



Randomization Tests and Spatially Auto-Correlated Data

Author(s): Marie-Josée Fortin and Geoffrey M. Jacquez

Reviewed work(s):

Source: *Bulletin of the Ecological Society of America*, Vol. 81, No. 3 (Jul., 2000), pp. 201-205

Published by: [Ecological Society of America](#)

Stable URL: <http://www.jstor.org/stable/20168439>

Accessed: 17/09/2012 21:39

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Ecological Society of America is collaborating with JSTOR to digitize, preserve and extend access to *Bulletin of the Ecological Society of America*.

<http://www.jstor.org>

- Deming, W. E. 1975. On probability as a basis for action. *American Statistician* **29**:146–152.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187–211.
- Illich, I. 1976. *Limits to medicine*. Penguin Books, Hammondsworth, UK.
- Johnson, D. H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* **63**:763–772.
- Johnson, R. R., K. F. Higgins, D. E. Naugle, and J. A. Jenks. 1999. A comparison of sampling techniques to estimate number of wetlands. *Wildlife Society Bulletin* **27**:103–108.
- Lawton, J. 1999. Size matters. *Oikos* **85**:19–21.
- Magnusson, W. E. 1997. Teaching experimental design in ecology, or how to do statistics without a bikini. *ESA Bulletin* **78**:205–209.
- Moore, D. S. 1997. Bayes for beginners? Some reasons to hesitate. *American Statistician* **51**:254–261.
- Petersen, J. E., J. C. Cornwell, and W. M. Kemp. 1999. Implicit scaling in the design of experimental aquatic ecosystems. *Oikos* **85**:3–18.
- Stern, M. J. 1998. Field comparisons of two rapid vegetation assessment techniques with permanent plot inventory data in Amazonian Peru. Pages 269–283 in F. Dallmeier and J. A. Comiskey, editors. *Forest biodiversity research, monitoring and modeling*. UNESCO and Parthenon Publishing, Paris, France.

William E. Magnusson
 Coordenação de Pesquisas
 em Ecologia
 Instituto Nacional De Pesquisas
 da Amazônia
 Caixa Postal 478
 69011-970 Manaus AM
 Brasil
 E-mail: bill@inpa.gov.br

EXPLORING DEMOGRAPHIC AND GENETIC STOCHASTICITY WITH DICE

I found the article by Donovan and Goetz (1999), "Exploring demographic and environmental stochasticity with spreadsheets," very interesting and would like to relate a class period I devoted in my Introduction to Ecology class in which we explored these concepts. A crucial feature in any "simulation" is a random number generator, and the shake of a die is an ideal generator of the numbers 1 through 6.

Using my class of 30 students, I pulled two of them to the front of the room. A simulation timestep is as follows. Each student shakes a die. If a student's die comes up 1 or 2, he or she reproduces and pulls another student up to the front of the room. If a student's die comes up 6, the student dies. (Hence the net reproductive rate is 7/6.) On the way back to his or her seat, each student places the appropriate chalk mark on developing histograms of offspring produced and lifetime (number of shakes). The simulation can continue until either all or no students are standing. I acted as director, telling students to shake. Students that died took care of themselves. Reproducing students raised a hand, which I counted, then I asked that many students to join the population. Then I counted the total number of students and wrote that number on the board. Once the students have the hang of it, a timestep takes about 20–30 seconds, allowing two replicate simulations in roughly 20 minutes. Before class, I performed computer simulation runs, (much like Donovan and Goetz 1999) and had a teaching assistant plot the class's population counts on a transparency to overlay onto the computer results.

This class simulation very clearly demonstrated to the students the concept of demographic stochasticity: Each student lived several lives of various duration and reproductive out-

put. Likewise, the population trajectories were very different from the deterministic expectation for these low population sizes (at least in my replicates), but in my computer simulations I ran replicates with initial population sizes of 200. These large populations were very close to the deterministic expectations.

Finally, I could discuss the importance of genetic stochasticity in small populations, with the students having seen some individuals producing no offspring, and others producing five or six offspring. Environmental stochasticity can be described as changing the die outcomes leading to birth and death each timestep.

Literature cited

- Donovan, T. M., and J. E. Goetz. 1999. Exploring demographic and environmental stochasticity with spreadsheets. *ESA Bulletin* **80**: 235–239.

Will Wilson
 Department of Zoology
 Duke University
 Durham, NC 27708

RANDOMIZATION TESTS AND SPATIALLY AUTO-CORRELATED DATA

Introduction

An attractive feature of randomization tests is that statistical significance is evaluated based on empirical distributions generated from the observed sample. This "distribution-free" alternative to parametric statistics is quite appealing to ecologists faced with small data sets that do not meet the assumed parametric distribution, and for which requisite assumptions of independence are violated by spatial and/or temporal autocorrelation. However, while randomization tests may involve fewer assumptions (Crowley 1992): "fewer

assumptions” does not mean “no assumptions.” In addition, the choice of randomization test determines the statistical null and alternative hypotheses, and test implementation can have subtle and often unintended impacts on the hypotheses tested. Depending on the test chosen, and its implementation, ecologists may—or may not—correctly test the ecological hypothesis they originally intended to analyze (Simberloff 1990, 1993).

With recent advances in the power of personal computers, reference distributions of statistics are increasingly generated using randomization methods. However, caution must be exercised because the permutation procedures often assume data independence, and this assumption is invalid when the data are spatially or temporally autocorrelated (Noreen 1989, Good 1993, Edgington 1995, Manly 1997). This has caused some authors (Sokal et al. 1993, Fortin et al. 1996) to use restricted randomization to account for the spatial structure. The aim of this paper is to show how randomization tests can be used with spatially autocorrelated data that violate the independence assumption of randomization procedures. We offer some ways to perform restricted randomization when analyzing spatially dependent data.

Ecological data

Ecological data taken in the field or arising from animal behavior studies are usually characterized by small sample sizes, lack of normality, and by spatial and temporal autocorrelation. Spatial autocorrelation refers to the pattern in which observations from nearby locations are more likely to have similar magnitude than by chance alone (Legendre and Fortin 1989), and its intensity and extent can be quantified using spatial statistics (Cressie 1991). Issues of small sample size and departures from normality are not new, and nonparametric methods have been developed to address these problems (Sprent 1993). The problem of accounting for spatial autocorrelation in the data when conducting statistical inference has received attention only recently (Legendre and Fortin

1989, Legendre et al. 1990, Legendre 1993, Sokal et al. 1993). In fact, almost all ecological data exhibit some degree of spatial dependence, depending on the scale at which the data were recorded and then analyzed. Furthermore, ecological phenomena are characterized by the fact that more than one ecological process can act upon them, and at more than one spatial scale. Ecological data are a composite of several spatial scales: trends at macroscales; patches, gradients and patterns at meso- and local scales; and random patterns at local and microscales. These different processes and patterns at different scales are not necessarily linear and additive, which contributes to the degree of spatial dependence in the data.

Spatial autocorrelation typically has four sources, of which two are of direct scientific interest. Griffith (1992) distinguished among *spurious*, *interpolation*, *true*, and *induced* autocorrelation. Even when the observations are statistically independent, spurious autocorrelation may be observed. Interpolation autocorrelation arises when spatial response surfaces are smoothed, interpolated, or extrapolated. True spatial autocorrelation arises from causal interaction among nearby sample locations. Finally, spatial autocorrelation may be induced in a dependent variable through a causal relationship with another spatially autocorrelated variable. Spurious and interpolation autocorrelation are a nuisance. True and induced spatial autocorrelation arise from space-time processes (e.g., diffusion, migration, etc.) that are of direct scientific interest. This motivated researchers to use spatial autocorrelation as a clue or “signature” left by the past action of space-time processes (Legendre and Fortin 1989).

Because geography underlies most ecological interactions, spatial autocorrelation is a property inherent in almost all ecological data. In ecology, spatial autocorrelation is a by-product of ecological processes that possess a geographic component such as dispersal, allelopathy, and spatial competition for resources. This inherent property has fundamental statistical

implications: it violates assumptions of independence required by many parametric inferential tests (Legendre et al. 1990). To insure independent samples, the use of a random sampling design is often recommended. This design, however, will not remove true or induced spatial autocorrelation (Fortin et al. 1989).

Temporal dependence often occurs at both ecological and evolutionary scales. Hence, ecological data often evince annual, seasonal, and daily cycles, so observations made at about the same time tend to be similar. In addition, observations on individuals, such as height and mass, typically depend on previously observed values. Finally, hypotheses regarding community structure and species co-occurrence necessarily imply some degree of spatiotemporal dependence. This property of ecological data generates substantial statistical problems. These have been recognized for some time by ecologists and are a source of on-going debate and research (see, among others, Caswell 1976, Jackson and Somers 1988, Simberloff 1993, Gotelli and Graves 1996).

A similar problem occurs in the analysis of evolutionary data: subspecies are more likely to have comparable genotypes and phenotypes, and genetic dependence must also be accounted for. This concern has been acknowledged among evolutionary biologists, and several methods of restricted randomization of species characters have been developed (among others, Gittleman and Kot 1990, Gittleman and Luh 1992).

Resampling techniques

Several terms are used interchangeably when referring to resampling techniques, including “computer-intensive methods,” “Monte Carlo,” “permutation,” “randomization,” and “bootstrap.” In fact, these techniques are not interchangeable, and differ fundamentally in terms of their statistical mechanics. Here, we identify the main purpose of each technique.

The generic terms “resampling techniques” and “computer-intensive methods” refer to all methods in

which observed data are used to generate a reference distribution by means of randomization (Simon 1995). This reference distribution is then used to assess the significance of a statistic calculated from the observed (not randomized) data. Significance is evaluated under the assumption that the statistic computed using the observed data is sampled from the reference distribution generated with randomization technique. Resampling techniques have fewer assumptions than traditional statistics, which are based on distribution and probability theories, and support the analysis of ecological data with an unknown or an analytical undefined distribution.

“Monte Carlo” is a general term that refers to any test using a random-number generator. Applications include risk assessment and the propagation of uncertainty. Specifically, Monte Carlo simulations draw random samples from a modeled population, with probability distributions estimated from a real population, and test whether the observed sample was randomly drawn from that real population. As such, “resampling” is accomplished in the process of generating samples from the modeled population. This framework supports inference at the population level, with the caveat that inference is premised on the appropriateness of the model. “Bootstrap” is also another resampling technique that assumes the observations can be independently exchanged, and performs a sampling with replacement from an infinite population (Noreen 1989). Unlike randomization tests, statistical inference for the bootstrap applies to the population. “Jackknife” is a technique whereby a reference distribution is enumerated by repeatedly leaving out one observation at a time and recalculating the test statistic.

Randomization tests

The terms “permutation test” and “randomization test” are often used interchangeably. Formally, Fisher (1935) used the term “permutation test” to refer to a method that performed population-level inference, whereas “randomization test” was applied to meth-

ods for sample-based inference. The term “permutation test,” or “exact randomization,” is also used when the test provides the exact significance levels by exhaustive computation of all possible rearrangements (permutations) of the data. In practice, even with the current powerful computers, it is possible to perform the permutation test only when the number of observations, n , is small because the number of permutations increases as the factorial of the sample size ($n!$). For example, a sample size of seven implies 5040 possible permutations; with only three more observations, a sample size of 10 implies 3,628,800 permutations!

When the number of observations precludes an exact test, an “approximate randomization” test is used instead. This test generates a subset of the possible permutations. Because only a subsample of all possible permutations is calculated, many authors recommend that $\geq 10,000$ randomizations be used when constructing the reference distribution (Crowley 1992, Manly 1997). The reference distributions of any statistics are obtained using a five-step randomization procedure (Noreen 1989) that repeatedly reallocates the value of the observations over the sample, and then recalculates the statistic to generate the null reference distribution:

- 1) define the ecological question or hypothesis;
- 2) choose a statistic (already existing or design a new one);
- 3) compute the statistic for the observed data (call this T^*);
- 4) generate the null reference distribution by:
 - a) rearranging the order of observed data over the entire sample by shuffling it randomly (i.e., values of the response variable are shuffling over all the samples where each sample keeps its spatial, temporal or phylogenetic identity);
 - b) compute the statistic for data so randomized; and
 - c) repeat this step k times (e.g., 10,000 or more times according to the number of observations analyzed); and
 - 5) compare the observed statistic to the reference distribution.

Statistical inference in this framework is based on a comparison of the T^* statistic to the reference distribution. The statistical decision of accepting or rejecting the null statistical hypothesis has a corresponding ecological interpretation, and the final step is to accomplish this ecological interpretation. The number of randomizations, k , determines the significance level that can be resolved by the test. For example, a reference distribution based on $k = 100$ values can resolve a $P \leq 0.01$; for 1000 values $P \leq 0.001$; and so on.

With small sample size, randomization tests can produce an odd reference distribution. For example, at sample size $n = 5$, there are only 120 possible permutations. Hence, when the reference distribution is built using 10,000 randomizations, it can lead to the uneven repetition of some permutations, resulting in misleading probability values.

Randomization tests have several advantages, including flexibility and relative ease of implementation. They support significance testing without distributional assumptions (e.g., normality) and complex designs for which parametric tests do not exist. For comparable statistics, randomization tests are as powerful as parametric tests when the number of randomizations is large.

Some authors limit use of the term “randomization” to the initial randomization of treatment assignments in an experimental design, and use the term “re-randomization” for resampling techniques. Here, we use the term “randomization test” in reference to sample-based tests where observed data are used to generate the reference distribution (i.e., “resampling” techniques). Knowledge of the population from which the sample was taken is not used, and statistical inference, therefore, is limited to the sample.

Restricted randomization for spatially autocorrelated data

There are always two sides to a coin, and the flexibility of randomization tests is offset by the relatively

high level of understanding required to access this flexibility. For example, these randomization tests assume that data are exchangeable, so rearrangements (reordering, shuffling) of the samples are equally likely (Good 1993). This implies, therefore, that the observations are independent. Such assumption of exchangeability is violated, however, when data show an inherent structure, as is the case with spatially or temporally autocorrelated data and phylogenetic data. With such structured data, "restricted randomization" can be performed, by which the randomization procedure is restricted to take into consideration the inherent dependence (spatial, temporal, or phylogenetic) that might exist among the values of the samples.

Because ecological data inherently are spatially autocorrelated (Legendre 1993), what are the implications of blindly using randomization tests? The statistical null hypothesis,

under randomization tests implied by random shuffling, is that the observations are spatially independent. Translated in ecological terms, this means that under the test's null hypothesis, there is no "horizontal" spatial pattern, i.e., no spatial autocorrelation at the variable or species level (Fig. 1a), such that the observations can be shuffled independently of one another. Hence, randomization tests based on complete random shuffling are not appropriate for evaluating ecological hypotheses for which spatial autocorrelation is present. Generally, these kinds of spatial interactions must be incorporated into the null hypothesis, and thus in the statistical analysis, because they form the fabric against which specific ecological hypotheses are evaluated. This is why restricted randomization procedures have been developed (Manly 1997), and they are proposed here as an alternative to

randomization tests per se. Here, we present some of the restricted randomization procedures for spatially autocorrelated data.

Regional partitions.—When it is natural to partition the spatial field into several regions, one may randomize within but not across regions. This procedure assumes that an observation from one region is equally likely to have occurred at any other location within that region, and that it has zero probability of occurring at locations in other regions (see, for example, Sokal et al. 1993). Restricted randomization within regions implies some degree of spatial structure at the regional scale (horizontal interactions), but no spatial pattern within regions, as well as no vertical interactions among variables or species (Fig. 1b). In population genetics, this is a natural choice when a population is divided into spatially distinct subpopulations that do not interbreed.

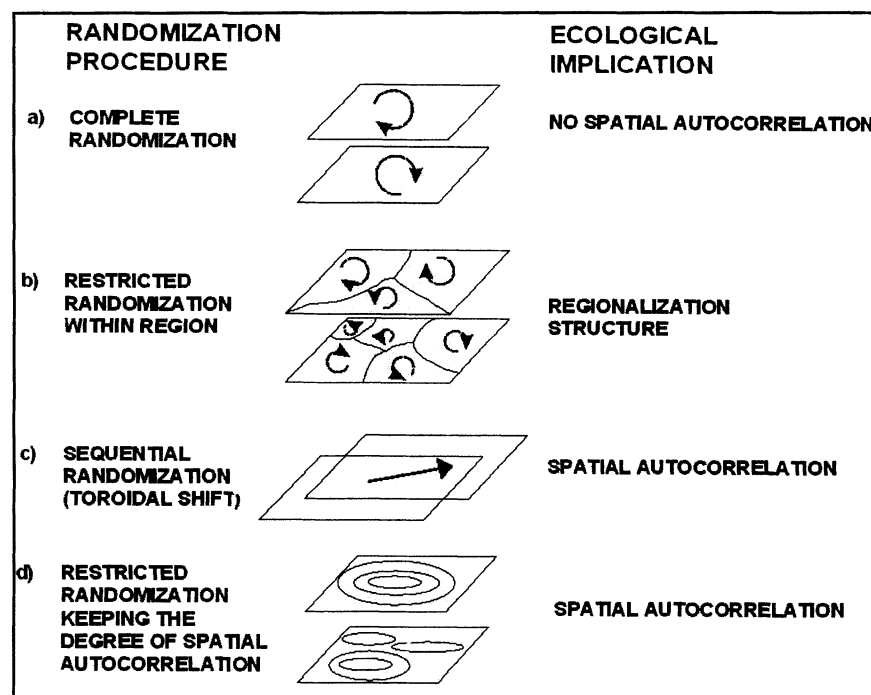


Fig. 1. Ecological implications of the different randomization procedures for the correlation between two spatially autocorrelated variables. Here each variable sampled over a two-dimensional area is represented by a layer. (a) Randomization that implies no spatial structure among variables or species; (b) Restricted randomization within region, which implies some degree of spatial structure at the regional scale but no spatial structure within regions; (c) Restricted randomization keeping the sequence of the data fixed by doing a toroidal shift (i.e., the spatial pattern at the variable or species is preserved); and (d) Restricted randomization based on the degree of spatial autocorrelation of the observed data.

Toroidal shift.—Another type of restricted randomization procedure is to keep the sequence, the spatial or temporal structure, of the data fixed and then slide the "map" of one variable over the other. For spatial data, this is achieved using a two-dimensional torus constructed by connecting the map margins and then sliding one variable map over the other (Fig. 1c). This procedure maintains the horizontal interactions, and therefore the spatial autocorrelation at the variable or species level, but not the vertical structure. Depending on the number of restricted randomizations performed given the shape and size of the study area, this test can be too liberal (Fortin et al. 1996).

To maintain spatial autocorrelation.—A better way to insure that the horizontal spatial structure is preserved is to perform a restricted randomization based on the degree of spatial autocorrelation in the observed data, using geostatistics' conditional simulation algorithm (Cressie 1991). This will generate restricted randomizations having the same degree of spatial autocorrelation as the

observed one (Fig. 1d). Potential disadvantages of this method include excessive computing time.

Conclusion

Randomization tests open a wide range of possibilities for testing specific hypotheses with complex, dependent ecological data. However, depending on the type of randomization test employed, an ecologist may or may not be testing the ecological hypothesis that he or she intended. This can result in arrival at the wrong statistical and ecological interpretations. The above discussion and classification of randomization procedures illustrates that randomization tests in ecology are particularly prone to mis-specification of the null hypothesis, primarily because the null hypothesis is embedded in the randomization procedure and is not self-evident. For example, complete randomization such that all observations are equiprobable corresponds to a statistical null hypothesis of Complete Spatial Randomness (CSR). In many instances, such a null hypothesis is ecologically ridiculous, and restricted randomization may be used instead. However, care must be exercised to assure that the null hypothesis implied under restricted randomization is ecologically tenable. A firm understanding of the null and alternative hypotheses of the chosen randomization test is required in order to assure that the biological and ecological questions under study are correctly addressed.

Acknowledgments

We thank the Statistical Ecology Section of ESA for sponsoring the Symposium, "Permutation tests in ecology: how should we test ecological hypotheses?" at the 1996 ESA Annual Meeting in Providence, Rhode Island. This research was funded by an NSERC grant to M.-J. Fortin, and by SBIR grants R43 CA69864 and R44 CA65366 and a grant from the National Cancer Institute to G. M. Jacquez.

Literature cited

- Caswell, H. 1976. Community structure: a neutral model analysis. *Ecological Monographs* 46:327–354.
- Cressie, N. A. C. 1991. *Statistics for spatial data*. Wiley, New York, New York, USA.
- Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecology and Systematics* 23:405–447.
- Edgington, E. S. 1995. *Randomization tests*. Third edition. Marcel Dekker, New York, New York, USA.
- Fisher, R. A. 1935. *The design of experiments*. Oliver and Boyd, Edinburgh, UK.
- Fortin, M.-J., P. Drapeau, and G. M. Jacquez. 1996. Quantification of the spatial co-occurrences of ecological boundaries. *Oikos* 77:51–60.
- Fortin, M.-J., P. Drapeau, and P. Legendre. 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio* 83:209–222.
- Gittleman, J. L., and M. Kot. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. *Systematic Zoology* 39:227–241.
- Gittleman, J. L., and H. K. Luh. 1992. On comparing comparative methods. *Annual Review of Ecology and Systematics* 23:383–404.
- Griffith, D. A. 1992. What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *L'Espace géographique* 3:265–280.
- Good, P. 1993. *Permutation tests: a practical guide to resampling methods for hypothesis testing*. Springer-Verlag, New York, New York, USA.
- Gotelli, N. J., and G. R. Graves. 1996. *Null models in ecology*. Smithsonian Institution Press, Washington, D.C., USA.
- Jackson, D. A., and K. M. Somers. 1988. Are probability estimates from the permutation model of Mantel's test stable? *Canadian Journal of Zoology* 67:766–769.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–1673.
- Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80:107–138.
- Legendre, P., R. R. Sokal, N. L. Oden, A. Vaudor, and J. Kim. 1990. Analysis of variance with spatial autocorrelation in both the variable and the classification criterion. *Journal of Classification* 7:53–75.
- Manly, B. F. J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. Second edition. Chapman and Hall, London, UK.
- Noreen, E. 1989. *Computer-intensive methods for testing hypotheses*. Wiley, New York, New York, USA.
- Simberloff, D. 1990. Hypotheses, errors, and statistical assumptions. *Herpetologica* 46:351–357.
- Simberloff, D. 1993. Competition theory, hypothesis-testing, and other community ecological buzzwords. *American Naturalist* 122:626–635.
- Simon, J. 1995. *Resampling: the new statistics*. Resampling Stats, Arlington, Virginia, USA.
- Sokal, R. R., N. L. Oden, B. A. Thomson, and J. Kim. 1993. Testing for regional differences in means: distinguishing inherent from spurious spatial autocorrelation by restricted randomization. *Geographical Analysis* 25:199–210.
- Sprent, P. 1993. *Applied nonparametric statistical methods*. Second edition. Chapman and Hall, London, UK.

Marie-Josée Fortin and
Geoffrey M. Jacquez
School of Resource and
Environmental Management
Simon Fraser University
Burnaby, British Columbia,
Canada V5A 1S6
and
BioMedware
516 North State Street
Ann Arbor, MI 48104–1236 USA