Assessing the Significance of Global and Local Correlations under Spatial Autocorrelation; a Nonparametric Approach

Júlia Viladomat[1], **Rahul Mazumder**[2], **Alex McInturff**[3], **Douglas J. McCauley**[4], and **Trevor Hastie**[1]

Júlia Viladomat: juliaviladomat@gmail.com

[1]Department of Statistics, Stanford University, Stanford 94305, U.S.A

[2]Department of Statistics, Columbia University, New York 10027, U.S.A

[3]Department of Biology, Stanford University, Stanford 94305, U.S.A

[4]Department of Ecology, Evolution, and Marine Biology, UC Santa Barbara, Santa Barbara 93106, U.S.A

Summary

We propose a method to test the correlation of two random fields when they are both spatially auto-correlated. In this scenario, the assumption of independence for the pair of observations in the standard test does not hold, and as a result we reject in many cases where there is no effect (the precision of the null distribution is overestimated). Our method recovers the null distribution taking into account the autocorrelation. It uses Monte-Carlo methods, and focuses on permuting, and then smoothing and scaling one of the variables to destroy the correlation with the other, while maintaining at the same time the initial autocorrelation. With this simulation model, any test based on the independence of two (or more) random fields can be constructed. This research was motivated by a project in biodiversity and conservation in the Biology Department at Stanford University.

Keywords

Geostatistics; Monte-Carlo methods; Resampling; Spatial autocorrelation; Spatial statistics; Variogram

1. Motivation

Assessing significance of the correlation coefficient is not straightforward if the values of the variables involved vary smoothly with location (throughout the paper smooth refers to spatial autocorrelation, and it may be the case that the variable changes abruptly over a short

---

distance). With spatially autocorrelated data, nearby points may provide almost identical information. Hence there is a tension between sample size, resolution and the number of independent measurements; i.e. at some level, more data, meaning sampling the process at a higher resolution, does not mean more information. As a result, classical tests Fisher (1915) tend to incorrectly reject the null (larger type II errors). Some work has been done, particularly in the field of Geostatistics, to overcome this problem. For instance, Clifford, Richardson, and Hemon (1989) propose a method that estimates an effective sample size $M < N$ to be used in such tests, in an attempt to capture the real uncertainty. The correlation coefficient is thus evaluated with a Student's t distribution with M degrees of freedom (distribution with larger variance), which accounts for the loss of precision due to the underlying spatial component. The method, however, is developed for Gaussian random fields and in reality smoothed processes tend to be non-Gaussian.

We propose a Monte-Carlo method to test the correlation of two random fields that takes account of the spatial autocorrelation. By

1. randomly permuting the values of one of the fields across space we eliminate the dependence between them, and

2. smoothing and scaling the permuted field we approximately recover, with the help of the variogram, the spatial structure suppressed in (1).

In the same spirit, Allard, Brix, and Chadoeuf (2001) propose a method that is based on random local rotations, but applied to the characteristics of the spatial structure of point processes, where the intensity (rate parameter) is assumed to be constant at small scales and varies at large scales, the opposite situation we encounter with random fields, were the autocorrelation usually fades away as the distance increases.

By repeating steps 1) and 2) many times, we can obtain approximate realizations of the null distribution of interest. In fact, with this simulation model, it is possible to examine the null distribution of a larger variety of statistics. In particular any test based on the independence of two (or more) random fields can be constructed, the simplest example being the test for a single Pearson's correlation coefficient between the two random fields. We will call it *global correlation*.

Other spatially local tests may be of interest — for example, which regions show strong correlation between the two fields — the question that our collaborators posed (McCauley et al., 2012). They mapped the locations of sites over the world using two criteria:

- amount of species richness — biodiversity, and

- travel time in days needed to reach the nearest city — remoteness.

Figures 1(a) and 1(b) represent both these fields, and we see that they are spatially very smooth. Are remoteness and biodiversity correlated with one another? i.e. are there more species in remote areas that are better insulated from human disturbance? To succinctly communicate the strength of these correlations, the authors were interested in reporting a p-value map for the areas where overlap between remoteness and biodiversity occurs. Initially they used Geographically Weighted Regression methods (Fotheringham, Brunsdon, and

Charlton, 2002), a set of regression techniques that tackle spatially varying relationships. This book has captured considerable attention in the Geostatistics community. However, these methods focus on comparing coefficients for different spatial areas, and identifying the areas with stronger relationships, but with no assessment to whether the coefficients in the model are significant or not.

Given the map of correlations in Figure 1(c), where each value corresponds to the correlation between biodiversity and remoteness in a given neighborhood (we will call them *local correlations*), we will apply our method and produce a map of p-values, where each p-value assesses significance of the correlation in that particular location. See Appendix 2 for details on how to calculate the local correlations.

We organize the paper as follows. In section 2 we show the results of applying our method to the biodiversity dataset. Section 3 illustrates the limitations of the standard test under spatial autocorrelation. Section 4 describes in detail the algorithm proposed in this paper, and in Section 5 we study the behavior of the method by performing power and type I error analyses, and compares it to the approach in Clifford et al. (1989). We conclude and summarize our findings in the discussion.

2. Biodiversity Data

Protecting remote ecosystems is the future of global diversity, WWF ecologists divided the world into 16 unique regions (Biomes) based on land cover and climate. In this paper, and for simplicity, we will focus on the part of the biodiversity dataset that corresponds to the American region of Biome 1 (Tropical and subtropical moist broadleaf forests) to illustrate our methodology.

Biodiversity (X) is the result of estimating the number of species of plants, amphibians, birds and mammals in an area of 100×100 km and centered at location s. The estimates for the 4 groups is normalized to a maximum score of 10, with X being the average of those normalized counts. Remoteness (Y) combines a number of data sets that influence speed of travel: road networks, angle of slope, density of vegetation, river courses, etc. It takes values between $1 - 8$ and indicates the travel time in days needed to reach the nearest city larger than $50, 000$ inhabitants from location s, where 8 represents any travel time larger than 7 days. Our sample is denoted by $(\mathbf{X_s}, \mathbf{Y_s}) = [(X_{s_1}, Y_{s_1}), \ldots, (X_{s_N}, Y_{s_N})]$, $\mathbf{s} = (s_1 \ldots s_N)$, where $s_i \in \mathbb{R}^2$ are the longitude and latitude coordinates of observation i and $N = 19, 926$.

2.1 The empirical and smoothed variogram for Biodiversity

The *theoretical variogram* is a function describing the degree of spatial dependence of a random field X_s. It is defined as the variance of the difference between field values at two

locations s_i and s_j across realizations of the field: $\gamma(s_i, s_j) = \frac{1}{2}\mathrm{Var}(X_{s_i} - X_{s_j})$, see Cressie (1993). The *empirical variogram* for the sample $X_{s1}, \ldots, X_{sN}$ is the collection of pairs of distances $u_{ij} = \|s_i - s_j\|$ between s_i and s_j, and their corresponding variogram ordinates $v_{ij} = \frac{1}{2}(X_{s_i} - X_{s_j})^2$. The empirical variogram for biodiversity is plotted in Figure 2(a). Since γ is expected to be a smooth function of distance, it is common to smooth the

empirical variogram to improve its properties as an estimator for γ. In our algorithm, we will use the *smoothed variogram* $\hat{\gamma}$, which is obtained using a kernel smoother and defined in Appendix 1. The black curve in Figure 2(a) is the smoothed variogram for biodiversity with bandwidth $h = 0.746$, and truncated at distance $u = 10$, which corresponds to the 25% percentile of the distribution of pairs of distances. We truncate the variogram because the precision of the estimate is expected to decrease as the distance increases, since a decreasing number of pairs are involved.

When pairs of distances are small, the variance of $X_{s_i} - X_{s_j}$ is small, indicating that observations close to each other are very similar. As pair distances increase, the autocorrelation dies off and the variogram grows; as farther apart observations are less similar. The increasing nature of the variogram as the distance increases is a common behavior that corroborates the spatial autocorrelation present in the variable biodiversity.

2.2 Applying our Method to the Biodiversity Dataset

In this section we apply to the biodiversity dataset the algorithm that we describe in Section 4. Given $\mathbf{X_s}$ and $\mathbf{Y_s}$, we will test significance of 1) the global correlation coefficient [H_0 : $\rho_{\mathbf{X_s}, \mathbf{Y_s}} = 0$], and 2) the set of local correlations $\hat{r}^{\lambda}_{\mathbf{X_s}, Y_s}(s_1), \ldots, \hat{r}^{\lambda}_{\mathbf{X_s}, Y_s}(s_N)$ plotted in Figure 1(c) [H_0 : $\rho_{\mathbf{X_s}, \mathbf{Y_s}}(s_j) = 0, \forall j$], and calculated using a Gaussian kernel with neighborhood window truncated at bandwidth $\lambda = 5.281$ (see Appendix 2).

We will see that the algorithm returns $\hat{\mathbf{X}}_s^1, \ldots, \hat{\mathbf{X}}_s^B$ as a result of randomizing and smoothing $\mathbf{X_s}$ $B = 1000$ times, with bandwidths $\in (0.005, 0.785)$. These are proxies for $\mathbf{X_s}$ with approximately the same autocorrelation but with the characteristic of being independent of $\mathbf{Y_s}$. The pairs $(\hat{\mathbf{X}}_s^1, \mathbf{Y}_s), \ldots, (\hat{\mathbf{X}}_s^B, \mathbf{Y}_s)$ will be used to test the situations 1) and 2) above.

We chose to randomize $\mathbf{X_s}$ (biodiversity), as we are free to pick the most convenient one, since the purpose is to break the dependence between both variables.

2.2.1 Testing the Global Correlation—The observed global correlation coefficient between biodiversity and remoteness is $\hat{\gamma}_{\mathbf{X_s}, \mathbf{Y_s}} = 0.224$. We use $\mathbf{r} = (r_1, \ldots, r_B)$ as an estimate of the sampling distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$ under the null hypothesis of independence, where $r_i = \text{cor}(\hat{\mathbf{X}}_s^i, \mathbf{Y}_s)$; see Figure 2(b). Using this null distribution to test $H_0 : \rho_{\mathbf{X_s}, \mathbf{Y_s}} = 0$, the p-value is $p_{\hat{r}} = P(|r_{\mathbf{X_s}, Y_s}| > |\hat{r}_{\mathbf{X_s}, Y_s}|) = \frac{1}{B} \sum_{i=1}^{B} I[|r_i| > |\hat{r}_{\mathbf{X_s}, Y_s}|] = 0.057$ (indicated with a black line in Figure 2(b)). Had we assumed that the sample pairs were independent, and used instead a Student's t with $N - 2$ degrees of freedom, the p-value would have been effectively zero, since the variance of the Student's t distribution is significantly smaller.

2.2.2 Testing Local Correlations—Each pair of random fields $(\hat{\mathbf{X}}_s^i, \mathbf{Y}_s), i = 1, \ldots, B$, can be used to calculate a new map of local correlations under the null hypothesis ($\hat{\mathbf{X}}_s^i$ and $\mathbf{Y}_s$ are constructed to be independent). Hence we can compute local p-values in exactly the same was as was done globally above. A sample distribution of the statistic $r_{\mathbf{X_s}, \mathbf{Y_s}}(s_j)$ under

the null is $\mathbf{r}(s_j) = (r_1(s_j), \ldots, r_B(s_j))$, where $r_i(s_j) = r^\lambda_{\hat{\mathbf{X}}^i_s, \mathbf{Y_s}}(s_j)$. The p-value for testing H_0 : $\rho_{\mathbf{X_s}, \mathbf{Y_s}}(s_j) = 0$ at location s_j is

$$p_{s_j} = P(|r_{\mathbf{X_s}, \mathbf{Y_s}}(s_j)| > |\hat{r}^\lambda_{\mathbf{X_s}, \mathbf{Y_s}}(s_j)|) = \frac{1}{B}\sum_{i=1}^{B} I[|r_i(s_j)| > |\hat{r}^\lambda_{\mathbf{X_s}, \mathbf{Y_s}}(s_j)|].$$ The resulting p-value map is plotted in Figure 3, where we can identify the regions that are strongly correlated. For comparison, we also plot the map of p-values had we used the standard test.

3. Behavior of $r_{\mathbf{X_s}, \mathbf{Y_s}}$ under Spatial Autocorrelation

If $(x_1, y_1), \ldots, (x_N, y_N)$ is an independent and normally distributed sample, the distribution for the Pearson's correlation coefficient under $\rho_{X,Y} = 0$ is

$$f_N(r) = \frac{(1-r^2)^{\frac{N-4}{2}}}{B\{\frac{1}{2}, \frac{1}{2}(N-2)\}}, \|r\| \le 1 \quad (1)$$

The statistic $t = \dfrac{(N-2)^{\frac{1}{2}} r}{(1-r^2)^{\frac{1}{2}}}$ is used to test for $H_0 : \rho_{X,Y} = 0$, which follows a Student's t distribution with $N - 2$ degrees of freedom, see Kenney and Keeping (1951).

In the previous section we have seen that even though the sample size of the biodiversity dataset was very large ($N = 19,296$), the sample distribution for $r_{\mathbf{X_s}, \mathbf{Y_s}}$ was somewhat wide (Figure 2(b)).

This emphasizes the following point, which we demonstrate via a simulation. For a pair of spatially correlated random fields, the sample size or, more precisely, the resolution at which the fields are sampled, can play less of a role in the behavior of the distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$; rather it is the extent of the spatial autocorrelation that determines this distribution (see Walther (1997) for an equivalent problem with time series).

Let W_s be a stationary and isotropic Gaussian random field, $s \in \mathbb{R}^2$, with Matérn auto-correlation function $\phi(u) = [2^{\kappa-1}\Gamma(\kappa)]^{-1}(u/\phi)^\kappa K_\kappa(u/\phi)$ where $\phi > 0$ is the scale, and the shape parameter $\kappa > 0$ determines the smoothness of the process, $\text{var}(W_s) = \sigma^2$. For $\kappa = 0.5$, the Matérn autocorrelation function reduces to the exponential, and when $\kappa \to \infty$ to the Gaussian.

Suppose X_{s_i} is generated by a stationary process

$$X_{s_i} = W_{s_i} + Z_i, \quad (2)$$

with Z_i being mutually independent and identically distributed with zero mean and nugget variance τ^2 (measurement error), see Diggle and Ribeiro (2007). The theoretical variogram of X_{s_i} under stationarity is described in Appendix 1 and illustrated in Web Figure 4.

Figures 4(a) and 4(b) are $\mathbf{X_s}$ and $\mathbf{Y_s}$, two independent realizations of this process in the grid $[0, 1] \times [0, 1]$ with resolution $N = 101 \times 101 = 10,201$, and parameters $\kappa = 0.5$ and $\phi = 0.30$.

We have used the R package RandomFields to simulate these processes (R Core Team, 2013; Schlather et al., 2013).

In Web Figure 1, Figure (a) is the scatter plot of $\mathbf{X_s}$ and $\mathbf{Y_s}$, whereas Figure (d) is the scatter plot of two independent samples, each of them mutually independent (non-spatially correlated) and normally distributed. The correlation coefficient is much larger for $\mathbf{X_s}$ and $\mathbf{Y_s}$ ($\hat{r_{\mathbf{X_s}, \mathbf{Y_s}}}$ = 0.3). Figure (b) and (c) are two new sets of independent simulated processes $\mathbf{X_s}$ and $\mathbf{Y_s}$, showing negative strong correlation and a correlation closer to zero respectively. Thus, as we have seen, a consequence of the spatial autocorrelation is a larger variance for the distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$. In fact, the larger κ, the larger the variance. Intuitively, consider the extreme case of two very smooth one-dimensional fields (i.e lines) with randomly chosen orientations (slopes); the correlation of sampled pairs will either be -1 and $+1$ depending on the randomly chosen orientation; a maximal variance situation. The same intuition applies to higher-dimensional fields.

The null distribution for $r_{\mathbf{X_s}, \mathbf{Y_s}}$ under $\rho_{\mathbf{X_s}, \mathbf{Y_s}} = 0$ can be estimated by generating the independent pairs $(\mathbf{X}_{\mathbf{s}}^{i}, \mathbf{Y}_{\mathbf{s}}^{i})$. With this distribution, the probability of observing values as extreme or more than 0.3 is 0.16, with no evidence against H_0.

Although $\mathbf{X_s}$ and $\mathbf{Y_s}$ have been constructed to be independent of each other, if we use the Student's t distribution to test $H_0 : \rho_{\mathbf{X_s}, \mathbf{Y_s}} = 0$, the p-value for $\hat{r_{\mathbf{X_s}, \mathbf{Y_s}}}$ = 0.3 is 0.

4. The Algorithm

We propose a method that approximately recovers the null distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$, or any other statistic based on the independence of $\mathbf{X_s}$ and $\mathbf{Y_s}$. The following scheme summarizes the main steps of the algorithm.

Let $\mathbf{X_s}$ and $\mathbf{Y_s}$ be a realization of two random fields. Repeat the following two steps B times:

1. Randomly permute the values of $\mathbf{X_s}$ over $\mathbf{s}$, which we denote by $\mathbf{X}_{\pi(\mathbf{s})}$; this means $\mathbf{X}_{\pi(\mathbf{s})}$ and $\mathbf{Y_s}$ are independent.

2. Smooth and scale $\mathbf{X}_{\pi(\mathbf{s})}$ to produce $\hat{\mathbf{X_s}}$, such that its smoothed variogram $\hat{\gamma}$ approximately matches $\gamma(\mathbf{X_s})$; i.e. the transformed variable $\hat{\mathbf{X_s}}$ has approximately the same autocorrelation structure as $\mathbf{X_s}$.

The random fields $\hat{\mathbf{X}}_{\mathbf{s}}^{1}, \ldots, \hat{\mathbf{X}}_{\mathbf{s}}^{B}$ have approximately the same autocorrelation structure as $\mathbf{X_s}$, but are each independent of $\mathbf{Y_s}$. Hence, the pairs $(\hat{\mathbf{X}}_{\mathbf{s}}^{1}, \mathbf{Y_s}), \ldots, (\hat{\mathbf{X}}_{\mathbf{s}}^{B}, \mathbf{Y_s})$ are the ingredients for the calculation of a null distribution. For instance, the distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$ under $H_0 : \rho_{\mathbf{X_s}, \mathbf{Y_s}} = 0$ can be approximated by $\mathbf{r} = (r_1, \ldots, r_B)$, where $r_i = \mathrm{cor}(\hat{\mathbf{X}}_{\mathbf{s}}^{i}, \mathbf{Y_s})$, but we could proceed equivalently with any other statistic based on the independence of $\mathbf{X_s}$ and $\mathbf{Y_s}$.

Note that we do not pose restrictions on which random field to permute, but the algorithm assumes that $\mathbf{X_s}$ is stationary.

Step (2) of the algorithm is described in detail in the following section.

4.1 Step 2: Matching Variograms

This step focuses on recovering the intrinsic spatial structure of $\mathbf{X_s}$ that was eliminated with the random permutation. As we have seen, the null distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$ is mainly determined by the amount of autocorrelation, and this step will determine how well we are able to recover it. Our approach is non-parametric, which implies that the variogram matching does not rely on model assumptions, such as choosing a parametric model for the variogram.

Formally, the problem reduces to choosing a variogram from the family

$$\beta\hat{\gamma}(\mathbf{X_s^{\delta}})+\alpha \quad (3)$$

that best approximates $\widehat{\gamma(\mathbf{X_s})}$, the smoothed variogram of $\mathbf{X_s}$.

Let $\mathbf{X}_{\pi(\mathbf{s})}$ be the permuted random field obtained in step (1) above. Choose   to be a set of values for the proportion of neighbors to consider for the smoothing step.

1. Calculate the smoothed variogram $\widehat{\gamma(\mathbf{X_s})}$ by smoothing the empirical variogram of $\mathbf{X_s}$ (see Appendix 1).

2. For each $\delta \in$ repeat:

    a. Construct the smoothed variable $\mathbf{X_s^{\delta}}$ using a kernel smoother that fits a constant regression to $\mathbf{X}_{\pi(\mathbf{s})}$ at each location s_j (see Appendix 2).

    b. Calculate $\hat{\gamma}(\mathbf{X_s^{\delta}})$ as in (1).

    c. Fit a linear regression between $\hat{\gamma}(\mathbf{X_s^{\delta}})$ and $\widehat{\gamma(\mathbf{X_s})}$, where $(\hat{\alpha_{\delta}}, \hat{\beta_{\delta}})$ are the least-squares estimates.

3. Choose $\delta^* \in$ such that the sum of squares of the residuals of the fit in (2) (c) is minimized.

4. Transform $\mathbf{X_s^{\delta^*}}$ by $\hat{\mathbf{X}}_{\mathbf{s}}^{\delta^*} = |\hat{\beta}_{\delta^*}|^{\frac{1}{2}}\mathbf{X_s^{\delta^*}} + |\hat{\alpha}_{\delta^*}|^{\frac{1}{2}}\mathbf{Z}$, where $\mathbf{Z}$ is a vector of mutually independent and identically distributed Z_i's with zero mean and unit variance.

By varying the tuning parameter or proportion of neighbors δ in step (2) we obtain a family of variograms with different shapes, choosing the one more similar to the variogram of the original variable $\mathbf{X_s}$.

The linear transformation in Step (4) ensures that the scale and intercept of $\hat{\gamma}(\hat{\mathbf{X}}_{\mathbf{s}}^{\delta^*})$ match those of $\widehat{\gamma(\mathbf{X_s})}$, since the smoothing in step (2)(a) has changed the scale of $\mathbf{X_s}$ (the smoother $\mathbf{X_s^{\delta^*}}$, the smaller the variance), in addition to the intercept (nugget variance) of $\widehat{\gamma(\mathbf{X_s})}$.

The variogram of the random field $\hat{\mathbf{X}}_{\mathbf{s}}^{\delta^*}$ is a member of the family in (3), and $\hat{\mathbf{X}}_{\mathbf{s}}^{\delta^*}$ has been constructed to match $\widehat{\gamma(\mathbf{X_s})}$ in shape, scale, and intercept.

Note that $|\hat{\beta}_{\delta*}|\mathrm{var}(\mathbf{X}_{\mathbf{s}}^{\delta*})$ is an estimate of σ^2 in (A.2), $|\hat{\alpha_{\delta*}}|$ is an estimate of τ^2, and correspondingly $\mathrm{var}(\hat{\mathbf{X}}_{\mathbf{s}}^{\delta*})=|\hat{\beta}_{\delta*}|\mathrm{var}(\mathbf{X}_{\mathbf{s}}^{\delta*})+|\hat{\alpha}_{\delta*}|$ is an estimate of $\sigma^2 + \tau^2$.

For an illustration of the variogram matching see Figure 5. Different δs result in different shapes for $\hat{\gamma}(\mathbf{X}_{\mathbf{s}}^{\delta})$, and the best match between $\hat{\gamma(\mathbf{X}_{\mathbf{s}})}$ and $\hat{\gamma}(\mathbf{X}_{\mathbf{s}}^{\delta})$ (grey) is reached with $\delta = 0.085$. The residual sum of squares of linearly regressing $\hat{\gamma}(\mathbf{X}_{\mathbf{s}}^{\delta})$ on $\hat{\gamma(\mathbf{X}_{\mathbf{s}})}$ for different values of δ are plotted in Web Figure 2.

5. Simulation Results

In this section we study the performance of our method for testing $H_0 : \rho_{\mathbf{X_s, Y_s}} = 0$ by comparing the estimate of the null distribution for $r_{\mathbf{X_s, Y_s}}$ provided by our algorithm with an estimate of the true null distribution.

In addition to that, we carry out power and type I error analyses (Section 5.2), and compare the results to Clifford's approach.

Let $\mathbf{X_s}$ and $\mathbf{Y_s}$ be two independent Gaussian random fields in the grid $[0, 1] \times [0, 1]$ with resolution $N = 101 \times 101 = 10, 201$, following model (2) with $\sigma^2 = 1$, $\mu = 0$, and:

1. Gaussian autocorrelation function $\phi(u) = \exp[-(u/\varphi)^2]$ with scale parameter $\varphi = 0.3$.

2. exponential autocorrelation function $\phi(u) = \exp(-u/\varphi)$ with $\varphi = 0.3$.

We apply our algorithm to one realization of the pair $(\mathbf{X_s, Y_s})$ in situations (1) and (2), with $B = 1000$ and bandwidths $\mathrm{gauss} = (0.1, 0.2, \ldots, 0.9)$ and $\mathrm{exp} = (0.07, 0.080, \ldots, 0.18)$ respectively.

The null distributions that the algorithm returns are plotted respectively in Figures 4(c) and 4(d), together with the true nulls obtained by simulating 1000 times the pairs $(\mathbf{X}_{\mathbf{s}}^i, \mathbf{Y}_{\mathbf{s}}^i)$ from the models above. The results are slightly better for (1), but we manage to recover the null in both situations.

5.1 Clifford's Effective Sample Size

We briey review Clifford's procedure prior to using it as a comparator. Clifford et al. (1989) propose to use $f_M(r)$ in (1) as the distribution of reference under the null. The variance of $f_M(r)$ is $\dfrac{1}{M-1}$. Their suggestion is to estimate the effective sample size M by $\hat{M}=\left\lfloor 1+\dfrac{1}{\hat{\sigma}_r^2}\right\rfloor$, as a result of equating $\dfrac{1}{M-1}$ to σ_r^2, the variance of the sample correlation. Hence, an estimate for the null distribution of $r_{\mathbf{X_s, Y_s}}$ is $f_M(r)$.

How to Estimate σ_r^2—In Clifford et al. (1989)'s Appendix they prove that $\sigma_r^2=\dfrac{var(S_{\mathbf{X_s Y_s}})}{E(S_{\mathbf{X_s}}^2)E(S_{\mathbf{Y_s}}^2)}$ to the first order, and under the assumption of normality, where $S_{\mathbf{X_s Y_s}}$

is the sample covariance, $S^2_{\mathbf{X_s}}$ and $S^2_{\mathbf{Y_s}}$ are the sample variances of $\mathbf{X_s}$ and $\mathbf{Y_s}$, $\mathrm{var}(S_{\mathbf{X_s Y_s}}) =$ trace$(\Sigma_{\xi_s}\Sigma_{\eta_s})$, $\Sigma_{\xi_s} = P\Sigma_{\mathbf{X_s}}P$, $\Sigma_{\eta_s} = P\Sigma_{\mathbf{Y_s}}P$, $\Sigma_{\mathbf{X_s}}$ and $\Sigma_{\mathbf{Y_s}}$ are the covariance matrices of $\mathbf{X_s}$ and $\mathbf{Y_s}$ respectively, $P = I - \dfrac{1}{N}\mathbf{11}'$ and $\mathbf{1}$ is a vector of 1's of dimension N.

An estimate for $\mathrm{var}(S_{\mathbf{X_s Y_s}})$ is obtained by imposing a stratified structure on $\Sigma_{\mathbf{X_s}}$ and $\Sigma_{\mathbf{Y_s}}$. More precisely, they assume that the set of all ordered pairs of elements of $\mathbf{s}$ can be divided into strata $S_0, S_1, S_2, \ldots$, and that the covariances within strata are constant. Then,

$$\hat{\sigma}^2_r = \frac{\sum_k N_k \hat{C}_{\mathbf{X_s}}(k)\hat{C}_{\mathbf{Y_s}}(k)}{N^2 S^2_{\mathbf{X_s}} S^2_{\mathbf{Y_s}}},$$ where N_k is the number of pairs in stratum S_k and

$\hat{C}_{\mathbf{X_s}}(k) = \dfrac{1}{N_k}\sum_{(i,j)\in S_k}(X_{s_i} - \bar{X}_k)(X_{s_j} - \bar{X}_k)$ is an autocovariance estimate for stratum S_k. They use the sample variogram to choose the number of strata.

5.2 Power Analysis and Type I Error Estimates

In this section we estimate the power and type I error of the test for our procedure, and compare it to Clifford's approach.

Web Figure 3 summarizes different scenarios for the null and alternative distributions, as a function of the effect ρ we would like to detect and the amount of spatial autocorrelation φ. The power to detect a fixed effect ρ decreases as the spatial autocorrelation increases. Equivalently, for a fixed φ the power decreases with the effect size. These data have been simulated assuming we know the truth, and indicates how well we can do in each situation. The grid size g is set to 0.05 ($N = 441$), but identical results were obtained for grid size $g = 0.01$ ($N = 10, 201$).

The simulation experiment goes as follows. For each combination of autocorrelation in $\varphi \in (0.05, 0.1, 0.3)$ and resolution in $g \in (0.05, 0.01)$, and for the power calculations, for each size effect in $\rho \in (0.2, 0.5, 0.8)$, we simulate 10 independent pairs $(\mathbf{X}^j_{\mathbf{s}}, \mathbf{Y}^j_{\mathbf{s}})$ from the usual model in the grid $[0, 1] \times [0, 1]$, with grid size g and Gaussian autocorrelation function with scale parameter φ. Then we do the following:

1. apply our algorithm to each pair $(\mathbf{X}^j_{\mathbf{s}}, \mathbf{Y}^j_{\mathbf{s}})$ with $B = 1000$, $\varphi_{=0.3} = (0.1, 0.2, \ldots, 0.9)$, $\varphi_{=0.1} = (0.03, 0.034, \ldots, 0.074)$, and $\varphi_{=0.05} = (0.013, 0.014, \ldots 0.027)$ respectively. Let $(\mathbf{r}_1, \ldots, \mathbf{r}_{10})$ be the resulting 10 nulls.

2. apply Clifford's method to each pair $(\mathbf{X}^j_{\mathbf{s}}, \mathbf{Y}^j_{\mathbf{s}})$ by generating B independent and normally distributed random vectors $\mathbf{X}_i$ and $\mathbf{Y}_i$ of dimension $\hat{M}$. Let $(\mathbf{r}^{\mathrm{Cl}}_1, \ldots, \mathbf{r}^{\mathrm{Cl}}_{10})$ be the resulting 10 nulls, where $\mathbf{r}^{\mathrm{Cl}}_j = (r^{\mathrm{Cl}}_{1j}, \ldots, r^{\mathrm{Cl}}_{Bj})$ and $r^{\mathrm{Cl}}_{ij} = \mathrm{cor}(\mathbf{X}_i, \mathbf{Y}_i)$.

In addition to that, $(\mathbf{r}^{\mathrm{Tr}}_1, \ldots, \mathbf{r}^{\mathrm{Tr}}_{10})$ are 10 nulls under the truth, by simulating B independent pairs $(\mathbf{X}^i_{\mathbf{s}}, \mathbf{Y}^i_{\mathbf{s}})$ from the same model, where $\mathbf{r}^{\mathrm{Tr}}_j = (r^{\mathrm{Cl}}_{1j}, \ldots, r^{\mathrm{Tr}}_{Bj})$ and $r^{\mathrm{Tr}}_{ij} = \mathrm{cor}(\mathbf{X}^i_{\mathbf{s}}, \mathbf{Y}^i_{\mathbf{s}})$.

The type I error of the test should be equal to the significance level $\alpha = 0.05$. We use the nulls $(\mathbf{r}^{\mathrm{Tr}}_1, \ldots, \mathbf{r}^{\mathrm{Tr}}_{10})$, $(\mathbf{r}_1, \ldots, \mathbf{r}_{10})$ and $(\mathbf{r}^{\mathrm{Cl}}_1, \ldots, \mathbf{r}^{\mathrm{Cl}}_{10})$ to estimate the type I errors associated to

both methods and the truth. We draw 100 samples from $(\mathbf{X}_{\mathbf{s}}^{k}, \mathbf{Y}_{\mathbf{s}}^{k})$ under $H_0 : \rho_{\mathbf{X_s}, \mathbf{Y_s}} = 0$, and use respectively $\mathbf{r}_j^{\mathrm{Tr}}, \mathbf{r}_j$ and $\mathbf{r}_j^{\mathrm{Cl}}$ to assess significance of $\hat{r}_k = \mathrm{cor}(\mathbf{X}_{\mathbf{s}}^{k}, \mathbf{Y}_{\mathbf{s}}^{k})$, for $k = 1, \ldots, 100$. Out of the 100 samples, the proportion of times the p-values are smaller than $\alpha = 0.05$ is an estimate of the type I error. We repeat the same process for 100 replicates, obtaining 100 type I error estimates. We then repeat the process again for each null $j = 1, \ldots, 10$ and average the 10×100 type I error estimates, which are shown in Table 1. Estimates of the standard errors are obtained by computing the standard deviation of the 10×100 estimates, and appear in brackets in the table.

The power of the test is calculated using a similar sampling scheme. We generate pairs $(\mathbf{X}_{\mathbf{s}}^{k}, \mathbf{Y}_{\mathbf{s}}^{k})$ as before, but then transform $\mathbf{Y}_{\mathbf{s}}^{k} \leftarrow \eta \mathbf{X}_{\mathbf{s}}^{k} + \mathbf{Y}_{\mathbf{s}}^{k}$ with $\eta = \rho / \sqrt{(1 - \rho^2)}$ (leading to a population correlation of ρ). The r_k^*'s are thus obtained under the alternative hypotheses $H_1 : \rho_{\mathbf{X_s}, \mathbf{Y_s}} = \rho$. We then follow the same procedure as was done for the type-I errors, namely steps (1) and (2) above, leading to the power estimates in Table 2.

We highlight the values in Table 1 that most deviate from α. Type I error estimates for our method are close to α, which is less of a case for Clifford's method, except for $\varphi = 0.05$.

Note also in Table 2 that power results for our method are typically closer to the power obtained under the truth, and that better power results in Clifford's case go along with larger type I error estimates, these values are highlighted in the table.

Finally, observe that there are no obvious differences between both resolutions, where the grid sizes $g = 0.01$ and $g = 0.05$ correspond to $N = 101 \times 101 = 10201$ and $N = 21 \times 21 = 441$ respectively. This is important, as it highlights the fact that the power of the analysis is mainly driven by the autocorrelation, and even if we increase significantly the resolution, the power stays the same as it is constrained by φ.

6. Discussion

This papers addresses the consequences of spatial autocorrelation on the distribution of statistics like correlation between random fields. We develop a nonparametric approach for sampling from the null-hypothesis of independence, that involves three steps:

1. pick one of the fields and estimate the spatial autocorrelation structure via its variogram

2. randomly permute the values in this field

3. apply a local smoothing to the permuted values, using a bandwidth and rescaling so that its resulting variogram matches the original in step (1).

Steps (2) and (3) are repeated B times to obtain B realizations from the null. These realizations can be used to measure many different kinds of independence, including measures of local dependence.

A referee of an earlier draft of this paper made a useful suggestion for an alternative and more direct approach. Their suggestion was to simulate a Gaussian random field according

to a model of the variogram, with parameters estimated by weighted least square, maximum likelihood or composite likelihood, with non-Gaussian data transformed to normal scores prior to the estimation of the variogram. One of the advantages of our method, though, is that it does not rely on model assumptions, which turns it into a more flexible, non-parametric approach.

Allard et al. (2001) address a related problem in the context of pairs of point processes, where the focus is on local correlations between pairs of realizations of counts. They randomize one of the processes using local rotations, under the assumption that locally the process rates are constant. Our randomization scheme treats broader range correlations, so the locally constant assumption is not appropriate.

Another approach is proposed by Clifford et al. (1989), where they estimate an effective sample size that takes the autocorrelation into account. We compare this approach to our method through simulations, which show that our algorithm behaves well in practice, with type I error estimates close to α, and power estimates close to the maximum power for each combination of φ and ρ.

Since correlation may exist simultaneously at a number of different geographical scales, the method can be used to calculate a p-value for the global correlation, as well as p-values for the local correlations (p-values map), summarizing the strength of the relationship between both fields in a particular location.

One of the important consequences of autocorrelation is that increasing the sample size does not necessarily increase the power to find significance. There is no concept of sample size, since what we observe is one realization of the random field, and the amount of data that we have is the resolution. The information that the sample provides is limited by the spatial autocorrelation. Consequently, in practice it may be more important to focus on using methods that adjust for autocorrelation, than on collecting more data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Allard D, Brix A, Chadoeuf J. Testing Local Independence between Two Point Processes. Biometrics. 2001; 57:508–517. [PubMed: 11414577]

Clifford P, Richardson S, Hemon D. Assessing the Significance of the Correlation Between Two Spatial Processes. Biometrics. 1989; 45:123–134. [PubMed: 2720048]

Cressie, NA. Statistics for Spatial Data. John Wiley; 1993.

Diggle, PJ.; Ribeiro, PJ. Model-based Geostatistics. Springer; 2007.

Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika. 1915; 10:507–521.

Fotheringham, AS.; Brunsdon, C.; Charlton, M. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley; 2002.

Kenney, JF.; Keeping, ES. Mathematics of Statistics (Part Two). New York: D. Van Nostrand Cmpany, Inc.; 1951.

Loader C. locfit: Local Regression, Likelihood and Density Estimation. R package version 1.5–9.1. 2013

McCauley DJ, McInturff A, Nuñez TA, Young HS, Viladomat J, Mazumder R, et al. Nature's last stand: Identifying the world's most remote and biodiverse ecosystems. In review. 2012

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

Schlather M, Menck P, Singleton R, Pfaff B, team RC. RandomFields: Simulation and Analysis of Random Fields. R package version 2.0.66. 2013

Walther G. Absence of Correlation between the Solar Neutrino Flux and the Sunspot Number. Physical Review Letters. 1997; 79:4522–4524.

Appendix 1

The smoothed variogram $\hat{\gamma}$ for $\mathbf{X_s}$ is defined by smoothing the empirical variogram using a kernel smoother:

$$\hat{\gamma}(u_0) = \frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} w_{ij} v_{ij}}{\sum_{i=1}^{N}\sum_{j=i+1}^{N} w_{ij}}. \quad \text{(A.1)}$$

It assigns weights that die off smoothly as distance to u_0 increases, with $w_{ij} = K_h(\|u_0 - u_{ij}\|)$, $u_{ij} = \|s_i - s_j\|$, $v_{ij} = \frac{1}{2}(X_{s_i} - X_{s_j})^2$ and $K_h(x) = \exp\{\frac{-(2.68x)^2}{2h^2}\}$, the Gaussian kernel is scaled so that their quartiles are at $\pm 0.25h$, with h being the bandwidth (R function ksmooth). The variogram $\hat{\gamma}$ is obtained by evaluating (A.1) at distances $\mathbf{u} = (u_1, \ldots, u_{100})$ uniformly chosen within the range of pair distances u_{ij}.

The theoretical variogram of a stationary random field $X_{s_i} = W_{s_i} + Z_i$ in (2) is:

$$\gamma(u) = \sigma^2 \{1 - \varphi(u)\} + \tau^2, \quad \text{(A.2)}$$

where $\varphi(u)$ is the autocorrelation function of W_{s_i}, typically a monotone decreasing function with $\varphi(0) = 1$ and $\varphi(u) \to 0$ as $u \to \infty$. Its most important feature is its behavior near $u = 0$, and how quickly it approaches zero when u increases, which reflects the physical extent of the spatial autocorrelation in the process. When $\varphi(u) = 0$ for u greater than some finite value, this value is known as the range of the variogram. The intercept τ^2 corresponds to the nugget variance, the conditional variance of each measured value X_{s_i} given the underlying signal value W_{s_i}. The sill is the asymptote $\tau^2 + \sigma^2$ and corresponds to the variance of the observation process X_{s_i}. Web Figure 4 gives a schematic illustration.

Appendix 2

We smooth a given random field $\mathbf{X_s}$ by fitting a local constant regression at locations $s_1, \ldots, s_N$ using the R package locfit (Loader, 2013):

$$\hat{f}_\delta(s) = \frac{\sum_{\|s-s_i\| \leq \lambda_s} w_{s_i} X_{s_i}}{\sum_{\|s-s_i\| \leq \lambda_s} w_{s_i}}, \quad \text{(A.3)}$$

where $w_{s_i} = K_{\lambda_s}(\|s - s_i\|)$, $K_{\lambda_s}(x) = \exp(\frac{-2.5x^2}{2\lambda_s^2})$ is a Gaussian kernel, and λ_s controls the smoothness of the fit. For a fitting point s, λ_s is such that the neighborhood contains the $k = \lfloor N\delta \rfloor$ nearest points to s in Euclidean distance, where $\delta \in (0, 1)$ is a tuning parameter that indicates the proportion of neighbors. Each $\hat{f}_\delta(s)$ is estimated with the k observations that fall within the ball $B_{\lambda_s}(s)$ centered at s and of radius λ_s (the kernel truncates at one standard deviation). We use a varying λ_s because it reduces data sparsity problems by increasing the radius in regions with fewer observations.

The smoothed variable $\mathbf{X_s^\delta}$ in Section 4.1 in step (2)(a) of the algorithm is the result of fitting the function (A.3) to $\mathbf{X}_{\pi(\mathbf{s})}$.

The local correlations in Figure 1(c) are calculated using this same approach. The local correlation at location s is defined as:
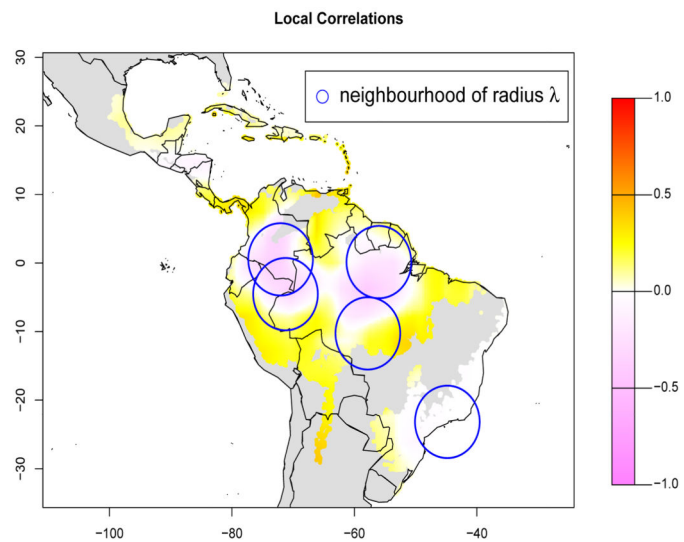
$$\hat{r}_{\mathbf{X_s}, Y_{\mathbf{s}}}^\lambda(s) = \frac{\sum_{\|s-s_i\| \leq \lambda} w_{s_i}(X_{s_i} - \bar{X}_{\mathbf{s}})(Y_{s_i} - \bar{Y}_{\mathbf{s}})}{\sqrt{\sum_{\|s-s_i\| \leq \lambda} w_{s_i}(X_{s_i} - \bar{X}_{\mathbf{s}})^2 \sum_{\|s-s_i\| \leq \lambda} w_{s_i}(Y_{s_i} - \bar{Y}_{\mathbf{s}})^2}} \quad \text{(A.4)}$$

where $\bar{X}_{\mathbf{s}} = \frac{\sum w_{s_i} X_{s_i}}{\sum w_{s_i}}$ and $\bar{Y}_{\mathbf{s}} = \frac{\sum w_{s_i} Y_{s_i}}{\sum w_{s_i}}$. We then compute (A.4) by breaking it down and separately evaluating the quantities $\Sigma w_{s_i} X_{s_i}$, $\Sigma w_{s_i} Y_{s_i}$, $\Sigma w_{s_i} X_{s_i} Y_{s_i}$, $\sum w_{s_i} X_{s_i}^2$ and $\sum w_{s_i} Y_{s_i}^2$ using locfit as above.
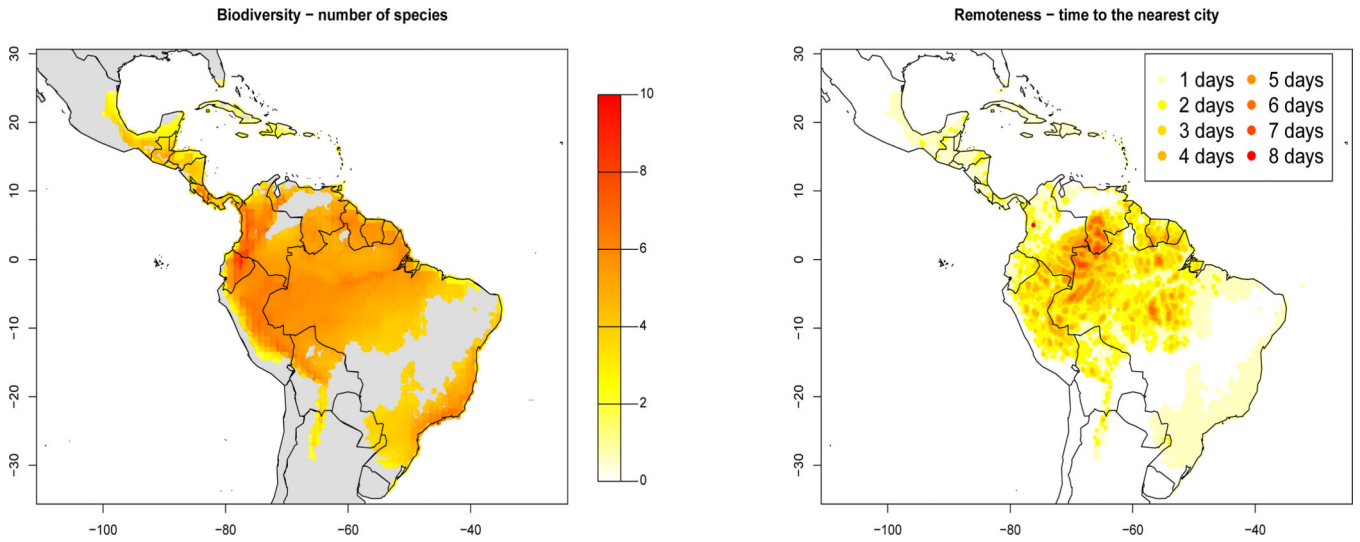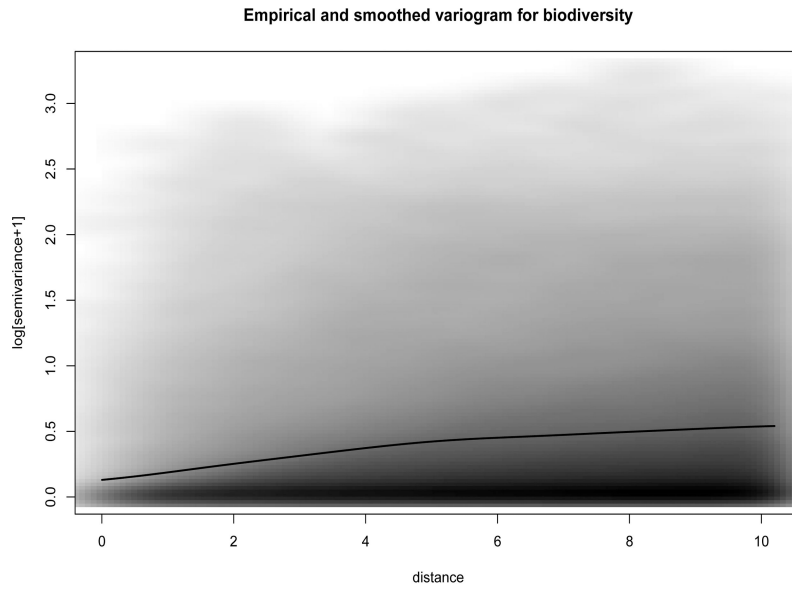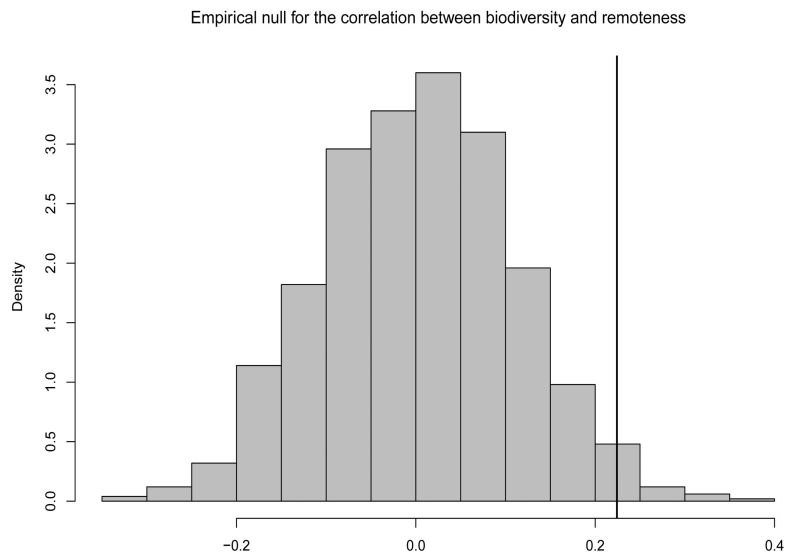
(a) Biodiversity

(b) Remoteness

(c) Local correlations

Figure 1.
Top right: Biodiversity as a function of domain. Top left: Remoteness as a function of domain. Bottom: Local correlations between biodiversity and remoteness using a Gaussian kernel with bandwidth $\lambda = 5.281$ (see Appendix 2), where the blue circles indicate the extent of the neighborhood for 5 locations at random. The grey areas correspond to areas with no data.

Empirical and smoothed variogram for biodiversity

(a) Empirical and smoothed variogram for biodiversity

Empirical null for the correlation between biodiversity and remoteness

(b) Null distribution estimated with our algorithm

Figure 2.
Top: empirical (scatter plot in grey scale) and smoothed (black line) variogram for biodiversity in a logarithmic scale. The variograms are truncated at distance 10 and the smoothed variogram is calculated using a Gaussian kernel with bandwidth $h = 0.746$ (see Appendix 1). The intensity of the grey color in the empirical variogram indicates the density of the data. Bottom: Null distribution of $r_{\mathbf{X_s}, \mathbf{Y_s}}$ for the biodiversity dataset estimated with our algorithm. The black line corresponds to the observed correlation $\hat{r}_{\mathbf{X_s}, \mathbf{Y_s}} = 0.224$. The p-value is 0.057.
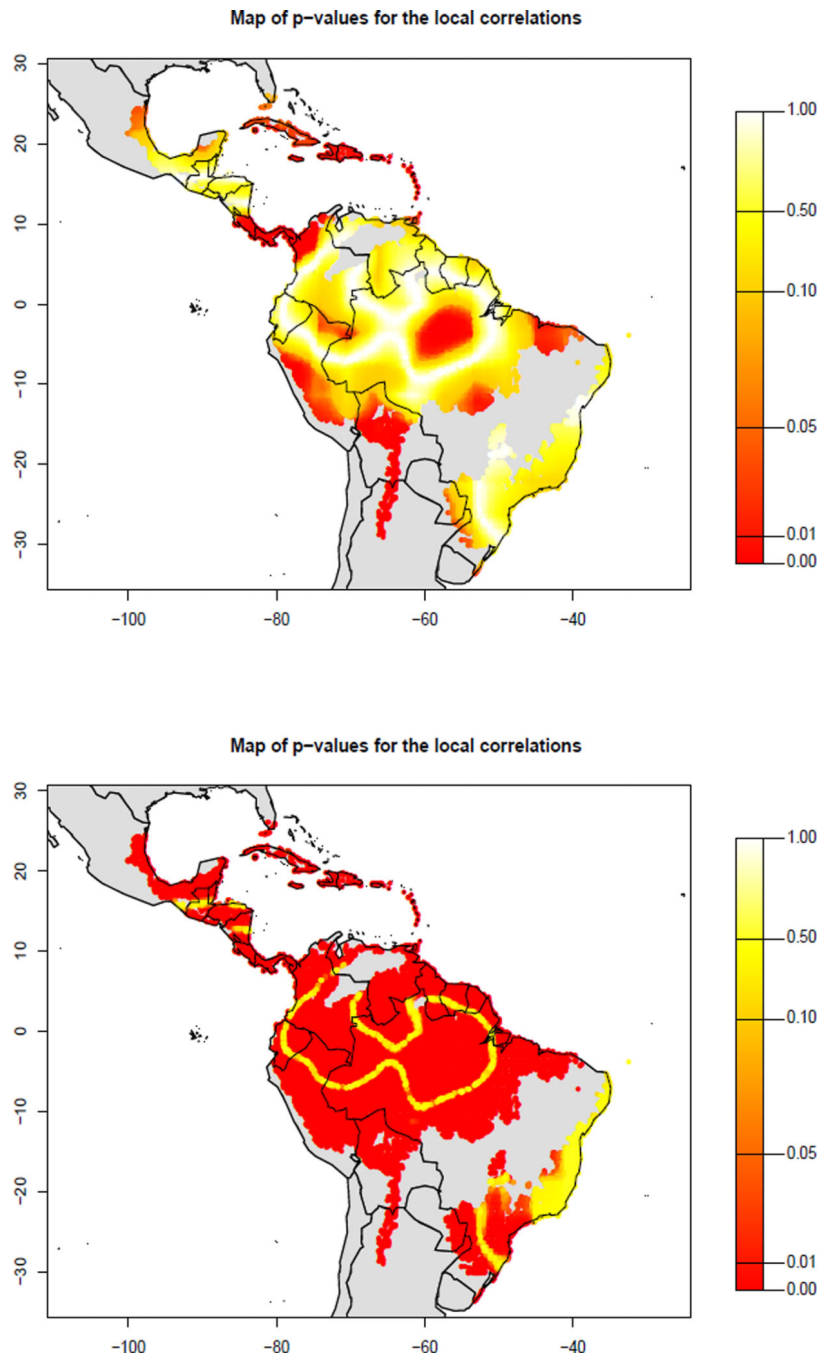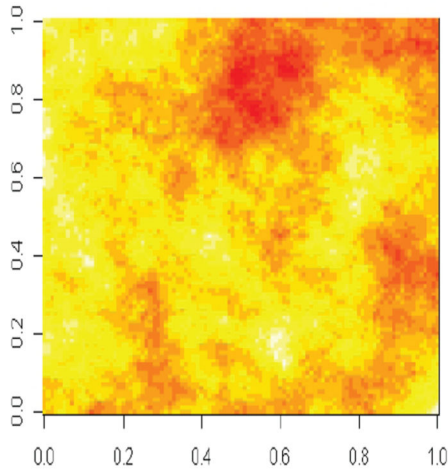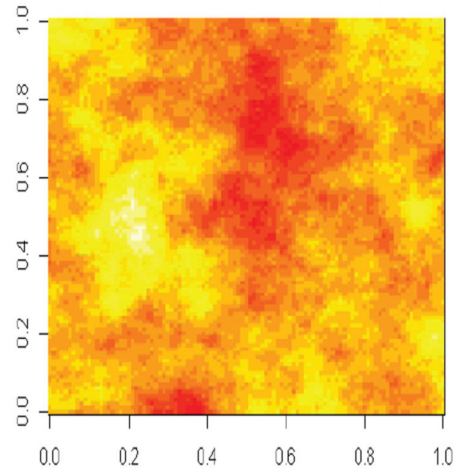
Figure 3.
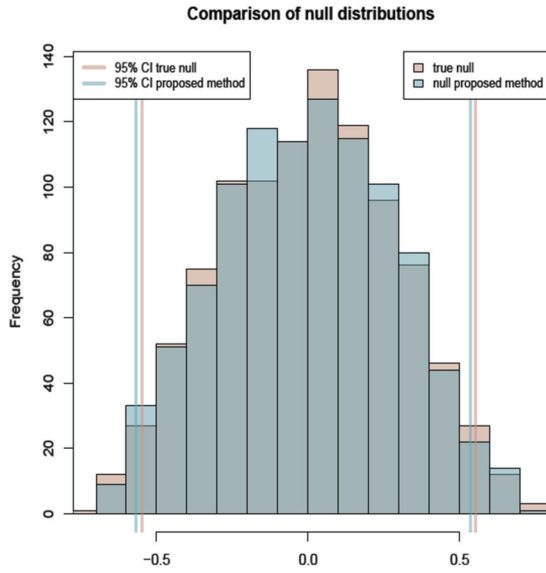Top: map of p-values for the correlations between biodiversity and remoteness in Figure
1(c), obtained with our method with $B = 1000$ and $\in (0.005, 0.785)$. Bottom: p-values map
if we assume that there is no spatial autocorrelation and use the standard test to assess the
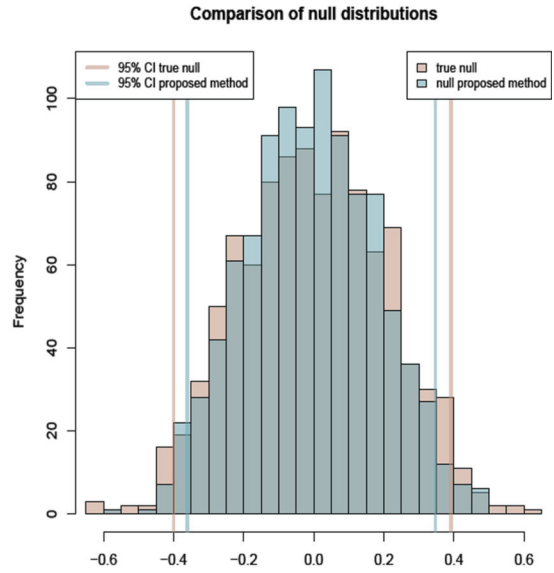local correlations.

(a) Gaussian random field $\mathbf{X_s}$

(b) random field $\mathbf{Y_s}$, independent of $\mathbf{X_s}$

(c) Gaussian autocorrelation function.

(d) exponential autocorrelation function.

Figure 4.
Top, left and right: Illustration of $\mathbf{X_s}$ and $\mathbf{Y_s}$, two independent realizations of a Gaussian random field with exponential autocorrelation function with $\varphi = 0.03$ and grid size $g = 0.01$. Bottom, left and right: null distribution for $r_{\mathbf{X_s}, \mathbf{Y_s}}$ returned by the algorithm (blue), for a Gaussian and exponential autocorrelation function respectively, together with the true null (pink) obtained by simulating from model (2). The corresponding 95% Confidence Intervals are added to the plot.
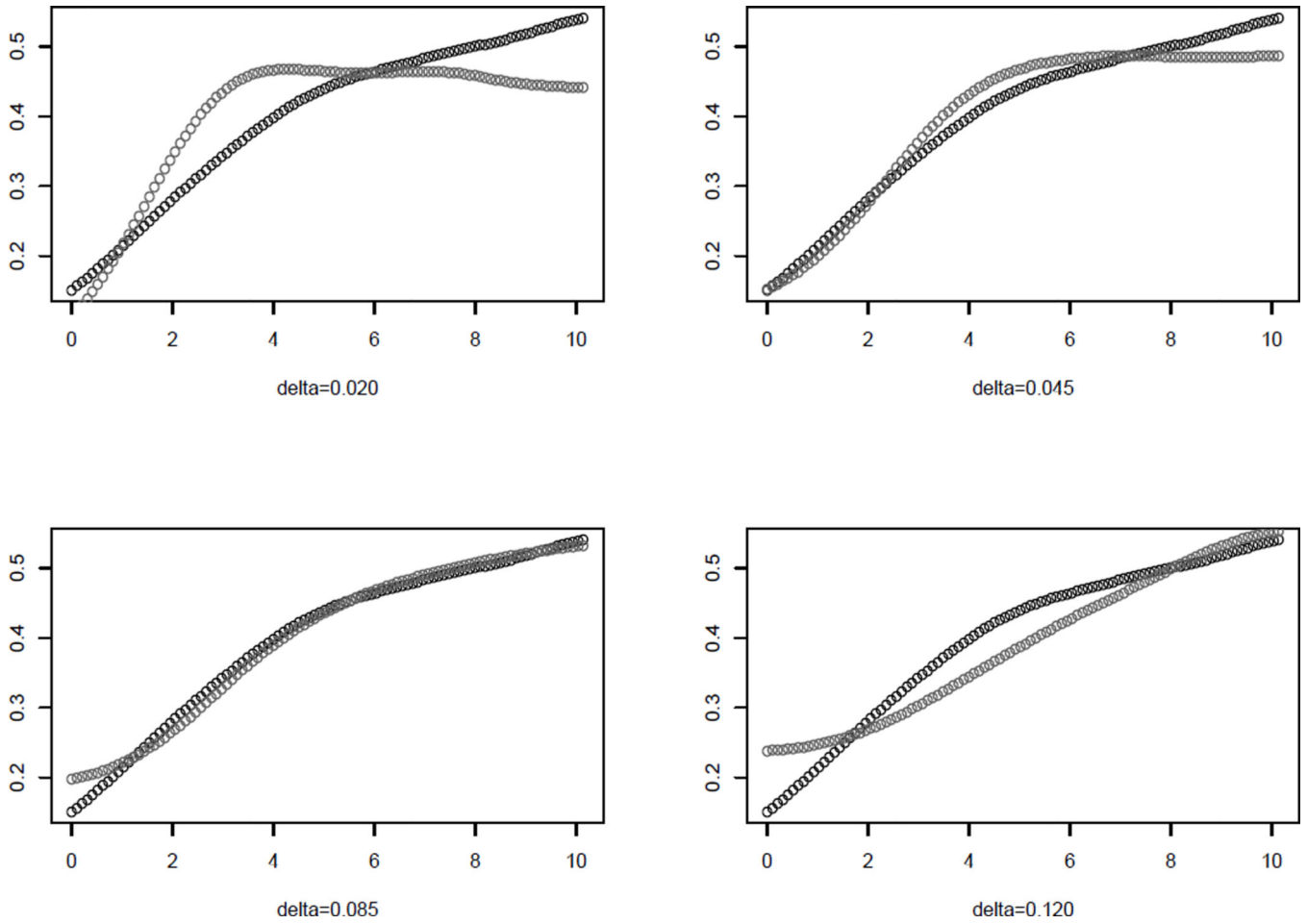
Figure 5.
Variogram matching between the original biodiversity variable $\mathbf{X_s}$ (black) and the transformed variable $\hat{\mathbf{X}}_{\mathbf{s}}^{\delta}$ (grey) for 4 different values of bandwidth: $= (0.020, 0.045, 0.085, 0.120)$. The best match is reached with $\delta = 0.085$.

Table 1

Type I error estimates obtained by contrasting pairs simulated under $H_0 : \rho_{X_S, Y_S} = 0$, to the null distribution returned by our algorithm and Clifford's method, in addition to the true null obtained by simulating independent pairs under model (2). Results are presented for different levels of autocorrelation and resolution: $\varphi \in (0.05, 0.1, 0.3)$ and $g \in (0.05, 0.01)$. The standard errors are in brackets.

	Type I Error Estimates			
	$\varphi = 0.05$	$\varphi = 0.1$	$\varphi = 0.3$	
grid size = 0.05	0.052 (0.024)	0.046 (0.022)	0.048 (0.024)	true
	0.059 (0.029)	0.056 (0.035)	0.051 (0.041)	ours
	0.055 (0.027)	0.055 (0.029)	**0.093** (0.045)	Clifford
grid size = 0.01	0.053 (0.027)	0.049 (0.023)	0.051 (0.022)	true
	0.075 (0.030)	0.051 (0.027)	0.050 (0.045)	ours
	0.055 (0.027)	0.058 (0.026)	**0.086** (0.044)	Clifford

Table 2

Power estimates obtained by contrasting pairs simulated under $H_1 : \rho_{X_s, Y_s} = \rho$, to the null distribution returned by our algorithm and Clifford's method, in addition to the true null obtained by simulating correlated pairs under model (2). Results are presented for different levels of autocorrelation, effect size and resolution: $\varphi \in (0.05, 0.1, 0.3)$, $\rho \in (0.2, 0.5, 0.8)$ and $g \in (0.05, 0.01)$.

		Power Estimates			
		φ = 0.05	**φ = 0.1**	**φ = 0.3**	
grid size = 0.05	ρ = 0.2	0.921 (0.026)	0.418 (0.059)	0.103 (0.034)	true
		0.926 (0.031)	0.444 (0.087)	0.089 (0.042)	ours
		0.926 (0.030)	0.446 (0.077)	**0.153** (0.045)	Clifford
	ρ = 0.5	1 (0)	0.997 (0.005)	0.426 (0.055)	true
		1 (0)	0.997 (0.005)	0.420 (0.089)	ours
		1 (0)	0.997 (0.005)	**0.558** (0.100)	Clifford
	ρ = 0.8	1 (0)	1 (0)	0.951 (0.021)	true
		1 (0)	1 (0)	0.879 (0.127)	ours
		1 (0)	1 (0)	0.950 (0.045)	Clifford
grid size = 0.01	ρ = 0.2	0.915 (0.033)	0.405 (0.052)	0.104 (0.030)	true
		0.935 (0.031)	0.404 (0.073)	0.124 (0.069)	ours
		0.927 (0.030)	0.415 (0.064)	**0.169** (0.046)	Clifford
	ρ = 0.5	1 (0)	0.996 (0.007)	0.424 (0.059)	true
		1 (0)	0.996 (0.007)	0.391 (0.154)	ours
		1 (0)	0.996 (0.007)	**0.510** (0.111)	Clifford
	ρ = 0.8	1 (0)	1 (0)	0.937 (0.022)	true
		1 (0)	1 (0)	0.949 (0.030)	ours
		1 (0)	1 (0)	**0.972** (0.018)	Clifford