

Article

## An Entropy-Based Contagion Index and Its Sampling Properties for Landscape Analysis

Bernard R. Parresol \* and Lloyd A. Edwards

USDA Forest Service, Southern Research Station, Asheville, NC 28806, USA;

E-Mail: laedwards@fs.fed.us

\* Author to whom correspondence should be addressed; E-Mail: bparresol@fs.fed.us;  
Tel.: +1-828-667-5261; Fax: +1-828-667-9097.

Received: 6 July 2012; in revised form: 18 March 2014 / Accepted: 19 March 2014 /

Published: 26 March 2014

---

**Abstract:** Studies of spatial patterns of landscapes are useful to quantify human impact, predict wildlife effects, or describe various landscape features. A robust landscape index should quantify two components of landscape diversity: composition and configuration. One category of landscape index is the contagion index. Some landscape ecologists promote the use of relative contagion indices. It is demonstrated, using simulated landscapes, that relativized contagion indices are mathematically untenable. A new entropy contagion index ( $\Gamma$ ) is developed. Distributional properties of  $\hat{\Gamma}$  are derived. It is shown to be asymptotically unbiased, consistent, and asymptotically normally distributed. A variance formula for  $\hat{\Gamma}$  is derived using the delta method. As an application, the pattern and changes in forest types across four soil-geologic landform strata were analyzed on the 80,000 ha Savannah River Site in South Carolina, USA. One-way analysis of variance was used for hypothesis testing of contagion among strata. The differences in contagion across the strata provide insight to managers to meet structural objectives.

**Keywords:** concentration; delta method; expected value; forest types; geometric distribution; hypothesis testing; simulated landscapes

**MSC Codes:** 62E20; 62F03; 62H35

---

## 1. Introduction

Landscape diversity refers to the various ecosystems (including human, e.g., cities, farm country, *etc.*) within a large area. The structure observed in landscapes result from complex interactions between physical, biological, and social forces [1]. Agencies involved with natural resources have been expanding current management approaches to address landscape-level concerns and issues [2]. With the recognition of such problems as loss of biodiversity, climate change, and ecosystem degradation, a necessary evolution in management scale has come about, based on watersheds or other ecological landscape scale units rather than political boundaries. A quantitative basis for measuring spatial structure is a prerequisite to implementing landscape management. Without such, structural objectives cannot be established nor can the understanding of spatial dynamics necessary to achieve structural objectives be realized.

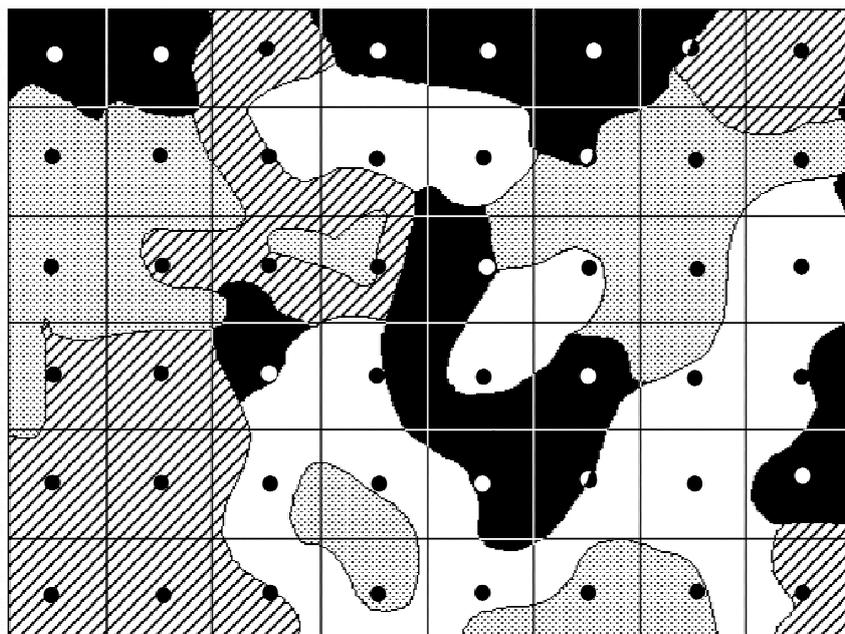
A landscape contagion index is a quantitative metric, a single statistic applied on a broad spatial scale in which two distinct components are confounded: composition and configuration. Composition refers to both the total number of land cover categories or “patch” types and their relative proportions in the landscape, whereas configuration refers to the spatial pattern of patches in the landscape [3]. Contagion, as defined by O’Neill *et al.* [4], measures the extent to which landscape elements are aggregated or clumped. Higher values of contagion generally result from landscapes with a few large, contiguous patches, whereas lower values usually characterize landscapes with many small patches. Also, holding the number of categories more or less constant, contagion values, in general, should decrease as category proportions become more even.

As a statistic, a contagion index is limited if it cannot be used in making comparisons of diversities among different landscapes or the same landscape through time based on sample data. If the land cover categories are the same between two landscapes, then contingency table analysis could be used to test if relative proportions of categories are the same between the landscapes. This would answer the composition question, but not the configuration question. Spatial sampling is typified by systematic sampling at lattice intersection points. As one moves along transects from point to point, transitions are made from patch type to patch type (Figure 1). The sequences of transitions can be regarded as realizations of a Markov chain. A matrix of transition probabilities can be constructed for each landscape and the null hypothesis that the matrices of transition probabilities are the same can be tested with a log likelihood ratio test [5]. However, if the number or kinds of patch types differ among landscapes, how then does one test for differences in landscape diversity? This same problem faced ecologists with species diversity indices. Hutcheson [6] devised an analysis of variance (ANOVA) type test for the Shannon species diversity index by investigating the distributional properties and sampling variance of the index. Moran [7] gives the sampling variance of a join-count statistic for testing spatial dependence from simple binary classifications of landscapes. The sampling properties of a landscape index must be known before one can construct an appropriate test. Is the index unbiased? Is it consistent? Can the variance be computed or approximated? Is the sample distribution normal? A formal investigation is necessary to answer these questions.

This paper develops a correctly specified entropy-based contagion index and provides a statistical basis for its use in hypothesis testing. There were a number of specific objectives of this research. First, to show using simulated landscapes, that the currently used relativized contagion index possesses

mathematically undesirable behavior. Second, to derive an index that exhibits desirable mathematical behavior. Third, study the sampling properties of the new index. And fourth, apply the new index to forested landscapes looking at contagion of forest types on the four soil-geologic landform strata of the Savannah River Site in South Carolina.

**Figure 1.** A four-category landscape mosaic sampled with a lattice of plots (circles) and overlaid with a grid to aid in computation of join counts.



## 2. Relative Contagion

Li and Reynolds [3] defined relative contagion ( $RC$ ) as:

$$RC = 1 - EE / EE_{\max} \quad (1)$$

where  $EE$  denotes the entropy value. The well-known measure of entropy for categorical data, derived by Shannon [8], is  $-\sum p_i \ln(p_i)$ . Based on Equation (1), Li and Reynolds gave the following index:

$$RC = 1 + \frac{\sum_{i=1}^n \sum_{j=1}^n p_{ij} \ln(p_{ij})}{2 \ln(n)} \quad (2)$$

where  $p_{ij} = p_i \cdot p_{j|i}$ ,  $p_i$  is the proportion in land cover  $i$  (see Figure 1) and  $n$  is the total number of land cover categories (patch types) in a particular landscape,  $p_{j|i} = N_{ij} / N_i$  is the conditional probability,  $N_{ij}$  is the number of joins or adjacencies between grid cells (pixels) of patch types  $i$  and  $j$  (using the 4 orthogonal neighbor rule, preserving the order of pixels [9]), and  $N_i$  is the total number of joins between pixels of patch type  $i$  and all patch types (including patch  $i$  itself). With this definition of  $p_{ij}$  Li and Reynolds proved  $EE_{\max} = 2 \ln(n)$ . The index  $RC$  ranges from 0 to 1. Other (relativized) contagion indices are reviewed in [10]. The index  $RC$  has been used to detect changes in spatial patterns and structure across a variety of landscapes around the globe [11-13].

### 3. The Problem with Relativizing Contagion

For an index to be appropriate, its behavior must follow the precepts of its definition. In the case of relativized contagion, it is easy to demonstrate illogical behavior. Simulated landscapes can be used to formally investigate behavior.

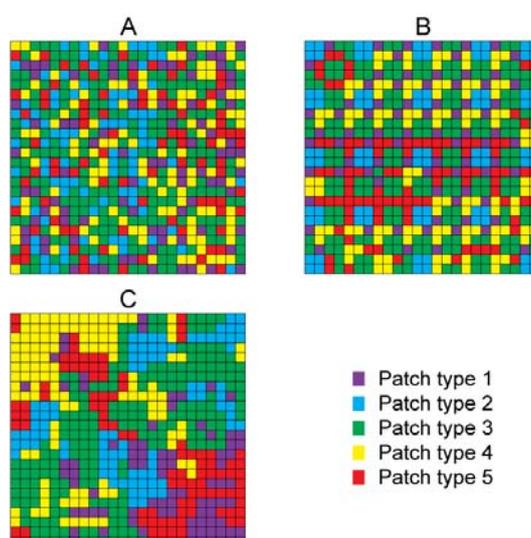
#### 3.1. Simulated Landscapes

We generated a series of stochastically simulated landscapes with different spatial configurations (random, uniform, and aggregated) and numbers of patch types (from 2 to 10), with an increasing gradient of evenness of the proportions of patch types (see Figure 2 for examples of simulated landscapes). Relative to Figure 3a, as the number of patch types were increased (2-10) the relative proportion of land per patch type was adjusted the same for each of the three simulated landscape configurations (i.e. random, uniform, and aggregated) to reflect an increasing gradient of evenness (i.e. with 2 patch types  $RE \approx 0$  and as patch types were progressively added  $RE \approx 1$ ). In Figure 3b as the number of patch types increased (2-10) the proportion of land per patch types remained constant, (i.e.  $RE \approx 1$ ). The rationale behind this scheme is to have coverage of points between the extremes of high contagion to low contagion. There were 27 simulated landscapes, nine for each spatial configuration. As a second approach, another series of landscapes were constructed for random, uniform, and aggregated configurations, and numbers of patch types (from 2 to 10), but with completely even proportions of patch types. Again, 27 landscapes were created, nine for each spatial configuration. A simple 0 to 1 scaled measure of evenness is given by [14]:

$$RE = \frac{-\ln\left(\sum_{i=1}^n p_i^2\right)}{\ln(n)} \quad (3)$$

where  $p_i$  is the proportion of the landscape in patch type  $i$  and  $RE$  (relative evenness) approaching 0 means increasing unevenness of the  $n$  categories and  $RE = 1$  means all categories occur in equal proportion.

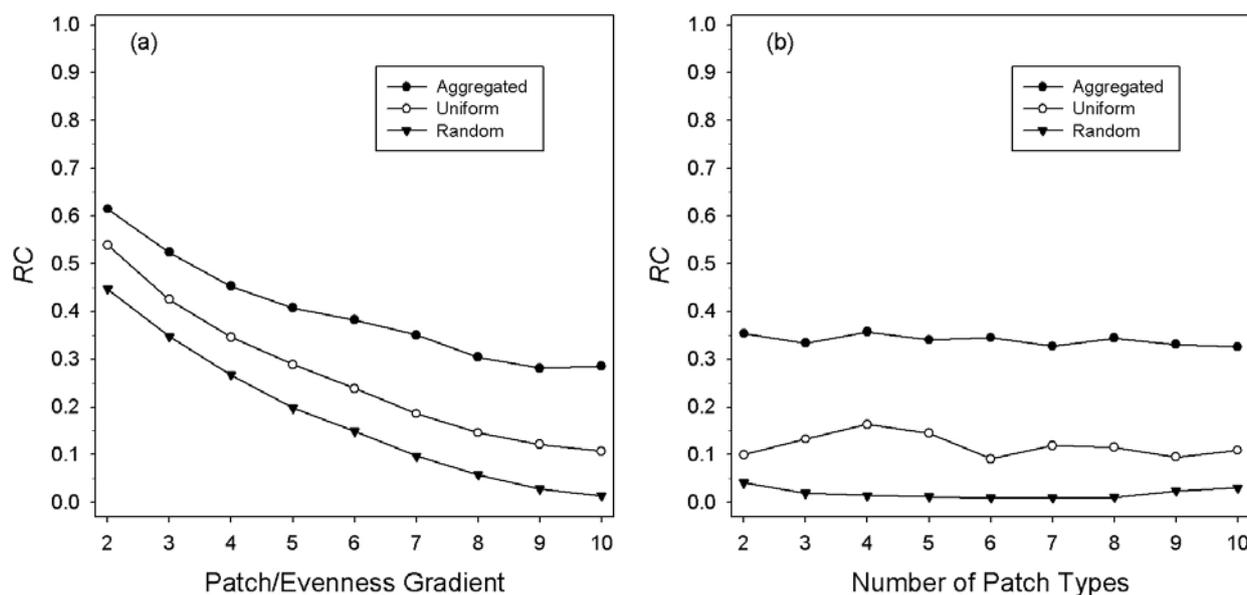
**Figure 2.** Examples of (A) random, (B) uniform, and (C) aggregated landscapes used to investigate contagion index behavior.



3.2. Deficiencies of Relative Contagion

Figure 3 displays two graphs of the *RC* values. In (a), under the increasing gradient of evenness, the three spatial configurations separate out very cleanly and logically, but the values only fall between 0 and 0.6. Thus the *RC* values do not cover the upper portion of the 0 to 1 range (recall the simulated landscapes were designed to cover high to low contagion), showing some insensitivity to composition. In (b), under the same degree of evenness, as before, the three spatial configurations separate out very cleanly, but the three lines are essentially flat, showing that *RC* is completely insensitive to changes in number of patch types on the landscape when category proportions are equal. What does this mean? *RC* is measuring something other than contagion. For each value along the *x*-axis *RE* is constant, at its maximum value of 1. There is a direct parallel between the constancy of *RE* and the constancy of *RC* for each spatial configuration. It appears *RC* is measuring evenness.

**Figure 3.** Trend lines showing relationship between *RC* and the two controlled variables: spatial pattern and number of patch types. (a) There is an increasing gradient of evenness along the *x*-axis. (b) Relative evenness (*RE*)  $\approx 1$  means all categories occur in equal proportion at each point along the *x*-axis.



To strengthen this point, we constructed three test uniform landscapes, where landscape A has four categories, landscape B has five, and landscape C has six. Conceptually, as more patch types are included on a landscape, contagion should decrease. We changed patch type proportions on landscape A slightly to create landscapes B and C with one and then two additional patch types. There should be a small decrease in contagion from landscape A to B, and then from B to C. Table 1 lists the breakdown of category proportions on the three landscapes and gives the index values computed on these landscapes. The relative contagion index increases from A to B to C. The relativized index is not reflecting changes in contagion, but rather it maybe reflecting changes in evenness, as shown by *RE*.

**Table 1.** Illustrative example showing illogical behavior of relative contagion (*RC*) and proper behavior of the entropy index ( $\Gamma$ ).

LANDSCAPE	PATCH TYPE PROPORTIONS						INDEX VALUES		
	1	2	3	4	5	6	<i>RE</i>	<i>RC</i>	$\Gamma$
A	0.50	0.25	0.15	0.10			0.77	0.20	-2.22
B	0.47	0.25	0.15	0.10	0.03		0.71	0.25	-2.41
C	0.47	0.23	0.15	0.10	0.03	0.02	0.66	0.30	-2.51

The effect of scaling contagion relative to the maximum contagion possible creates indices with mathematically undesirable qualities. Such indices are overly sensitive to small variation in composition. Sampling error could easily sway the result one direction or the other, so if contagion indices are to be meaningful they should be relatively insensitive to such a change. This revelation is not new. Quantitative ecologists [15, 16] argued against relativizing species diversity indices and they demonstrated that such indices are mathematically untenable.

#### 4. Entropy Index

The maximum-entropy principle [8, 17] provides a means to obtain least-biased statistical inference under uncertainty. The rule is to choose the probabilities so as to maximize the uncertainty when one has only partial information about the possible outcomes. One can view the adjacency of patch type *i* and *j* as a general problem in waiting times [18] for the encounter of state *ij*, which follows a geometric distribution. We will show that by defining contagion as a generalized function and inserting expected values of random variables based on the uncertainty of the encounter of state *ij*, a theoretically proper entropy index results.

##### 4.1. Contagion Generalized

In a contagious landscape the typical patch type is relatively concentrated. Therefore contagion can be viewed as a function of concentration:

$$\Gamma(L) = \varphi - \sum \sum p_{ij} C(p_{ij}) \tag{4}$$

where  $\Gamma(L)$  is the contagion on landscape *L* associated with the measure of concentration *C*,  $\varphi$  is any real constant, and  $p_{ij} = p_i \cdot p_{j|i}$  (proportion of patch type *i* times the conditional probability). Because the meaning of contagion is the inverse of the meaning of species diversity, it is necessary to subtract the quantity  $\sum \sum p_{ij} C(p_{ij})$  from some constant  $\varphi$  to reverse the scale and provide a contagion formulation. Recall that  $p_{ij}$  is the probability that two randomly chosen adjacent pixels belong to type *i* and *j* out of *n* patch types. Define *X* + 1 as the number of random picks of adjacent pixels up to and including the first encounter of state *ij*. This scheme is a general problem in waiting times. Under this scheme *X* has a geometric distribution:

$$P(X = x | p_{ij}) = p_{ij} (1 - p_{ij})^x, \quad x = 0, 1, 2, \dots \tag{5}$$

The ratio  $X/(X + 1)$  provides a reasonable measure of concentration.

4.2. The New Index

Let  $C(p_{ij}) = E[X | p_{ij}] \times E[1 / (X + 1) | p_{ij}]$ . Because  $E[X | p_{ij}] = (1 - p_{ij}) / p_{ij}$  and  $E[1 / (X + 1) | p_{ij}] = -p_{ij} \ln(p_{ij}) / (1 - p_{ij})$  (see Appendix A for derivations),  $C(p_{ij}) = -\ln(p_{ij})$ . Use of this result in Equation (4) with  $\varphi = 0$  gives:

$$\Gamma = \sum_{i=1}^n \sum_{j=1}^n p_{ij} \ln(p_{ij}). \tag{6}$$

The mathematical arguments used with the concentration function results in the entropy information value [8]. It has already been established that  $EE_{\max} = 2\ln(n)$  using the 4-neighbor double count adjacency rule. Therefore  $\Gamma$  is bounded between  $-2\ln(n)$  and 0. Expanding the  $p_{ij}$  term to its computational form, Equation (6) can be rewritten as:

$$\Gamma = \sum_{i=1}^n \sum_{j=1}^n p_i \frac{N_{ij}}{\sum_{j=1}^n N_{ij}} \ln \left( p_i \frac{N_{ij}}{\sum_{j=1}^n N_{ij}} \right). \tag{7}$$

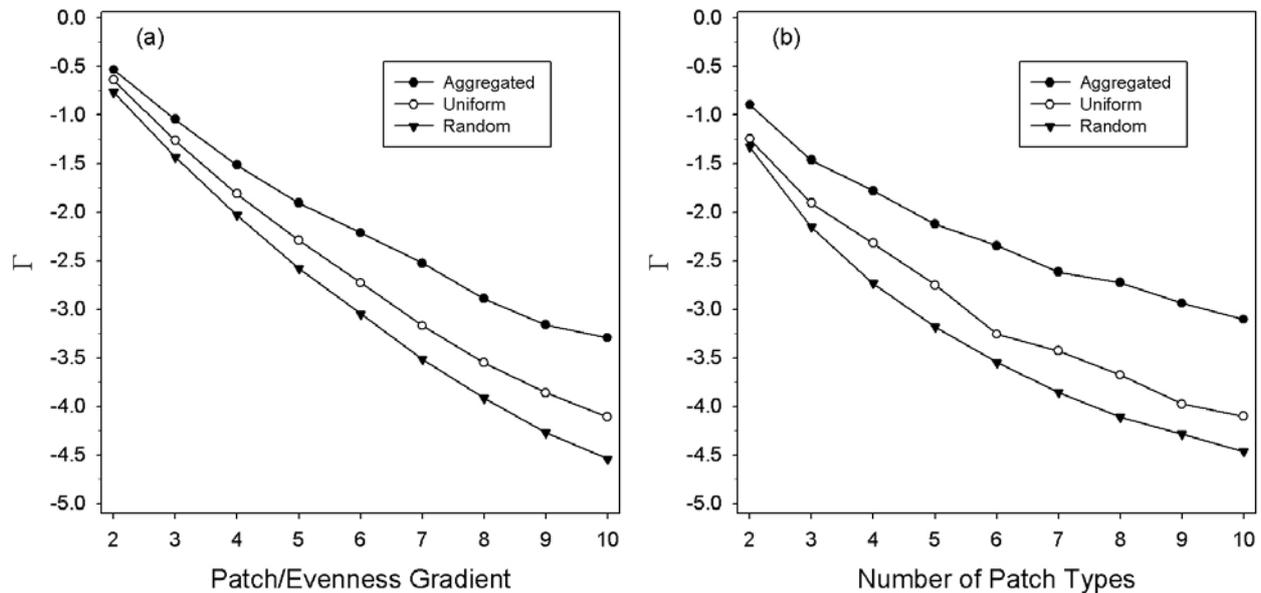
The index  $RC$  is readily obtained from  $\Gamma$  by dividing by  $EE_{\max}$  and adding 1, thus  $RC$  is  $\Gamma$  scaled to the interval  $[0,1]$ .

4.3. Behavior of  $\Gamma$

As a final check on the properties of  $\Gamma$ , let us apply it to the three previous simulated landscapes, (*i.e.*, random, uniform, and aggregated) and compare to Section 3, Figures 3(a-b).

A graph of the  $\Gamma$  values computed on the simulated landscapes generated with an increasing gradient of evenness is displayed as Figure 4(a). As is readily seen,  $\Gamma$  is sensitive to both composition and configuration. The three spatial configurations separate out logically with aggregated landscapes having the highest values, followed by uniform landscapes, and then the randomly arranged landscapes having the lowest values. This is expected since random landscapes have little spatial autocorrelation whereas uniform and aggregated landscapes have increasing spatial autocorrelation of patches. There is a sharp decrease in all three curves with increasing number of patch types and evenness, covering nearly the full range of index values. This meets with the conceptual definition of contagion. A graph of the  $\Gamma$  values from the 27 landscapes with  $RE \approx 1$  is displayed in Figure 4(b). Again we see that  $\Gamma$  distinguishes between the three spatial configurations and decreases with increases in number of land cover categories. Its behavior is consistent with changes in composition and configuration. Contrast this with the behavior of  $RC$  on the same landscapes (Figure 3(b)). The index  $RC$  failed to be sensitive to changing composition.

**Figure 4.** Trend lines showing relationship between  $\Gamma$  and the two controlled variables: spatial pattern and number of patch types. (a) There is an increasing gradient of evenness along the  $x$ -axis. (b) Relative evenness (RE)  $\approx 1$  means all categories occur in equal proportion at each point along the  $x$ -axis.



### 5. Sampling Properties of the Entropy Index

To obtain an estimate of the contagion on the basis of a given sampling lattice, the unknown *a priori* probabilities, the  $p_{ij}$ s, are replaced by estimated probabilities. In this connection, the properties of the random variable:

$$\hat{\Gamma} = \Gamma(\hat{p}_{11}, \hat{p}_{12}, \dots, \hat{p}_{mn}) \tag{8}$$

are what we wish to determine.

#### 5.1. Bias and Variance of $\hat{\Gamma}$

To estimate the first moment of Equation (8), it is necessary to expand the function in a Taylor series about the point  $(p_{11}, \dots, p_{mn})$  and take the expected value. We write the expansion as follows:

$$\begin{aligned} \Gamma(\hat{p}_{11}, \dots, \hat{p}_{mn}) = & \Gamma(p_{11}, \dots, p_{mn}) + \sum \sum \Gamma'(p_{ij})(\hat{p}_{ij} - p_{ij}) + \frac{1}{2} \sum \sum \Gamma''(p_{ij})(\hat{p}_{ij} - p_{ij})^2 \\ & + \frac{1}{6} \sum \sum \Gamma'''(p_{ij})(\hat{p}_{ij} - p_{ij})^3 + \frac{1}{24} \sum \sum \Gamma^{(4)}(p_{ij})(\hat{p}_{ij} - p_{ij})^4 + \dots \end{aligned} \tag{9}$$

The expectation of Equation (9) involves the central moments of the random variables  $\hat{p}_{ij}$ , which follow a multinomial distribution. Let  $T$  be the total number of lattice plots, and let  $\varepsilon_{ij} = \hat{p}_{ij} - p_{ij}$ . The moment-generating function for the multinomial distribution [19] is:

$$E[e^{\alpha \varepsilon_{ij}}] = \left( E[e^{\alpha(x_{ij}/T - p_{ij}/T)}] \right)^T \tag{10}$$

where  $x_{ij}$  takes the value 1 with probability  $p_{ij}$ , and 0 with probability  $1 - p_{ij}$ . Using Equation (10) to compute expected values of the terms in Equation (9) and simplifying, we obtain the first moment or mean of Equation (8):

$$\begin{aligned} E[\hat{\Gamma}] &= \Gamma(p_{11}, \dots, p_{mm}) + o(T^{-1}) + o(T^{-2}) + o(T^{-3}) + \dots \\ &= \Gamma + o(T^{-1}). \end{aligned} \tag{11}$$

From Equation (11) we can infer that  $\hat{\Gamma}$  is biased, but for any reasonable size  $T$  the bias is very small, and in fact is less than  $T^{-1}$ .

The estimated  $\hat{\Gamma}$  equals the true  $\Gamma$  plus a random error  $\xi$ , *i.e.*  $\hat{\Gamma} = \Gamma + \xi$ . By definition,  $E[\xi^2] = E[(\hat{\Gamma} - \Gamma)^2]$ , and the variance of  $\hat{\Gamma}$  is  $E[\xi^2] - E^2[\xi]$ . From Equation (11) we know that  $E[\xi] = o(T^{-1})$ , consequently  $E^2[\xi] = o(T^{-2})$ , which is negligible. Therefore, the variance of  $\hat{\Gamma}$  is:

$$\text{var}(\hat{\Gamma}) = E[\xi^2] = E[(\hat{\Gamma} - \Gamma)^2] \tag{12}$$

A common approach for establishing the variance of a scalar statistic that is a function of many variables is the delta method [20,21]. The delta method uses the first order multivariate Taylor series expansion to produce an estimated variance, *i.e.*,  $\text{var}(\hat{\Gamma})$ . The variance derivation for  $\hat{\Gamma}$  is given in Appendix B. For  $\hat{\Gamma}$  we have:

$$\text{var}(\hat{\Gamma}) = \frac{\sum_{i=1}^n \sum_{j=1}^n \hat{p}_{ij} \ln^2(\hat{p}_{ij}) - \left[ \sum_{i=1}^n \sum_{j=1}^n \hat{p}_{ij} \ln(\hat{p}_{ij}) \right]^2}{T} \tag{13}$$

### 5.2. Consistency of $\hat{\Gamma}$

The property of consistency ensures that an estimate is close to the true parameter value with a high probability if the sample size is sufficiently large. Let  $\hat{\theta}_T$  be an estimator of  $\theta$  based on a sample of size  $T$ . Sufficient conditions for an estimator  $\hat{\theta}_T$  to be consistent for  $\theta$  [22] are:

$$\lim_{T \rightarrow \infty} E[\hat{\theta}_T] = \theta \tag{14}$$

$$\lim_{T \rightarrow \infty} \text{var}(\hat{\theta}_T) = 0. \tag{15}$$

An estimator that satisfies (14) is said to be asymptotically unbiased. Thus an estimator is consistent if any bias it has goes to zero as the sample size increases and if its variance goes to zero as  $T \rightarrow \infty$ . The limit of Equation (11) is  $\Gamma$  so condition (14) is met; and the limit of Equation (13) is zero so condition (15) is met. Thus  $\hat{\Gamma}$  is a consistent estimator.

### 5.3. Asymptotic Normal Distribution of $\hat{\Gamma}$

To prove that as  $T \rightarrow \infty$  the distribution of  $\hat{\Gamma}$  converges to the normal distribution with mean value as in Equation (11) and variance in Equation (13), write the Taylor expansion shown in Equation (9) in the form:

$$\sqrt{T}(\hat{\Gamma} - \Gamma) = \sqrt{T} \left[ \sum \sum (\hat{p}_{ij} - p_{ij}) [1 + \ln(p_{ij})] + \frac{1}{2} \sum \sum \frac{(\hat{p}_{ij} - p_{ij})^2}{p_{ij}(1-\theta) + \theta \hat{p}_{ij}} \right], \tag{16}$$

where  $0 < \theta < 1$ , and apply the convergence in distribution theorem (see definition 4.1 in [23], also see theorem 28.4 in [24]) to the random variable  $\sqrt{T}(\hat{\Gamma} - \Gamma)$ . The theorem as stated in [24] says that if you have a function  $H(m_\nu, m_\rho)$  of two central moments, then “If, in some neighbourhood of the point  $m_\nu = \mu_\nu, m_\rho = \mu_\rho$ , the function  $H(m_\nu, m_\rho)$  is continuous and has continuous derivatives of the first and second order with respect to the arguments  $m_\nu$  and  $m_\rho$ , the random variable  $H(m_\nu, m_\rho)$  is asymptotically normal, . . . .” It follows from this theorem that any sample characteristic based on moments is, for large values of  $T$ , approximately normally distributed about the corresponding population characteristic.

## 6. Hypothesis Testing with the Entropy Index

We have just computed the mean value and variance for the statistical estimate  $\hat{\Gamma}$  of the contagion of a landscape. The estimate was shown to be asymptotically unbiased, consistent, and asymptotically normal. A direct and very important use of the asymptotic normality and variance of the given estimator is in hypothesis testing. Tests against an *a priori* null can be accomplished via a one-sample *t*-test or construction of a confidence interval about the point estimate. Likewise, two point estimates can be compared using a two-sample *t*-test. Three or more estimates can be compared using analysis of variance. For the entropy index, a simple one-way ANOVA can be constructed as follows: let  $\hat{\Gamma}_i$  = contagion value for the *i*th group, *t* = number of groups, and  $T_i$  = number of observations or patches for the *i*th group, then:

$$\begin{aligned} \text{among groups variance} &= \frac{\sum_{i=1}^t (\hat{\Gamma}_i - \bar{\Gamma})^2}{t-1}, \quad \bar{\Gamma} = \frac{\sum_{i=1}^t T_i \hat{\Gamma}_i}{\sum_{i=1}^t T_i} \\ \text{within groups variance} &= \frac{\sum_{i=1}^t T_i \widehat{\text{var}}(\hat{\Gamma}_i)}{\sum_{i=1}^t T_i} \tag{17} \\ F &= \frac{\text{among groups variance}}{\text{within groups variance}} \text{ with } t-1, \sum T_i - t \text{ degrees of freedom.} \end{aligned}$$

Often, hypotheses of interest can be expressed in terms of linear combinations of the estimates as  $\mathbf{R}\hat{\Gamma} = \mathbf{r}$  where  $\mathbf{R}$  and  $\mathbf{r}$  are a matrix and a vector of known elements that specify the hypotheses of interest. For example, if the hypothesis is that the elements of  $\hat{\Gamma}$  sum to unity,  $\mathbf{R} = [1, 1, \dots, 1]$  and  $\mathbf{r} = 1$ .

The Wald, Lagrange multiplier, and quasi-likelihood ratio tests can be used to test the null hypothesis  $\mathbf{R}\hat{\Gamma} = \mathbf{r}$  versus the alternative  $\mathbf{R}\hat{\Gamma} \neq \mathbf{r}$ .

## 7. Application: Analysis of Contagion of Forest Types on a South Carolina Landscape

### 7.1. Study Site and Data

The landscape under study was The United States Department of Energy Savannah River Site (SRS), an 80,000 ha National Environmental Research Park located on the Upper Coastal Plain and Sandhills physiographic provinces in South Carolina, USA (Figure 5). The Site naturally groups into four major strata based on a gradient in soil-geologic landform conditions from the northeast representing the Sandhills to the southwest representing the Savannah River swamp. The tail extending from the southeast was aggregated with the Savannah River swamp as these areas occupy bottomlands and wetlands containing bottomland hardwood stands and bald-cypress-tupelo (*Taxodium distichum-Nyssa aquatic*) stands. The SRS today contains approximately 74,000 ha of forested landscape divided into over 8,800 stands or management units. When the SRS was established in 1951, approximately 33,000 ha were in old-fields and the balance consisted of cutover forest land with low stocking [25]. The old fields and cutover forests were planted with loblolly pine (*Pinus taeda*), longleaf pine (*P. palustris*) and slash pine (*P. elliottii*).

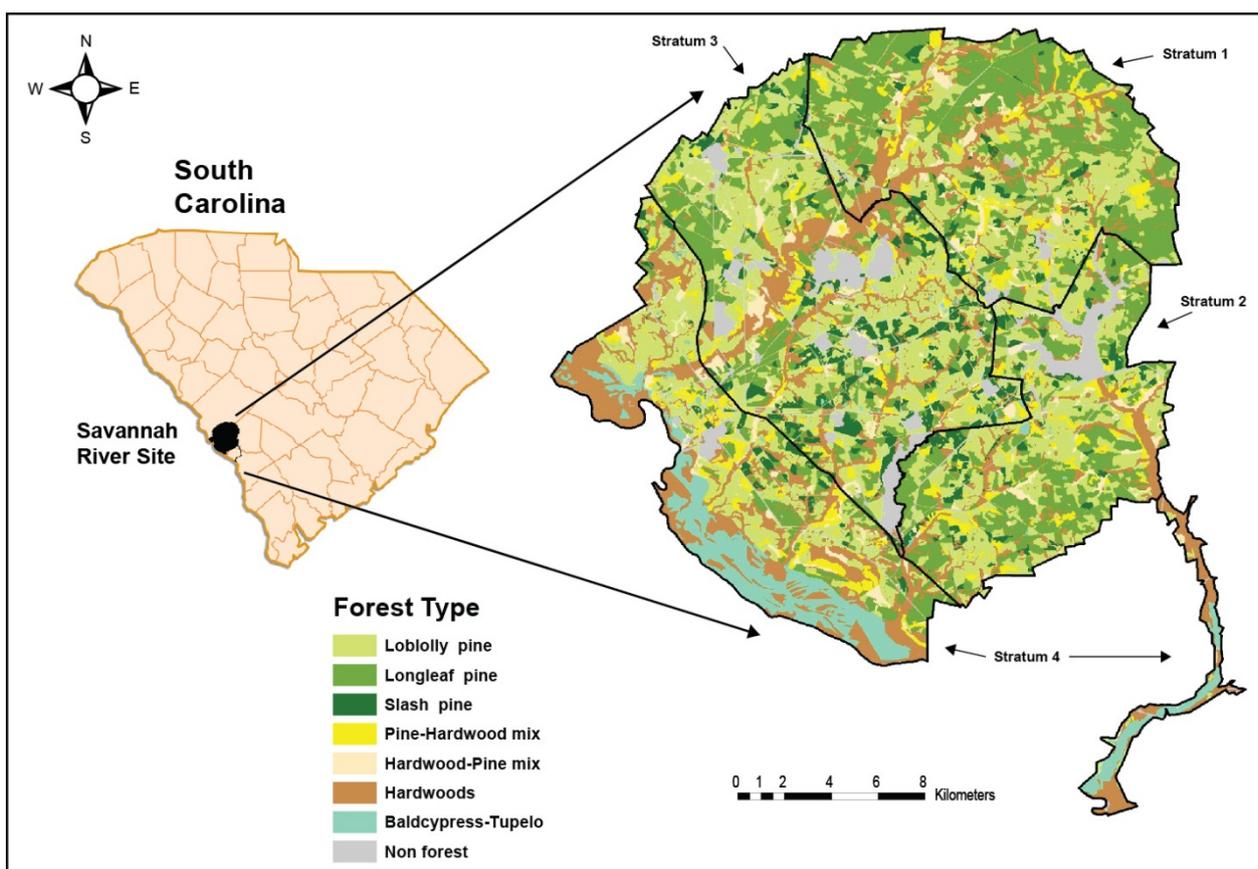
Seven species of federally threatened or endangered plants and animals occur on the SRS. The Site supports populations of 44 plant species designated as “sensitive.” These are species that are not federally protected but are of local concern. In addition, the SRS has designated 32 species of animals as sensitive—one insect, eight mollusks, one fish, seven reptiles and amphibians, eight birds, and seven mammals [25]. Management for the endangered red-cockaded woodpecker (*Picoides borealis*) predominates on 67,000 acres in the eastern portion of the Site. The species requires large pine trees and an open understory. Many aspects of management on the SRS are geared toward providing nesting and foraging habitat for red-cockaded woodpeckers (RCW) [25]. Areas suitable for RCW would have reasonably high contagion and be predominantly composed of pine forests. Other species of concern, such as the reptiles and amphibians, generally prefer wetter more diverse habitats and should do well in moderate to low contagion areas [26]. Many mammals and some birds are generalists that can adapt to a variety of conditions that exhibit a range of contagion. The investigation of contagion will provide insight about structural conditions [27].

The Site maintains a forest stands database recording species composition, age, fire history and other relevant information. Stand composition is determined from operational records, field observations, infrared color photography and periodic surveys. The last survey was completed in 2010 and looked at 1,680 points in the field. This information was used to update the stands database. For analysis purposes, stands are assigned to one of seven broad forest types [28]. Forest types and their proportions on each stratum are given in Table 2. Stands of the same forest type were colored the same to create a landscape level view of forest types (Figure 5).

**Table 2.** Proportion of each forest type by strata.

FOREST TYPE	STRATA			
	1	2	3	4
Loblolly pine	0.3016	0.3287	0.3400	0.2232
Longleaf pine	0.3776	0.2804	0.1665	0.0733
Slash pine	0.0381	0.0571	0.1045	0.0243
Pine-Hardwood	0.0615	0.0609	0.0495	0.0633
Hardwood-Pine	0.0336	0.0311	0.0364	0.0327
Hardwoods	0.1521	0.1202	0.1516	0.3423
Baldcypress-Tupelo	–	0.0036	0.0039	0.1856
Non forest	0.0356	0.1181	0.1475	0.0553

**Figure 5.** The Savannah River Site in South Carolina, southeastern USA. Major forest vegetation groups are shown within each of the four large scale strata corresponding to soil-geologic landform gradients.



7.2. Results and Discussion

We used FRAGSTATS [29] to compute the adjacency matrices for each stratum based on a 10-m image resolution. The forest type adjacencies counted in stratum 3 are shown in Table 3. The adjacency values, along with the proportion values in Table 2, were used to compute contagion [Equation (7)] and its concomitant variance [Equation (13)] for each stratum. Table 4 lists the strata  $\hat{\Gamma}$

values, their variances, the number of patches, the results of the ANOVA using Equation (17), and how the index values separated out using Tukey’s multiple comparison procedure. The ANOVA was highly significant with  $P < 0.001$ . We see that stratum 1 has a significantly larger contagion than the other strata, with stratum 2 being significantly larger than strata 3 and 4. Strata 3 and 4 were not significantly different from each other.

**Table 3.** Adjacency matrix of forest types for stratum 3.

FOREST TYPE	1	2	3	4	5	6	7	8
<b>1- Loblolly pine</b>	3281710	14774	11181	7147	7389	35207	542	44300
<b>2- Longleaf pine</b>	14774	1596132	3981	3281	2069	11488	336	21094
<b>3- Slash pine</b>	11181	3981	995896	2045	1045	9231	540	19003
<b>4- Pine-Hardwood</b>	7147	3281	2045	469776	1358	6787	365	6316
<b>5- Hardwood-Pine</b>	7389	2069	1045	1358	345660	4172	251	3210
<b>6- Hardwoods</b>	35207	11488	9231	6787	4172	1433772	833	11910
<b>7-Baldcyp.-Tupelo</b>	542	336	540	365	251	833	36088	222
<b>8- Non forest</b>	44300	21094	19003	6316	3210	11910	222	1305048

**Table 4.** Contagion values, variances, ANOVA  $F$ -test, and Tukey’s test on the four large scale strata corresponding to soil-geologic landform gradients on the Savannah River Site.

STRATA	$\hat{\Gamma} (*)$	$\text{var}(\hat{\Gamma})$	$T$	ANOVA
1	-1.760 <sup>a</sup>	0.000718	2140	$F = 24.15$
2	-1.927 <sup>b</sup>	0.000888	1659	$P < 0.001$
3	-2.020 <sup>c</sup>	0.000468	2983	
4	-2.038 <sup>c</sup>	0.000754	2122	

(\*) Tukey’s test, values followed by the same letter are not significantly different using  $\alpha = 0.05$ .

Recall that contagion is bounded between  $-2\ln(n)$  and 0. Three of the strata have eight forest type categories, and stratum 1 has seven categories, meaning the minimum contagion is about -4.2 to -3.9. Thus the contagion values in Table 4 are a little above the midrange telling us contagion is generally moderate on this landscape. A number of forest types, such as the Pine-Hardwood and Hard-Pine types which range from 3% to 6% (Table 2), are minor components interspersed on the landscape (Figure 5). These minor interspersed components have the effect of lowering the contagion.

To understand the differences in contagion across the strata let us examine composition and configuration. From Table 2 it is apparent that the Pine-Hardwood and Hardwood-Pine types are consistently minor components of all the strata. Slash pine is also a minor component on strata 1, 2 and 4, but occupies 10% of stratum 3. Loblolly pine is a major component of all strata but is much lower in stratum 4. Longleaf is a major component in strata 1 and 2, a prominent component in stratum 3, but a small component in stratum 4. Hardwood stands are a prominent feature along the riparian areas and lowlands of strata 1, 2, and 3 but is the largest component of stratum 4 which has substantial bottomlands along the Savannah River swamp. Non forest areas occupy lakes and facilities and vary in proportion across the strata.

In examining Figure 5, we see large aggregates of longleaf and loblolly pine stands, with high proportions of these forest types, in strata 1 and 2. This is why these two strata have the largest contagion. Strata 3 and 4 have a bit more intermixing of categories compared to strata 1 and 2, hence their lower contagion. Compositionally strata 3 and 4 have similar proportional makeup but in different categories. For example, 34% of stratum 3 is loblolly but 34% of stratum 4 is hardwoods; 17% of stratum 3 is longleaf whereas about 19% of stratum 4 is in bald-cypress-tupelo. This is indicative of why strata 3 and 4 did not have significantly different contagion. It is no accident that the RCW has its greatest concentration of cluster sites in strata 1 and 2, with the high compositional abundance of longleaf and loblolly stands and the high aggregation of these types, ideal nesting and foraging conditions.

## 8. Concluding Remarks

The concept of diversity links the notions of richness and evenness [30]. It is well established in the ecological literature that the evenness or equitability component of diversity can be measured independent from richness in two ways which converge for large sample sizes: evenness =  $D / D_{\max}$  and evenness =  $(D - D_{\min}) / (D_{\max} - D_{\min})$  where  $D_{\min}$  and  $D_{\max}$  refer to the minimum and maximum value that a diversity index  $D$  can attain [31, 32]. Hence an index of evenness is a relativized diversity index. Contagion subsumes richness, evenness, and spatial pattern. Relativizing contagion has the effect of creating an evenness index while retaining the configuration component (see Figure 3b and Table 1). Even Li and Reynolds [3] state: “ $RC$  is in essence a function of an evenness index,  $EE / EE_{\max}$ ; thus, it has all the advantages and disadvantages of an evenness index.” Apparently they were not fully aware of the mathematical implications of relativizing contagion. Since  $RC$  is not a true measure of contagion, and evenness is better quantified using  $RE$ , there seems little justification for use of relative contagion. The relationship between  $\Gamma$  and  $RC$  is now clear. The entropy-based Shannon species diversity index, when scaled by its maximum possible value, becomes an index of evenness, and is no longer an index of diversity. The contagion analogue of the diversity index is  $\Gamma$ . Therefore, the true entropy index of contagion is  $\Gamma$ ; and  $RC$ , the scaled or relativized version of  $\Gamma$ , is its corresponding “evenness/configuration” index.

In examining landscapes, for whatever purpose, it is important to consider the objectives of the examination and to design the analysis to provide the needed information. This typically will involve many metrics, never just a single one. In this paper, we focused on contagion because of its extensive use in landscape analysis and the need for a course correction to steer users away from relative contagion. This paper was not meant to be a thorough landscape ecological analysis of the Savannah River Site, only a demonstration of the use of the entropy contagion index. For a recent perspective paper on the use and misuse of landscape metrics see [33].

## Acknowledgments

We would like to thank Harbin Li for use of his landscape simulation program, SHAPC. We would also like to thank Lisa Parresol for her graphic of Figure 1. Funding was provided by the USDA Forest Service Southern Research Station and the Department of Energy-Savannah River Operations Office through the U.S. Forest Service Savannah River under interagency Agreement DE-AI09-00SR22188.

## Author's Contribution

B Parresol was responsible for the conception and design and drafting of manuscript. B. Parresol and L Edwards were both responsible for acquisition, analysis and interpretation of data. Primary author Dr. Parresol passed away during publication process. Therefore Edwards, L. responsible for all manuscript revisions and final proof reading.

## Conflicts of Interests

The authors declare no conflict of interest.

## References

1. Burgess, R.L., Sharpe, D.M., Eds. *Forest Island Dynamics in Man-Dominated Landscapes*; Springer-Verlag: New York, NY, USA, 1981.
2. Salwasser, H. Conserving biological diversity: A perspective on scope and approaches. *For. Ecol. Manage.* **1990**, *35*, 79–90.
3. Li, H.; Reynolds, J.F. A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecol.* **1993**, *8*, 155–162.
4. O'Neill, R.V.; Krummel, J.R.; Gardner, R.H.; Sugihara, G.; Jackson, B.; DeAngelis, D.L.; Milne, B.T.; Turner, M.G.; Zygmunt, B.; Christensen, S.W.; Dale, V.H.; Graham, R.L. Indices of landscape pattern. *Landscape Ecol.* **1988**, *1*, 153–162.
5. Kullback, S.M.; Kupperman, M.; Ku, H.H. Tests for contingency tables and Markov chains. *Technometrics* **1962**, *4*, 573–608.
6. Hutcheson, K. A test for comparing diversities based on the Shannon formula. *J. Theor. Biol.* **1970**, *29*, 151–154.
7. Moran, P.A.P. The interpretation of statistical maps. *J. Roy. Stat. Soc. B* **1948**, *10*, 243–251.
8. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949.
9. Riitters, K.H.; O'Neill, R.V.; Wickham, J.D.; Jones, K.B. A note on contagion indices for landscape analysis. *Landscape Ecol.* **1996**, *11*, 197–202.
10. Parresol, B.R. Sample-Based Forest Landscape Diversity Indices. Ph.D. Thesis, Louisiana State University, Baton Rouge, LA, USA, 1998.
11. Sachs, D.L.; Sollins, P.; Cohen, W.B. Detecting landscape changes in the interior of British Columbia from 1975 to 1992 using satellite imagery. *Can. J. For. Res.* **1998**, *28*, 23–36.
12. Shi, X.; Hao, F.; Zhao, C.; Wang, D.; Ouyang, W. Landscape dynamics analysis of guide wetlands in Yellow River watershed by landsat series data. *Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci. B2* **2008**, *37*, 247–252.
13. Thapa, R.B.; Murayama, Y. Spatial structure of land use dynamics in Kathmandu Valley. *Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci. B8* **2008**, *37*, 11–16.
14. Turner, M.G. Landscape ecology: The effect of pattern on process. *Annu. Rev. Ecol. Syst.* **1989**, *20*, 171–197.
15. Sheldon, A.L. Equitability indices: Dependence on the species count. *Ecology* **1969**, *50*, 466–467.

16. Peet, R.K. Relative diversity indices. *Ecology* **1975**, *56*, 496–498.
17. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
18. Feller, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1968; Volume 1, pp. 47–50.
19. Shenton, L.R.; Hutcheson, K. Moments of moments and frequencies: Expectation of power-sums related to categorized data. In *Technical Reprint Series 17*; University of Georgia Computer Center: Athens, GA, USA, 1969; pp. 7–39.
20. Rao, C.R. *Linear Statistical Inference and Its Applications*; John Wiley & Sons: New York, NY, USA, 1965; pp. 321–322.
21. Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. *Discrete Multivariate Analysis: Theory and Practice*; MIT Press: Cambridge, MA, USA, 1975; pp. 492–497.
22. Judge, G.G.; Hill, R.C.; Griffiths, W.E.; Lütkepohl, H.; Lee, T. *Introduction to the Theory and Practice of Econometrics*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1988; p. 85.
23. White, H. *Asymptotic Theory for Econometricians*; Academic Press: Orlando, FL, USA, 1984.
24. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946; p. 366.
25. Kilgo, J.C., Blake, J.I., Eds. *Ecology and Management of a Forested Landscape: Fifty Years on the Savannah River Site*; Island Press: Covelo, CA, USA, 2005.
26. Wigley, T.B.; Lancia, R.A. Wildlife communities. In *Southern Forested Wetlands: Ecology and Management*; Messina, M.G., Conner, W.H., Eds.; Lewis Publishers: Boca Raton, FL, USA, 1998; pp. 205–236.
27. Wallgren, M.; Bergström, R.; Danell, K.; Skarpe, C. Wildlife community patterns in relation to landscape structure and environmental gradients in a Swedish boreal ecosystem. *Wildl. Biol.* **2009**, *15*, 310–318.
28. Parresol, B.R.; Shea, D.; Ottmar, R. Creating a fuels baseline and establishing fire frequency relationships to develop a landscape management strategy at the Savannah River Site. In *Fuels Management—How to Measure Success*, Proceedings RMRS-P-41; Andrews, P.L., Butler, B.W., Comps.; USDA Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2006; pp. 351–366.
29. McGarigal, K.; Marks, B.J. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*. General Technical Report PNW-GTR-351; USDA Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 1995.
30. Pielou, E.C. *Mathematical Ecology*; John Wiley & Sons: New York, NY, USA, 1977.
31. Pielou, E.C. *Ecological Diversity*; John Wiley & Sons: New York, NY, USA, 1975; pp. 14–17.
32. Magurran, A.E. *Ecological Diversity and its Measurement*; Princeton University Press: Princeton, NJ, USA, 1988; pp. 36–37.
33. Li, H.; Wu, J. Use and misuse of landscape indices. *Landscape Ecol.* **2004**, *19*, 389–399.

Appendices

Appendix A: Expected Value Calculations

The contagion formula for  $\Gamma$  [Equation (6)] comes from particular expected values based on a geometric random variable (see Equation (5)). Let  $q_{ij} = 1 - p_{ij}$ , the required expected values are:

$$\begin{aligned}
 E[X | p_{ij}] &= \sum_{x=0}^{\infty} x p_{ij} q_{ij}^x \\
 &= p_{ij} q_{ij} + 2 p_{ij} q_{ij}^2 + 3 p_{ij} q_{ij}^3 + \dots \\
 &= p_{ij} q_{ij} (1 + 2 q_{ij} + 3 q_{ij}^2 + \dots) \\
 &= p_{ij} q_{ij} (1 - q_{ij})^{-2} = (1 - p_{ij}) / p_{ij},
 \end{aligned}
 \tag{18}$$

$$\begin{aligned}
 E[1/(X+1) | p_{ij}] &= \sum_{x=0}^{\infty} \frac{1}{x+1} p_{ij} q_{ij}^x \\
 &= p_{ij} + \frac{1}{2} p_{ij} q_{ij} + \frac{1}{3} p_{ij} q_{ij}^2 + \frac{1}{4} p_{ij} q_{ij}^3 + \dots \\
 &= p_{ij} (1 + \frac{1}{2} q_{ij} + \frac{1}{3} q_{ij}^2 + \frac{1}{4} q_{ij}^3 + \dots) \\
 &= p_{ij} (q_{ij} + \frac{1}{2} q_{ij}^2 + \frac{1}{3} q_{ij}^3 + \frac{1}{4} q_{ij}^4 + \dots) / q_{ij} \\
 &= p_{ij} [-\ln(1 - q_{ij})] / q_{ij} \\
 &= -p_{ij} \ln(p_{ij}) / (1 - p_{ij}).
 \end{aligned}
 \tag{19}$$

Appendix B: Delta Method for Variance of  $\hat{\Gamma}$

Let  $\varepsilon_{ij} = (\hat{p}_{ij} - p_{ij})$  and  $(\partial\Gamma / \partial p_{ij})|_{p_{ij}=\hat{p}_{ij}}$  be the partial derivative of  $\Gamma$  with respect to  $p_{ij}$  evaluated at  $p_{ij} = \hat{p}_{ij}$ . The first order multivariate Taylor series expansion of  $\hat{\Gamma}$  is

$$\begin{aligned}
 \hat{\Gamma} &\approx \Gamma + \varepsilon_{11} \left( \frac{\partial\Gamma}{\partial p_{11}} \right)_{|p_{11}=\hat{p}_{11}} + \dots + \varepsilon_{1n} \left( \frac{\partial\Gamma}{\partial p_{1n}} \right)_{|p_{1n}=\hat{p}_{1n}} \\
 &+ \varepsilon_{21} \left( \frac{\partial\Gamma}{\partial p_{21}} \right)_{|p_{21}=\hat{p}_{21}} + \dots + \varepsilon_{2n} \left( \frac{\partial\Gamma}{\partial p_{2n}} \right)_{|p_{2n}=\hat{p}_{2n}} + \dots \\
 &+ \varepsilon_{n1} \left( \frac{\partial\Gamma}{\partial p_{n1}} \right)_{|p_{n1}=\hat{p}_{n1}} + \dots + \varepsilon_{nn} \left( \frac{\partial\Gamma}{\partial p_{nn}} \right)_{|p_{nn}=\hat{p}_{nn}}.
 \end{aligned}
 \tag{20}$$

The Taylor series expansion in Equation (20) provides the following linear approximation:

$$\xi = \hat{\Gamma} - \Gamma \approx \sum_{i=1}^n \sum_{j=1}^n \varepsilon_{ij} \left( \frac{\partial\Gamma}{\partial p_{ij}} \right)_{|p_{ij}=\hat{p}_{ij}}.
 \tag{21}$$

By squaring Equation (21) we get

$$\begin{aligned} \xi^2 &\approx \left[ \sum_{i=1}^n \sum_{j=1}^n \varepsilon_{ij} \left( \frac{\partial \Gamma}{\partial p_{ij}} \right) \Big|_{p_{ij}=\hat{p}_{ij}} \right] \left[ \sum_{k=1}^n \sum_{l=1}^n \varepsilon_{kl} \left( \frac{\partial \Gamma}{\partial p_{kl}} \right) \Big|_{p_{kl}=\hat{p}_{kl}} \right] \\ &\approx \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \varepsilon_{ij} \varepsilon_{kl} \left( \frac{\partial \Gamma}{\partial p_{kl}} \right) \Big|_{p_{kl}=\hat{p}_{kl}} \left( \frac{\partial \Gamma}{\partial p_{ij}} \right) \Big|_{p_{ij}=\hat{p}_{ij}} . \end{aligned} \tag{22}$$

From Equations (12) and (22), the  $\text{var}(\hat{\Gamma})$  is estimated as

$$\widehat{\text{var}}(\hat{\Gamma}) = \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial \Gamma}{\partial p_{ij}} \right) \Big|_{p_{ij}=\hat{p}_{ij}} E[\varepsilon_{ij} \varepsilon_{kl}] \sum_{k=1}^n \sum_{l=1}^n \left( \frac{\partial \Gamma}{\partial p_{kl}} \right) \Big|_{p_{kl}=\hat{p}_{kl}} . \tag{23}$$

The covariances  $E[\varepsilon_{ij} \varepsilon_{kl}]$  in Equation (23) can be determined using Equation (10). Resolving the covariances and partial derivatives and simplifying gives

$$\widehat{\text{var}}(\hat{\Gamma}_2) = \frac{\sum_{i=1}^n \sum_{j=1}^n \hat{p}_{ij} \ln^2(\hat{p}_{ij}) - \left[ \sum_{i=1}^n \sum_{j=1}^n \hat{p}_{ij} \ln(\hat{p}_{ij}) \right]^2}{T} . \tag{24}$$

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).