithm, which maintains the same degree of spatial autocorrelation, but the locations of minima and maxima are random (bottom panel).

1 https://www.birdlife.org/, accessed on 30/11/2014

To isolate the effect of spatial structure in the pseudo-climatic variables, we also fully randomized each pseudo-climate variable, destroying its spatial structure. We related species richness to both the spatially structured pseudo-climate variables and to the fully randomized variables.

We also calculated the pairwise correlations between environmental variables. These correlations reflect the effect of induced spatial structure in the absence of any inherent spatial autocorrelation (which may be present in species richness).

Richness model

Next, we modeled richness as functions of the environmental variables in North America, using the following model and its subsets:

$$log_{10}\left(SR + 0.5\right) = c_0 + c_1 * T +$$
$$c_2 * P^{1/3} + c_3 *\left(T * P^{1/3}\right) + \varepsilon \qquad (2)$$

where SR= species richness, T=mean annual temperature, P= total annual precipitation, and ε is a normally distributed error term. Regression coefficients are represented by c_0, c_1, c_2, and c_3. We fitted this model using the real climate data (T and/or $P^{1/3}$) as well as each set of pseudo-temperature ($\mathcal{T}$) and/or pseudo-precipitation ($\mathcal{P}^{1/3}$), and we noted the coefficients of determination (R^2) for each fit.

The fits to the 1000 pseudo-climate datasets yield a distribution of the R^2 between richness and spatially structured environmental variables in the absence of any causal links among them. We then compared the R^2 using the real environmental data to the distribution generated using pseudo-environmental variables. We could have examined other models (e.g., models with higher-order polynomial terms or other environmental variables), but simple models suffice to evaluate the effect of spatial dependence in the dependent and independent variables.

Test of predictive ability

For any fitted model, R^2 provides a measure of explained variance within a dataset but not a measure of the model's predictive accuracy. We tested the predictive capacity of each of the 1000 models fitted to the North American pseudo-climate data, and the 1 model fitted to the real data. We did this in two ways. First, we examined within-sample predictive capacity using a jackknife technique. Using the North American data, we excluded one site, and we calculated a regression model with the remaining data. We then used that model to predict richness at the hold-out site. We repeated this process for each observation in the data set.

Second, we tested out-of-sample predictive ability. For each data set, we fitted a model in North America, and we used the environmental data from South America to predict richness in each quadrat in South America. We carried out this procedure using both the

real climatic data and pseudo-climate. We restricted the test cases in South America to quadrats whose temperature and precipitation both fall within the ranges of those variables in North America.

Residual effects of pseudo-variables

We also tested whether the residual variation in species richness, after controlling for the real environmental variables, is related to pseudo-temperature ($\mathcal{T}$) or pseudo-precipitation ($\mathcal{P}^{1/3}$). To do this, using the North American data, we fitted each of the following subsets of the original model, adding pseudo-temperature:

$$log_{10}\left(SR + 0.5\right) = c_0 + c_1 * T + c_2 * \mathcal{T} + \varepsilon$$

$$log_{10}\left(SR + 0.5\right) = c_0 + c_1 * P^{1/3} + c_2 * \mathcal{T} + \varepsilon$$

$$log_{10}\left(SR + 0.5\right) = c_0 + c_1 * T + c_2 * P^{1/3} +$$
$$c_3 *\left(T * P^{1/3}\right) + c_4 * \mathcal{T} + \varepsilon \qquad (3)$$

We then repeated these analyses, substituting $\mathcal{P}^{1/3}$ for $\mathcal{T}$.

All statistics were done using R version 3.4.1, with the exception of confidence intervals around coefficients of determination, which were calculated using VassarStats[2].

Results

How well the algorithm works

Chapman's (2010) algorithm yielded spatially autocorrelated pseudo-environmental variables that vary spatially in ways that resemble the real temperature and precipitation data, but whose gradients appear slightly less regular (Figure 1). Chapman's algorithm assumes isotropic autocorrelation (i.e., equally strong in all directions), whereas the Earth's geometry and rotation most strongly constrain temperature in the N–S direction and precipitation in the E–W direction (in the Americas). Chapman's algorithm also does not prevent variables from having multiple minima or maxima along any particular transect, whereas this is uncommon in real climatic data. The pseudo-climate variables have the same mean spatial autocorrelation as the real data (by design), but the autocorrelation in the simulated data is slightly stronger at short distances, and slightly weaker at long distances, than in the real data (Figure S1).

Correlations among simulated environmental data

Spatially structured variables (here, pseudo-temperature or pseudo-precipitation) tend to be collinear with one another (Table 1), despite the absence of any causal connection. When simulated environmental data *are not* spatially structured (because their values have been randomized through space), the pairwise correlations among simulations of the pseudo-climatic data are very small (Table 1). When randomly generated variables *are*

Table 1. Pearson correlations between pairwise combinations of replicate simulations of pseudo-temperature (which has the same Moran's I as observed temperature across the Americas, n=1000), pseudo-precipitation (which has the same Moran's I as observed precipitation, n=1000), and observed temperature and precipitation. Note that temperature is more strongly spatially structured than precipitation (Fig. 2). The expected r values are all near 0, but most correlations involving spatially structured data are much stronger (positively or negatively). Consequently, expected r^2 values are nearly always significantly greater than 0. The number of pairwise correlations used to calculate the quantiles is given by n. In all cases, the underlying data consisted of observations of climate or pseudo-climate in 3744 quadrats across the Americas.

Data	statistic	2.5% quantile	Median	97.5% quantile	n
Pseudo-temperature x pseudo-temperature, randomized through space	r	-0.003	0.000	0.003	499500
	r^2	0.0000	0.0001	0.0001	499500
Pseudo-temperature x pseudo-temperature, (spatially structured)	r	-0.800	-0.009	0.854	499500
	r^2	0.000	0.183	0.750	499500
Pseudo-precipitation x pseudo-precipitation (spatially structured)	r	-0.423	-0.006	0.461	499500
	r^2	0.000	0.026	0.244	499500
Pseudo-temperature x pseudo-precipitation (spatially structured)	r	-0.616	0.003	0.576	1000000
	r^2	0.000	0.057	0.423	1000000
Observed temperature x pseudo-temperature	r	-0.770	0.058	0.830	1000
	r^2	0.000	0.208	0.702	1000
Observed temperature x pseudo-precipitation	r	-0.606	-0.014	0.568	1000
	r^2	0.000	0.065	0.422	1000
Observed precipitation x pseudo-temperature	r	-0.682	0.045	0.617	1000
	r^2	0.000	0.135	0.465	1000
Observed precipitation x pseudo-precipitation	r	-0.482	-0.030	0.529	1000
	r^2	0.000	0.046	0.298	1000

spatially structured, the expected correlation among them is still r≈0. However, the variation is huge, with some r^2>0.7 (Figure 2). Among these correlations, 96.2% are statistically significant at α=0.05 (statistical power is high: n=1000). Similarly, 89.5% of the pairwise correlations among the pseudo-precipitation simulations are significant at α=0.05. It is important to note that the median coefficient of determination between replicate, causally unrelated, but spatially structured, pseudo-temperature simulations is r^2=0.183. We will return to this point below.

Spatially structured variables need not share the same spatial structure to be strongly collinear (Figure 2). Of the pairwise pseudo-temperature – pseudo-precipitation correlations, 94.0% are statistically significant at α=0.05. The strength of collinearities between causally unrelated variables appears to depend upon how strongly spatially structured the two variables are, not how similarly structured the variables are (Table 1). Pseudo-temperature ($\mathcal{T}$) is more strongly spatially structured than pseudo-precipitation ($\mathcal{P}$). We observed that median $\mathcal{T}$-$\mathcal{T}$ collinearity (i.e., between pairs of $\mathcal{T}$ simulations) > $\mathcal{T}$-$\mathcal{P}$ collinearity > $\mathcal{P}$-$\mathcal{P}$ collinearity (Table 1, Figure 2). These results are qualitatively similar to those of Lennon (2000).

Correlations among pseudo-climate variables are relatively insensitive to the spatial extent of the data, at least over continental to hemispheric extents (Table 2). If anything, the correlations were slightly stronger over smaller spatial extents.

Correlations between simulated and observed environmental data

Observed temperature and precipitation are spatially structured. Not surprisingly, nearly all $\mathcal{T}$ and $\mathcal{P}$ simulations were collinear with both observed temperature and precipitation (Table 1 and Figure 3).

Richness correlations with real environmental data

The observed variation of species richness across the Americas is strongly related to climate. A simple linear regression of $\log_{10}$(richness) as a function of temperature statistically accounts for 68% and 76% of the spatial variation for mammals and birds, respectively, in the Americas (Figure 4). Multiple regressions that include both temperature, precipitation, and their interaction increase explained variation by another ~3%-7% (Table 3). These results are similar to other published results (Wright et al. 1993; Field et al. 2009). We will return to the differences between continents (Figure 4) below.

Richness correlations with simulated environmental data

Richness is also related to the simulated pseudo-environmental variables, even though there is (by design) no causal link between the two. The median correlations between richness and the pseudo-variables are much weaker than the correlations with the real environmental variables (Table 3): $r^2=0.15$ to $r^2=0.31$ for $\log_{10}$(richness) and pseudo-temperature, and $r^2=0.06$ to $r^2=0.07$ for pseudo-precipitation. However, some
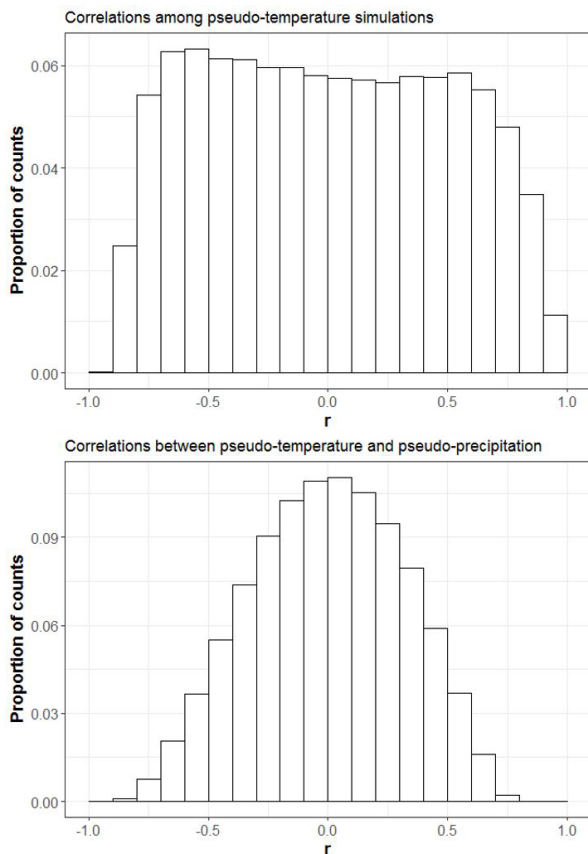

Figure 2. Top: The frequency distribution of pairwise Pearson correlations between randomly generated pseudo-temperature gradients across the Americas. All simulations have the same spatial autocorrelation. Median r=-0.009, 95% c.i.= [-0.80 − +0.85]. n=1000 simulations, yielding 499500 paired comparisons. Of these, 96.2% are statistically significant at α=0.05. Bottom: Pearson correlations between pseudo-temperature and pseudo-precipitation, which have different spatial structures. Median r=-0.003, 95% c.i=[-0.62 − +0.58]
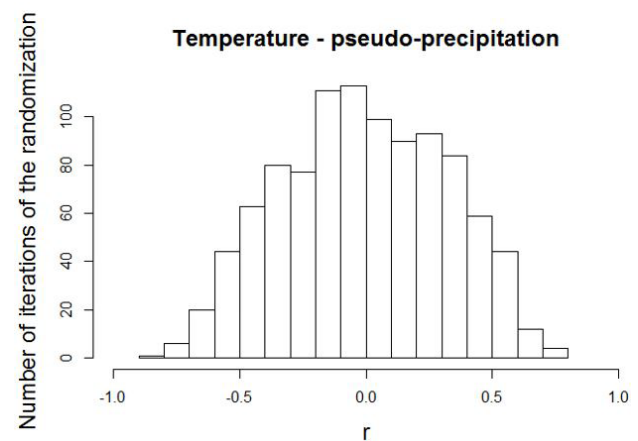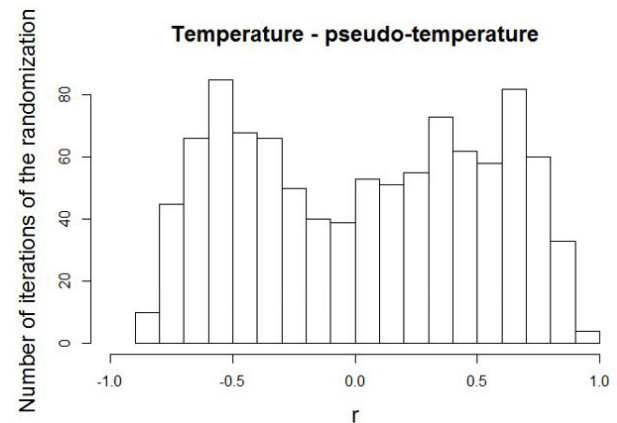
Figure 3. The correlations between observed temperature and 1000 simulations of pseudo-temperature (top panel) and of pseudo-precipitation (bottom panel). The only way in which these two sets of data differ is the strength of the spatial autocorrelation in the two simulated variables, which influences their tendency to be collinear with observed temperature.

Table 2. Does the correlation between spatially structured variables depend upon the spatial extent of the data? The underlying data represent 499500 pairwise comparisons between 1000 randomly generated, but spatially structured, pseudo-temperature data surfaces.

Data	statistic	2.5% quantile	Median	97.5% quantile	N quadrats
North and South America	r	-0.800	-0.009	0.854	3742
	r^2	0.000	0.183	0.750	3742
North America	r	-0.849	-0.005	0.982	1978
	r^2	0.001	0.207	0.826	1978
South America	r	-0.869	-0.013	0.910	1764
	r^2	0.001	0.244	0.855	1764

Mammal richness =f(Temperature in the Americas)
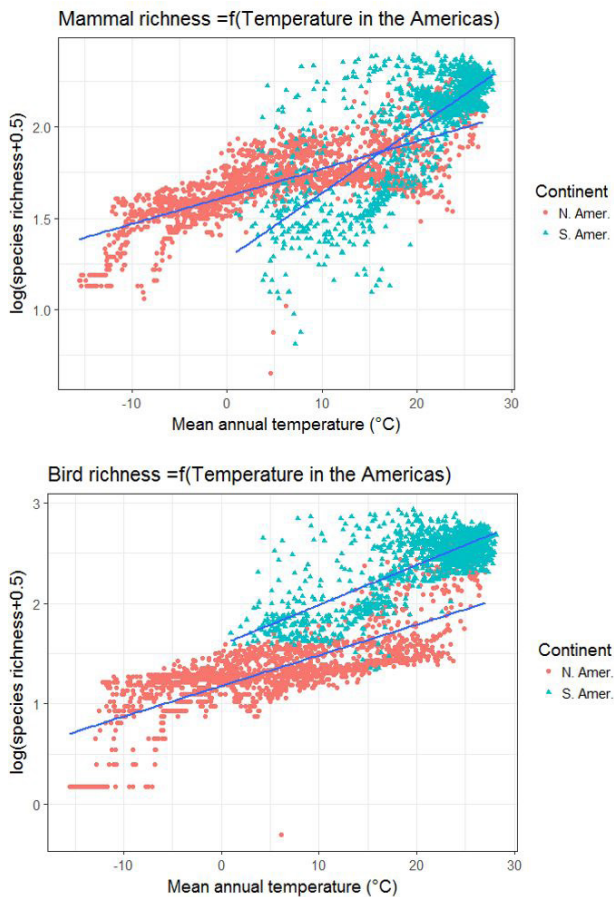
Bird richness =f(Temperature in the Americas)

Figure 4. Log$_{10}$ mammal (top panel) and bird (bottom panel) species richness in North and South America as linear functions of temperature. The lines represent th e lines of best fit of simple linear regressions in the two subsets of the data.

simulations yielded pseudo-climates that have much stronger correlations with richness. The one-tailed non-parametric 95% confidence limits (obtained from the 1000 simulations) included simple linear correlations between observed richness and pseudo-temperature as strong as $r^2=0.70$, and multiple-correlations as strong as $R^2=0.74$ (Table 3 and Figure 5).

The richness – pseudo-climate coefficient of determination has three components (Figure 6). First, as statistical theory predicts, there is usually some small, but non-zero correlation between richness and fully randomized pseudo-climate variables (represented by the + in the centre of Figure 6). Approximately 5% of the correlations with fully randomized pseudo-temperature were significant at α=0.05 (n=1978), as theory predicts.

Second, as shown in Figure 2, spatially structured variables tend to be collinear with one another, even with no causal link between them. Consequently, 96% of the pairwise correlations between richness and pseudo-temperature are statistically significant (p≤0.05). The more strongly pseudo-temperature or pseudo-precipitation is collinear with observed temperature, the more strongly richness is correlated with that pseudo-climate variable(Figure 6, also Figure S2). In contrast, richness is not strongly correlated with pseudo-climate variables that are collinear with pseudo-precipitation (Figure S2).

Third, even when pseudo-temperature is not collinear with temperature, it shows an elevated probability of being correlated with richness by chance. When pseudo-temperature was not collinear with true temperature (viz., the points directly above and below the + in the centre of Figure 6), correlations between richness and pseudo-temperature ranged between -0.3 < r < +0.3. This correlation represents

Table 3. The shaded lines in the table show the observed coefficient of determination (r^2 or R^2) in simple regressions of log$_{10}$(species richness) as a function of mean annual temperature (T), or of the cube root of total annual precipitation ($P^{1/3}$). Some multiple regressions also include the interaction T* $P^{1/3}$. n=3742 quadrats across North and South America. Unshaded lines shows the same models using simulated environmental variables — pseudo-temperature ($\mathcal{T}$) and pseudo-precipitation ($\mathcal{P}$) — whose ranges and spatial autocorrelation are the same as those of the real T and P data. N=1000 simulations. Values in parentheses are two-tailed, parametric 95% confidence intervals around the r^2 or R^2 values. Values in square brackets represent the two-tailed 95% confidence interval estimated from the repeated simulations.

Model: log$_{10}$(species richness) ~	Climate data	r^2 or R^2 Mammals	r^2 or R^2 Birds
T	Observed	0.675 (0.658 - 0.691)	0.781 (0.769 - 0.792)
$\mathcal{T}$	Simulated	0.146 [0.001 - 0.594]	0.309 [0.001 - 0.736]
T + $\mathcal{T}$	Both	0.682 [0.675 - 0.748]	0.814 [0.781-0.879]
T + $\mathcal{P}^{1/3}$	Both	0.682 [0.675 − 0.736]	0.790 [0.781-0.847]
$P^{1/3}$	Observed	0.485 (0.462-0.507)	0.478 (0.454-0.501)
$\mathcal{P}^{1/3}$	Simulated	0.056 [0.000-0.365]	0.074 [0.000-0.492]
$P^{1/3}$ + $\mathcal{T}$	Both	0.518 [0.486 − 0.687]	0.607 [0.478-0.831]
$P^{1/3}$ + $\mathcal{P}^{1/3}$	Both	0.501 [0.486 − 0.607]	0.507 [0.478-0.677]
T + $P^{1/3}$+ T* $P^{1/3}$	Observed	0.741 (0.727 - 0.753)	0.815 (0.804 - 0.824)
$\mathcal{T}$+ $\mathcal{P}^{1/3}$ + $\mathcal{T}$* $\mathcal{P}^{1/3}$	Simulated	0.282 [0.028 - 0.646]	0.434 [0.044 - 0.779]
T + $P^{1/3}$+ T* $P^{1/3}$ + $\mathcal{T}$	Both	0.746 [0.741 - 0.790]	0.844 [0.815-0.905]
T + $P^{1/3}$+ T* $P^{1/3}$ + $\mathcal{P}^{1/3}$	Both	0.745 [0.741 − 0.780]	0.822 [0.815-0.865]

the effect of spatial structure *per se*, plus (potentially) collinearity with other environmental drivers that are not collinear with temperature.

Simulated environmental data in multiple regressions

Adding a spatially structured pseudo-environmental variable to a regression nearly always increases explained variance. On average, addition of pseudo-temperature or pseudo-precipitation to a simple regression of $\log_{10}$(richness) as a function of temperature, or to a multiple regression in which $\log_{10}$(richness) is a function of temperature + precipitation$^{1/3}$ + temperature*precipitation$^{1/3}$, increased the explained variance by ~3% (Table 2). This would be significant at α=0.05 in any regression with >91 observations.

In practical terms, this means that one should expect that *any* independent variable that is spatially structured will be statistically significant in a simple regression, or in a multiple regression, where the dependent variable is spatially structured, provided that there is at least moderate statistical power.
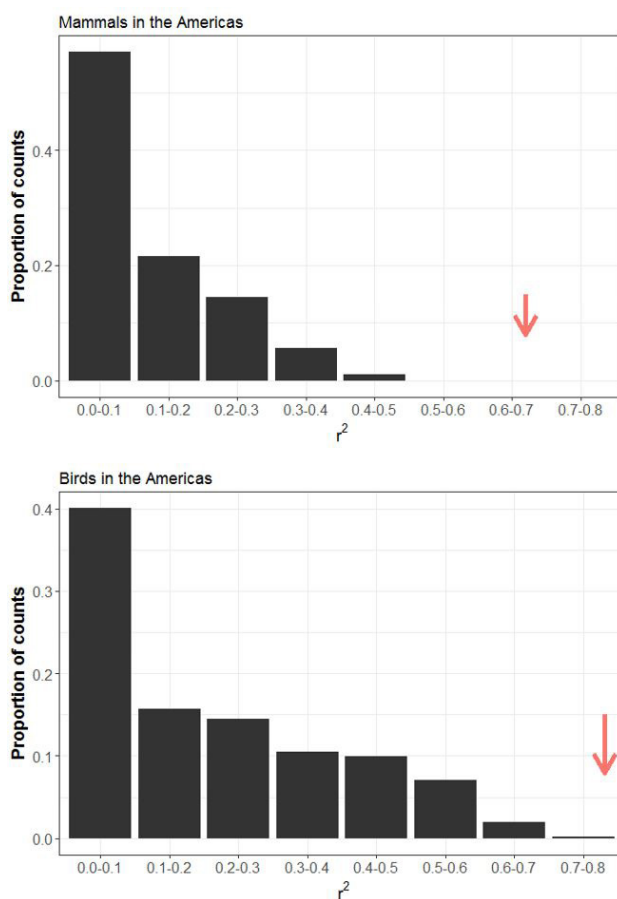

Figure 5. The frequency distributions of observed coefficients of determination (r^2) of $\log_{10}$(species richness) as a linear function of pseudo-temperature in 1000 simulations for mammals (upper panel) and birds (lower panel). Pseudo-temperature has the same degree of spatial autocorrelation as observed temperature. The correlations using the real climatic data are indicated by the red arrows.

Does within-sample prediction accuracy unmask spurious correlations between spatially structured variables?

If a statistical model captures the true driver(s) of a dependent variable, then the model should accurately predict new instances of the dependent variable. Many published studies test the "predictive accuracy" of a model by randomly dividing a dataset into two subsets. One subset is used to fit the statistical model, and the other serves as the "test" data set. However, random assignment of data to two groups means that the error distribution and the spatial structure, are the same in the training and the test data sets. This suggests that within-sample, hold-out tests of predictive accuracy
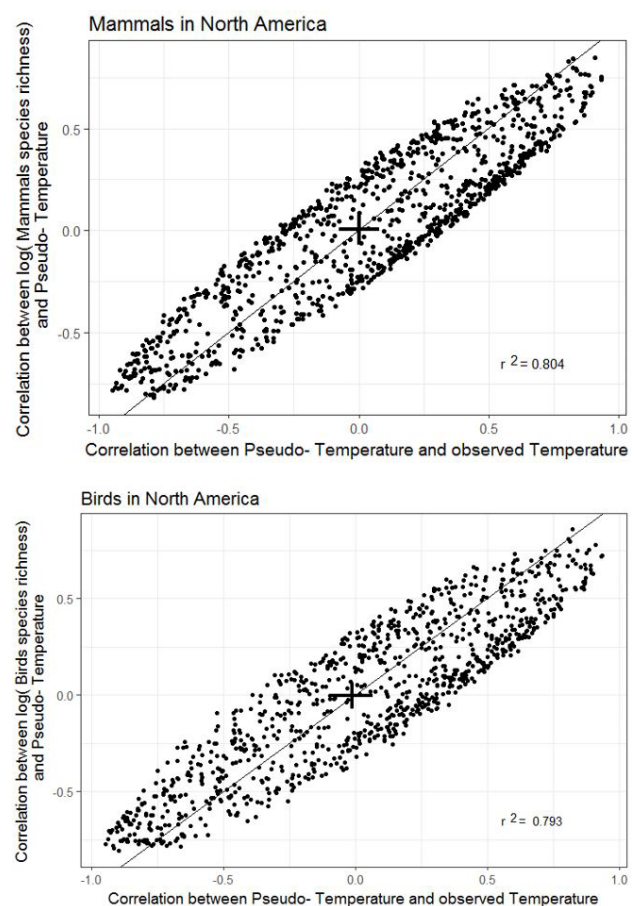

Figure 6. The correlations between $\log_{10}$(species richness) and pseudo-temperature, shown as a function of the correlation between pseudo-temperature and observed temperature. The data are from North America. Top panel: mammals; bottom panel: birds. The + in the centre of the cloud of points represents the range of correlations observed in completely randomized data (1000 sets of pseudo-climate data; its non-parametric 95% confidence interval is half of its total range). Thus, the + represents the range of correlations that occur by chance in spatially unstructured data. All of the points more extreme than these bars show correlations that result from the spatial structure in the two variables. The diagonal line represents the 1:1 relationship.

should be equivalent to measures of explained and residual variance.

When we examined the "predictive accuracy" of the log(richness) – temperature relationship in North America using a jackknife procedure, we found that the relationship between observed and predicted richness had a slope of 1 and an intercept of 0, as expected (Figure 7). Moreover, the shape and the r^2 of the observed–predicted relationship were very close to the fit of the original model (cf. Figures 4 and 7). A within-sample, hold-one-out procedure using real data tells nothing that the standard statistics of the original model (F and r^2) did not already tell.

Using pseudo-climatic data, we found the same result (Figure 8). Even though the relationship between richness and pseudo-climate is entirely spurious and relatively weak, within-sample predictions mirror the original fit. When the richness– pseudo-climate relationship happens to be strong (even though it is entirely spurious), within-sample predictive accuracy is high.
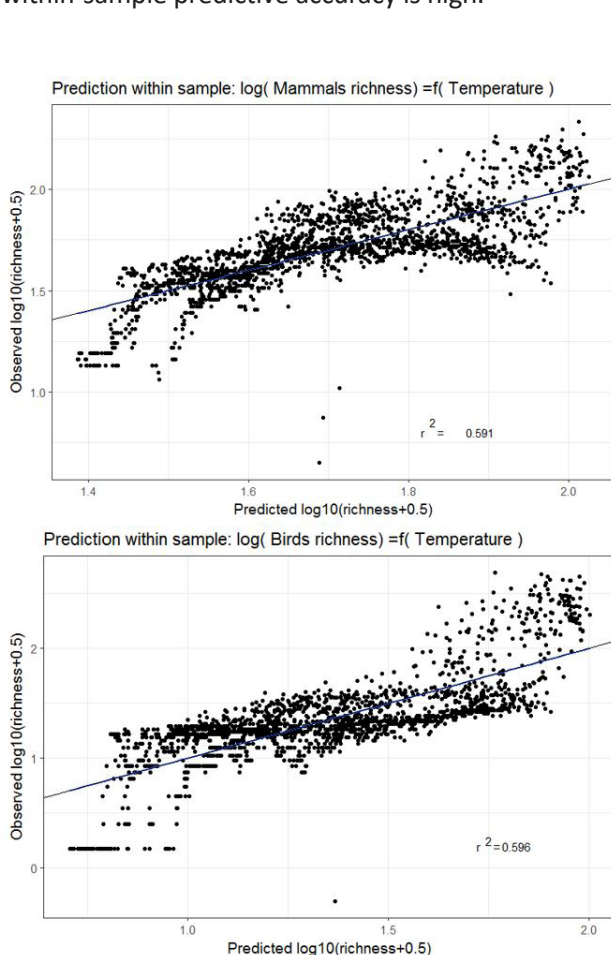
Does out-of-sample prediction accuracy unmask spurious relationships?

A stronger test for the hypothesis that a statistical model captures the true driver of a dependent variable (as opposed to a spurious correlation due to spatial autocorrelation) is that the model should also accurately predict the dependent variable in geographic regions outside the region used to calibrate the original model. Because environmental variables are unlikely to be spatially structured in the same way in different regions, out-of-sample prediction should fail using the pseudo-climate data. In contrast, if there is a causal link between richness and climate, out-of-sample prediction should work using the real data.

Using the real data (Figure S3), we found that observed richness in South America is strongly correlated with richness predicted from the richness–temperature
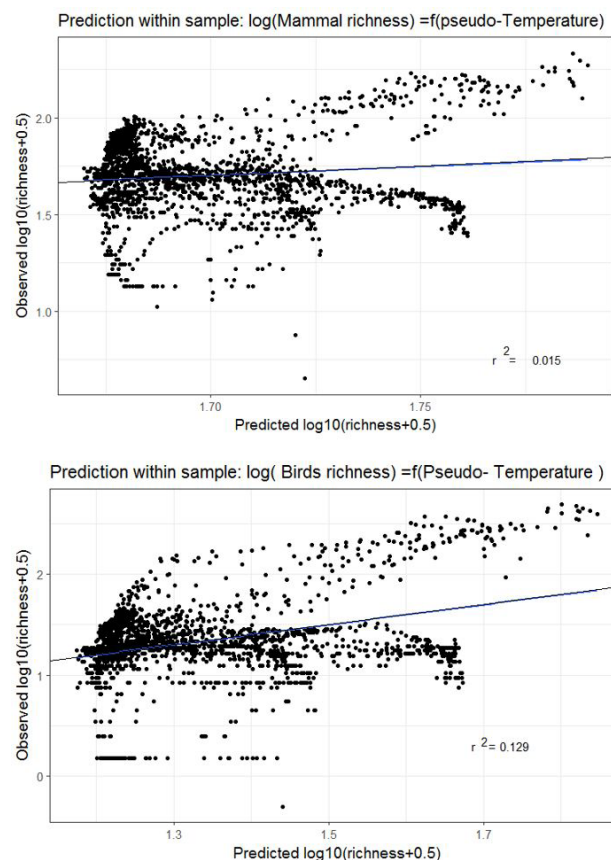

Figure 7. Test of the *within-sample* predictive ability of a regression model using a jack-knife procedure: using the North American data, hold one observation out, fit the model, and predict the value of the hold-out. Model: log(species richness) as a linear function of observed temperature. Top panel: mammals; bottom panel: birds. The black line represents the predicted 1:1 relationship. The blue line is a fitted linear relationship. The two lines are nearly exactly superimposed.

Figure 8. Test of the *within-sample* predictive ability of a regression model using a jack-knife procedure: using the North American data, hold one observation out, fit the model, and predict the value of the hold-out. Model: log (species richness) as a linear function of one simulation of pseudo-temperature. In this particular simulation of pseudo-temperature, temperature and pseudo-temperature were positively collinear (r=0.677). Top panel: mammals; bottom panel: birds. The black line represents the predicted 1:1 relationship. The blue line is a fitted linear relationship. The two lines are nearly exactly superimposed.

relationship fitted using the North America data. This is true for both mammals and birds. Moreover, these relationships are nearly as strong as the observed richness–temperature correlations in North America. However, for both taxonomic groups, the slopes and intercepts of the observed–predicted relationship differed from the expected values of 1.0 and 0.0. This difference, which is also evident in the original scatterplots (Figure 4), indicates that the linear function of temperature alone is an insufficient explanation of the richness patterns across the Americas, despite its impressive predictive power. Something is missing from the model.

Out-of-sample prediction using the pseudo-climatic data can yield quite variable results. Consider, first, a simulation in which — by chance — pseudo-temperature was weakly negatively collinear with observed temperature in North America (r=-0.17), but strongly positively collinear in South America (r=+0.68). The North American model predicts that richness should also be weakly negatively correlated with pseudo-temperature in South America. Instead, richness in South America is positively correlated with pseudo-temperature. Consequently, observed richness in South America is *negatively* related to predicted richness (Figure 9). This would lead one (correctly) to reject the hypothesis that the correlation reflects a causal relationship.

Consider, now, a second pseudo-temperature dataset in which temperature and pseudo-temperature are weakly positively correlated on both continents (r=+0.198 and +0.172). Observed richness in South America was essentially unrelated to richness predicted from the North American model (r=0.006 for birds, r=0.000 for mammals; not shown). Again, the lack of relationship between predicted and observed richness in the test region would lead one (correctly) to reject the hypothesis that the fitted linear model reflects a causal link.

Finally, consider, a third set of pseudo-climate data in which temperature and pseudo-temperature happened to be strongly positively collinear in both the calibration and test data sets (r=+0.604 and +0.791). In this case, out-of-sample prediction of richness is surprisingly strong (r^2=0.71; Figure S4) even though there is no causal relationship. The slope and intercept again differ significantly from the expected values of 1.0 and 0.0. However, the strong out-of-sample predictive ability would probably be viewed as "support" for a causal link between richness and this pseudo-climatic variable even though (by design) there was no causal link.

Let us summarize this section. Within-sample tests of predictive ability say nothing beyond the original measures of goodness of fit. Unsuccessful out-of-sample prediction can reject the hypothesis of causation. Successful out-of-sample prediction means that the hypothesis of a causal link has survived a strong test. Evidence that is consistent with a hypothesis is often viewed as "supporting" the hypothesis, but one cannot infer that the causal hypothesis is true. Popper and Miller (1983) pointed out (somewhat counter-intuitively) that evidence that is consistent with a hypothesis does not increase the probability that the hypothesis is true. Our results illustrate why: a hypothesis may survive a test, despite being false.
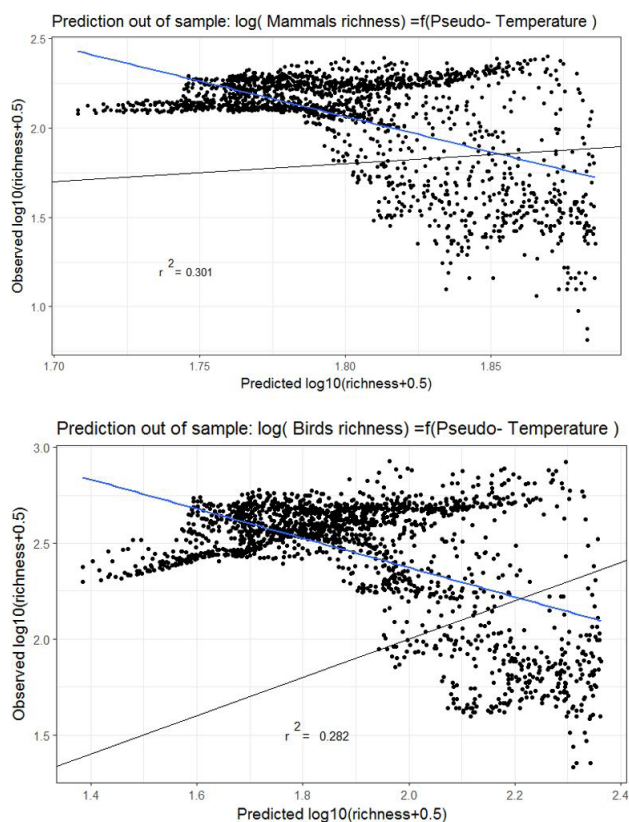

Figure 9. Test of the *out-of-sample* predictive ability of a regression model (as in Figure S3), except that the independent variable here is pseudo-temperature. In this particular simulation of pseudo-temperature, real temperature and pseudo-temperature are weakly negatively collinear in North America (r=-0.166) and strongly positively collinear in South America (r=+0.677). Top panel: mammals; bottom panel: birds. The black line represents the predicted 1:1 relationship. The blue line is a fitted linear relationship.

Discussion

We began with two questions: how strongly correlated are causally unrelated, spatially structured, variables likely to be? Are correlations between species richness and environmental variables likely to be artefacts of the fact that both variables are spatially structured? Our simulation study supports the following conclusions:

1) The best-known effect of inherent spatial autocorrelation (i.e., spatially structured residuals) is that significance tests become too liberal (Dutilleul et al. 1993; Dutilleul and Legendre 1993). Here, we show that pseudo-climatic variables with exogenously induced spatial structure and no inherent spatial autocorrelation have a dramatically increased probability of being significantly correlated, even with no causal link whatsoever between the pairs of pseudo-climatic variables.

2) Spatial structure does not induce bias. The expected correlation between randomly generated, spatially structured variables is zero.

3) Spatial structure does greatly increase the variance in correlations between causally unrelated variables. Some correlations are very strong, purely by chance. Consequently, the expected r^2 between spatially structured variables is much greater than the r^2 between variables that are not spatially structured. In the case of pseudo-temperature across the Americas, spatial structure increased the median r^2 from 0.000 to 0.183.

4) "The First Law of Geography" (Tobler 1970) states that "everything is related to everything else, but near things are more related than distant things." This is a consequence of conclusion 2), above. However, we propose a reformulation of Tobler's law as three new laws: #1) Everything in geography *that is spatially structured* will be collinear. #2) Near things are more related than distant things. #3) The more strongly spatially structured two variables are, the stronger the collinearity between them will be.

5) The strength of the expected r^2 between two spatially structured, but causally unrelated, variables:

   a. increases with the strength of the spatial structure in each of the two variables;

   b. depends weakly, or not at all, on the spatial extent of the data, at least over broad spatial extent;

   c. appears not to depend upon how similarly spatially structured the two variables are.

6) Many studies have hypothesized that there is a causal link between species richness and contemporary environmental variables (Wright 1983; Currie 1991; O'Brien 1998; Field et al. 2009). Treating our simulations as a null model (Beale et al. 2008; Chapman 2010), one prediction of this hypothesis is that the observed richness–environment correlations should be stronger than those obtained using random data that are similarly spatially-structured. The observed richness–temperature correlations ($r^2 \approx 0.7$) were much stronger than the median correlations between richness and pseudo-temperature (0.15 – 0.31). However, the one-tailed 95% confidence interval for richness–pseudo-temperature correlations included correlations as strong as $r^2 = 0.70$. The richness–observed temperature correlation was stronger than that (Table 1), but only marginally so (but see point 12 below). Spatial structure alone is unlikely to generate richness–temperature correlations as strong as those observed here.

7) A second observation is inconsistent with the hypothesis that the richness–temperature relationship is an artefact of spatial structure. If this were true, then strength of the richness–pseudo-temperature correlation should be unrelated to collinearity between temperature and pseudo-temperature. The data in Fig. 6 should describe a circle. In contrast, we found that, richness is strongly related to pseudo-temperature only when pseudo-temperature is strongly collinear with observed temperature. When temperature and pseudo-temperature are not collinear, the correlation between richness and pseudo-temperature is similar to the expected correlation among replicate pseudo-temperature simulations.

8) A practical consequence of conclusion 4 (above) is that, for *nearly any* case in which spatial variation in biotic variable Y is hypothesized to be driven by variation in environmental variable X, the two variables will be significantly correlated whether there is any real causal link or not. As a preliminary rule-of-thumb, we suggest that the expected correlation is likely to be on the order of $0.15 < r^2 < 0.30$. It remains to be determined how sensitive this expectation is to the study design (i.e., the spatial structure of the variables, spatial extent, spatial grain, shape of the domain, the algorithm used to simulate spatial structure, etc.).

9) Within-sample prediction provides no additional evidence that a correlation between Y and X reflects a causal link beyond what standard goodness-of-fit statistics already provide.

10) Out-of-sample prediction provides a stronger test of a hypothesized, general causal link between two variables Y and X_1. For example, if there is a causal link between richness and environment, and if that causal link acts globally, then a richness–environment relationship derived in one part of the world should accurately predict the variation in richness in other parts of the world, provided that ranges of environmental variables in the training region overlap the ranges in the test region (Francis and Currie 2003). If out-of-sample prediction is poor, one can correctly reject the hypothesis of a general causal link between Y and X_1. However, out-of-sample predictions can be surprisingly good, even in the absence of a causal link. This can happen if the spatial variation of Y is driven by variation in X_2, and if X_1 and X_2 are similarly collinear in both the calibration and the test data sets. Thus, out-of-sample prediction accuracy is a necessary, but not a sufficient, test of a causal link.

11) In the specific case of bird and mammal richness in the Americas, out-of-sample prediction is only partly consistent with the prediction that richness is controlled by a linear function of contemporary temperature. Observed richness in South America is strongly correlated with richness predicted from the richness–environment model fitted in North America. However, there are clear quantitative differences between the patterns in North and South America that are unrelated to temperature and precipitation. A linear function of these variables is insufficient to account for the variation of richness across the Americas. More complete models for species richness have been investigated elsewhere; the objective here

is to demonstrate how out-of-sample predictions can allow the investigator to test hypotheses and come up with new ones.

12) In principle, one could use a randomization study as a null model (as we have done here) to test whether a correlation between spatially structured variables is stronger than would be expected due to spatial structure alone (Beale et al. 2008, Chapman 2010). However, our study highlights a serious drawback to this approach. Suppose that biotic variable Y is strongly correlated with environmental variable X_1. An uncensored, spatially-structured randomization of X_1 will yield many pseudo-X_1 surfaces that are highly correlated with true X_1 (Figure 6). Biotic variable Y will be strongly correlated with those randomizations. Consequently, the randomization study provides an excessively conservative test of the significance of the $Y \sim X_1$ relationship because the randomizations have an elevated probability of being collinear with the true X_1 data. Thus, it is probably too strong to say that there are "no macroscale associations with climate in European birds" (Beale et al. 2008). Rather, it would be more accurate to say that those associations with climate may be real, but randomizations cannot exclude the possibility that the associations result from spatial structure. Using the subset of randomizations in which pseudo-X_1 is not collinear with X_1 may provide a more reasonable null distribution.

13) Spurious correlations resulting from spatially structured variables are likely to be rampant in the biogeographic and macroecological literatures. Other evidence must be adduced before suggesting that a correlation reflects a causal link.

Acknowledgements

Supplementary Materials

The following materials are available as part of the online article from https://escholarship.org/uc/fb

Figure S1. The decay of spatial autocorrelation (quantified by Moran's I) as a function of distance in North America.

Figure S2. The correlation between $\log_{10}$(species richness) and pseudo-precipitation.

Figure S3. Test of the out-of-sample predictive ability of a regression model.

Figure S4. Exactly the same as the preceding supplementary figure, using (simulated) pseudo-temperature, rather than observed temperature.

References

Adams, J. (2009) Species Richness: Patterns in the Diversity of Life. Springer, Chichester UK.

Araújo, M.B., Ferri-Yáñez, F., Bozinovic, F., Marquet, P.A., Valladares, F. & Chown, S.L. (2013) Heat freezes niche evolution. Ecology Letters, 16, 1206-1219.

Bahn, V. & McGill, B.J. (2013) Testing the predictive performance of distribution models. Oikos, 122, 321-331.

Beale, C.M., Lennon, J.J. & Gimona, A. (2008) Opening the climate envelope reveals no macroscale associations with climate in European birds. Proceedings of the National Academy of Sciences USA, 105, 14908-14912.

Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. & Elston, D.A. (2010) Regression analysis of spatial data. Ecology Letters, 13, 246-264.

Bini, L.M., Diniz, J.A.F., Rangel, T., Akre, T.S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araujo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Bohning-Gaese, K., Borges, P.A.V., Castro-Parga, I., Chey, V.K., Chown, S.L., de Marco, P., Dobkin, D.S., Ferrer-Castan, D., Field, R., Filloy, J., Fleishman, E., Gomez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D., Kitching, I.J., Leon-Cortes, J.L., Lobo, J.M., Montoya, D., Morales-Castilla, I., Moreno, J.C., Oberdorff, T., Olalla-Tarraga, M.A., Pausas, J.G., Qian, H., Rahbek, C., Rodriguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R. & Hawkins, B.A. (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography, 32, 193-204.

Chapman, D.S. (2010) Weak climatic associations among British plant distributions. Global Ecology and Biogeography, 19, 831-841.

Clifford, P., Richardson, S. & Hemon, D. (1989) Assessing the significance of the correlation between two spatial processes. Biometrics, 45, 123-134.

Currie, D.J. (1991) Energy and large scale patterns of animal and plant species richness. American Naturalist, 137, 27-49.

Currie, D.J. (2019) Where Newton might have taken ecology. Global Ecology and Biogeography, 28, 18-27.

Dale, M.R.T. & Fortin, M.-J. (2002) Spatial autocorrelation and statistical tests in ecology. Ecoscience, 9, 162-167.

Diniz-Filho, J.A.F., Nabout, J.C., de Campos Telles, M.P., Soares, T.N. & Rangel, T.F.L.V.B. (2009) A review of techniques for spatial modeling in geographical, conservation and landscape genetics. Genetics and Molecular Biology, 32, 203-211.

Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16, 129-138.

Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30, 609-628.

Dutilleul, P. & Legendre, P. (1993) Spatial heterogeneity against heteroscedasticity - An ecological paradigm versus a statistical concept. Oikos, 66, 152-171.

Dutilleul, P., Clifford, P., Richardson, S. & Hemon, D. (1993) Modifying the t test for assessing the correlation between two spatial processes. Biometrics, 49, 305-314.

Field, R., Hawkins, B.A., Cornell, H.V., Currie, D.J., Diniz-Filho, J.A.F., Guégan, F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M. & Turner, J.R.G. (2009) Spatial species-richness gradients across scales: a meta-analysis. Journal of Biogeography, 36, 132-147.

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

Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography, 27, 245-256.

Francis, A.P. & Currie, D.J. (2003) A globally-consistent richness-climate relationship for angiosperms. American Naturalist, 161, 523-536.

Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. Oikos, 121, 171-180.

Hawkins, B.A. & Porter, E.E. (2003) Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America. Global Ecology and Biogeography, 12, 475-482.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965-1978.

Hortal, J., Diniz, J.A.F., Bini, L.M., Rodriguez, M.A., Baselga, A., Nogues-Bravo, D., Rangel, T.F., Hawkins, B.A. & Lobo, J.M. (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. Ecology Letters, 14, 741-748.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology Letters, 12, 334-350.

Kühn, I. (2007) Incorporating spatial autocorrelation may invert observed patterns. Diversity and Distributions, 13, 66-69.

Legendre, P. (1993) Spatial autocorrelation - trouble or new paradigm. Ecology, 74, 1659-1673.

Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. Ecography, 23, 101-113.

O'Brien, E.M. (1998) Water-energy dynamics, climate, and prediction of woody plant species richness: an interim general model. Journal of Biogeography, 25, 379-398.

Patterson, B.D., Ceballos, G., Sechrest, W., Tognelli, M.F., Brooks, T., Luna, L., Ortega, P., Salazar, I. & Young, B.E. (2007) Digital distribution maps of the mammals of the Western Hemisphere, version 3. Available at: http://www.natureserve.org/conservation-tools/digital-distribution-maps-mammals-western-hemisphere (accessed Jan. 2015).

Peres-Neto, P.R. & Legendre, P. (2010) Estimating and controlling for spatial structure in the study of ecological communities. Global Ecology and Biogeography, 19, 174-184.

Popper, K. & Miller, D. (1983) A proof of the impossiblility of inductive probability. Nature, 302, 687-688.

Rigler, F.H. (1982) Recognition of the possible: an advantage of empiricism in ecology. Canadian Journal of Fisheries and Aquatic Sciences, 39, 1323-1331.

Rosenzweig, M.L. (1995) Species Diversity in Time and Space. In Cambridge University Press, Cambridge UK.

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region. Economic Geography, 46, 234-240.

Wallace, A.R. (1876) The Geographical Distribution of Animals. New York, Hafner.

Wright, D.H. (1983) Species-energy theory: an extension of species-area theory. Oikos, 41, 496-506.

Wright, D.H., Currie, D.J. & Maurer, B.A. (1993) Energy supply and patterns of species richness on local and regional scales. Species Diversity in Ecological Communities: Historical and Geographical Perspectives (ed. by R.E. Ricklefs and D. Schluter), pp. 66-74. University of Chicago Press, Chicago.