Spring 2010

Nested subsets, scale, and the distribution of abundance: A macroecological approach

Elita Baldridge
Fort Hays State University

Recommended Citation

Baldridge, Elita, "Nested subsets, scale, and the distribution of abundance: A macroecological approach" (2010). *Master's Theses*. 162.
DOI: 10.58809/HNDQ9275
Available at: https://scholars.fhsu.edu/theses/162

NESTED SUBSETS, SCALE, AND THE DISTRIBUTION OF ABUNDANCE: A MACROECOLOGICAL APPROACH

being

A Thesis Presented to the Graduate Faculty
of the Fort Hays State University in
Partial Fulfillment of the Requirements for
the Degree of Master of Science

by

Elita Baldridge
B.S., Kansas State University

Date_________________________ Approved__________________________
Major Professor

Approved__________________________
Chair, Graduate Council

This Thesis for
the Master of Science Degree

by

Elita Baldridge

has been approved

Chair Supervisory Committee

Supervisory Committee

Supervisory Committee

Supervisory Committee

Chair, Department of Biological Sciences

NOTES

This manuscript has been formatted in the style of the journal Ecography.

The title page of this thesis, with lines representing sites and letters representing species, is nested when sorted by incidence or abundance (Randomizations = 10,000, sites = 10, species = 22, Incidence: $N_0 = 50$, $p = 0.0017$, Abundance: $N_0 = 48$, $p < 0.001$).

ABSTRACT

While the canonical nested subset pattern suggests that less species-rich areas will contain a proper subset of the species observed in richer areas, actual data sets do not show perfect nestedness; however, they show a pattern with more structure than would be expected by chance. Biological processes like immigration and extinction have traditionally been thought to produce the nested subset pattern. These processes acting indirectly could cause variation in the distribution and abundance of species that could produce nestedness at a variety of scales. Determining at what scales the pattern is observed might allow inference of processes that are more likely to be acting at those levels.

My hypotheses were 1) species will be observed at more sites in the matrix when the matrix is constructed from sites in the center of the geographic range and fewer sites when the matrix is constructed from sites near the edge of the geographic range, 2) the matrix will be nested when sorted by incidence or abundance, 3) nestedness will be observed at intermediate scales, but not at the smallest or largest scales, and 4) the most species-rich site within the matrix will be the site of first occurrence for species.

I used the 2003 North American Breeding Bird Survey data set (BBS) for the conterminous United States to construct the matrices used in analysis. An evenly spaced grid was used in a stratified sampling design to identify seed points. The geographic extent of analysis started within the route closest to the seed point and increased to include routes from multiple seed points. Two data matrices were constructed for each seed point at each scale; one sorted by incidence, and one sorted by total abundance of species as the independent variable.

Due to limitations of existing software, I wrote a new program called Cudgel to test my hypotheses. Cudgel has the ability to sort the matrix by abundance or other independent variables. The metric $N_{0,}$ which calculates the number of unexpected absences, and null model Random 1, which preserves species presences while allowing presences to vary across sites, was used to calculate nestedness. The number of randomizations performed during the Monte Carlo simulation was 10,000.

My results did not support the hypothesis that species will be observed at more sites in the matrix when the matrix is created from sites in the center of the geographic range. I did observe a significant correlation between incidence and abundance, with significant matrices showing a greater correlation ($t = 2.976$, $df = 231$, $p = 0.003$). Through graphical analysis, I determined that significance increased with increasing scale, but the metric values were lowest at intermediate spatial extents. While the species with the highest incidence were present at the most species-rich site first, species with low incidence showed greater variability in the site of first occurrence.

Because there was a relationship between incidence and abundance, it is possible that the nested subset pattern could be caused by distribution and abundance. If the nested subset pattern is caused by distribution and abundance, research should be focused on the factors that are producing the observed variation in distribution and abundance.

ACKNOWLEDGEMENTS

While many individuals provided assistance that allowed me to complete this thesis, there are a few who should receive special acknowledgement. My advisor, Dr. Rob Channell, has helped me not only with the completion of this thesis, but has helped me to grow tremendously as a scientist and as a person. My fellow graduate student and member of the Channell lab, Georgina Jacquez, provided me with valuable insights, a lasting friendship, and popcorn. The members of my committee, Dr. Greg Farley, Dr. Elmer Finck, and Dr. John Heinrichs, also improved the quality of this thesis.

The volunteers of the Breeding Bird Survey do an excellent job of providing researchers with long-term, large-scale data, and I could not have written this thesis without them. I am also grateful to the Department of Biological Sciences at Fort Hays State University for providing me with a teaching assistantship and the Fleharty Fellowship, for which I also thank the Fleharty family.

I also thank Dr. Wirt Atmar, Dr. Steven Chown, Dr. Javier Galeano, Dr. Paulo Guimarães, Dr. Miguel Rodriguez-Gironés, and Dr. Werner Ulrich for graciously providing me with code for their nestedness analysis programs.

Finally, I thank my husband and family for their incredible patience and support in cheerfully accepting the long hours I spent doing research. G. Baldridge was particularly helpful in her willingness to go on long walks and listen to me while I soliloquized about research possibilities and frustrations while I developed my project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

NESTED SUBSETS, SCALE, AND THE DISTRIBUTION OF ABUNDANCE: A MACROECOLOGICAL APPROACH

INTRODUCTION

Macroecology is a discipline that studies patterns and processes at large spatial scales to explain statistical patterns in ecological data (Brown 1995). Macroecology tends to use preexisting data that have been collected on a large spatial or temporal scale by multiple observers. With the development of more effective computer technology, researchers can use data mining techniques to synthesize information from a variety of sources to address questions that have not been previously addressed (Barkely 1993). The macroecological approach is particularly useful for pattern analysis, which examines patterns to infer processes that could produce the patterns. The macroecological approach of examining patterns across a variety of scales might allow inferences of processes that are more likely to be acting at all scales of analysis. It might also uncover information suggesting that processes important at one scale of analysis are not important on a larger or smaller scale (Brown 1995). Since the relatively recent development of macroecology, it has been used to study patterns and infer processes that structure and assemble communities (Brown and Maurer 1989).

Communities are assembled from overlapping species occurrences. Species occurrence is determined by four factors: immigration; extinction; speciation; and emigration (Figure 1). Speciation and immigration can be considered together because they are both processes that increase species richness (Figure 1). On an ecological time scale, speciation is not a major process, but on an evolutionary time scale, speciation could have a major effect (MacArthur and Wilson 1967). Emigration and extinction can be considered together because they are processes that decrease species richness (Figure 1). Emigration is not considered to be an important factor in species occurrence because most emigrants die in the process of emigration, and thus would be indistinguishable from extinction. Immigration and extinction are considered to be the dominant processes controlling species occurrence on an ecological time scale (MacArthur and Wilson 1967).

Assembly rules, which are patterns of species occurrence that indicate processes that could contribute to community structure, have been viewed as having both deterministic and stochastic elements (Diamond 1975, M'Closkey 1978, Fox 1990). Assembly rules are deterministic or mechanistic in that they indicate regular patterns of species occurrences that lead to community structure. However, they are stochastic in that the species in the community can vary according to chance, but the pattern can still be observed. Nested subsets, another type of assembly rule, occur when the most species rich site contains all species in the study area, while less species-rich sites form a proper subset of the species found in all richer sites (Patterson and Atmar 1986) (Figure 2). While this is the canonical pattern, data sets show departures from perfect nestedness (Patterson and Atmar 1986). Both unexpected presences and unexpected absences of species are observed in actual data sets.

While immigration and extinction (Patterson and Atmar 1986) are considered to be the two dominant processes that could produce the nested subset pattern, a variety of other processes have also been proposed to explain the cause of the pattern, including nested habitats (Simberloff and Martin 1991), periodic disturbance (Bloch et al. 2007), passive sampling (Cutler 1994), habitat heterogeneity (Sfenthourakis et al. 2004), stress tolerance (Cook et al. 2004) and speciation (Cook and Quinn 1995). Using the hypothetical scenario presented in Figure 2, extinction could generate the pattern if all the species were originally present at all sites, species D was able to persist only at site 1, species C was able to persist at sites 1 and 2, species B was able to persist at sites 1, 2, and 3, and species A was able to persist at all sites. Alternatively, immigration could create the pattern if all species were present in a regional species pool, but had not yet colonized the sites. In Figure 2, the nested subset pattern could be observed if species A could immigrate to all sites, species B could immigrate to sites 1, 2, and 3, species C could immigrate to sites 1 and 2, and species D could only immigrate to the closest site, site 1.

Some of the alternative processes that have been proposed can be lumped under immigration and extinction, while other ideas are illogical with closer examination. Periodic disturbances would result in a series of extinctions and immigrations. Stress tolerance is likewise related to extinction, as the different abilities of species to survive and reproduce under different environmental conditions results in the different abilities of species to persist in an area (Hutchinson 1957). Habitat heterogeneity and nested habitats are not good explanations of the nested subset pattern because both habitat heterogeneity and nested habitats must be produced by other processes, particularly because the word “habitat” is frequently used as a substitute for “local plant community”. There is no logical reason why processes like immigration and extinction should act differently on plants than on any other group of organisms. In addition, the idea that the nested subset pattern exists due to nested habitats is a tautology.

Passive sampling results when incomplete sampling causes variation that creates a pattern in the incomplete dataset (Cutler 1994, Ulrich and Gotelli 2007). With passive sampling, only the most abundant species would be sampled, causing rare species to be consistently underrepresented in the dataset. This would result in errors of omission where rare species were not sampled, even though they were present in the local community. Errors in the data produce the pattern because the errors occur in a regular manner. If the data were complete, the pattern would not be observed. Because biological data are always undersampled, it is important to consider whether errors or random variation present in the data could be producing the pattern in question. The idea that the nested subset pattern could be produced by a statistical process rather than a biological process was introduced by the proposal of passive sampling as a process. However, it is important to remember that patterns do not cease to exist simply because a process is not biological. This idea is applicable not only to the nested subset pattern, but also potentially other patterns that might be observed.

Traditional explanations for the nested subset pattern, particularly those used for conservation, have formed a direct link between pattern and process. However, the link between the traditional processes thought to produce the nested subset pattern could be indirect. With this scenario, immigration and extinction might be acting to produce patterns in the distribution and abundance of species, which then produces the nested subset pattern in an emergent link. In this case, although the traditionally cited processes could be indirectly contributing to pattern structure, they are not the proximate causes of the pattern.

The nested subset pattern has a great deal of generality, and has been observed for many taxa across several continents (Table 1). Because of the generality of this pattern and the link to extinction as a possible process causing the pattern, it has been proposed as having conservation utility (Patterson 1987, Hansson 1998, Fleishman et al. 2002). The nested subset pattern has been treated for conservation applications as if it is produced by a mechanistic community level process. However, nestedness might be an emergent pattern produced by processes acting at the population level.

I will examine if there is a pattern associated with the site at which species first occur within the nested subset matrix. The site of first occurrence is the most species-rich site at which a species is present in the matrix. If there is a pattern, this could infer that the process producing nestedness acts mechanistically. If there is no pattern to species entering the matrix, then the process does not act mechanistically. The canonical nested subset pattern makes the prediction that the most species-rich site is the one that is the most likely to acquire a new species, because the most species-rich site will contain all species, and less species rich sites will have a subset of the species observed at the most species-rich site. Therefore, I hypothesize that the most species-rich site within the matrix will be the site of first occurrence for species.

A macroecological approach at different scales could provide insights into the processes that are producing the nested subset pattern. I will examine nestedness at different scales to determine at what scales nestedness is observed. Determining at what scales the pattern is observed could allow inference of processes that are more likely to be acting at those scales. At small spatial extents, I expect that local variation will obscure the pattern, and I expect that the nested subset pattern will not be observed at large spatial extents because matrices are made up of species from different regional species pools. Much of the work on the nested subset pattern has probably been at intermediate scales, such as island archipelagos, and most of these studies have observed nestedness (Table 1). In addition, random variation should have less of an effect at intermediate scales, and intermediate spatial scales tend to consist of a single species pool. I hypothesize that nestedness will be observed at intermediate scales, but not at the smallest or largest scales.

If differences in abundance could produce the nested subset pattern, this would indicate that nestedness might be indirectly caused by processes like immigration and extinction because differences in distribution and abundance are present at all scales, while immigration and extinction only occur at local scales. The combined local extinctions of a species at sites throughout the entire geographic range of the species result in the global extinction of a species. Because nestedness has been observed when matrices have traditionally been sorted by incidence (Table 1), incidence and abundance tend to be highly correlated (Wright 1991), and previous researchers have observed positive correlations between abundance and incidence (Brown 1984, Gaston and Lawton 1990). I hypothesize that the matrix will be nested when sorted by incidence or abundance.

If an indirect link exists between the nested subset pattern and the processes traditionally thought to cause the pattern, i.e. immigration and extinction, examining how position in the geographic range influences the number of sites at which a species is present in the matrix could indicate if the variation in distribution and abundance is structured or unstructured. An indirect link between the nested subset pattern and immigration and extinction could be inferred if matrices are nested by incidence and abundance. If a pattern is observed between matrix position in the geographic range and the number of sites at which a species is present, this would indicate structured variation, while no pattern would suggest unstructured variation.

Brown (1984) described the abundant center hypothesis, a pattern between position in the geographic range of a species and abundance of that species. According to the abundant center hypothesis, species tend to be the most abundant in the center of the geographic range and least abundant at the edge of the geographic range (Brown 1984). If species occur at more sites in the matrix when the matrix is assembled from sites near the center of the geographic range of the species, this would indicate that smaller scale patterns of abundance within the geographic ranges of species could contribute to the nested subset pattern. I hypothesize that species will be observed at more sites in the matrix when the matrix is constructed from sites in the center of the geographic range of the species, and fewer sites when the matrix is constructed from sites near the edge of the geographic range of the species.

METHODS AND MATERIALS

DATA

I used the 2003 North American Breeding Bird Survey data set (BBS) for the conterminous United States to create the matrices used in the analyses (USGS Patuxent Wildlife Research Center, 2009). The BBS consists of routes 24.5 miles long with stops every 0.5 miles. The routes are surveyed each year by skilled volunteer observers. The BBS is administered so that the routes are run according to a standardized set of guidelines, which results in higher quality data. After the data were downloaded, I excluded all species from the dataset that were coded as unidentified, hybrids, or those that were only recorded at the generic level. Additionally, I recoded subspecies so that they were classified as the parent species. Only routes within the conterminous United States were used in the analyses because of problems with survey coverage associated with Alaska and Canada.

Although the BBS started in 1966, data became available for each stop starting in 1997. The year 2003 was arbitrarily selected from the years containing stop data (Figure 3). The North American Breeding Bird Survey has been noted to have a variety of potential error sources, including different abilities of recorders to appropriately identify species as well as sampling artifacts that could be present in the data (Bart and Shultz 1984, Bart et al. 1995, Kendall et al. 1996). In addition, some areas of the United State are poorly sampled, particularly in the western United States (Figure 3). However, the BBS is the most exhaustive large scale data set available, and the scale of the study will likely mitigate some of the noise present in the data. While a pattern might be obscured by noise with a small sample size, the pattern should be observed more clearly with a large sample size. Furthermore, more observers reduce the likelihood that all of the observers will have equivalent problems detecting the same species.

An evenly spaced grid was constructed in a stratified sampling design to provide seed points for the smallest extent of analysis (Figure 4). The 2003 BBS routes were apportioned to each seed point based on the Euclidean distance from the seed point. The geographic extent of the matrices was determined by the distance from the seed point. The initial extent was obtained by creating matrices from fifty stops within a single route with the shortest distance from the seed point. Larger geographic extents used the total incidence or abundance for each species summed over all stops along the route, causing each route to become a site within the matrix. Geographic extents at 50 km, 100 km, 150 km, and 200 km distance from the seed point contained both matrices with all the routes within a given radius from the seed point, and six randomly selected routes within a given radius from the seed point (Figure 5). Six routes were determined to be the minimum number of routes at a given scale in order to produce a matrix for analysis.

At extents greater than the 200 km radius from the seed point, routes were grouped together using a k-means clustering algorithm and a nearest neighbor algorithm (Figures 6 and 7). Both of these clustering techniques were used because they produce different clustering patterns, one forming a linear network, and the other forming tight clusters. The initial clustered matrices were constructed by using six randomly selected routes at a distance of 200 km from the seed point, and grouping those six together with six routes from another seed point. Clustered matrices for larger geographic extents were formed by grouping clusters from smaller geographic extents together (Figures 6 and 7). A matrix was not created for all of the seed points because it would have exceeded matrix dimension limits of Microsoft Excel. Two data matrices were created for each seed point at each scale; one sorted by incidence, and one sorted by total abundance of species as the independent variable.

SOFTWARE

Nested subset analyses have traditionally been performed with Monte Carlo simulations (Patterson and Atmar 1986). The observed matrix is sorted by species richness and incidence to create a tightly packed matrix that will minimize the number of unexpected absences and unexpected presences. A metric is then applied to the packed matrix to assess the observed number of unexpected absences, unexpected presences, or both. Monte Carlo simulations are performed by randomizing the observed data without replacement a set number of times. For each randomized matrix, the matrix is then sorted by incidence and species richness to produce a tightly packed matrix. From the packed matrix, a metric value is calculated, which is used to build a distribution of expected values. If the probability that the observed value, calculated from a packed matrix, is significantly different than the distribution of the expected values, then the observed matrix is determined to be significantly different from the null hypothesis (Manly 1990).

Existing software uses a variety of different metrics, which provide a method to assess the degree of nestedness, and null models, which provide a way to randomize data to build a distribution of values based on the expectation of no relationship between the dependant and independent variables. However, the existing software uses metrics and null models that provide a poor test of the nested subset pattern (Ulrich and Gotelli 2007), lack the ability to label rows and columns, or lack the ability to sort the matrix by independent variables. In order to properly test my hypotheses, I wrote a new program called Cudgel. Cudgel was written to meet the following criteria.

- Matrices are input into Cudgel with Microsoft Excel.
- Matrices can be run individually, or in batches.
- If run in a batch, the results are output in a batch summary, and as individual output files, also in Microsoft Excel.
- Species and site labels are preserved.
- The program sorts the matrices based on independent variables (Lomolino 1996), or by incidence and species richness, as in traditional analyses.

The Monte Carlo simulations were performed by randomizing the observed data matrix 10,000 times using a null model, which provided Cudgel with an algorithm to randomize the observed matrix, producing a distribution of expected matrices. The probability that the matrix was more nested than random chance was calculated for each expected matrix by dividing the number of times the metric value of the randomized matrices were less than or equal to the observed metric value by the total number of randomizations.

The program was validated by using test matrices of various dimensions and properties with known solutions to determine if Cudgel was producing the correct answers. The validation allowed me confidence that the program performed the analyses correctly. Although I would like Cudgel to ultimately have the ability to use all the metrics and null models currently in use for nestedness analysis, it is currently programmed with the metric and null model I used in my analyses.

The metric N_0 and the null model Random 1 were used in some of the original nestedness analyses by Patterson and Atmar (1987). N_0 calculates the total number of unexpected absences in the data matrix (Patterson and Atmar 1987). This is calculated by counting the number of sites where the species does not occur that are more species rich than the least species rich site where the species was present. Random 1, also called fixed-incidence proportional or PE, keeps the total number of presences of a species constant, but allows the presences to vary across sites (Patterson and Atmar 1987). This preserves the total number of times that a species was observed, but allows species richness to change from what was actually observed at any given site.

Although Ulrich and Gotelli (2007) concluded that the fixed-fixed null model in combination with the metrics N_1 or BR is a better choice for nestedness analyses, the fixed-fixed model has low statistical power and was rejected for this study in favor of the metric N_0 and the null model Random 1. N_0 and Random 1 performed well in combination to provide a good balance between type I and type II errors. Although N_0 is sensitive to matrix size (Ulrich and Gotelli 2007), the effect of matrix size can be statistically removed after analysis with linear regression so that the metric values from matrices of different size can be compared to each other.

FOCAL SPECIES

Focal species were used to test the hypothesis that species will be present at more sites in the matrix near the center of the geographic range than at the edge of the geographic range. Focal species were selected based on body size (mass), and geographic range size. The CRC Handbook of Avian Body Masses (Dunning 2007) was used to obtain data on the body masses of the potential focal species, while range maps were obtained from NatureServe (Ridgely et al. 2007) in collaboration with Robert Ridgely, James Zook, The Nature Conservancy - Migratory Bird Program, Conservation International - CABS, World Wildlife Fund - US, and Environment Canada – WILDSPACE, and geographic range sizes were calculated using Aerograph (Channell 2007). The geometric mean of body mass and of geographic range size was calculated. After the means were calculated, the birds were divided into four categories, large body size/large geographic range, small body size/small geographic range, large body size/small geographic range, and small body size/large geographic range.

For a species to be selected as a focal species, it had to meet the following criteria: native to North America, not be intensively managed (i.e. hunting or conservation), the center of the range and a range edge must fall within the borders of the contiguous United States, certain taxonomic status, and only one species per family could be included per category. Twenty species were to be randomly selected, five in each category. After application of the criteria, the large body size/ small geographic range category was eliminated because of an insufficient number of species, and the number of species was revised to twenty-one, seven in each remaining category (Table 2).

ANALYSES

Matrices were tested for nestedness using Cudgel, and statistical analyses were performed with SPSS version 12 (SPSS 2003). Figures were constructed using R, version 2.8.1 (R Development Core Team 2008). To determine if species enter and leave the nested subset matrix at the most species rich site, I determined the most species-rich site at which each species was present in the packed matrix at the within route scale and at the 200 km scale. Only significant matrices were used in this analysis.

Relative species position in the matrix was determined by counting the species with the highest incidence as the first rank, position 1, and the last rank as the species with the lowest incidence (Figure 8). The absolute species position in the matrix was determined by dividing the rank of the species in the matrix, counting from the species with the highest incidence to the species with the lowest incidence, by the total number of species in the matrix, so species that had higher incidences had values closer to zero, while the last species in the matrix had a value equal to one (Figure 8). I examined 36 significant matrices at the route level, using each stop on the route as a site in the matrix, for patterns in matrix position. I also analyzed 59 matrices at a geographic extent encompassing an area within a 200 km radius from the seed point. Based on the definition of nestedness, I expected that all species occur at the first site more frequently than other sites.

I tested if there was a correlation between the number of matrices at which the species was present, which was an index of regional commonness, and the average position of the species within a matrix, an index of local commonness. I expect that there will be a significant correlation between regional and local commonness.

To determine how nestedness was affected by geographic extent, a regression was performed between the metric value, N_0, and matrix size to provide residuals of N_0 and matrix size that could be used as a new metric value. This removed the effect of matrix size on the metric value, allowing comparison between matrices of different sizes. The residuals were graphed against geographic extent to determine how nestedness changed with scale. I expect that sites will be the most nested at intermediate scales, and the degree of nestedness will be equivalent between the large scale and small scale.

I used a Spearman correlation between incidence and abundance to test the hypothesis that more abundant species will also have greater incidence in the matrices than less abundant species. After the Spearman correlations were performed, a Mann-Whitney U was performed on the Spearman rho values to determine whether the correlation between incidence and abundance was greater in matrices that were significantly nested than in matrices that were not significantly nested (Zar 1999). Because I was using the correlation values as a metric, it was not necessary to perform a correction for multiple comparisons. I expect that more abundant species will have a greater relative rank (Figure 8) in the nestedness matrix, and that the correlation will be tighter for significant matrices vs. non-significant matrices.

I performed Spearman correlations to test if there was a correlation between geographic range size and presence or absence of species within the matrix. I selected the significant matrices from the smallest spatial extent and randomly selected an equal number of non-significant routes. I selected all of the non-significant 200 km scale routes, and randomly selected significant routes equal to the number of significant within route matrices. The result was 77 correlations, with a corrected α of 0.01. A B-Y method false discovery rate (FDR) correction was used instead of a Bonferroni correction to correct for multiple comparisons because the B-Y FDR has a good balance between type I and type II errors, while the Bonferroni correction is overly conservative and thus, has a greater probability of committing type II errors (Narum 2006).

Spearman correlations were used to test the hypothesis that species will be present at more sites in the matrix near the center of the range and fewer sites toward the edge of its range. The corrected B-Y FDR significance level was set at $\alpha = 0.015$. The 200 km scale was used for this analysis, and the average position within the geographic range was calculated for each seed at which each of the focal species was present. If the focal species was not present at a minimum of six seed points, it was excluded from analyses.

RESULTS

I graphed the absolute rank of species position in the matrix (Figure 8) against the rank of the most species-rich site at which the species was present at for two scales, within a route, and 200 km from the seed point (Figure 9, A and B). With the traditional explanation for nestedness, I would have expected to observe a horizontal scatter of points around zero. Species with high incidence had a low absolute position, low variability in initial position, and were first present at sites with higher species richness at both extents (Figure 9, A and B). Species with a high absolute position had low incidence, high variability in initial position, and were found first at sites that ranged from low to high species richness for both extents (Figure 9, A and B).

I graphed incidence of species in the matrix against initial site position at both scales, and observed a similar relationship (Figure 10, A and B). Species with higher incidences had low variability in initial position, and were first present at sites with higher species richness at both extents (Figure 10, A and B). Species with low incidence had high variability in initial position, and were found first at sites that ranged from low to high species richness at both extents (Figure 10, A and B).

To determine the relationship between the absolute species position in the matrix with incidence, I graphed incidence against absolute species position (Figure 11, A, B and C). Species with low incidence had higher absolute species position ranks, determined by dividing the rank of the species in the matrix counting from highest to lowest incidence by the total number of species in the matrix, while species with high incidence had low species position values (Figure 11, A, B and C). I expected to see a negative linear relationship between absolute species position and incidence if the hypothesis that species enter the nested subset matrix at the most species-rich site was supported (Figure 11, A, B and C). The non-linear relationship and the observation that the graphs do not have the same shape for all scales indicate that it is difficult to predict the position of species within the matrix because species position is highly dependent on the incidence of other species in the matrix (Figure 11, A, B and C). However, species with high incidences will tend to have a lower absolute rank in the matrix (Figure 8) while species with low incidences will tend to have a higher absolute rank in the matrix (Figure 8) (Figure 11, A, B and C).

I observed a significant correlation between regional commonness and local commonness ($rho = 0.619$, $n = 446$, $p < 0.001$). Species that are common locally also tend to be common regionally.

I hypothesized that the matrices would be nested at intermediate scales, but not at the smallest or largest scales. I found significant nestedness ($p < 0.05$) at all scales; however the proportion of sites that were significantly nested varied with scale (Figure 12). At the smallest scales, a majority of the routes were not nested, while at the largest scales, all routes were nested (Table 3). I had expected that matrices would be less nested at small extents and large extents than at intermediate extents. Although I observed a relationship between nestedness and scale, the observed relationship was different than I had initially hypothesized. To determine the degree of nestedness, I graphed the residuals of N_0 and matrix size against scale (Figure 13). Matrices that were not nested tended to have higher residual values than nested matrices (Figure 13). Residual values also tended to decrease as scale initially increased, and then increased at the largest scales (Figure 13).

I hypothesized that species with higher abundances will have an absolute rank in the matrix closer to zero when rank is counted from the species with the highest incidence or abundance to the lowest and divided by the total number of species in the matrix (Figure 8). I examined the matrices from within a route to 200 km distance from the seed point that contained all routes at each scale. These matrices were tested using a Spearman correlation to assess the relationship between incidence and abundance, relative to species position within the matrix. All matrices showed a significant correlation ($p<0.01$), allowing me to reject the null hypothesis that there would be no relationship between incidence and abundance. I performed a Mann-Whitney U comparing the Spearman rho values for significant matrices against the Spearman rho values of non-significant matrices, and observed a significant difference ($U = 7860$, $df = 231$, $p < 0.001$). Significant matrices showed a higher correlation between incidence and abundance than non-significant matrices. Because a closer relationship was observed for significant matrices, this indicates that distribution and abundance could contribute to the nested subset pattern.

I also examined if there was a correlation between species range size and incidence. The majority of Spearman correlations were not significant, indicating no relationship between geographic range size and incidence at these two scales (Table 4). I hypothesized that species would be present at more sites in the matrix near the center of the species geographic range. I used 200 km distance from the seed point matrices in this analysis. Anna’s Hummingbird (*Calypte anna*), Chuck-will’s-widow (*Caprimulgus* *carolinensis*), Chestnut-backed Chickadee (*Poecile rufescens*), and Tennessee Warbler (*Vermivora peregrina*) were excluded from this analysis because they were present in fewer than six matrices, the minimum needed to be able to perform Spearman correlations. I correlated absolute position in the matrix with average distance from the edge of the geographic range. Four species demonstrated a significant correlation between absolute matrix position and average distance from the edge of the geographic range, with two demonstrating a significantly negative correlation, and two demonstrating a positive correlation ($p < 0.015$) (Table 5).

DISCUSSION

The nested subset pattern states that the most species-rich site will contain all species, while less species-rich areas will contain a proper subset of the species found in all richer areas (Patterson and Atmar 1986). While the ideal pattern states that less species-rich areas will contain a proper subset of the species observed in richer areas, actual data sets do not show perfect nestedness; however, they show a pattern with more nestedness than would be expected by chance. Biological processes like immigration and extinction have traditionally been thought to produce the nested subset pattern; these processes acting indirectly could cause variation in the distribution and abundance of species that could produce nestedness at a variety of scales.

Macroecology, which studies patterns and processes at a variety of spatial and/or temporal scales to observe and explain statistical patterns in the data, provides an set of techniques that can be used to determine the scales over which the pattern is observed. Determining at what scales the pattern is observed might allow inference of processes that are more likely to be acting at those levels. A multi-scaled, macroecological approach could also provide insights into whether the pattern is being directly or indirectly produced by certain processes, as well as determining if the pattern is caused by structured or unstructured variation.

By analyzing the nested subset pattern, I attempted to identify its practical utility as a conservation tool, explore how patterns in commonness and rarity and the distribution of abundance influenced the nested subset pattern, and what those patterns in relation to the nested subset pattern revealed about community structure and assembly. I examined the utility of the nested subset pattern as a conservation tool by examining the relationship between species position in the matrix and site of first occurrence in the matrix, observing the scales over which nestedness was observed, and determining whether species were present at more sites in the matrix near the center of the geographic range of the species than the edge. To explore how patterns in commonness and rarity and the distribution of abundance influenced the nested subset pattern, I examined nestedness at a variety of scales, and determined whether species with greater abundances occurred higher in the nestedness matrix. Examining these different aspects of nestedness provided insights into community structure and assembly, and indicated future directions for research.

I did not observe a relationship between species position in the matrix and site of first occurrence in the matrix (Figure 8). Species that are rare have standardized species positions closer to one, which indicate they are present less often in the matrix. These species also have a great deal of variation in the site of first occurrence, the most species-rich site at which a species occurs. Common species do not have a great deal of variation in the site of first occurrence because they are constrained by having a greater number of occurrences. The distinction between common and rare species is also supported by the observation of the same relationship between incidence and site of first occurrence in the matrix. This indicates that the variation present in the matrices is not structured, because there is a great deal of unpredictable variation among the species richness of sites where rare species occur. The absence of points around the origin and in the upper right corner of the graph indicates that there are constraints to species position within the matrix relative to incidence (Figure 11, A, B and C), such that species with high incidences will be never have a high absolute ranking and be present at few sites, while species with low incidences will never have a low absolute ranking and be present at many sites (Figure 8).

If the pattern were being produced by a mechanistic process, this would mean that the nested subset matrix could be used predictively to assess what sites a species of interest would persist at the longest, and at what sites it would become extinct first (MacDonald and Brown 1992). Conversely, the pattern could be used to predict at what sites a species might immigrate to first, which could be used for invasive species monitoring or species reintroductions. However, if the processes producing the community structure are not mechanistic, then it would be wholly inappropriate to use this pattern as a conservation tool. The nested subset pattern is a community level pattern, and the pattern would be produced by mechanistic processes if the processes were acting at the community level. However, if the nested subset pattern is emergent at the community level and is produced by population level processes, then it would not be appropriate to use the pattern for conservation. In this case, conservation efforts should focus on the population level because the processes are acting at that level. In addition, questions have already been raised regarding the utility of the nested subset pattern in conservation when it is not perfectly nested (Fischer and Lindenmayer 2005).

The relationships observed between species position in the matrix and site of first occurrence have important implications for the utility of the nested subset pattern in conservation. Nestedness has been proposed as a conservation tool in the past (Patterson 1987, MacDonald and Brown 1992, Fleishman et al. 2002). Because it is known that species do not respond in the same manner to all stimuli, it is most likely inappropriate to use the nested subset pattern to model community responses to a given scenario as if all the species in the community responded in the same manner. Because conservation is performed at the population level, a community level pattern is at the wrong level in the ecological hierarchy, and should not be used for conservation.

Target species for conservation tend to be present at few sites and are present in lower abundances where they do occur. These species are rare, occur at fewer sites in the nested subset matrix, and have standardized species positions closer to one. It is inappropriate to use the nested subset pattern for conservation planning for two reasons. The nested subset pattern is not present in the majority of cases at small geographic extents, the scale at which conservation efforts are typically directed. If the pattern is not observed at a site, it cannot be used for conservation. In addition, conservation efforts target rare species that are present at few sites. Common species that are present at many sites are probably not in need of conservation. However, the pattern can accurately predict sites at which common species will be present, but fails to accurately predict the sites at which rare species occur (Figure 12). The high degree of variability in the site of first occurrence relative to species position in the matrix indicates that rare species do not consistently occur more frequently at sites that are more species-rich than sites that are less species-rich (Figure 7). Because the pattern cannot accurately predict the sites where rare species that could be in need of conservation might occur, the pattern has limited conservation utility.

The abundant center hypothesis describes a pattern in which species are most abundant and occur at more sites in the center of the geographic range than at the edge of their geographic range (Brown 1984). While many researchers initially documented this pattern of distribution and abundance (Brown et al. 1996, Gaston et al. 1997, Warren and

Gaston 1997, Thompson et al. 1998), more recent papers have questioned the generality of this pattern (Sagarin and Gaines 2002, Kolb et al. 2006, Symonds and Johnson 2006). Although Brown et al. (1996) considered their results to be consistent with the abundant center hypothesis, they did note that ranges were often characterized by “hotspots”, or areas within the geographic range that contained higher abundances than predicted by the pattern.

Using the same focal species as I used in the analysis between position in the geographic range and position in the matrix, I performed a test of the abundant center hypothesis and observed no pattern between position in the geographic range and abundance or presence at a site (Baldridge, unpublished). The results from the nested subset analysis also indicated no relationship between position in the geographic range and the number of sites at which a species was present in the matrix, which was contrary to my initial expectations, but supported more recent reservations about the generality of the abundant center hypothesis. However, these results are consistent with the idea that local communities are assembled based on the individual requirements of the species in the community. This is also consistent with the pattern being driven by the common species, in that common species are present at most sites throughout the geographic range of that species and rarer species are only present at a few sites according to individual requirements, regardless of position of the site in the range. This also has important conservation implications. Conservation sites are often prioritized based on position of the site within the geographic range using the abundant center hypothesis as an underlying assumption. Based on these results, position in the geographic range should no longer be a criterion for conservation prioritization. Species should be conserved where viable populations are located, regardless of the position of the site within the geographic range.

The signal from common species, which tend to have high incidence throughout their geographic range and have higher abundances, overwhelms the noise that is produced by the rare species, which could have higher incidences and abundances locally, but tend to have low incidences and abundances regionally (Figure 11). This indicates that the nested subset pattern is being driven by the presence and absence of the common species in these matrices, and will persist regardless of the position of the rare species in the matrix (Figure 11). In Figure 9, graphs A, B, and C exhibit a curved relationship between incidence and species position in the matrix. While I expected to see variation around a general trend, I had expected that variation to occur in a negative linear relationship. The deformation of the expected line indicates that the species with low incidences are present at lower positions in the matrix and are interchangeable to a large extent in the matrix, as indicated by the spread of points from an absolute position of 0.6 to 1 in graphs A and C in Figure 9 for species that were present at a single site. Because common species are present in most communities (MacArthur 1960) and would tend to have an absolute rank closer to zero (Figure 8), the nested subset pattern could be driven by the signal from the common species, which show less variability than the rare species (Figure 9, A, B, and C), thus allowing pattern detection.

I observed a significant correlation between local and regional abundance and incidence. How common or rare a species is related to the distribution and abundance of that species, with common species tending to have larger distributions, abundances, and incidences than rare species (Brown 1984). However, most species are rare (Preston

1948), and it is only a few species that are common throughout their entire range. Because the nested subset pattern appears to be driven by commonness and rarity, it could be suggested that passive sampling is the cause of the nested subset pattern because common species are detected more frequently than rarer species. With the BBS data, observer effects could produce sampling effects because certain species are more likely to be detected than others. It would be more likely for passive sampling to produce the nested subset pattern at smaller scales than larger scales. At a small geographic extent, within a route, passive sampling from observer effects could produce the nested subset pattern. However, because the majority of sites were not nested at the smallest scales, it is highly unlikely that passive sampling is producing the nested subset pattern in these analyses. Although passive sampling can produce the nested subset pattern, passive sampling would be unlikely to produce the pattern at all scales. As the spatial extent increases and routes are combined, it is unlikely that all observers will have problems detecting species in the same way. Because it is highly unlikely that observers will have detection problems in the same way, the effect should not be additive, and any signal produced by passive sampling will become noise at larger extents.

In order to determine how the degree of nestedness changed with spatial scale, I needed to compare the metric values of N_0 among matrices with different geographic extents. One problem associated with directly comparing N_0 values from one matrix to other matrices is the sensitivity of N_0 to matrix size. As a matrix becomes more nested, the metric value will decrease, assuming that the metric value is not related to size, which allows for the comparison of the degree of nestedness among matrices. However, as the spatial extent of my study increased, matrix size increased, and the metric values also increased linearly relative to matrix size, rather than to the degree of nestedness associated with increasing scale.

One way to approach the problem of matrix size inflating the metric values is to use a metric that is insensitive to matrix size. Ulrich et al. (2009) favored the metric NODF for its insensitivity to matrix size; however, NODF was not used because it was not included in the analyses by Ulrich and Gotelli (2007) to evaluate its performance with the various null models. In addition, other metrics that were supposed to be insensitive to matrix size, thus allowing for comparison of different sized matrices, have been subsequently found to be somewhat dependant on matrix size and fill (Ulrich et al. 2009).

Because some other metrics that were originally considered insensitive to size have been determined to be sensitive under further testing, statistically removing the effect of matrix size with linear regression seemed like a more practical approach. Using the residuals of N_0 and matrix size statistically removed the effect of matrix size on N_0, which provided new metric values of nestedness that were comparable between differently sized matrices. When examining the graph of the residuals against geographic extent, there is a discontinuity between the 200 km extent and the larger extents of grouped seed points, but this is likely caused by the method of matrix assembly, rather than anything of biological significance.

I had expected matrices to be most nested at intermediate extents, and had expected the smallest and largest spatial extents to be equally poorly nested. At small scales, I expected that local variation would obscure the pattern, and I thought that the nested subset pattern would no longer be observed at large spatial scales because matrices are made up of species from different regional species pools. Intermediate extents tended to have lower metric values than small extents or large extents, which followed my predictions. Contrary to my expectations, the largest extents had metric values that were substantially higher than the smallest extents. I also thought that I would have a larger proportion of significantly nested matrices at intermediate geographic extents, but less at large and small geographic extents. While the proportion of significantly nested matrices was the lowest at small spatial extents, the proportion of significant matrices increased with increasing extent. Although matrices at the largest scales were all significantly nested, the metric values were substantially greater than the non-significant matrices at the smallest scales. The discrepancy in metric values indicates a pattern that might be occurring at the largest geographic extents.

At the largest scales, the matrices began to contain species from more than one regional species pool. Because more of the species in the matrix do not co-occur, this inflates the metric values. If species do not co-occur at the same sites, more unexpected absences and unexpected presences will occur in the matrix, inflating metric values. In addition, matrix fill is more variable at smaller scales, and fill decreases at larger spatial scales (Figure 14). Decreased fill allows matrices with higher metric values to be nested because there are more randomizations possible. Matrix fill also indicates patterns in community structure. While there might be many species in the regional species pool, local communities are assembled from only a few of the species in the regional species pool. Additionally, local communities consist of common species, which are present in most of the communities (MacArthur 1960), while the rarer species vary from local community to local community (Hairston 1959, Hanski 1982). However, as noted earlier, the common species enhance pattern detectability, overwhelming the noise caused by the rarer species.

Matrices are nested when sorted by incidence or by abundance, although metric values indicate a more nested structure with incidence than with abundance. Metric values are more nested with incidence because hypotheses are not tested by sorting with incidence; rather it is a means to pack the matrix as tightly as possible and observe if the pattern observed is significantly different from random. Sorting with a different variable, e.g. abundance, will result in metric values that indicate a less nested structure. However, if the matrix is still nested when sorted by abundance or an independent variable, then there is a relationship between nestedness and the independent variable (Lomolino 1996). If not, the variable is not contributing to the pattern.

Significantly nested matrices demonstrate a tighter correlation between incidence and abundance. Because there is a tighter correlation between incidence and abundance in significantly nested matrices, this relationship indicates that differences in the distribution and abundance of species could be contributing to the nested subset pattern. If abundance and incidence were correlated equally between significant and non-significant matrices, it would have indicated that there was not a relationship between abundance and nestedness. Because the relationship between incidence and abundance is observed across a variety of spatial scales, the relationship indicates that the traditional processes (i.e. immigration and extinction) thought to produce the nested subset pattern are not likely to be directly producing the pattern. Extinction as a process only operates on local scales, at a site. At larger spatial scales, the sum of extinction events at several sites could produce the nested subset pattern, and the sum of extinction events is the distribution and abundance of species.

Although I would have liked to test to see if the matrices were more or less nested by extinction or immigration by considering extinction rate to be a function of area and immigration to be a function of isolation (MacArthur and Wilson 1967), it would have been impossible to determine the area and isolation for each of the sites in the matrix in an open, mainland system. Because this strategy could not be used, abundance was used to support the idea that differences in the distribution and abundance of species are directly producing the pattern, although immigration and extinction could be indirectly contributing to the pattern. Immigration and extinction could indirectly produce the pattern by causing differences in the distribution and abundance of species. Differences in distribution and abundance could produce the pattern directly at all spatial scales, while immigration and extinction could only produce the pattern directly on local scales, and indirectly at larger spatial scales.

The nested subset pattern could be related to the distribution of abundance and immigration and extinction in a variety of ways (Figure 14). While model A, which indicates a direct link between immigration and extinction and nestedness, has been the traditional explanation for the nested subset pattern, this model does not take into account the known relationship between distribution and abundance and immigration and extinction (Figure 14). Organisms with restricted distributions and large body sizes have higher extinction risks than organisms with smaller body sizes or broader distributions (Owens and Bennett 2000). In addition, species with lower abundances have higher extinction risks (Lawton and May 1995), and extinctions of local populations of a species will decrease regional abundance. Examining the nested subset pattern from a variety of scales provides insight into the processes that might be structuring the pattern. As a process, extinction only acts at the local scale, which makes model B, which indicates a direct link between nestedness and abundance and distribution, unable to operate at all scales. For model C, which suggests that the nested subset pattern is produced directly by both immigration and extinction and distribution and abundance, to produce the nested subset pattern, some of the species or sites in the nested subset matrix would have to be structured based solely on immigration or extinction, and some would have to be structured based on distribution and abundance, which is unlikely, due to the multiple scales over which nestedness is observed (Figure 14). Model D, indicating a direct link between distribution and abundance and nestedness, is the most likely scenario, because differences in distribution and abundance are observed at all spatial scales.

A nested subset matrix examines alpha diversity, the diversity of a site, and gamma diversity, the diversity of a region. The individual sites that create a nested subset matrix are at the local scale, while the entire matrix is at the regional scale. However, both alpha diversity and gamma diversity change with the scale of analysis. Ultimately, local communities can only be assembled from those species that are present in the regional species pool (Gaston 2000). On an ecological time scale at the regional level, the species pool is formed by processes that increase or decrease species richness, i.e. immigration and extinction. While the species pool is assembled through immigration and extinction, presence or absence of species at a given site is largely due to the niche of the species.

Because common species tend to be more tolerant of a wide variety of conditions and have higher abundances, they tend to be present at more sites than rarer species (Brown 1984). Rare species tend to be at more species-rich sites because they occur at those sites in addition to the common species already present, but occur in lower abundances than the common species where they do occur. Over the entire geographic extent, common species will tend to have higher abundances and larger distributions than rarer species (Brown 1984). This variation in the distribution and abundance of organisms at sites could result in the production of the nested subset pattern.

I did not observe a significant correlation between geographic range size and incidence. It has been assumed that species with large geographic ranges will have higher incidences than species with smaller geographic ranges (Brown 1984, Brown et al. 1996), i.e. species with larger ranges will be more common. Although I did not observe a pattern between geographic range size and incidence, it could have been due to the spatial extent of analysis. Because many species will have a geographic range size larger than the area encompassed by 200 km distance from a seed point, the extent might not have been large enough to detect a pattern. However, this indicates that a relationship between geographic range size and incidence is not likely to be structuring the nested subset pattern at smaller spatial extents. Because geographic range size and incidence are not factors at smaller extents, it is unlikely that this relationship is important in producing the nested subset pattern.

In addition, the majority of species are not common, as evidenced by the lognormal curve (Preston 1962a, b). Although the lognormal curve is extremely common in ecology and many models have been proposed as explanations of the pattern, the cause of the pattern is uncertain (May 1975, Sugihara 1980, Tokeshi 1990). The lognormal curve of abundance exhibits a right-skewed curve; the modal group has low abundances, and only a few species have very high abundances. Because the majority of species are rare, tending to have low abundances and incidences (Brown 1984), this creates noise in the analysis that prevents detection of a pattern at smaller spatial extents. At larger extents, common species tend to have larger geographic ranges, which will allow them to be present at more sites within the entire extent of the matrices, producing a stronger signal, while the rarer species will be present at comparatively fewer sites. Thus, commonness and rarity are caused by other factors, although commonness and rarity tend to be correlated with range size over larger geographic extents.

Since Patterson and Atmar (1986) first discussed the nested subset pattern, it has been used to describe a long list of regions and taxa where the pattern has been observed, to study patterns and processes producing community structure, used as a conservation tool, and provided endless debate about the appropriate metrics and null models that should be used to identify the pattern (Table 1). Simberloff and Martin (1991) indicated that the interesting part of nested subset analysis was not whether the pattern was found or not, but which species did not conform to the pattern and why.

In addition, because rare species do not occur predictably at the first sites in the matrix, as well as patterns of nestedness with scale, and the lack of pattern regarding position in the range and presence at sites, indicates that local communities are assembled based on the individual requirements of the species in the communities and the nested subset pattern is being produced by unstructured variation. If I had observed a relationship between position of the species in the matrix and site of first occurrence, this would have indicated that the variation in species composition was predictable, and the variation would therefore be structured. Because I did not observe a relationship between position of species in the matrix and site of first occurrence for all species, the variation was unpredictable and unstructured. In addition, because the pattern was not observed at small spatial extents in the majority of cases, but was observed at larger extents, this indicates that the variation of species present at each site was unstructured. If the variation was structured, I would have expected to have observed an equivalent proportion of nested sites at all geographic extents.

The nested subset pattern occurs when the most species-rich site contains all species, while less species-rich sites contain a subset of the species from the most species-rich site (Patterson and Atmar 1986). The canonical pattern suggests that nestedness occurs because the rare species follow a predictable pattern; namely that rare species are present at more species-rich sites. However, the relationship between species richness of a site and presence of rare species was not observed in this study, indicating that the pattern could be driven by the presence of common species. The dominant processes thought to cause the nested subset pattern are immigration and extinction. Because immigration and extinction are processes that occur at the local scale and nestedness was observed at all scales, although less at the local scale, it is unlikely that immigration and extinction could be producing the nested subset pattern directly. The distribution and abundance of species, created by immigration and extinction, is more likely to be producing the nested subset pattern.

REFERENCES

Armbrecht, I., et. al. 2001. Nested subsets and partition patterns in ant assemblages (Hymenoptera, Formicidae) of Colombian dry forest fragments. – Pan-Pacific Entomologist 77: 196-209.

Barkely, T. M. 1993. Synthesis: A historical perspective. – Annals of the Missouri Botanical Garden 80: 292-296.

Bart, J. and J. D. Shultz. 1984. Reliability of singing bird surveys: Changes in observer efficiency with avian density. – Auk 101: 307-318.

Bart, J. et al. 1995. Reliability of the breeding bird survey: Effects of restricting surveys to roads. – Auk 112: 758-767.

Blake, J. G. 1991. Nested subsets and the distribution of birds on isolated woodlots. – Conserv. Biol. 5: 58-66.

Bloch, C. P. et al. 2007. Effects of large-scale disturbance on metacommunity structure of terrestrial gastropods: Temporal trends in nestedness. – Oikos 116: 395-406.

Bolger, D. T. 1991. Occurrence patterns of bird species in habitat fragments: Sampling, extinction, and nested species subsets. – Am. Nat. 137: 155-166.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. – Am. Nat. 124: 255-279.

Brown, J. H. 1995. Macroecology. – University of Chicago Press, Chicago, IL.

Brown, J. H. and Maurer, B. A. 1989. Macroecology: The division of food and space among species on continents. –Science 243: 1145-1150.

Brown, J. H. et al. 1996. The geographic range: Size, shape, boundaries, and internal structure.

– Annu. Rev. Ecol. Evol. Syst. 27: 597-623.

Channell, R. 2007. Aerograph, version 7.0b: Biogeographer's Toolbox.

Cook, R. R. and J. F. Quinn. 1995. The influence of colonization in nested species subsets. – Oecologia 102: 413-424.

Cook, R. R. et al. 2004. Geographic variation in patterns of nestedness among local stream fish assemblages in Virginia. – Oecologia 140: 639-649.

Cutler, A. H. 1991. Nested faunas and extinction in fragmented habitats. – Conserv. Biol. 5: 496-505.

Cutler, A. H. 1994. Nested biotas and biological conservation: Metrics, mechanisms, and meaning of nestedness. – Landscape and Urban Planning 28: 73-82.

Diamond, J. M. 1975. Assembly of species communities. In: Cody, M.L, and Diamond, J.M. (eds.) Ecology and evolution of communities. Harvard University Press, pp. 342-444.

Dunning, J. B., ed. 2007. CRC Handbook of Avian Body Masses, 2nd ed. – CRC Press, Taylor and Francis Group.

Fernandez-Juricic, E. 2002. Can human disturbance promote nestedness? A case study with breeding birds in urban habitat fragments. – Oecologia 131:269-278.

Fischer, J. and Lindenmayer, D. B. 2005. Perfectly nested or significantly nested - an important difference for conservation management. – Oikos 109: 485-494.

Fleishman, E. and MacNally, R. M. 2002. Topographic determinants of faunal nestedness in Great Basin butterfly assemblages: Applications to conservation planning. – Conserv. Biol. 16: 422-429.

Fleishman, E. et al. 2002. Nestedness analysis and conservation planning: The importance of place, environment, and life history across taxonomic groups. – Oecologia 133: 78-89.

Fox, B. J. 1990. Small mammal community pattern in Australian heathland: A taxonomic rule for species assembly. In: Morris, D.W. et al (eds.) Patterns in the structure of mammalian communities. pp. 91-103. Special publications of the Museum, Texas Tech University.

Ganzhorn, J. U. and Eisenbeiß, B. 2001. The concept of nested species assemblages and its utility for understanding effects of habitat fragmentation. – Basic Applied Ecology 2: 87-95.

Gaston, K. J. 2000. Global patterns in biodiversity. – Nature 405: 220-226.

Gaston, K. J. and Lawton, J. H. 1990. Effects of scale and habitat on the relationship between regional disturbance and local abundance. – Oikos 58: 329-335.

Gaston, K. J. et al. 1997. Abundance-range size relationships of breeding and wintering birds in Britain: A comparative analysis. – Ecography 20: 569-579.

Greve, M. et al. 2005. Nestedness of Southern Ocean island biotas: Ecological perspectives on a biogeographical conundrum. – J. Biogeogr. 32: 155-168.

Guegan, J. F. and Hugueny, B. 1994. A nested parasite species subset pattern in tropical fish: Host as major determinant of parasite infracommunity structure. – Oecologia 100: 184-189.

Hairston, N. 1959. Abundance and community organization. – Ecology 40: 404-416.

Hanski, I. 1982. Dynamics of regional distribution: The core and satellite species hypothesis. – Oikos 38: 210 - 221.

Hansson, L. 1998. Nestedness as a conservation tool: Plants and birds of oak-hazel woodland in Sweden. – Ecol. Lett. 1: 142-145.

Hausdorf, B. and Hennig, C. 2003. Nestedness of north-west European land snail ranges as a consequence of differential immigration from Pleistocene glacial refuges. – Oecologia 135: 102-109.

Heino, J. et al. 2008. Temporal variability of nestedness and idiosyncratic species in stream insect assemblages. – Divers. Distrib. 15:198 - 206.

Hutchinson, G. E. 1957. Concluding remarks. – Cold Spring Harbour Symposium of Quantitative Biology 22: 415–427.

Hylander, K. et al. 2005. Differences in habitat quality explain nestedness in a land snail meta-community. – Oikos 108:351-361.

Jacquemyn, H. et al. 2007. Range size variation, nestedness and species turnover of orchid species along an altitudinal gradient on Reunion Island: Implications for conservation. – Biol. Conserv. 136:3 388-397.

Kadmon, R. 1995. Nested species subsets and geographic isolation: A case study. – Ecology 76: 458-465.

Kendall, W. L. et al. 1996. First-time observer effects in the North American Breeding Bird Survey. –Auk 113: 823-829.

Kolb, A. et al. 2006. Determinants of local abundance and range size in forest vascular plants. – Global Ecol. Biogeogr. 15: 237-247.

Krasnov, B. R. et al. 2005. Nested pattern in flea assemblages across the host’s geographic range. – Ecography 28: 475-484.

Lawton, J. H. and May, R. M., eds. 1995. Extinction Rates. – Oxford University Press, Oxford.

Lomolino, M. V. 1996. Investigating causality of nestedness of insular communities: Selective immigrations or extinctions? – J. Biogeogr. 23:699-713.

Loo, S. E. et al. 2002. An experimental examination of colonization as a generator of biotic nestedness. – Oecologia 132: 118-124.

MacArthur, R. H. 1960. On the relative abundance of species. – Am. Nat. 94: 25-36.

MacArthur, R. H. and Wilson, E. O. 1967. The Theory of Island Biogeography. – Princeton, N.J.: Princeton University Press.

MacDonald, K. A. and Brown, J. H. 1992. Using montane mammals to model extinctions due to global climate change. – Conserv. Biol. 6: 409-415.

MacNally, R. M. et al. 2002. Nestedness in fragmented landscapes: birds of the box-ironbark forest of south-eastern Australia. – Ecography 25: 651-660.

Manly, F. J. 1990. Randomization and Monte Carlo Method. – Chapman & Hall/CRC.

May, R. M. 1975. Patterns of species abundance and diversity. In: Cody, M. L and Diamond, J. R. (eds.) Ecology and evolution of communities. pp. 81-120.

M'Closkey, R. T. 1978. Niche separation and assembly in four species of Sonoran Desert rodents. – Am.Nat. 112: 683-694.

Meyer, C. F. J. and Kalko, E. K. V. 2008. Nestedness in fragmented landscapes: Birds of the box-ironbark forest of south-eastern Australia. – Ecography 25: 651-660.

Narum, S. R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. – Conservation Genetics 7: 783-787.

Owens, I. P. F. and Bennett, P. M. 2000. Ecological basis of extinction risk in birds: Habitat loss versus human persecution and introduced predators. – PNAS 97: 12144-1214.

Patterson, B. D. 1987. The principle of nested subsets and its implications for biological conservation. – Conserv. Biol. 1: 323-334.

Patterson, B. D. and Atmar, W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. – Biol. J. Linn. Soc. 28: 65-82.

Patterson, B. D. and Atmar, W. 1987. The principle of nested subsets and the structure of insular mammalian faunas and archipelagos. – Biol. J. Linn. Soc. 28:451-463.

Patterson, B. D. and Brown, J. H. 1991. Regionally nested patterns of species composition in granivorous rodent assemblages. – J. Biogeogr. 18: 395-402.

Preston, F. W. 1948. The commonness, and rarity, of species. – Ecology 29: 254-283.

Preston, F. W. 1962a. The canonical distribution of commonness and rarity, Part I. – Ecology 43: 185-215.

Preston, F. W. 1962b. The canonical distribution of commonness and rarity, Part II. – Ecology 43: 410-432.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. – ISBN 3-900051-07-0, <http://www.R-project.org>.

Rashleigh, B. 2008. Nestedness in riverine mussel communities: Patterns across sites and fish hosts. – Ecography 31: 612-619.

Ridgely, R. S. et al. 2007. Digital Distribution Maps of the Birds of the Western Hemisphere, version 3.0. – NatureServe, Arlington, Virginia, USA.

Sagarin, R. and Gaines, S. D. 2002. The ‘abundant centre’ distribution: To what extent is it a biogeographical rule? – Ecol. Lett. 5: 137-147.

Sfenthourakis, S. et al. 2004. From sampling stations to archipelagos: Investigating aspects of the assemblage of insular biota. – Global Ecol. Biogeogr. 13: 23-35.

Simberloff, D. and Martin, J. L. 1991. Nestedness of insular avifaunas: Simple summary statistics masking complex species patterns. – Ornis Fennica 68: 178-192.

Simonetti, J. A. 1994. Impoverishment and nestedness in caviomorph assemblages. – J. Mamm. 75: 979-984.

SPSS, version 12. 2003. – SPSS Inc., Chicago IL.

Sugihara, G. 1980. Minimal community structure: An explanation of species abundance patterns. – Am. Nat. 116: 770-787.

Summerville, K. S. et al. 2002. Does variation in patch use among butterfly species contribute to nestedness at fine spatial scales? – Oikos 97: 195-204.

Symonds, M. R. E. and Johnson, C. N. 2006. Range size-abundance relationships in Australian passerines. – Global Ecol. Biogeogr. 15:143-152.

Taylor, C. M. and Warren, M. L. 2001. Dynamics in species composition of stream fish assemblages: Environmental variability and nested subsets. – Ecology 82: 2320-2330.

Thompson, K. et al. 1998. Abundance-range size relationships in the herbaceous flora of central England. – J. Ecol. 86: 439-448.

Tokeshi, M. 1990. Niche apportionment or random assortment: Species abundance patterns revisited. – J. Anim. Ecol. 59: 1129-1146.

Ulrich, W. and Gotelli, N. 2007. Disentangling community patterns of nestedness and species co-occurrence. – Oikos 116: 2053-2061.

Ulrich, W. et al. 2009. A consumer's guide to nestedness analysis. – Oikos 118: 3-17.

USGS Patuxent Wildlife Research Center. 2009. North American Breeding Bird Survey, version 2009. 1. – <ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/>.

Warren, P. H. and Gaston, K. J. 1997. Interspecific abundance-occupancy relationships: A test of mechanisms using microcosms. – J. Anim. Ecol. 66: 730-742.

Wright, D. H. 1991. Correlations between incidence and abundance are expected by chance. J. Biogeogr. 18: 463-466.

Wright, D. H. et al. 2007. Changes in nestedness in experimental communities of soil fauna undergoing extinction. – Pedobiologia 50: 479-503.

Yiming, L. et al. 1998. Nested distribution of amphibians in the Zhoushan archipelago, China: Can selective extinction cause nested subsets of species? – Oecologia 113: 557-564.

Zar, J. H. 1999. Biostatistical analysis. – Prentice-Hall, Upper Saddle River.

Table 1: Literature review of studies on the nested subset pattern and results.

Region	Taxonomic Group	Results	Mainland or Island	Fragmented	Study
North America	Mammals	Nested	Mainland (isolated habitats)	Yes	Cutler 1991.
		Nested	Mainland (isolated habitats)	No	Patterson and Atmar 1986.
		Nested	Mainland	No	Patterson and Brown 1991.
	Birds	Nested	Mainland	Yes	Blake 1991.
		Nested	Mainland	Yes	Bolger 1991.
		Nested	Mainland (isolated habitats)	Yes	Cutler 1991.
		Nested	Mainland	No	Fleishman et al. 2002.
		Nested	Island	No	Simberloff and Martin 1991.
	Fish	Nested	Mainland	No	Cook et al. 2004.
		Nested	Mainland	No	Taylor and Warren 2001.
	Insects	Nested	Mainland (isolated habitats)	No	Fleishman and Mac Nally 2002.
		Nested	Mainland	No	Fleishman et al. 2002.
		Nested	Mainland	No	Summerville et al. 2002.
	Miscellaneous Invertebrates	Nested	Mainland	No	Bloch et al. 2007.
		Nested	Mainland	No	Rashleigh 2008.
		Nested	Mainland	No	Wright et al. 2007.
	Plants	Nested	Island	No	Kadmon 1995.
South America	Mammals	Nested	Mainland	No	Simonetti 1994.
		Nested	Island	No	Meyer and Kalko 2008.
	Insects	Nested	Mainland	Yes	Armbrecht et al. 2001.

Region	Taxonomic Group	Results	Mainland or Island	Fragmented	Study
Australia	Mammals	Not Nested	Mainland	Yes	Fischer and Lindenmeyer 2005.
	Birds	Nested	Mainland	Yes	Fischer and Lindenmeyer 2005.
		Nested	Mainland	Yes	MacNally et al. 2002.
	Reptiles	Not Nested	Mainland	Yes	Fischer and Lindenmeyer 2005.
	Miscellaneous Invertebrates	Nested	Mainland	No	Loo et al. 2002.
Africa	Mammals	Nested	Mainland	Yes	Ganzhorn and Eisenbeiß 2001.
	Miscellaneous Invertebrates	Nested	Mainland	No	Guegan and Hugueny 1994.
	Plants	Nested	Mainland	Yes	Jacquemyn et al. 2007.
Eurasia	Birds	Nested	Mainland	No	Hansson 1998.
		Nested	Mainland	Yes	Ganzhorn and Eisenbeiß 2001.
		Nested	Mainland	Yes	Fernandez-Juricic 2002.
	Amphibians	Nested	Island	No	Yiming et al. 1998.
	Insects	Nested	Mainland	No	Krasnov et al. 2005.
		Nested			Heino et al. 2008.
	Miscellaneous Invertebrates	Nested	Mainland	No	Hausdorf and Hennig 2003.
		Nested	Mainland	No	Hylander et al. 2005.
		Nested	Island	No	Sfenthourakis et al. 2004.
		Nested			Wright et al. 2007.
	Plants	Nested	Mainland	No	Hansson 1998.

Region	Taxonomic Group	Results	Mainland or Island	Fragmented	Study
Antarctic	Birds	Nested	Island	No	Greve et al. 2005.
	Insects	Nested	Island	No	Greve et al. 2005.
	Plants	Nested	Island	No	Greve et al. 2005.

Table 2. List of focal species by category, including body masses and geographic range sizes.

Species Category	Scientific name	Common name	Range Size (km²)	Body mass (g)
Large mass, large range	*Charadrius vociferus*	Killdeer	15800000	96.6
	Quiscalus quiscula	Common Grackle	8480000	106.1
	Caprimulgus carolinensis	Chuck-Will's-Widow	3280000	109.0
	Megaceryle alcyon	Belted Kingfisher	18100000	148.0
	Dryocopus pileatus	Pileated Woodpecker	5930000	287.3
	Podilymbus podiceps	Pied-billed Grebe	21100000	416.0
	Falco mexicanus	Prairie Falcon	4610000	734.0
Group average (148 species)			**8950000**	**203.6**
Small mass, large range	*Cistothorus platensis*	Sedge Wren	9500000	8.2
	Vermivora peregrina	Tennessee Warbler	12200000	8.9
	Melospiza lincolnii	Lincoln's Sparrow	16100000	16.6
	Icterus spurius	Orchard Oriole	6750000	19.5
	Sialia sialis	Eastern Bluebird	5900000	27.5
	Aeronautes saxatalis	White-throated Swift	3400000	32.1
	Picoides villosus	Hairy Woodpecker	13300000	63.3
Group average (216 species)			**8600000**	**20.1**
Small mass, small range	*Calypte anna*	Anna's Hummingbird	1270000	4.3
	Poecile rufescens	Chestnut-backed Chickadee	958000	9.4
	Sitta pygmaea	Pygmy Nuthatch	1950000	10.6
	Spinus lawrencei	Lawrence's Goldfinch	557000	11.0
	Vireo vicinior	Gray Vireo	654000	12.8
	Limnothlypis swainsonii	Swainson's Warbler	1750000	18.9
	Sphyrapicus thyroideus	Williamson's Sapsucker	1250000	47.6
Group average (130 species)			**1100000**	**12.7**

Table 3. Proportion of routes significantly nested at each scale of analysis, geographic extent increases from top to bottom of the table.

Extent	Grouping	Abundance	Incidence	# of Matrices	Total % Significant
Smallest extent	50 Stops	23	23	59	39%
	Scale 1	3	4	10	30%
↓	Scale 2	74	72	104	70%
	Scale 3	92	92	116	79%
	Scale 4	96	96	118	81%
	Scale 5	47	47	56	84%
	Scale 6	4	4	4	100%
	Scale 7	19	20	22	89%
	Scale 8	2	2	2	100%
	Scale 9	5	5	5	100%
	Scale 10	8	8	8	100%
	Scale 11	2	2	2	100%
	Scale 12	2	2	2	100%
	Scale 13	3	3	3	100%
	Scale 14	2	2	2	100%
	Scale 15	1	1	1	100%
	Scale 16	1	1	1	100%
Largest extent	Scale 17	1	1	1	100%

Table 4. Spearman correlation results between range size and incidence for nested and non-nested matrices at two geographic extents.

A. Correlation results at the 200 km scale.

B. Correlation results within route scale.

C. Correlation results for both extents.

A

	Nested	**Not nested**	**Total**
Significant	2	0	2
Not significant	21	8	29
Total	23	8	31

B

	Nested	**Not nested**	**Total**
Significant	1	1	2
Not significant	22	22	44
Total	23	23	46

C

	Nested	**Not nested**	**Total**
Significant	3	1	4
Not significant	43	30	73
Total	46	31	77

Table 5. Spearman’s rho values for 17 focal species. The Spearman correlation was between average position in the geographic range and species position within the matrix.

A. Significant correlations

B. Non-significant correlations

Species	Spearman's rho	n
Belted Kingfisher	0.289	49
Eastern Bluebird	-0.699	38
Pied-billed Grebe	0.92	28
Pileated Woodpecker	-0.49	31

A

Species	Spearman's rho	n
Chuck-Will's Widow	-0.441	17
Common Grackle	-0.118	43
Gray Vireo	0.143	6
Hairy Woodpecker	0.223	55
Killdeer	-0.203	57
Lincoln's Sparrow	-0.369	22
Orchard Oriole	0.068	36
Prairie Falcon	0.324	17
Pygmy Nuthatch	0.271	15
Sedge Wren	0.297	10
Swainson's Warbler	-0.067	9
White-throated Swift	-0.081	17
Williamson's Sapsucker	-0.273	11

B

Figure 1: Conceptual model of the factors contributing to species occurrence. Speciation and immigration positively affect species occurrence, extinction and emigration negatively affect species occurrence. Relative length of arrow indicates significance of process.

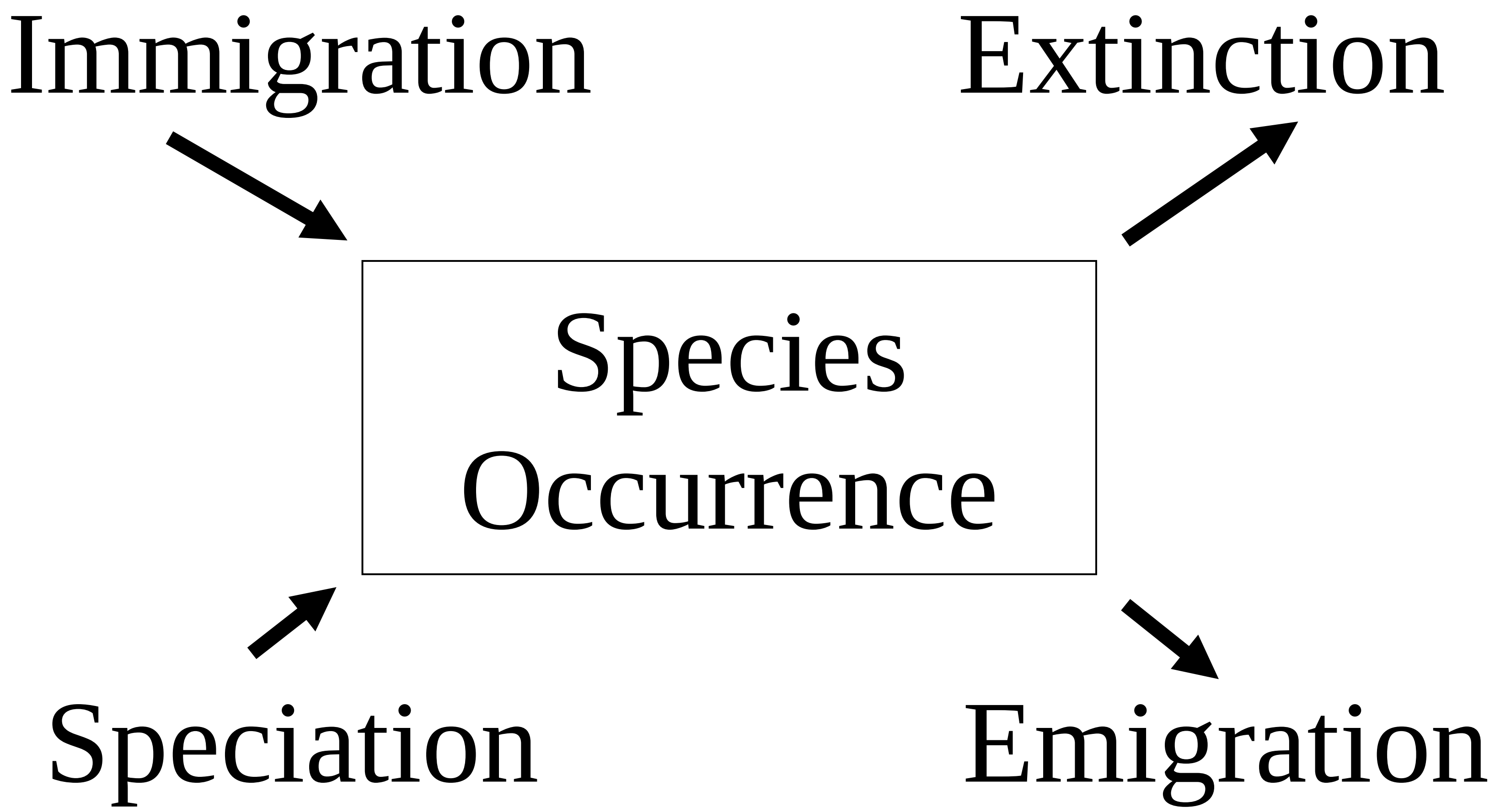
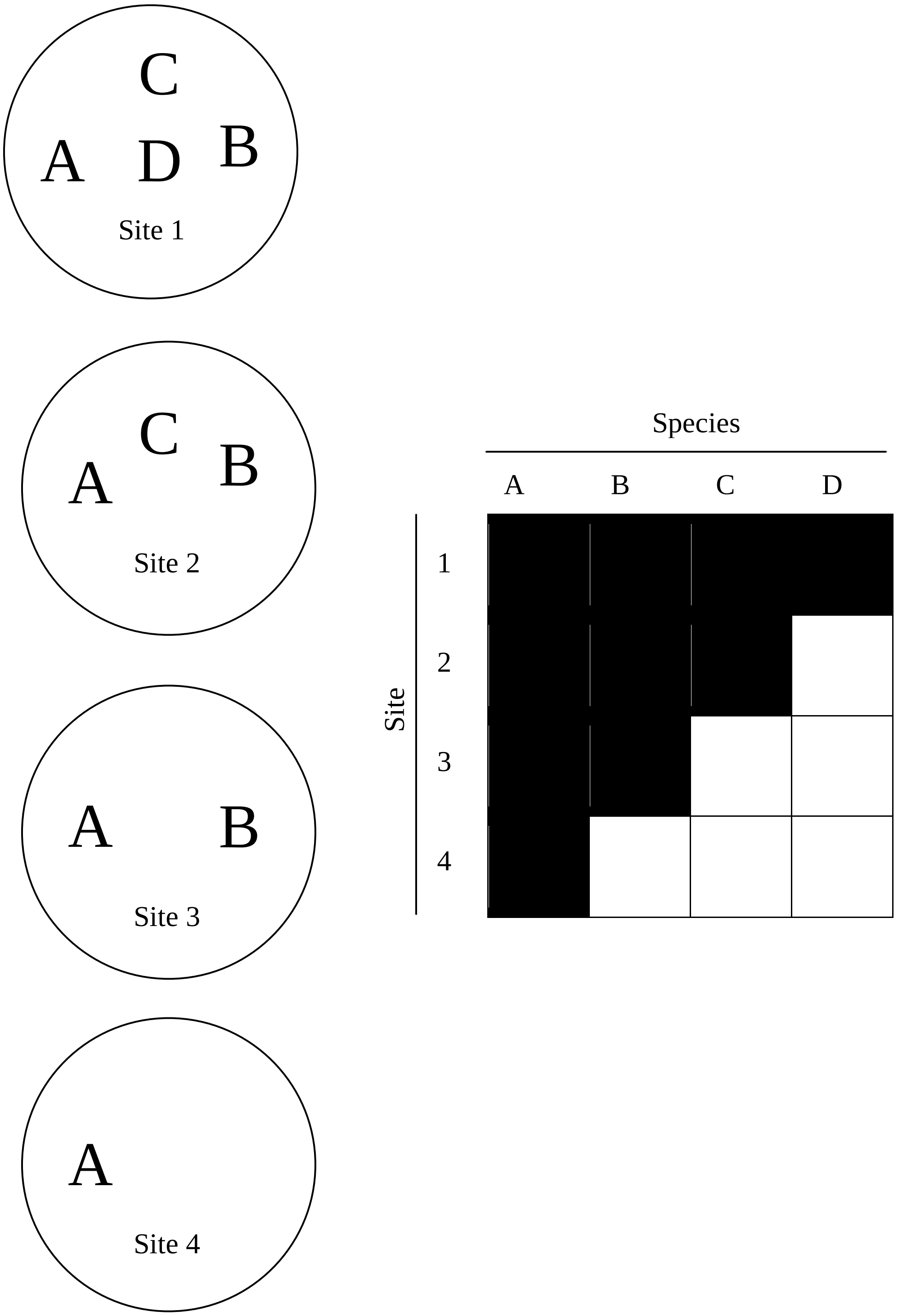

Immigration
Extinction
Species
Occurrence
Speciation
Emigration

Figure 2. Diagram of the nested subset pattern. The large circles indicate sites, while different species are indicated by letters. A nested subset matrix is created with species in columns and sites in rows.



Site	A	B	C	D
1	■	■	■	■
2	■	■	■	
3	■	■		
4	■			

Figure 3. Location of North American Breeding Bird Survey (BBS) routes for the conterminous United States, 2003.

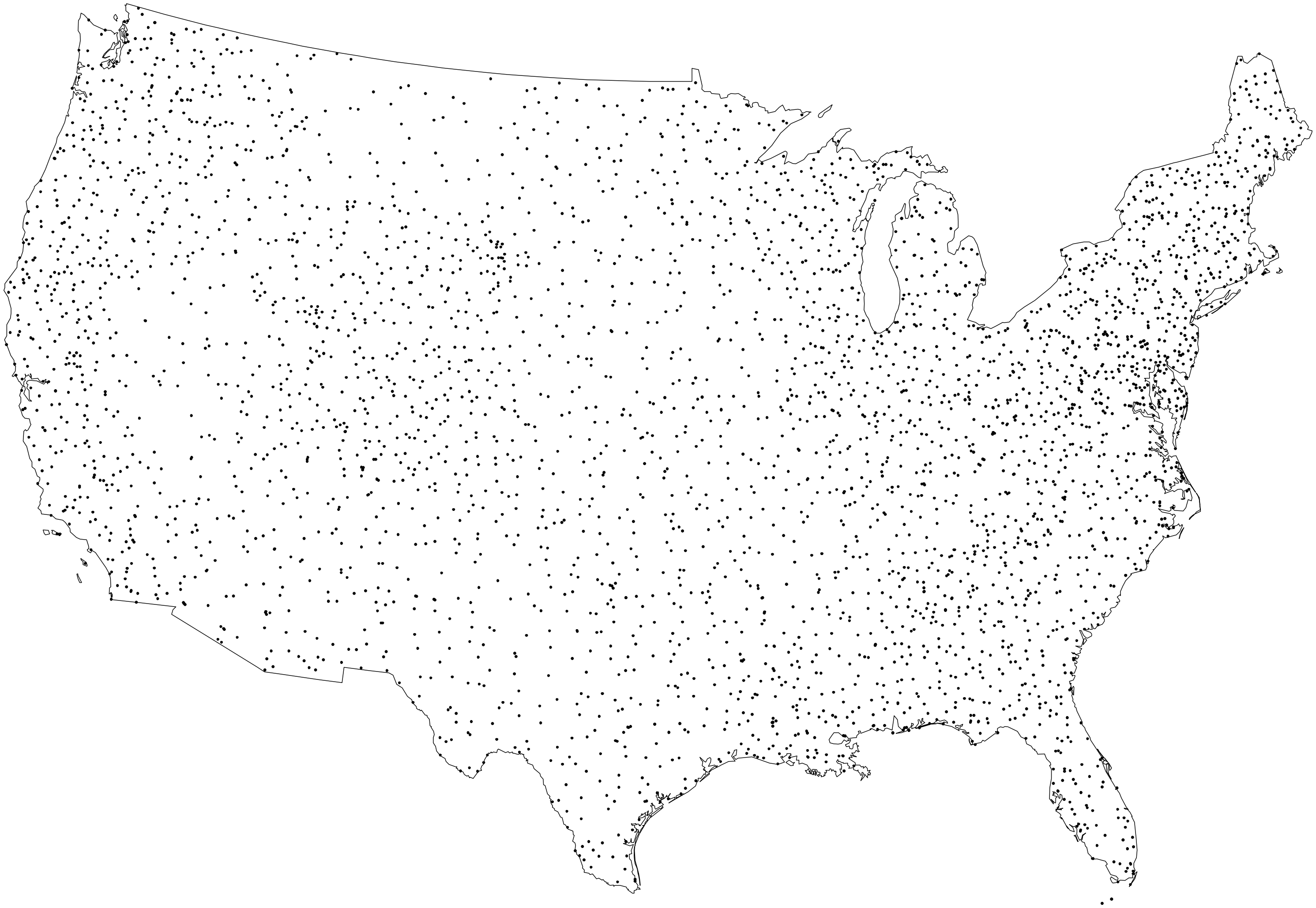


Figure 4. Grid of 59 equidistant seed points located within the conterminous United States.

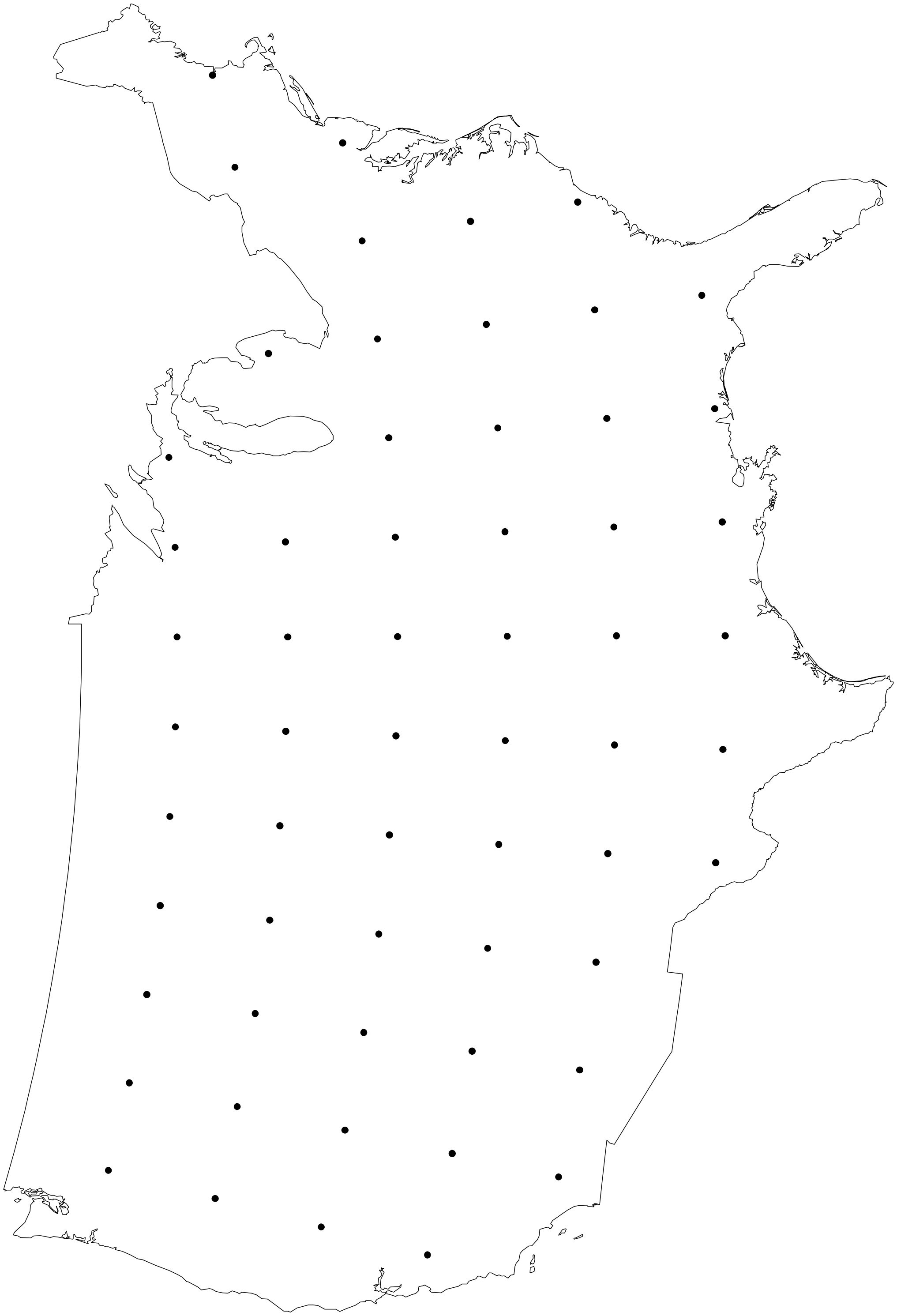


Figure 5. Seed points containing a minimum of six routes in surrounding areas at a given distance from the seed point.

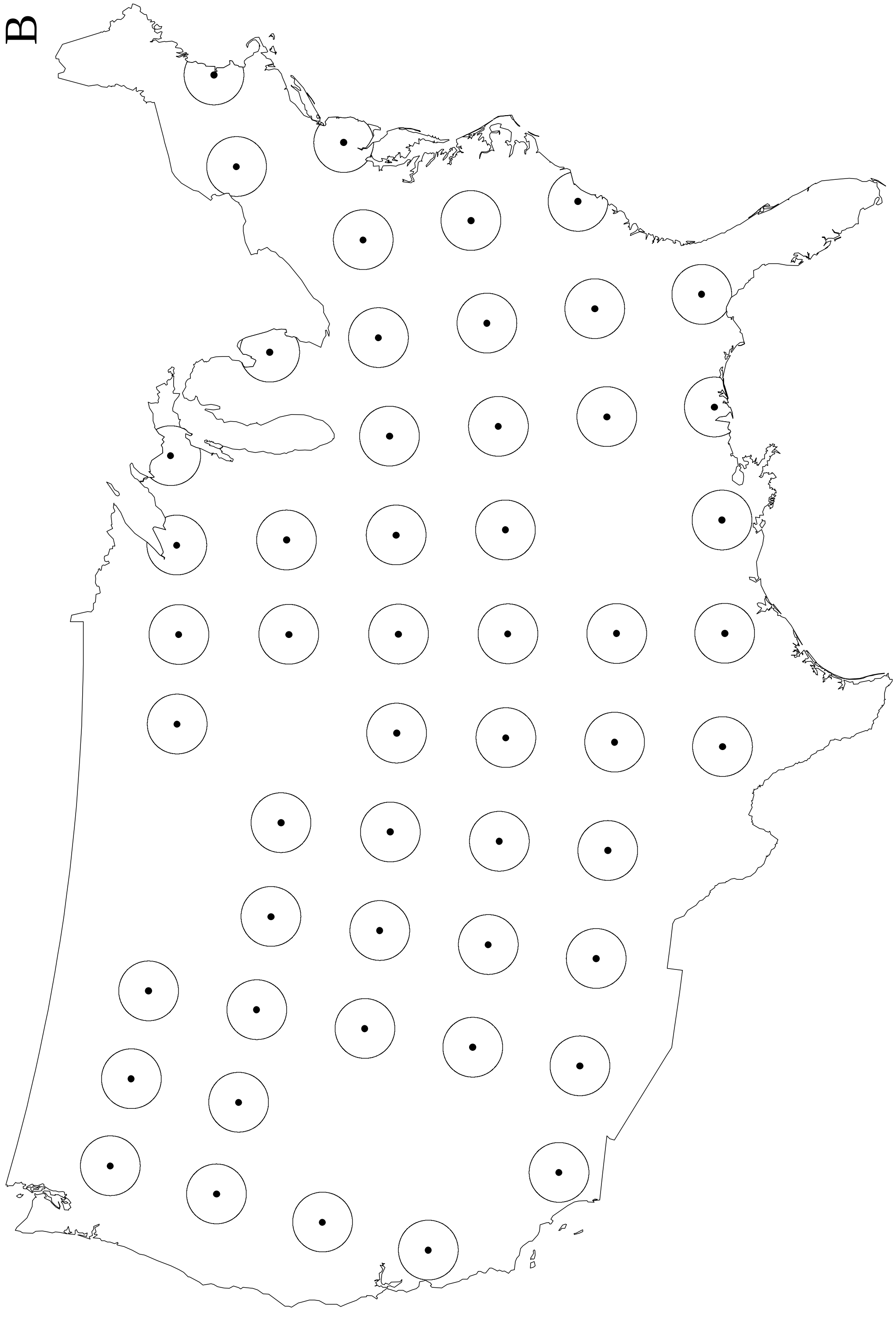
A. 50 km from seed point

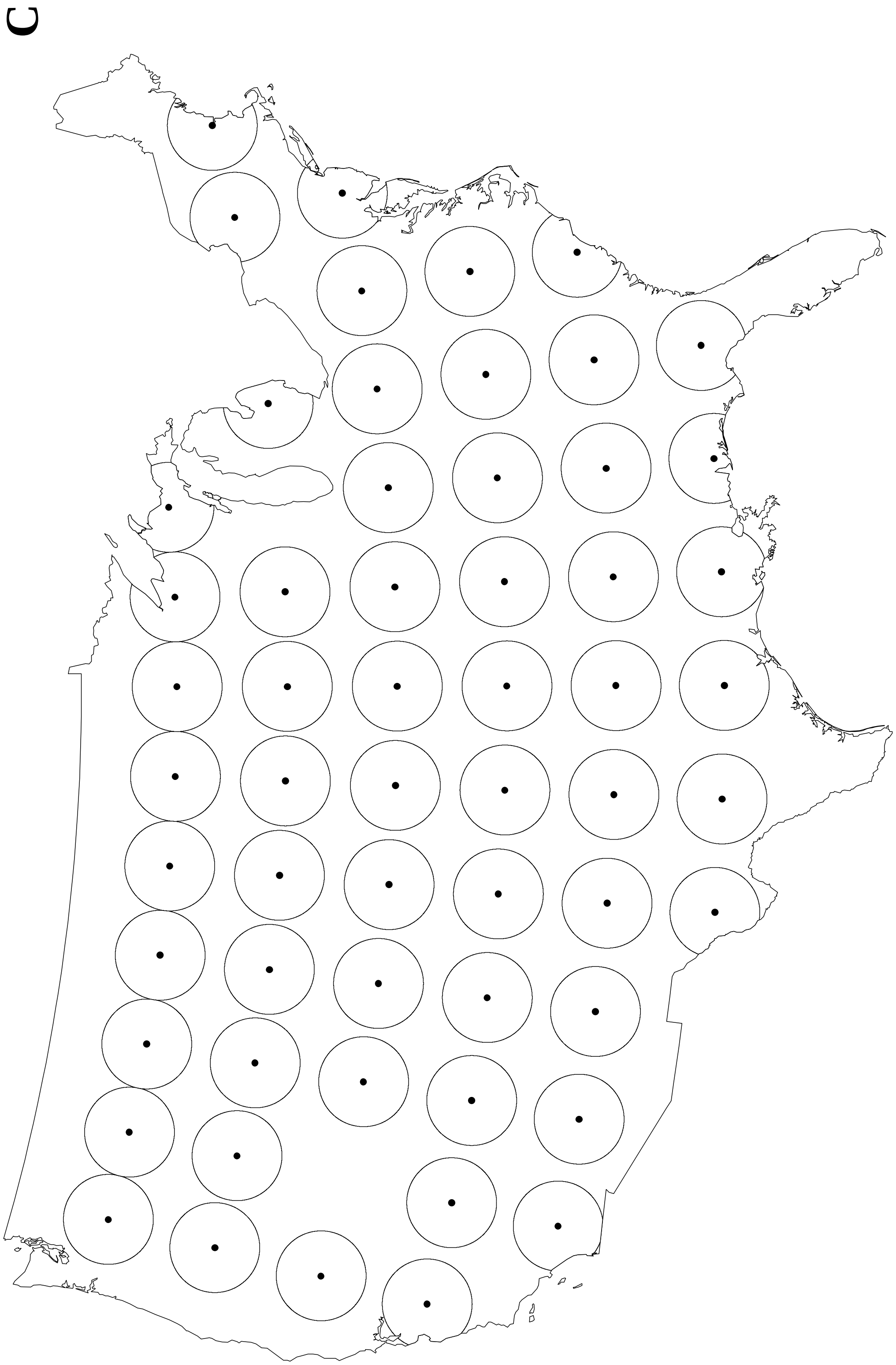
B. 100 km from seed point

C. 150 km from seed point

D. 200 km from seed point

A


B


C

D

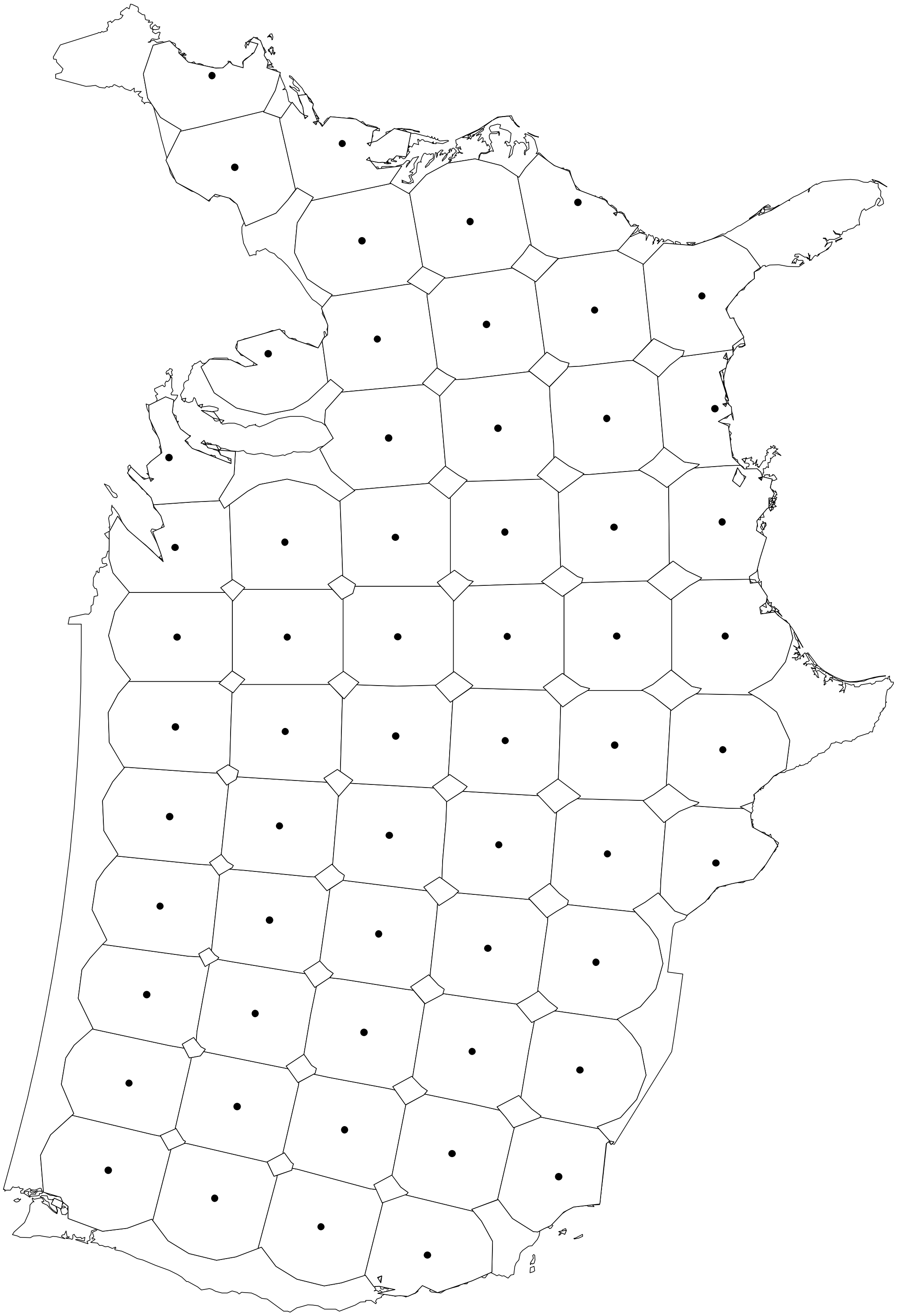


Figure 6. Groupings of seed points at each scale of analysis, created by applying a K-means clustering algorithm, from the smallest geographic extent to the largest extent. Shades of gray are used to delineate active groups, white areas were not used in analyses at that extent. The same shade of gray was used for more than one group; groups are defined as all contiguous areas of the same color.

A. First grouping: Smallest extent

B. Second grouping

C. Third grouping

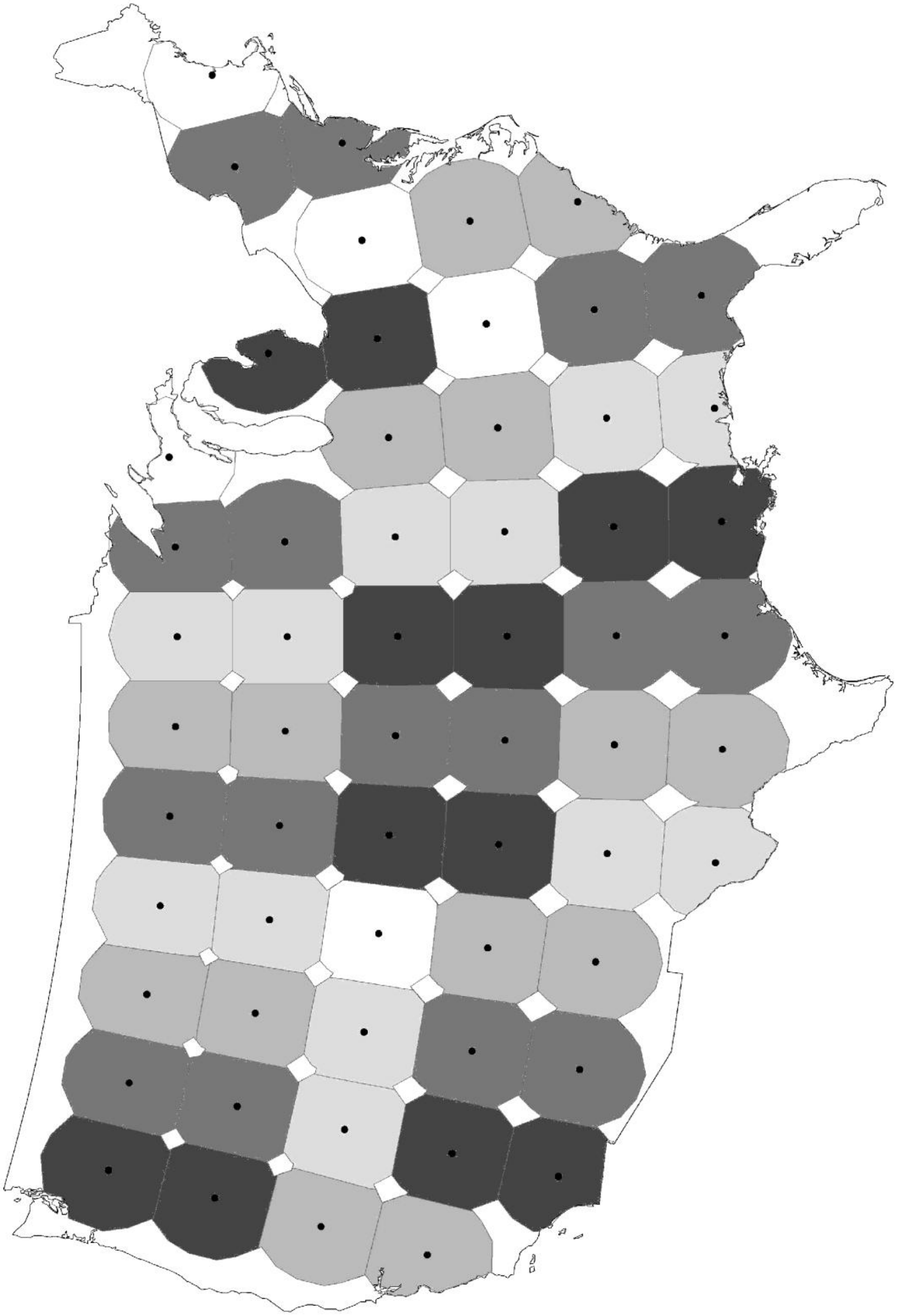
D. Fourth grouping

E. Fifth grouping

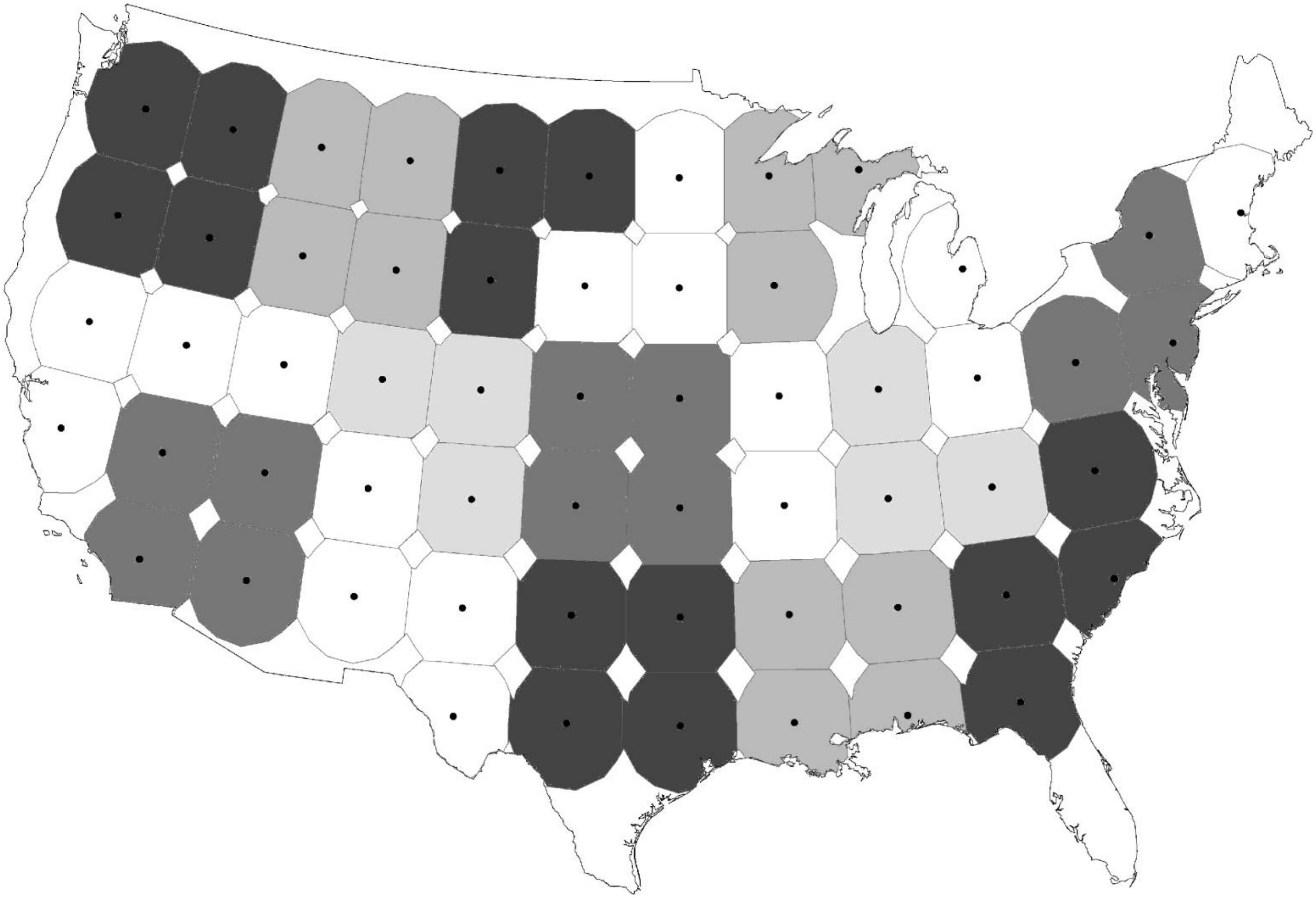
F. Sixth grouping

G. Seventh grouping: Largest extent

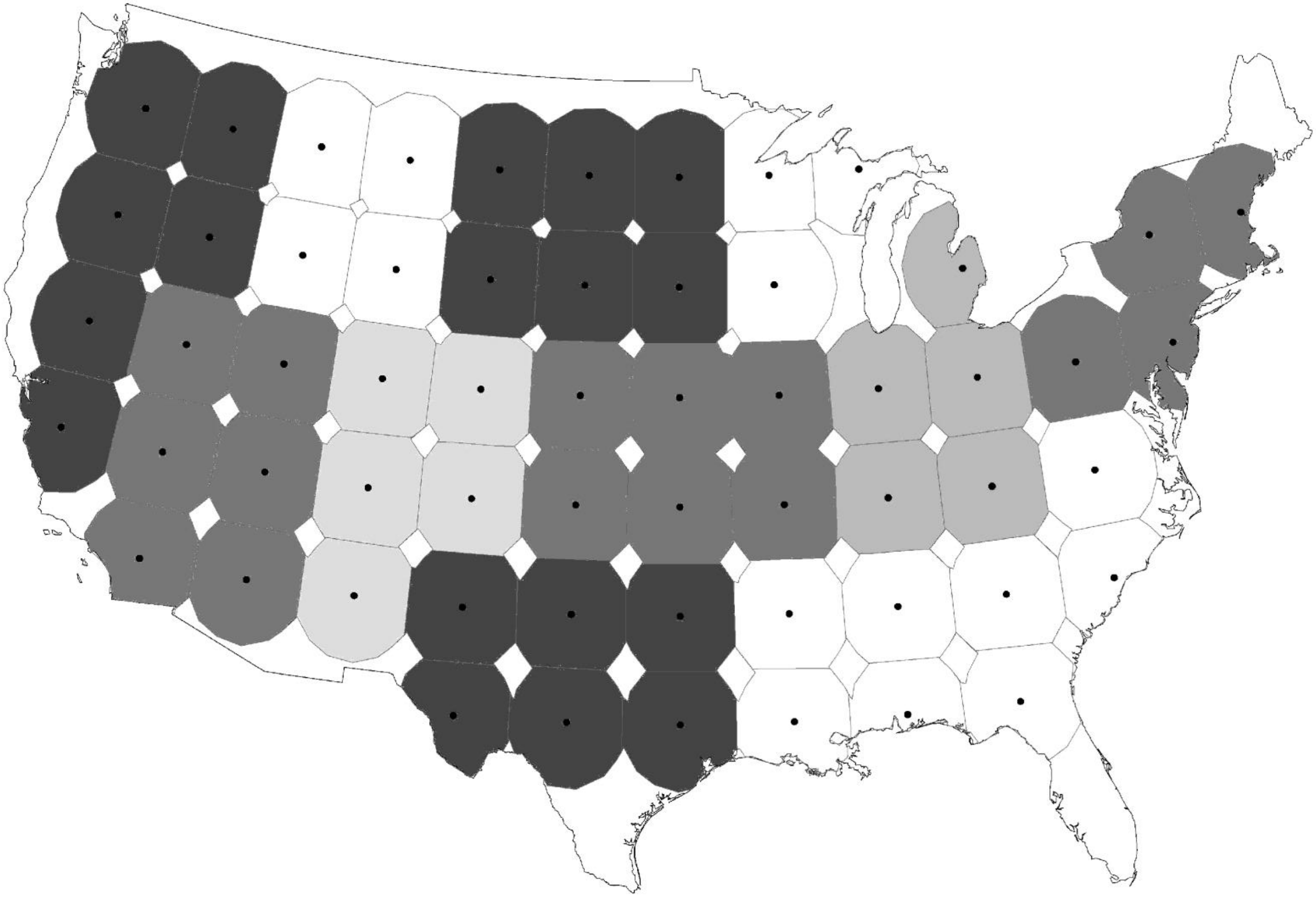
A

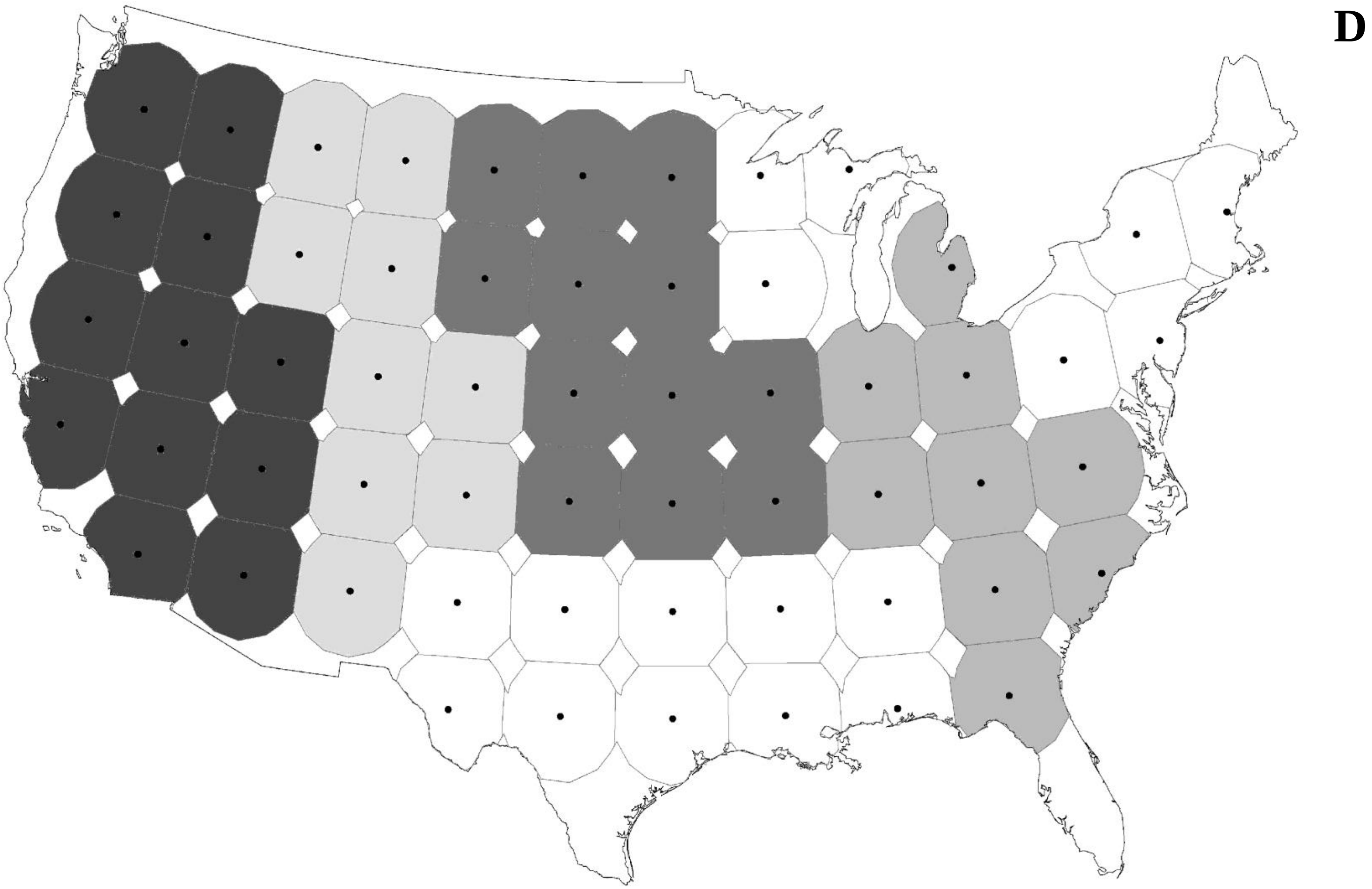


B

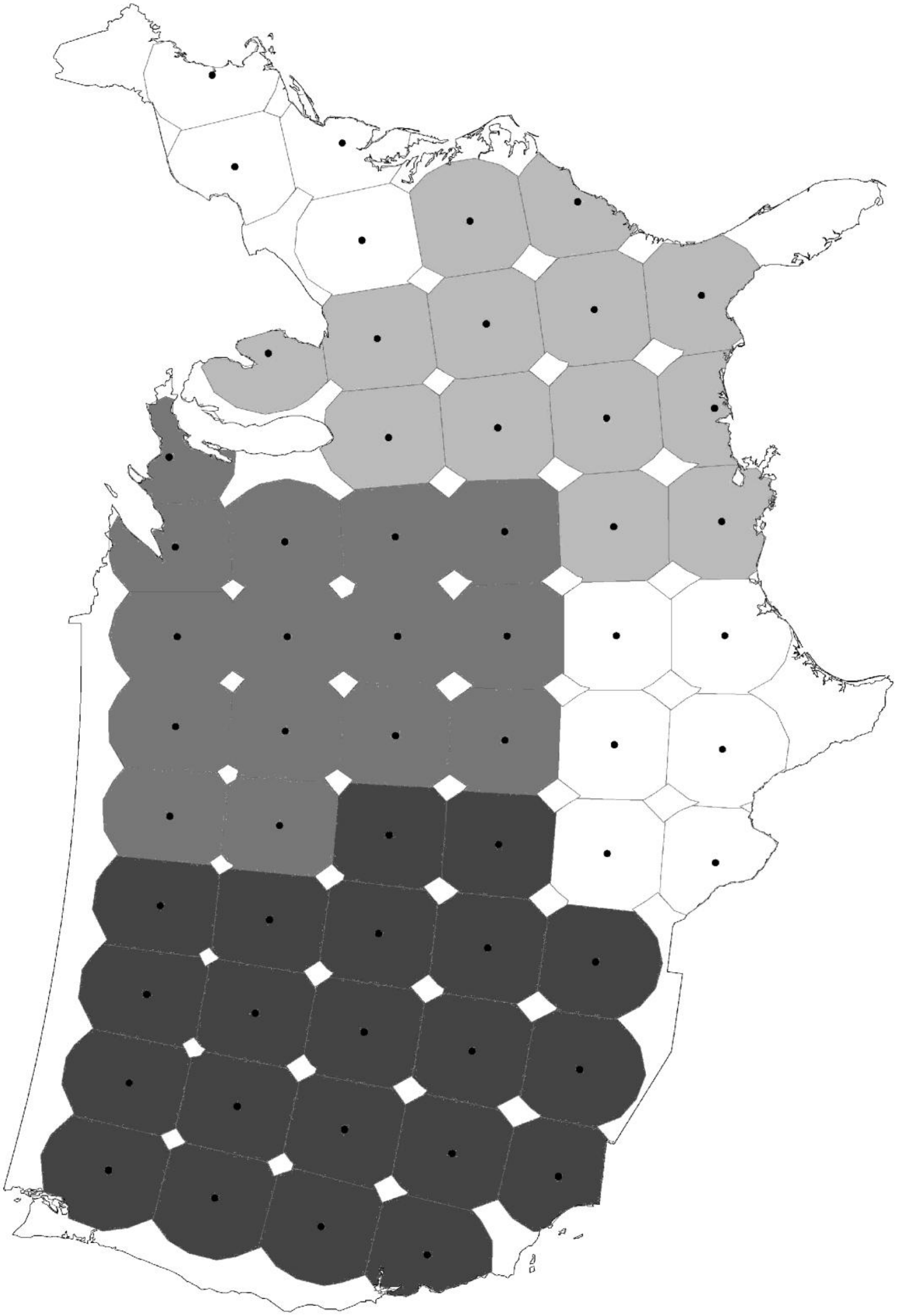


C

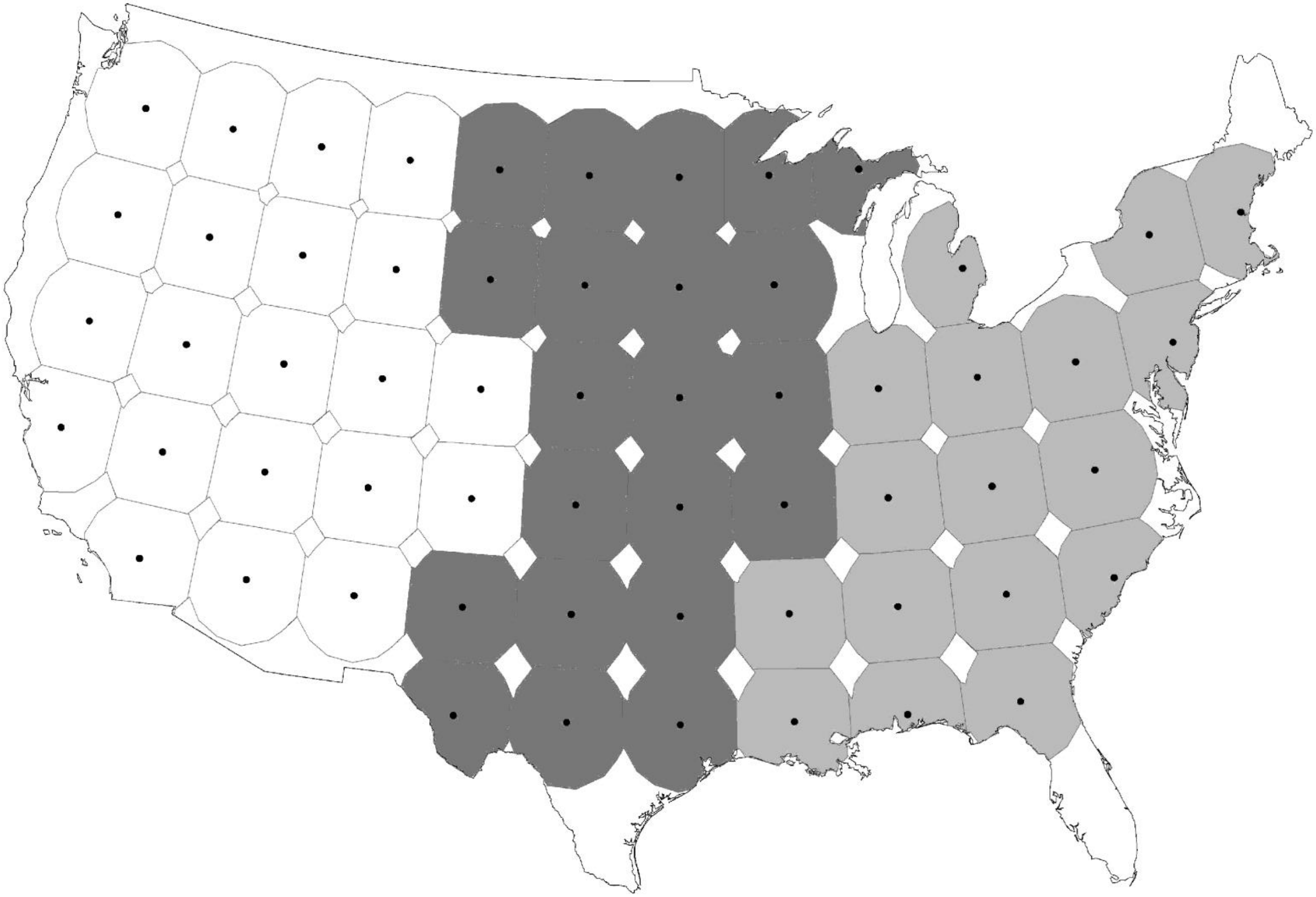



D

E



F



G

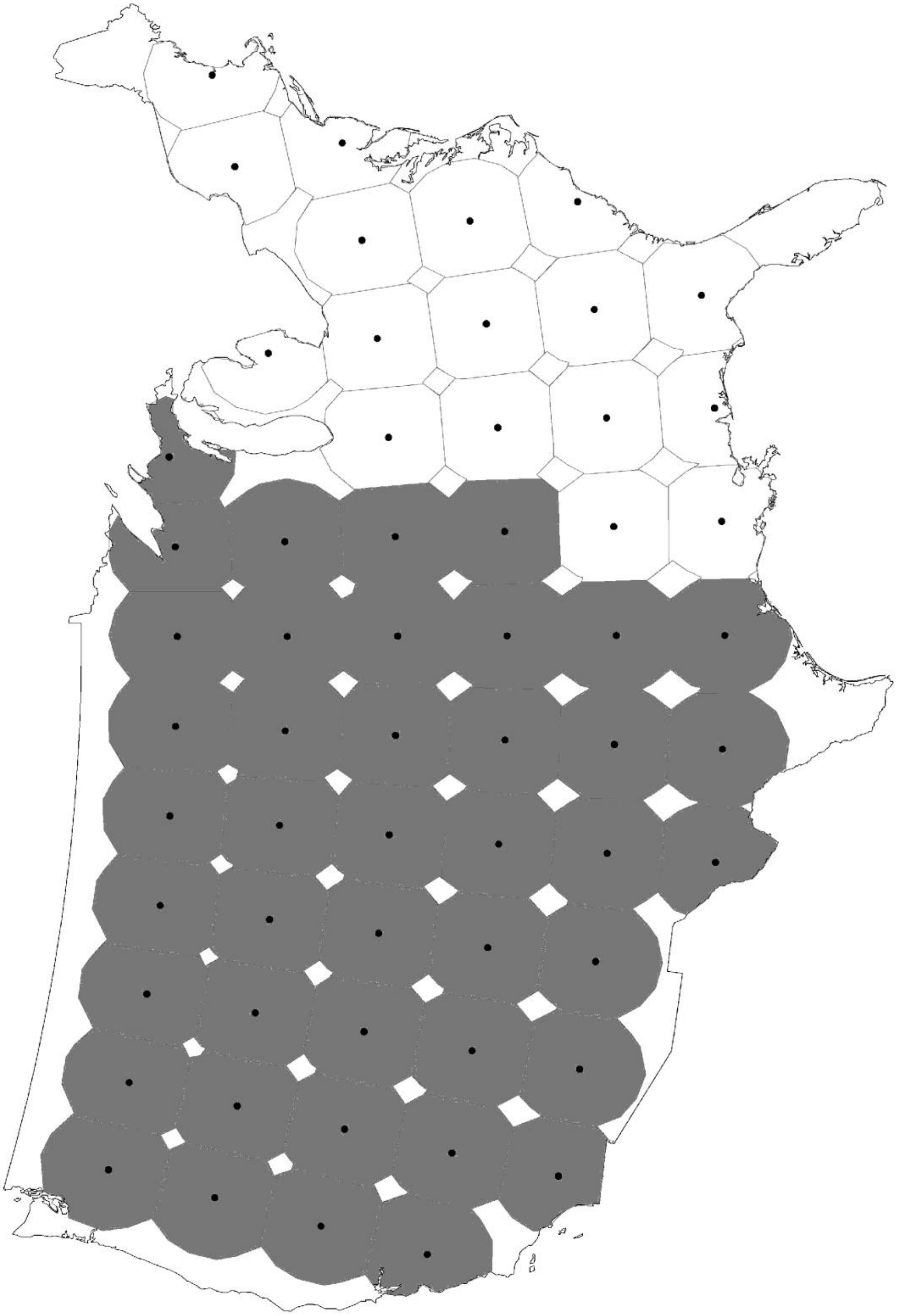


Figure 7. Groupings of seed points at each scale of analysis, created by applying a nearest neighbor clustering algorithm, from the smallest geographic extent to the largest extent. Shades of gray are used to delineate active groups, white areas were not used in analyses at that extent. The same shade of gray was used for more than one group, groups are defined as all contiguous areas of the same color.

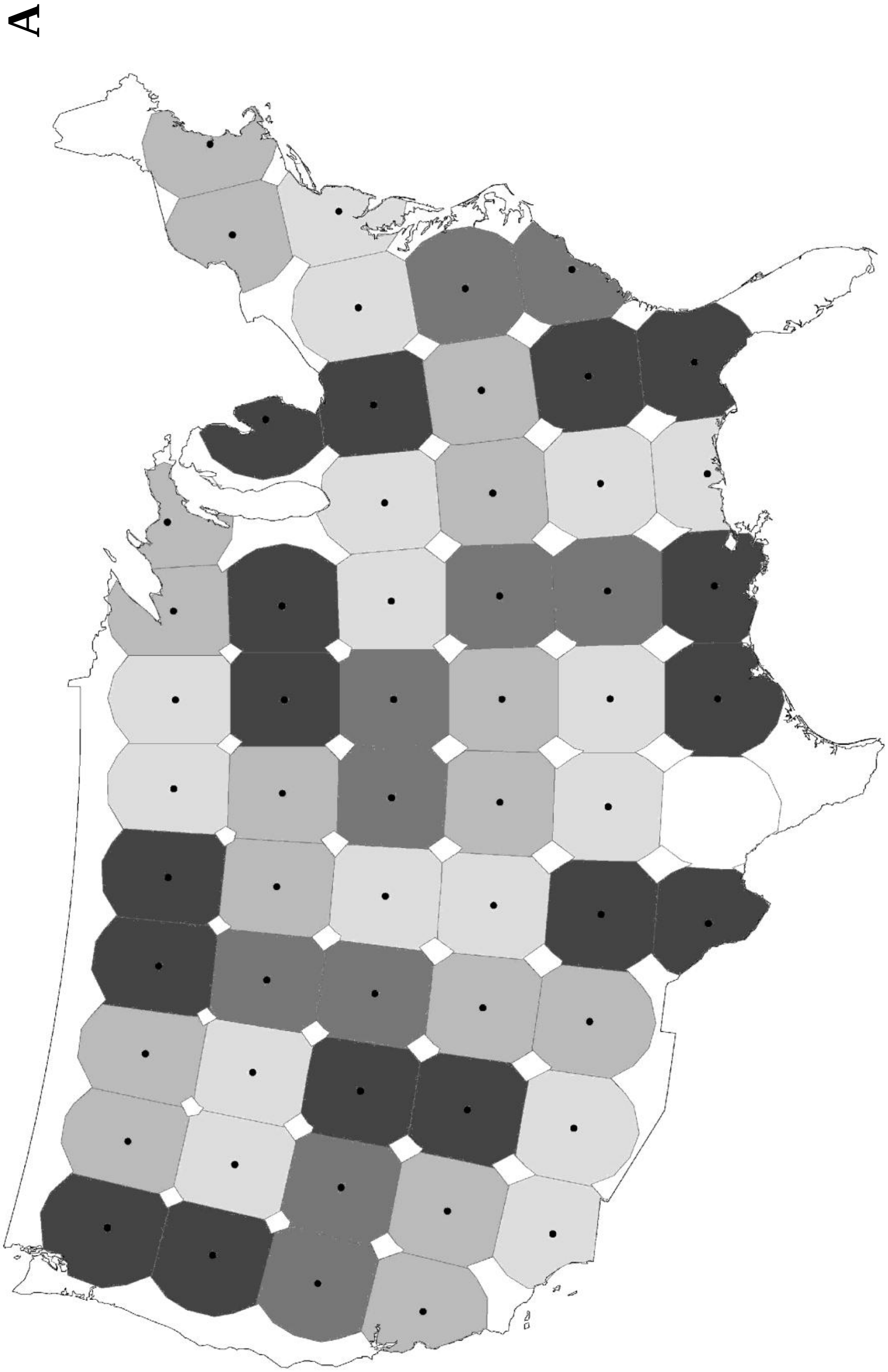
A. First grouping: Smallest extent

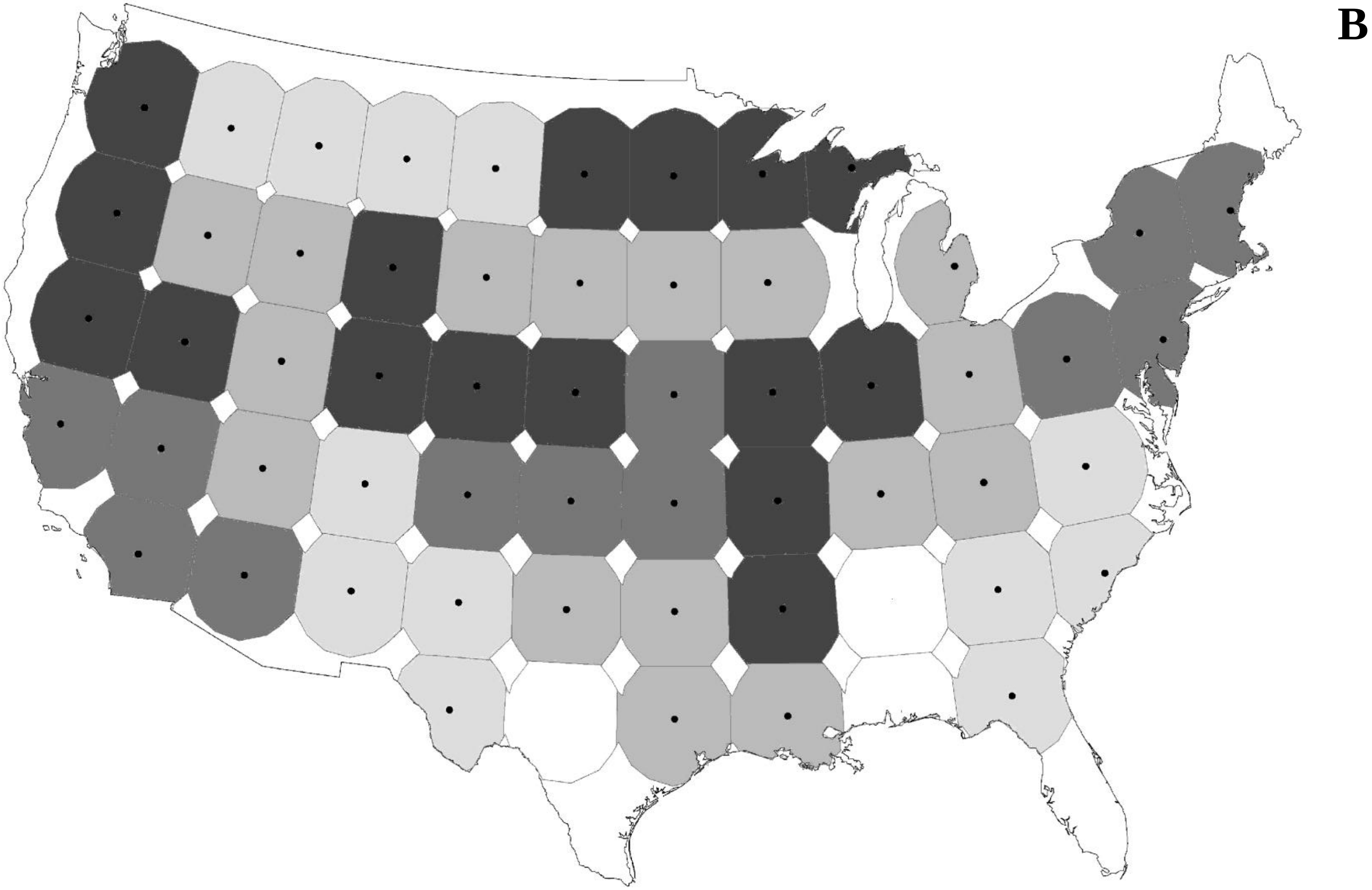
B. Second grouping

C. Third grouping

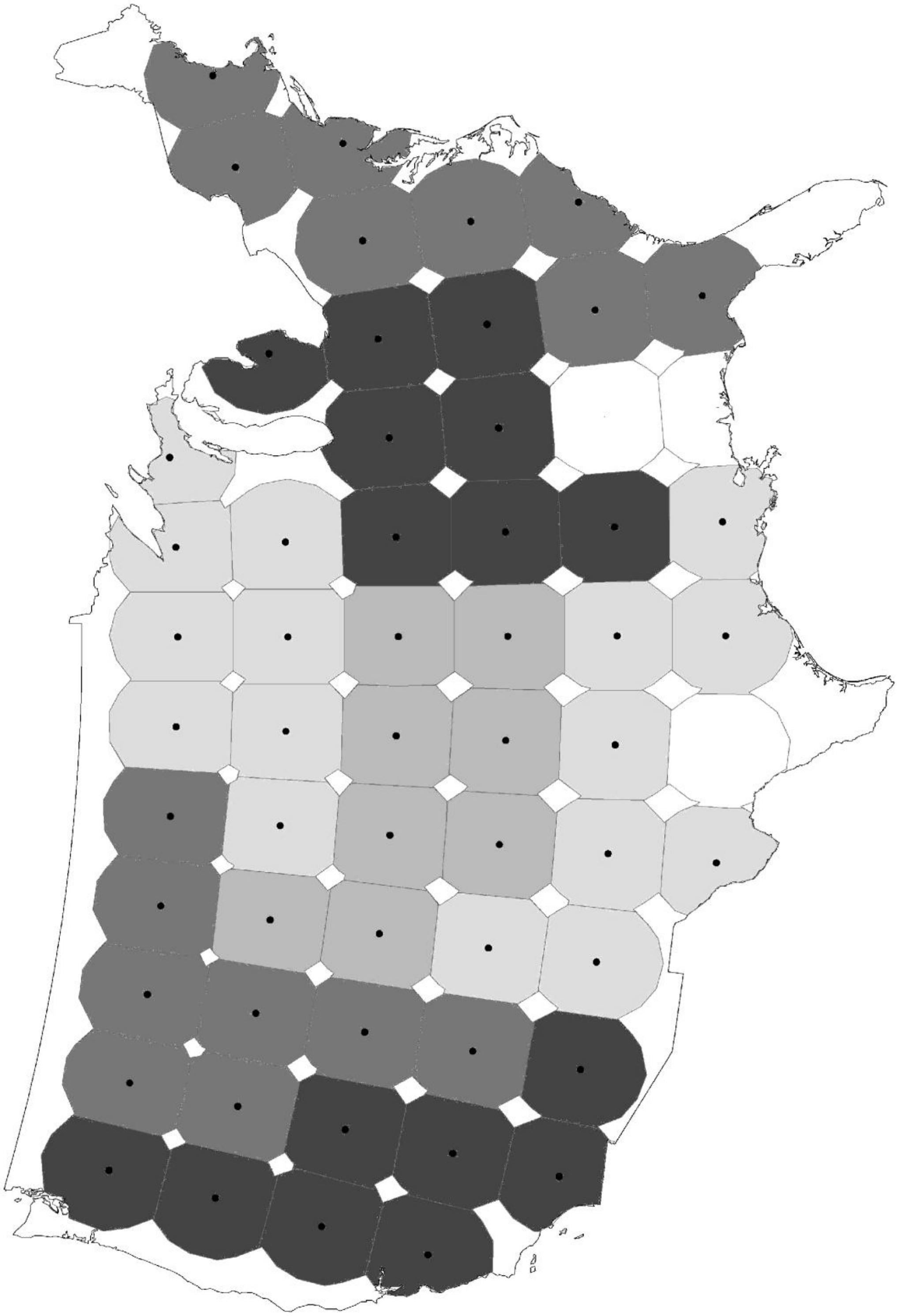
D. Fourth grouping

E. Fifth grouping: Largest extent

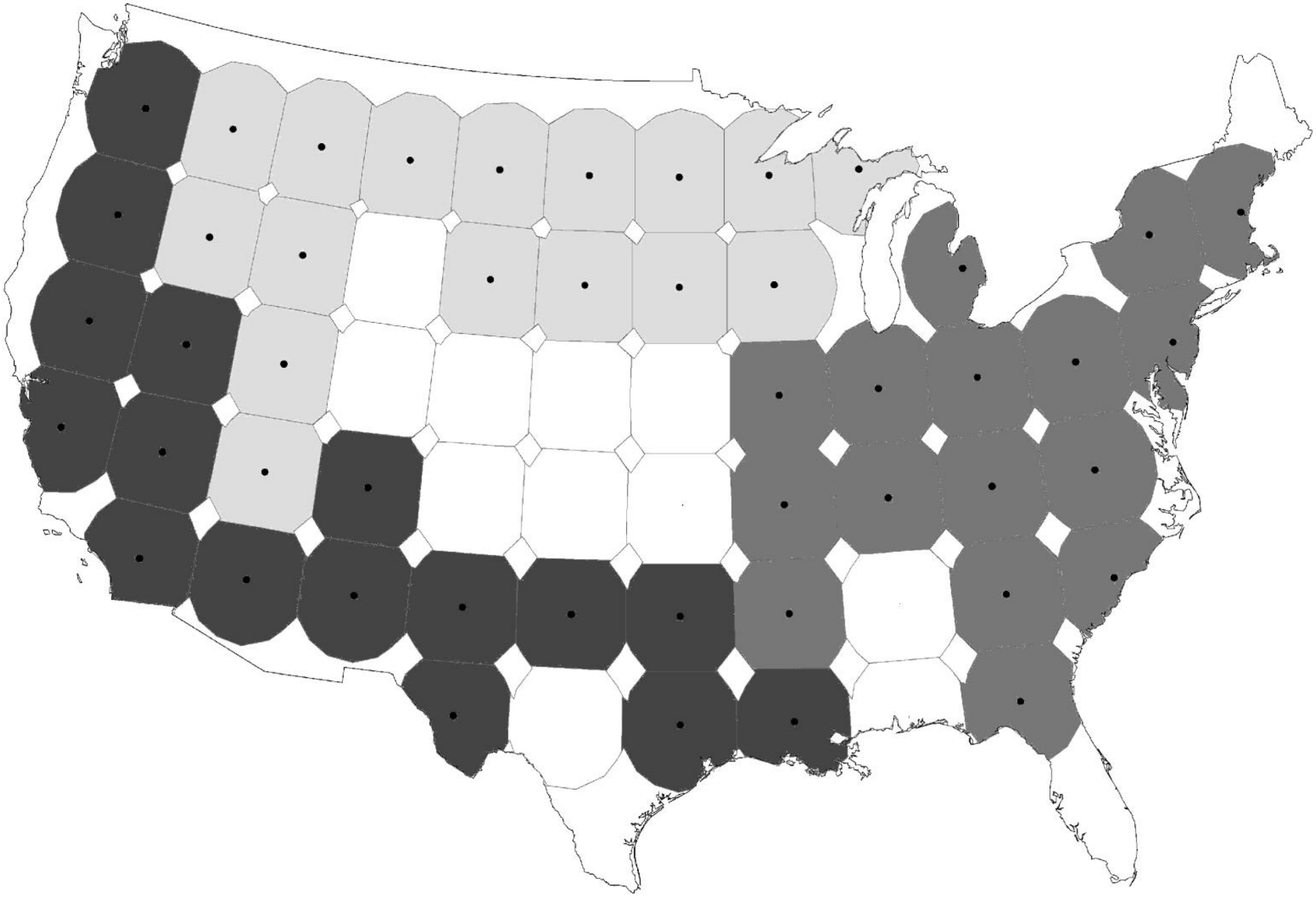

A


B

C



D



E

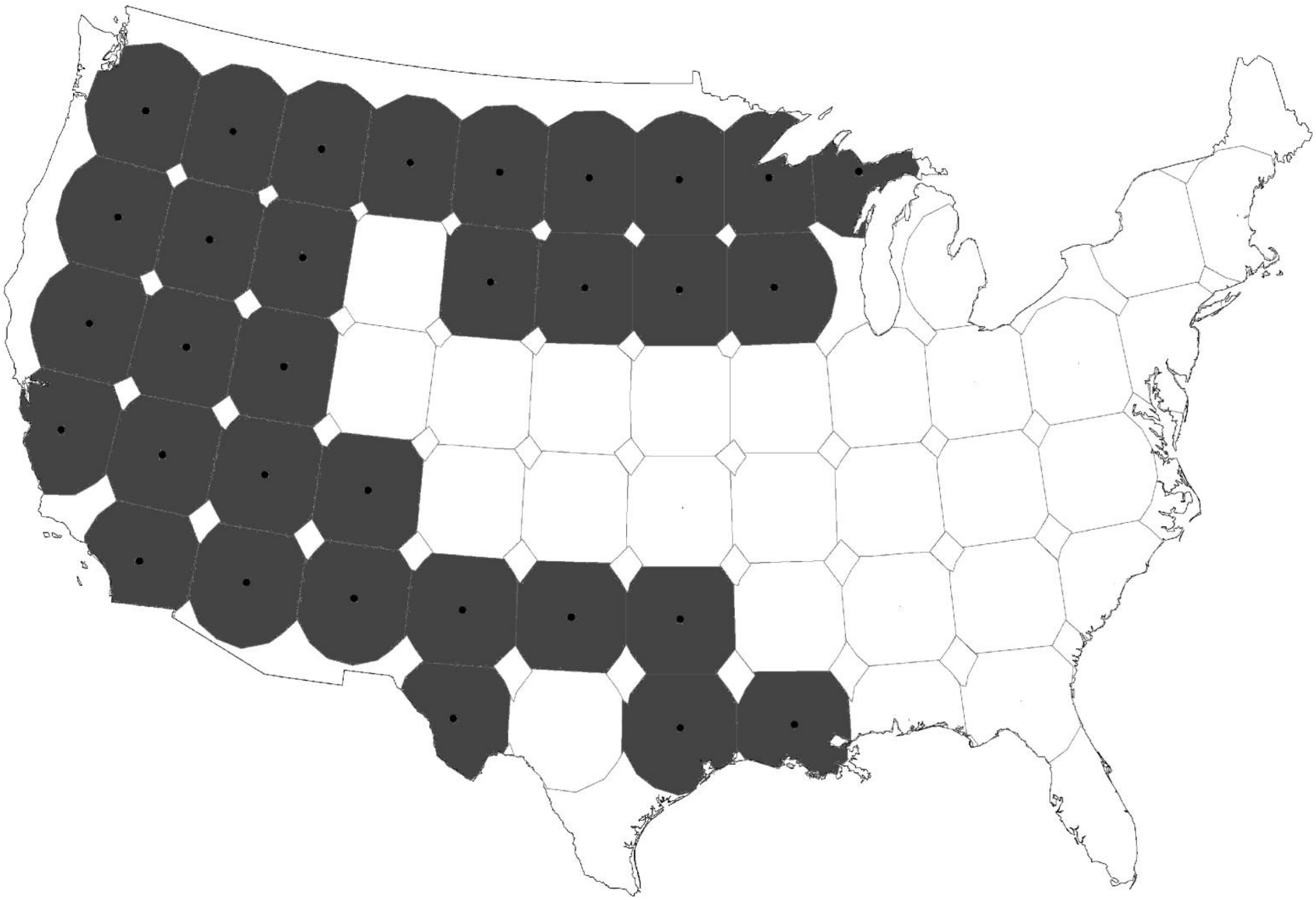


Figure 8. Relative and absolute site and species position within the matrix, counting from the most species-rich site to the least species-rich site and the highest incidence to the lowest incidence.

		Absolute rank	0.25	0.50	0.75	1.00
		Relative rank	1	2	3	4
Absolute rank	Relative rank	**Site \ Species**	**A** **Highest incidence**	**B**	**C**	**D** **Lowest incidence**
0.25	1	**1 Most species rich**	1	1	1	1
0.50	2	**2**	1	1	1	0
0.75	3	**3**	1	1	0	0
1.00	4	**4 Least species rich**	1	0	0	0

Figure 9. Absolute species position within the matrix plotted against initial site position for significantly nested within route matrices and 200 km from seed point matrices. Absolute species position is determined by dividing the relative species position, counting from the species with the highest incidence to the lowest, divided by the total number of species in the matrix. The 200 km seed point routes are distinguished with open circles and the within route level is identified with closed circles.

A. Results at the 200 km level, initial site standardized.

B. Results at the within route level, initial site unstandardized.

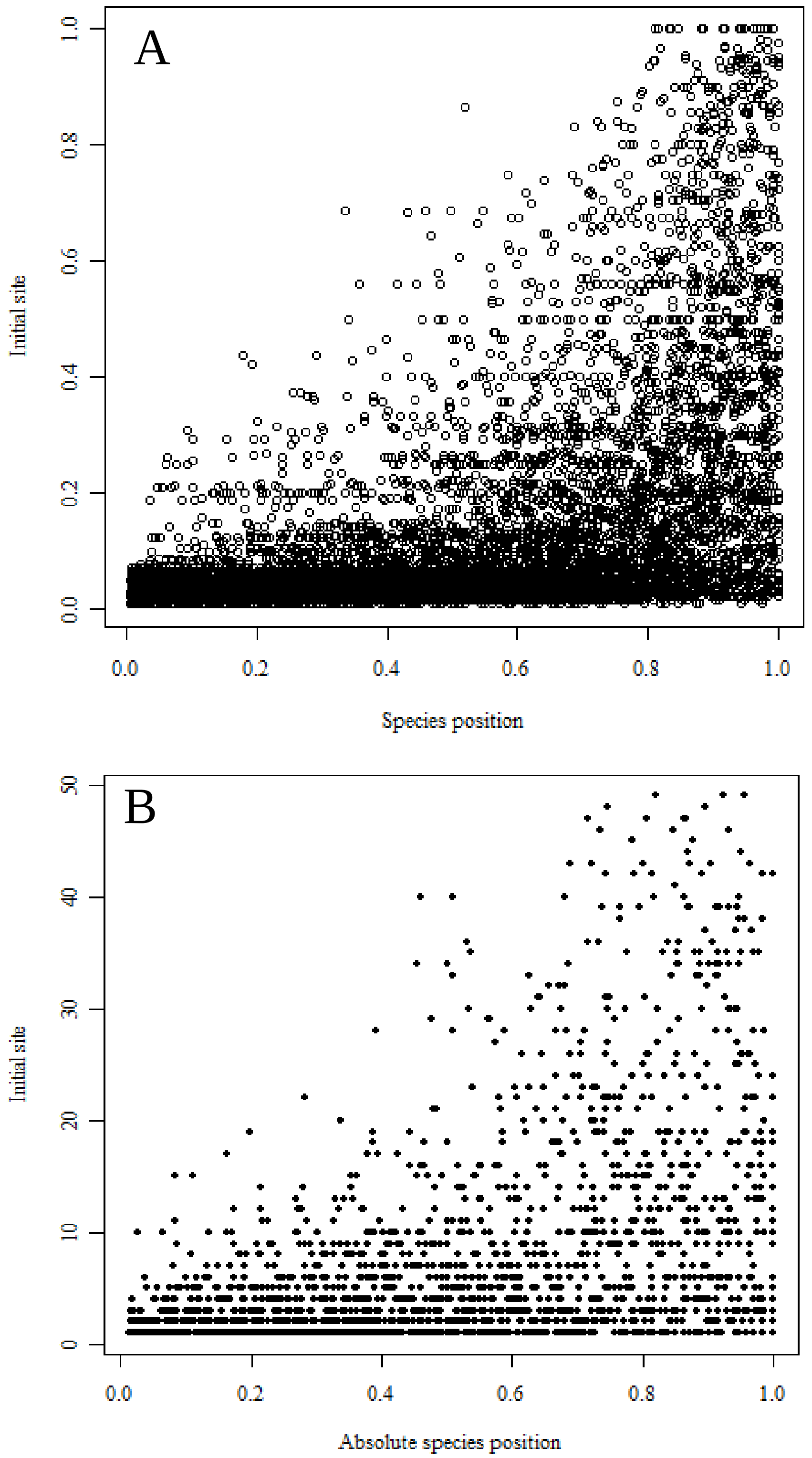
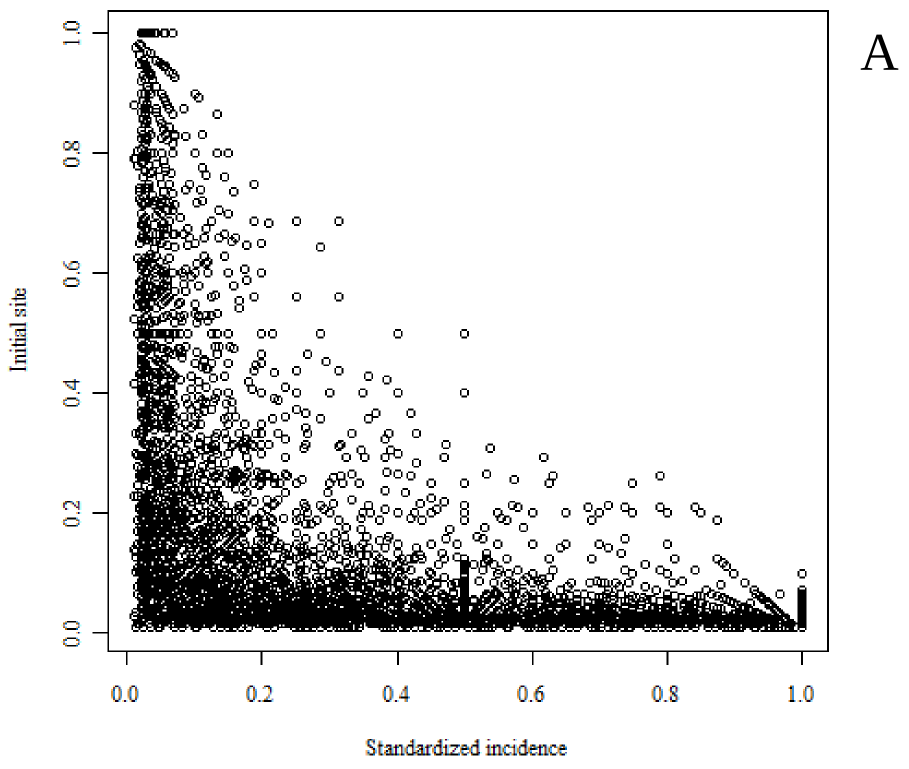

A
Initial site
0.0
0.2
0.4
0.6
0.8
1.0
Species position
0.0
0.2
0.4
0.6
0.8
1.0
B
Initial site
0
10
20
30
40
50
Absolute species position
0.0
0.2
0.4
0.6
0.8
1.0

Figure 10. Incidence plotted against initial site position for significantly nested within route matrices and 200 km from seed point matrices. The 200 km seed point routes are distinguished with open circles and the within route level is identified with closed circles. Incidence was standardized by dividing the incidence at a site with the total number of routes at a site.

A. Results at the 200 km level, incidence standardized.

B. Results at the within route level, incidence unstandardized.


A
Initial site
Standardized incidence
0.0
0.2
0.4
0.6
0.8
1.0

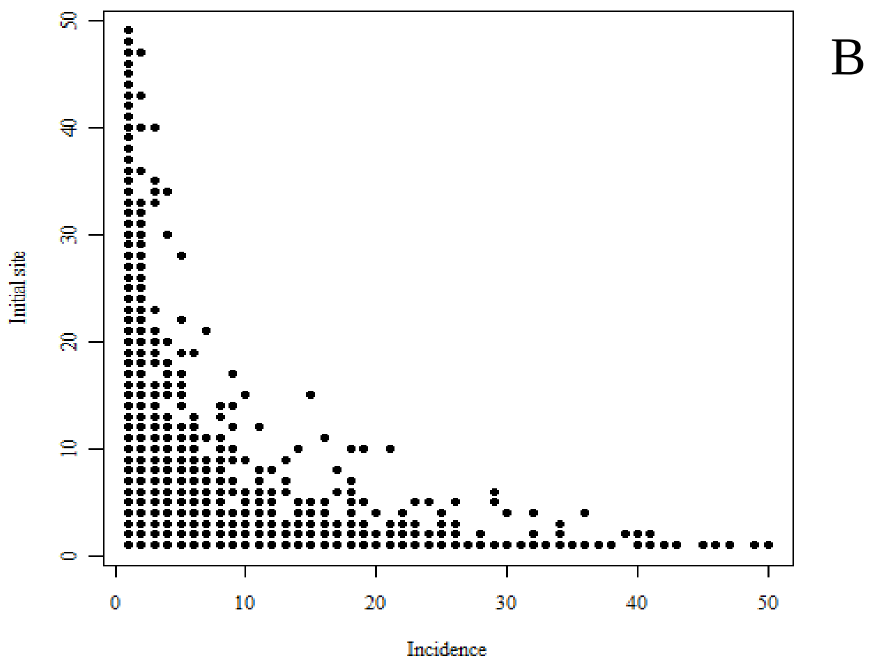
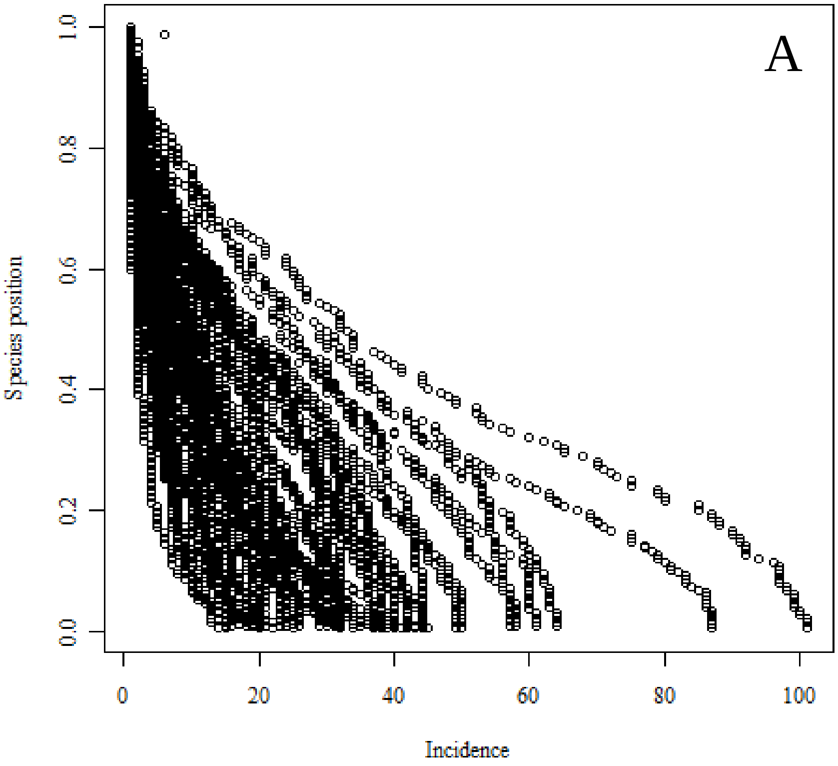

B
Initial site
Incidence
0
10
20
30
40
50

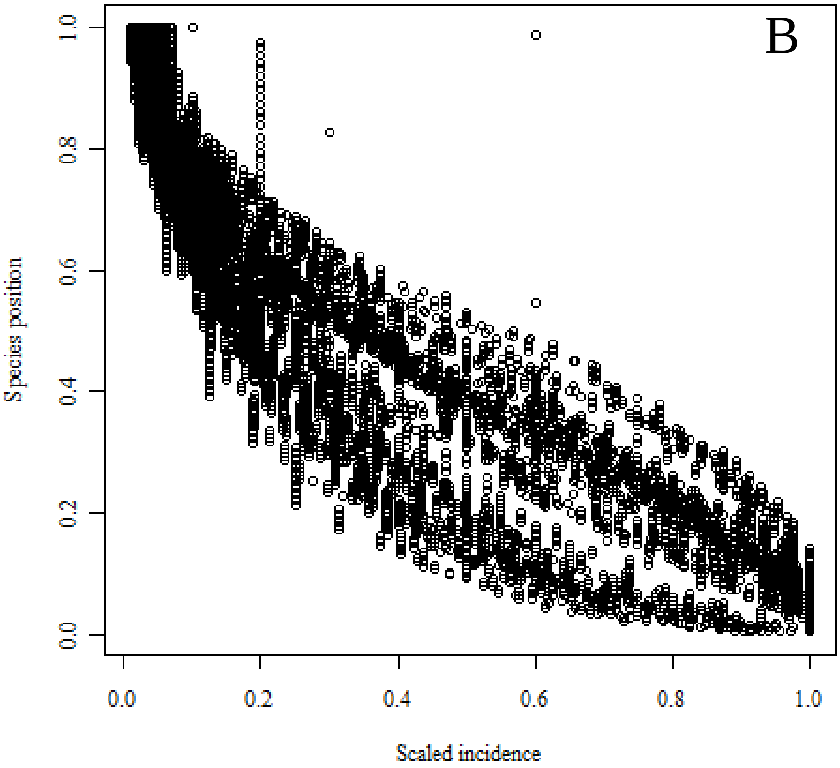
Figure 11. Incidence plotted against absolute species position for significantly nested within route matrices and 200 km from seed point matrices. The 200 km seed point routes are distinguished with open circles and the within route level is identified with closed circles. Incidence was standardized by dividing the incidence at a site with the total number of routes at a site.

A. Results at the 200 km level, incidence unstandardized.

B. Results at the 200 km level, incidence standardized.

C. Results at the within route level, incidence unstandardized.


A
Species position
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
Incidence


B
Species position
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Scaled incidence

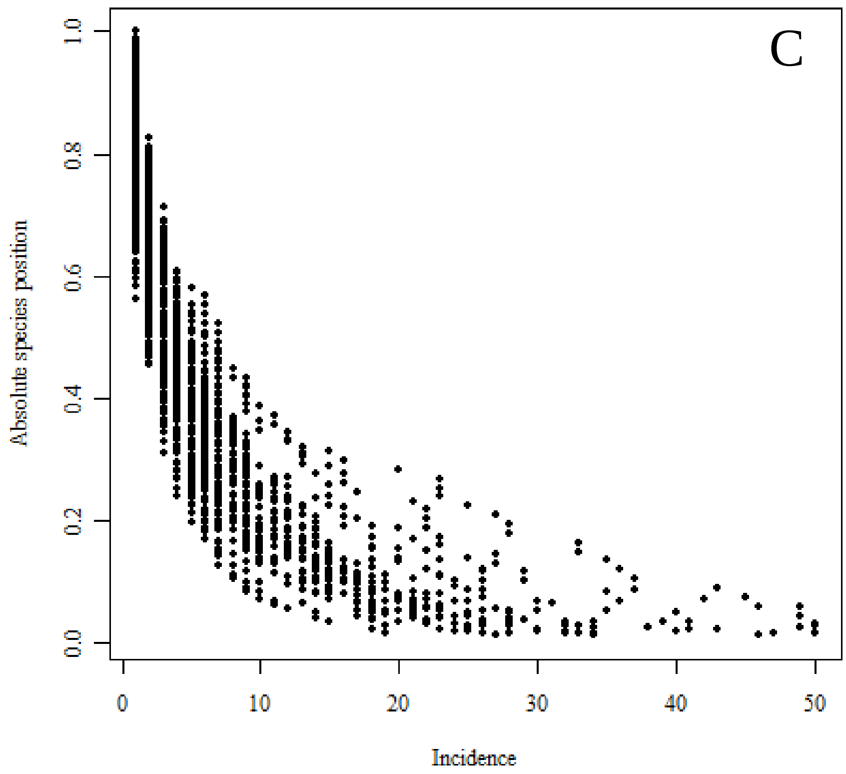
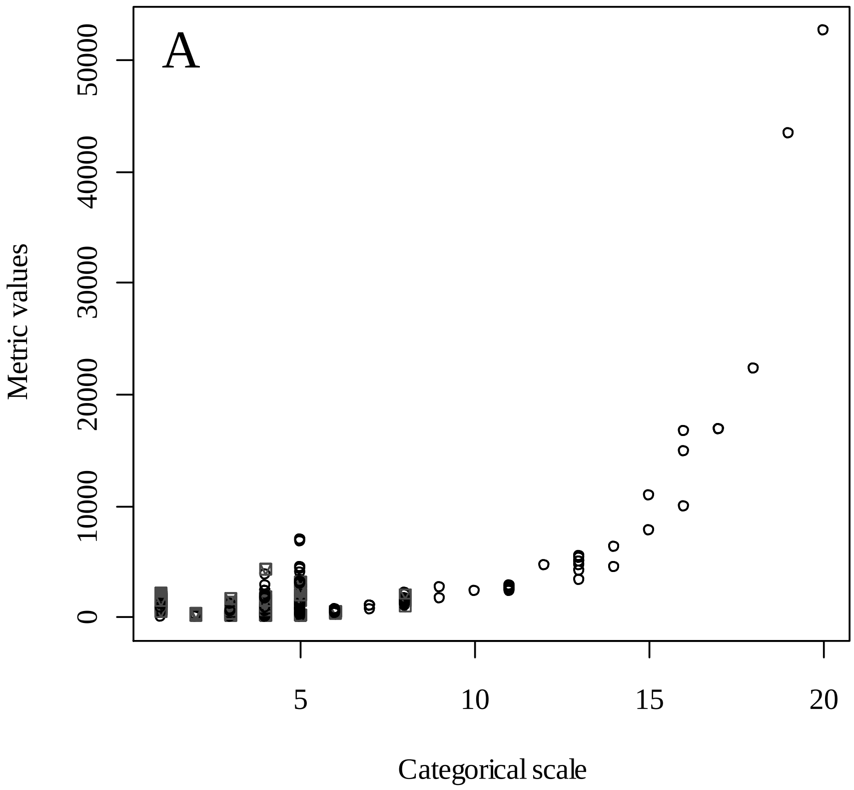

C
Absolute species position
0.0
0.2
0.4
0.6
0.8
1.0
Incidence
0
10
20
30
40
50

Figure 12. Metric value (N_0) graphed against two measures of scale. Significantly nested matrices are black and hollow circles, and non-significant matrices are gray and hollow squares.

A. Categorical scale

B. Area (km^2)


A
Metric values
0
10000
20000
30000
40000
50000
5
10
15
20
Categorical scale

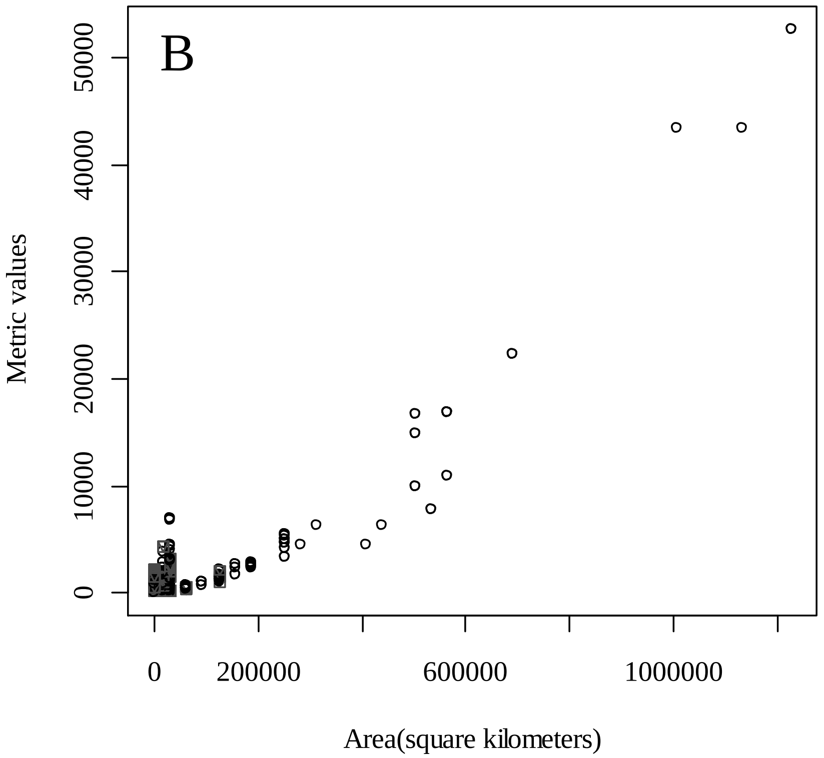
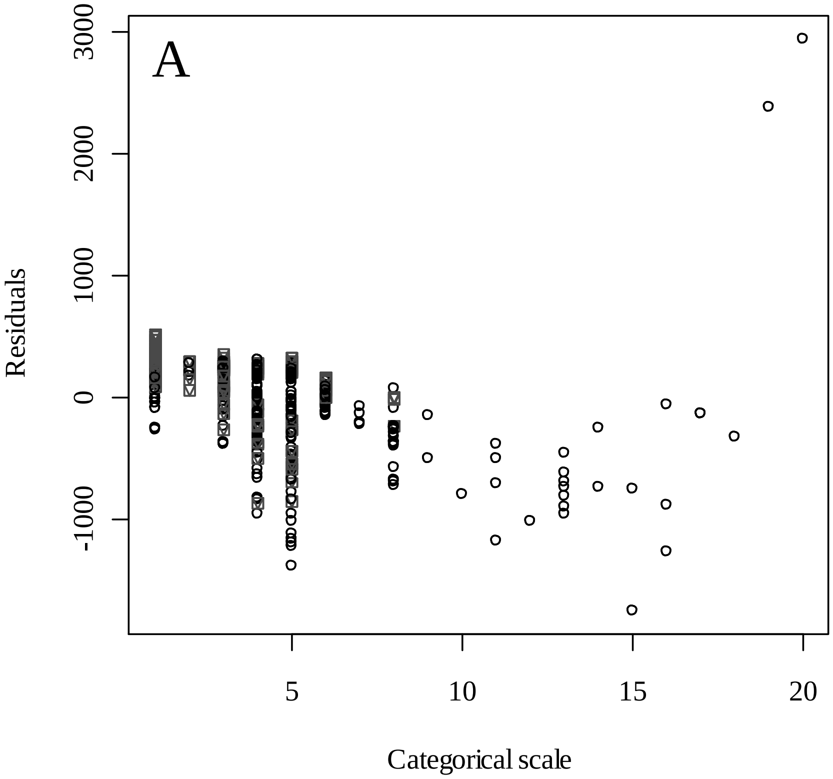

B
Metric values
0
10000
20000
30000
40000
50000
0
200000
600000
1000000
Area(square kilometers)

Figure 13. Residuals of metric value (N_0) and matrix size graphed against two measures of scale. Significantly nested matrices are black and open circles, and non-significant matrices are gray and open squares.

A. Categorical scale

B. Area (km^2)


A
Residuals
3000
2000
1000
0
-1000
5
10
15
20
Categorical scale

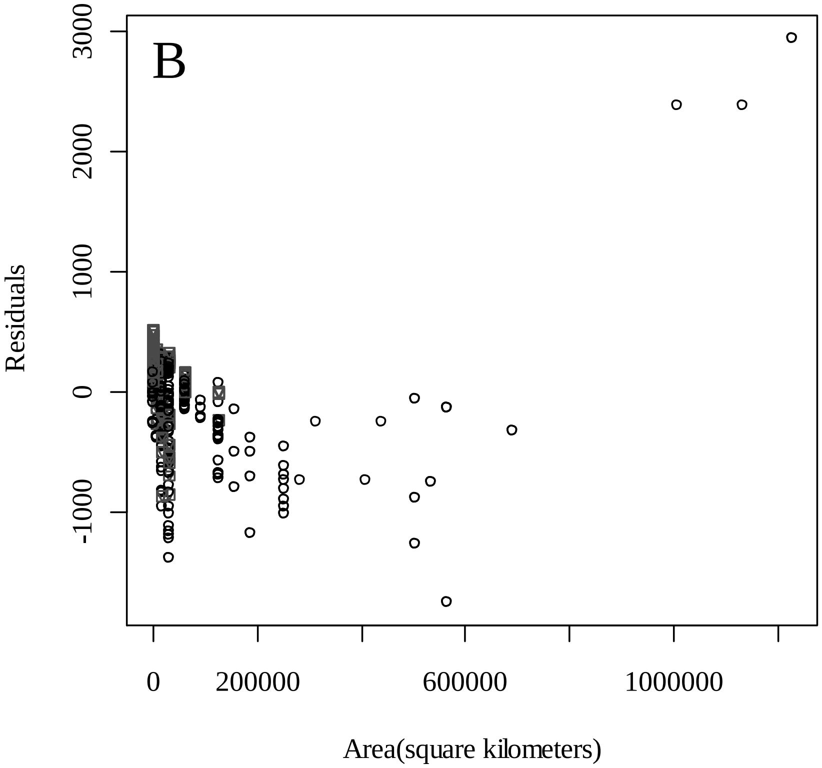
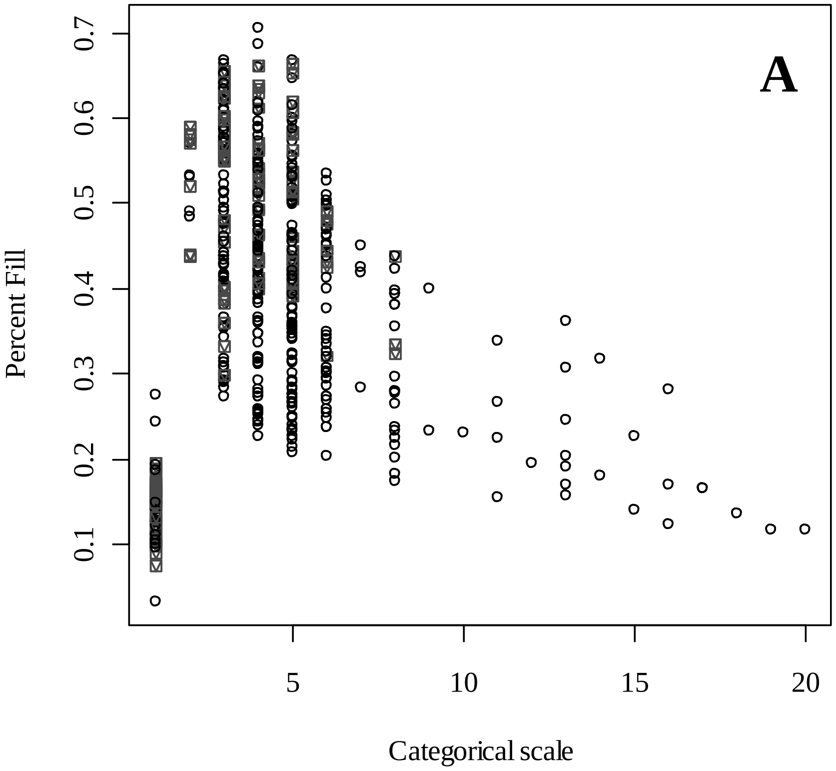

B
Residuals
3000
2000
1000
0
-1000
0
200000
600000
1000000
Area(square kilometers)

Figure 14. Matrix fill graphed against two measures of scale, significant matrices are black and hollow circles, non-significant matrices are gray and hollow squares.

A. Categorical scale.

B. Area (km^2).


A
Percent Fill
0.1
0.2
0.3
0.4
0.5
0.6
0.7
5
10
15
20
Categorical scale

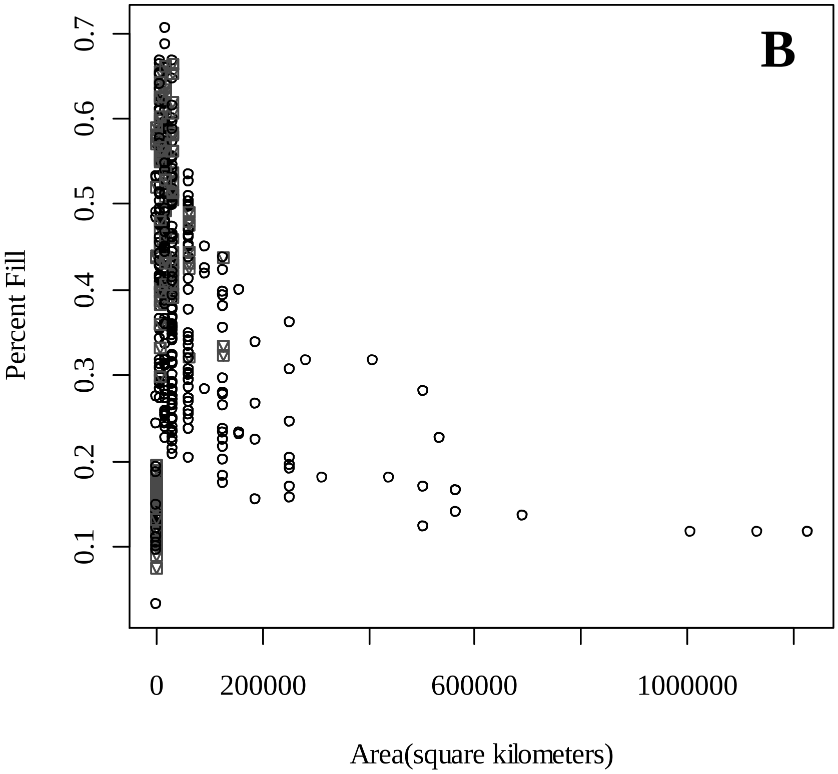

B
Percent Fill
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
200000
600000
1000000
Area(square kilometers)

Figure 15. Conceptual models of the possible interactions between immigration and extinction, distribution and abundance, and the nested subset pattern.

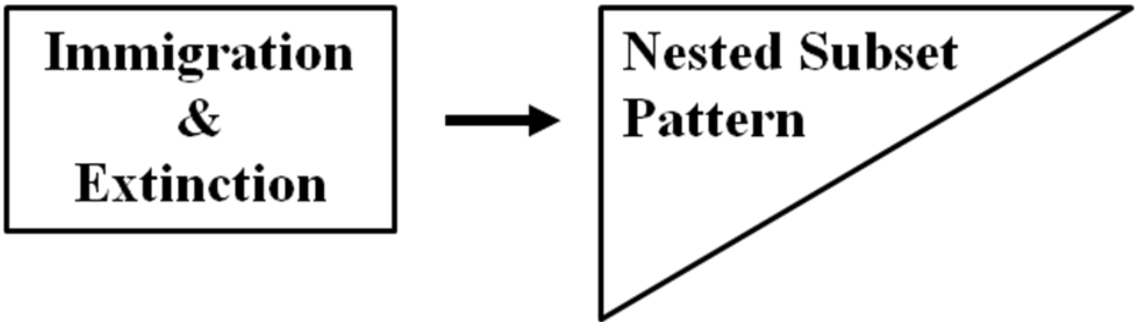
A. Immigration and extinction act directly on the nested subset pattern.

B. Immigration and extinction and distribution and abundance act directly on the nested subset pattern.

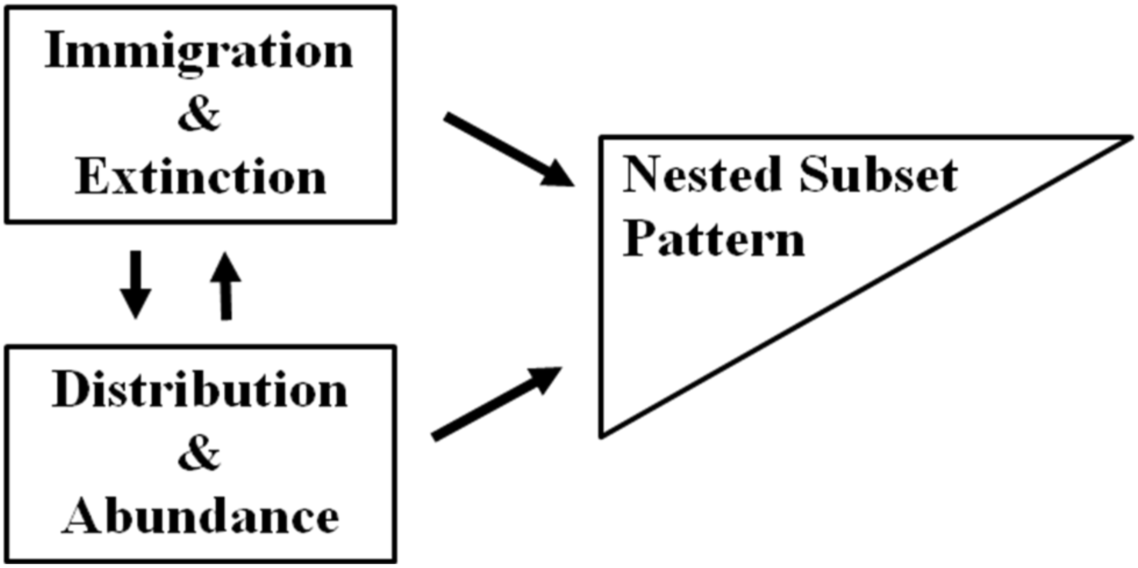
C. Immigration and extinction act directly on the nested subset pattern and distribution and abundance.

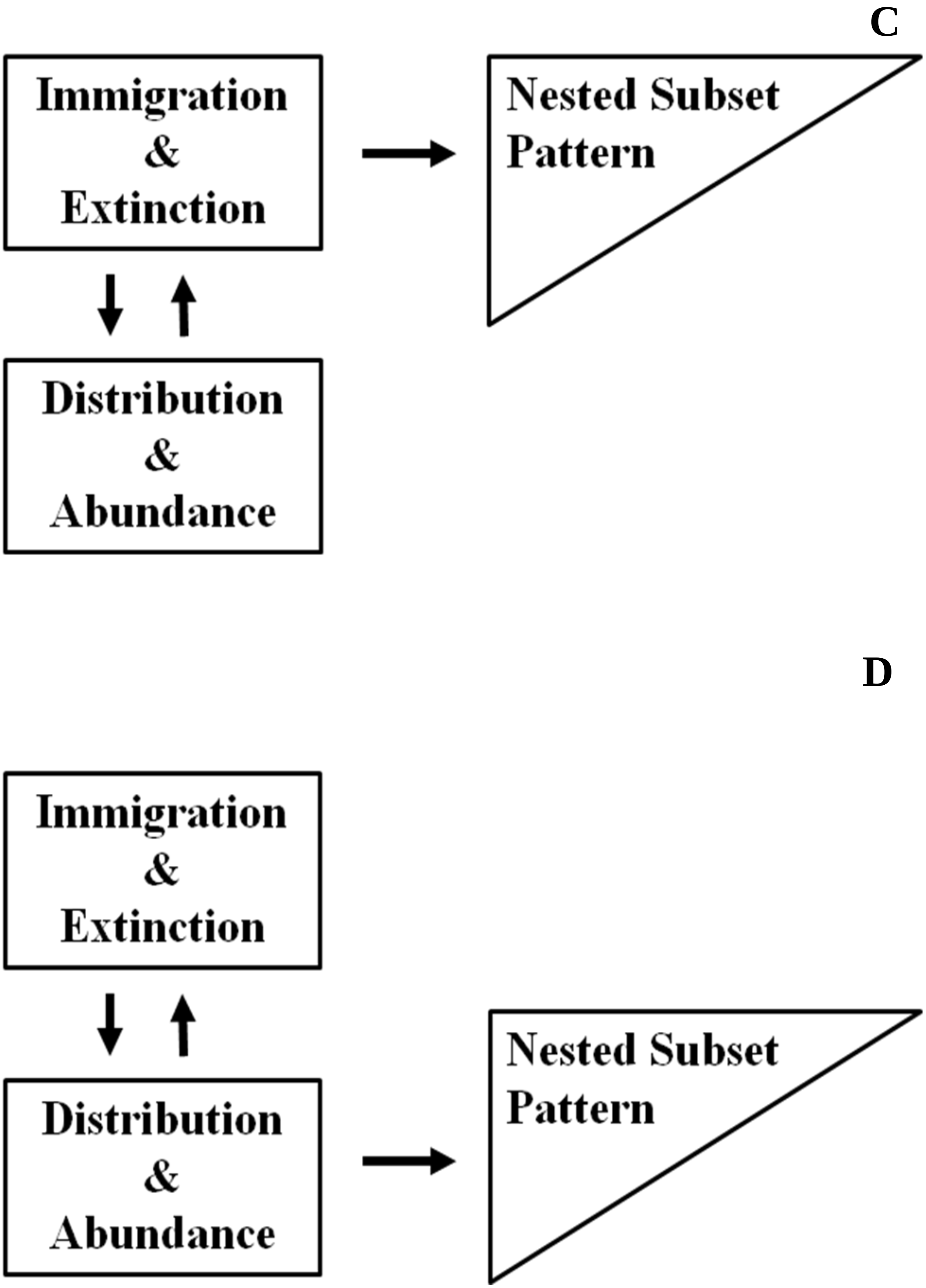
D. Immigration and extinction act indirectly on the nested subset pattern and directly on distribution and abundance.

A


Immigration
&
Extinction
Nested Subset
Pattern

B


Immigration
&
Extinction
Distribution
&
Abundance
Nested Subset
Pattern


C
Immigration & Extinction
Distribution & Abundance
Nested Subset Pattern
D
Immigration & Extinction
Distribution & Abundance
Nested Subset Pattern

Appendix A. Visual Basic 2008 code for the nested subset program used in analyses.

```
Option Explicit On

Public Class Cudgel

    'Declare variables
    Dim strInputFile As String
    Dim strOutputFile As String
    Dim ofdOpenFileDialog1 As OpenFileDialog
    Dim sfdSaveFileDialog1 As SaveFileDialog
    Dim intRandomizations As Integer 'Number of Randomizations
    Dim arrayMatrix(,) As Double
    Dim intMetricNumber As Integer
    Dim intNullNumber As Integer
    Dim boolInd As Boolean
    Dim boolOut As Boolean
    Dim boolNull As Boolean
    Dim arrayDataOut(7) As Double


    Private Sub Cudgel_Load(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles MyBase.Load

        Randomize()

        'Progess Bar limits
        'Try
        '    ProgressBar1.Minimum = 0
        '    ProgressBar1.Maximum = UBound(arrayMatrix)
        'Catch ex As Exception

        'End Try

    End Sub

    Private Sub Button1_Click_1(ByVal sender As System.Object, ByVal e
As System.EventArgs) Handles Button1.Click
        'Open File Dialog
        ofdOpenFileDialog1 = New OpenFileDialog
        'Locate input file
        With ofdOpenFileDialog1
            .AddExtension = True
            .AutoUpgradeEnabled = True
            .CheckFileExists = True
            .CheckPathExists = True
            .DefaultExt = ".xls,.txt, .xlsx"
            .Filter = "Excel files (*.xls)|*.xls|(*xlsx)|*.xlsx | Text
files (*.txt)|*.txt"
            .Multiselect = False
            .ShowHelp = False
            .Title = "Select input file."
        End With
```

```
        If ofdOpenFileDialog1.ShowDialog =
Windows.Forms.DialogResult.OK Then
            Try
                strInputFile = ofdOpenFileDialog1.FileName
            Catch ex As Exception


            End Try
        End If

        Label9.Text = strInputFile


    End Sub


    Private Sub Button2_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button2.Click

        sfdSaveFileDialog1 = New SaveFileDialog
        ' Must have a blank Excel workbook prepared to save data into.
        With sfdSaveFileDialog1
            .AddExtension = True
            .AutoUpgradeEnabled = True
            .CheckFileExists = False
            .CreatePrompt = True
            .CheckPathExists = True
            .DefaultExt = "xls, xlsx"
            .FileName = strOutputFile
            .ShowHelp = False
            .OverwritePrompt = False
            .Title = "Select output file."
        End With

        If sfdSaveFileDialog1.ShowDialog =
Windows.Forms.DialogResult.OK Then

        End If
        Try
            strOutputFile = sfdSaveFileDialog1.FileName
        Catch ex As Exception

        End Try

        Label10.Text = strOutputFile

    End Sub
    Public Sub Button3_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button3.Click
        If RadioButton1.Checked = True Then
            Batch.Show()
            Me.Hide()
        Else
```

```
            'Error handling protocols for radio buttons
            If Label9.Text = "Select input file." Then
                MessageBox.Show("Please specify an input file.")
                Exit Sub
            End If
            If Label10.Text = "Select output file." Then
                MessageBox.Show("Please specify an output file.")
                Exit Sub
            End If
            If RadioButton1.Checked = False And RadioButton2.Checked =
False Then
                MessageBox.Show("Please indicate the data format.")
                Exit Sub
            End If
            If RadioButton3.Checked = False And RadioButton4.Checked =
False Then
                MessageBox.Show("Please indicate whether species are in
rows or columns.")
                Exit Sub
            End If
            If RadioButton5.Checked = False And RadioButton6.Checked =
False Then
                MessageBox.Show("Indicate if the packed matrix should
be output or not.")
                Exit Sub
            End If
            If RadioButton7.Checked = False And RadioButton8.Checked =
False Then
                MessageBox.Show("Please indicate packing method.")
                Exit Sub
            End If
            If RadioButton9.Checked = False And RadioButton10.Checked =
False And RadioButton11.Checked = False And RadioButton12.Checked =
False And RadioButton13.Checked = False And RadioButton14.Checked =
False And RadioButton15.Checked = False And RadioButton16.Checked =
False And RadioButton17.Checked = False And RadioButton18.Checked =
False And RadioButton19.Checked = False Then
                MessageBox.Show("Please select a metric.")
                Exit Sub
            End If
            If RadioButton20.Checked = False And RadioButton21.Checked
= False And RadioButton22.Checked = False And RadioButton23.Checked =
False And RadioButton24.Checked = False And RadioButton25.Checked =
False And RadioButton26.Checked = False And RadioButton27.Checked =
False And RadioButton28.Checked = False Then
                MessageBox.Show("Please select a null model.")
                Exit Sub
            End If

            ''Error handling protocol for number of randomizations
            'If intRandomizations < 100 Then
            '    MessageBox.Show("Please enter a numerical value
between 100 and 10,000.")
            '    Exit Sub
```

```
            'End If
            'If intRandomizations > 10000 Then
            '    MessageBox.Show("Please enter a numerical value 
between 100 and 10,000.")
            '    Exit Sub
            'End If


            'Assign Randomizations Text Box value to variable
            Try
                intRandomizations = TextBox1.Text
            Catch ex As InvalidCastException

            End Try


            'Loading data into array
            Try

                'Make independant variable button boolean
                If RadioButton8.Checked = True Then
                    boolInd = True
                Else
                    boolInd = False
                End If

                'Make matrix output radio button boolean
                If RadioButton5.Checked = True Then
                    boolOut = True
                Else
                    boolOut = False
                End If

                'Put metric selection into a variable.
                If RadioButton9.Checked = True Then intMetricNumber = 1

                If RadioButton10.Checked = True Then intMetricNumber = 
2

                If RadioButton11.Checked = True Then intMetricNumber = 
3

                If RadioButton12.Checked = True Then intMetricNumber = 
4

                If RadioButton13.Checked = True Then intMetricNumber = 
5

                If RadioButton14.Checked = True Then intMetricNumber = 
6

                If RadioButton15.Checked = True Then intMetricNumber = 
7
```

```
                If RadioButton16.Checked = True Then intMetricNumber =
8

                If RadioButton17.Checked = True Then intMetricNumber =
9

                If RadioButton18.Checked = True Then intMetricNumber =
10

                If RadioButton19.Checked = True Then intMetricNumber =
11



                'Put null model selection into a variable.

                If RadioButton20.Checked = True Then intNullNumber = 1

                If RadioButton21.Checked = True Then intNullNumber = 2

                If RadioButton22.Checked = True Then intNullNumber = 3

                If RadioButton23.Checked = True Then intNullNumber = 4

                If RadioButton24.Checked = True Then intNullNumber = 5

                If RadioButton25.Checked = True Then intNullNumber = 6

                If RadioButton26.Checked = True Then intNullNumber = 7

                If RadioButton27.Checked = True Then intNullNumber = 8

                If RadioButton28.Checked = True Then intNullNumber = 9


                Call subloopInput(strInputFile, boolInd, boolOut,
strOutputFile, boolNull, intNullNumber, intMetricNumber,
intRandomizations, arrayDataOut)

            Catch ex As Exception

            End Try
        End If

        If RadioButton2.Checked = True Then
            MessageBox.Show("Run complete.")
        End If

    End Sub


    Private Sub Button4_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs)
```

```
        ' Exits program
        Application.Exit()


    End Sub

    Private Sub Button4_Click_1(ByVal sender As System.Object, ByVal e
As System.EventArgs) Handles Button4.Click
        'Exits Program
        Application.Exit()

    End Sub


End Class

Module Tantalus
    Public Function subloopInput(ByVal strInputFile As String, ByRef
boolInd As Boolean, ByRef boolOut As Boolean, ByVal strOutputFile As
String, ByVal boolNull As Boolean, ByVal intNullNumber As Integer,
ByVal intMetricNumber As Integer, ByRef intRandomizations As Integer,
ByRef arrayDataOut As Array)
        'Set up invisible Excel to handle input data
        Dim douMetricVal As Double

        Dim xlsObject1 As Microsoft.Office.Interop.Excel.Application
        xlsObject1 = New Microsoft.Office.Interop.Excel.Application

        Dim wbkInputFile As Microsoft.Office.Interop.Excel.Workbook
        Dim wksInputSheet As Microsoft.Office.Interop.Excel.Worksheet

        'Open input file
        xlsObject1.Application.Workbooks.Open(strInputFile)
        wbkInputFile = GetObject(strInputFile)
        wksInputSheet = wbkInputFile.Worksheets(1) 'Data must be in the
first worksheet

        'Get input data file parameters
        Dim intDataRows As Integer
        Dim intDataCols As Integer

        Dim rRows As Microsoft.Office.Interop.Excel.Range
        Dim rCols As Microsoft.Office.Interop.Excel.Range

        rRows = wksInputSheet.Range("A:A")
        rCols = wksInputSheet.Range("1:1")

        intDataRows = xlsObject1.WorksheetFunction.CountA(rRows)
        intDataCols = xlsObject1.WorksheetFunction.CountA(rCols)

        'Set up data array
        Dim intArrRows As Integer
        Dim intArrCols As Integer
        Dim intSpeciesRows As Integer
```

```
Dim intSitesCols As Integer

intArrRows = intDataRows
intArrCols = intDataCols

intSpeciesRows = intDataRows
intSitesCols = intDataCols



'Setup data array

Dim arraySpecies(intSpeciesRows) As String

Dim arraySites(intSitesCols) As String

Dim arrayNull(intArrRows, intArrCols) As Double
Dim arrayMatrix(intArrRows, intArrCols) As Double


'Load data array
Dim rCell As Microsoft.Office.Interop.Excel.Range

Dim intRowLoop As Integer
Dim intColLoop As Integer

Dim intXlsRowPos As Integer
Dim intXlsColPos As Integer


'Load species name array

intRowLoop = 0
intColLoop = 0

Do While intRowLoop < intDataRows

    rCell = wksInputSheet.Cells(intRowLoop + 1, 1)

    arraySpecies(intRowLoop) = rCell.Value

    intRowLoop = intRowLoop + 1

Loop


'Load sites array
intRowLoop = 0
intColLoop = 0

Do While intColLoop < intDataCols

    rCell = wksInputSheet.Cells(1, intColLoop + 1)
```

```
            arraySites(intColLoop) = rCell.Value

            intColLoop = intColLoop + 1

        Loop

        'Load array matrix
        intRowLoop = 0
        intColLoop = 0


        Do While intRowLoop < intDataRows
            intColLoop = 0
            intXlsRowPos = intRowLoop + 2

            Do While intColLoop < intDataCols
                intXlsColPos = intColLoop + 2

                rCell = wksInputSheet.Cells(intXlsRowPos, intXlsColPos)

                arrayNull(intRowLoop, intColLoop) = rCell.Value
                arrayMatrix(intRowLoop, intColLoop) = rCell.Value

                intColLoop = intColLoop + 1

            Loop

            intRowLoop = intRowLoop + 1

        Loop


        'Cleanup and close Excel application
        wbkInputFile.Close()
        xlsObject1.Application.Quit()


        Dim intRandomCounter As Integer
        Dim arrayMatrixVals(intRandomizations) As Double

        intRandomCounter = 0

        Do While intRandomCounter < intRandomizations + 1


            'Call null model function

            If intNullNumber = 5 Then FIProp(arrayNull, arrayMatrix,
intArrRows, intArrCols, intRandomCounter)


            'Call summarize function
            Call MatrixSummarize(arrayMatrix, boolInd, intArrRows,
intArrCols, intRandomCounter)
```

```
            'Call packing function

            Call PackMatrix(arrayMatrix, arraySpecies, arraySites,
intArrRows, intArrCols, intSpeciesRows, intSitesCols)

            If intRandomCounter = 1 And boolOut = True Then
                OutputMatrix(arrayMatrix, arraySpecies, arraySites,
intArrRows, intArrCols, strOutputFile)
            End If



            'Call metric function
            If intMetricNumber = 1 Then No(arrayMatrix, intArrRows,
intArrCols, douMetricVal)

            arrayMatrixVals(intRandomCounter) = douMetricVal


            intRandomCounter = intRandomCounter + 1

        Loop

        Call subpVal(arrayMatrixVals, intRandomizations, douMetricVal)

        Dim intSize As Integer

        intSize = ((intArrRows - 2) * (intArrCols - 2))

        Dim deciFill As Decimal

        Call subFill(intArrRows, intArrCols, arrayMatrix, intSize,
deciFill)

        Call DataOutput(boolNull, arrayMatrixVals, strOutputFile,
intRandomizations, douMetricVal, deciFill, intSize)

        'Fill batch output array
        arrayDataOut(0) = douMetricVal
        arrayDataOut(1) = arrayMatrixVals(0)
        arrayDataOut(2) = deciFill
        arrayDataOut(3) = intSize
        arrayDataOut(4) = intArrRows - 2
        arrayDataOut(5) = intArrCols - 2

        subloopInput = arrayDataOut(6)

    End Function
End Module

Module ModloopFIProp
```

```
    Public Function FIProp(ByRef arrayNull As Array, ByRef arrayMatrix
As Array, ByRef intArrRows As Integer, ByRef intArrCols As Integer,
ByRef intRandomCounter As Integer)

        Dim arrayRandom(intArrCols - 2) As Double
        Dim arrayTempRandom(intArrRows - 2, intArrCols - 2) As Double
        Dim intRowCounter As Integer
        Dim intColCounter As Integer
        Dim intCopyValue As Double
        Dim intPosRef As Integer
        Dim intValue As Double
        Dim intRandomCols As Integer
        Dim intRandomRows As Integer
        Dim intRandom As New Random

        ' ' Fixed- incidence proportional null model- Random 1 ' ' '

        '' Send observed matrix through without randomization
        If intRandomCounter = 0 Then

            FIProp = arrayMatrix

        Else

            'Shuffling protocol
            Dim intShuffle1 As Integer
            Dim intShuffle2 As Integer
            intRowCounter = 0
            intColCounter = 0
            intRandomCols = 0
            intRandomRows = 0

            Do While intRandomCols < intArrCols - 2

                arrayRandom(intRandomCols) = intRandomCols + 1

                intRandomCols = intRandomCols + 1
            Loop


            'Randomize and fill final array
            Do While intRowCounter < intArrRows - 2
                intRandomCols = 0
                intColCounter = 0

                Do While intRandomCols < intArrCols - 2

                    Randomize()

                    intPosRef = intRandom.Next(0, intArrCols - 3)
                    intValue = arrayRandom(intPosRef)
                    intCopyValue = arrayRandom(intRandomCols)
```

```
                    intShuffle1 = intValue
                    intShuffle2 = intCopyValue

                    arrayRandom(intRandomCols) = intShuffle1
                    arrayRandom(intPosRef) = intShuffle2

                    intRandomCols = intRandomCols + 1

                Loop

                'Fill final matrix using randomized key array
                Do While intColCounter < intArrCols - 2

                    intPosRef = arrayRandom(intColCounter)
                    intCopyValue = arrayNull(intRowCounter + 1,
intPosRef)
                    arrayMatrix(intRowCounter + 1, intColCounter + 1) =
intCopyValue

                    intColCounter = intColCounter + 1

                Loop

                intRowCounter = intRowCounter + 1
            Loop
            intRowCounter = 0
            intColCounter = 0

            FIProp = arrayMatrix
        End If
    End Function
End Module

Module ModloopSummarize
    Public Function MatrixSummarize(ByRef arrayMatrix As Array, ByRef
boolInd As Boolean, ByVal intArrRows As Integer, ByVal intArrCols As
Integer, ByVal intRandomCounter As Integer)

        Dim intCountRows As Integer
        Dim intCountCols As Integer

        'Copy independant variable and species richness to blank row
and column
        If intRandomCounter = 0 And boolInd = True Then
            'MessageBox.Show(" Matrix summarize.")
            Dim intRowCopy As Double
            Dim intColCopy As Double

            intCountRows = 0
            intCountCols = 0
            'Copy independant variable row to blank row
            Do While intCountCols < intArrCols - 1

                intRowCopy = arrayMatrix(0, intCountCols)
```

```
            arrayMatrix(intArrRows - 1, intCountCols) = intRowCopy

            intCountCols = intCountCols + 1

        Loop


        'Copy independant variable column to blank column

        intCountCols = 0
        intCountRows = 0

        Do While intCountRows < intArrRows - 1

            intColCopy = arrayMatrix(intCountRows, 0)

            arrayMatrix(intCountRows, intArrCols - 1) = intColCopy

            intCountRows = intCountRows + 1
        Loop



    Else
        'MessageBox.Show("Null summarize")

        Dim intRowSum As Integer
        Dim intColSum As Integer
        Dim intCount As Integer
        intCountCols = 1
        intCountRows = 1
        intCount = 0

        'Sum columns
        Do While intCountCols < intArrCols - 1
            intCountRows = 1
            intColSum = 0

            Do While intCountRows < intArrRows - 1


                If arrayMatrix(intCountRows, intCountCols) > 0 Then

                    intColSum = intColSum + 1

                Else

                    intColSum = intColSum

                End If

                intCount = intColSum
```

```
                    intCountRows = intCountRows + 1


                Loop

                arrayMatrix(intArrRows - 1, intCountCols) = intCount

                intCountCols = intCountCols + 1
            Loop


            'Sum rows

            intCountCols = 1
            intCountRows = 1

            Do While intCountRows < intArrRows - 1
                intCountCols = 1
                intRowSum = 0

                Do While intCountCols < intArrCols - 1

                    If arrayMatrix(intCountRows, intCountCols) > 0 Then

                        intRowSum = intRowSum + 1

                    Else

                        intRowSum = intRowSum

                    End If

                    intCount = intRowSum

                    intCountCols = intCountCols + 1

                Loop

                arrayMatrix(intCountRows, intArrCols - 1) = intCount

                intCountRows = intCountRows + 1

            Loop

        End If

        MatrixSummarize = arrayMatrix

    End Function

End Module

Module ModloopPack
```

```
    Public Function PackMatrix(ByRef arrayMatrix As Array, ByRef
arraySpecies As Array, ByRef arraySites As Array, ByVal intArrRows As
Integer, ByVal intArrCols As Integer, ByVal intSpeciesRows As Integer,
ByVal intSitesCols As Integer)

        'Declare new variables
        Dim arraySpeciesKey(intSpeciesRows) As String
        Dim arraySitesKey(intSitesCols) As String
        Dim arrayRowKey(intArrRows - 2, 1) As Double
        Dim arrayTempRows(intArrRows - 2, intArrCols) As Double
        Dim arrayTempMatrix(intArrRows, intArrCols) As Double
        Dim arrayColKey(1, intArrCols - 2) As Double
        Dim arrayTempCols(intArrRows, intArrCols - 2) As Double
        Dim intRowCount As Integer
        Dim intColCount As Integer
        Dim intCellCopy As Double
        Dim strKeyValue As String
        Dim intCompare1 As Double
        Dim intCompare2 As Double
        Dim intSwitch1 As Double
        Dim intSwitch2 As Double
        Dim intPosRef As Integer
        Dim intPosCount As Integer


        ' ' ' Sorting array rows ' ' '
        ''Fill row key array ''
        'Fill row key with sorting values
        intRowCount = 0
        intColCount = 0
        Do While intRowCount < intArrRows - 2

            intCellCopy = arrayMatrix(intRowCount + 1, intArrCols - 1)
            arrayRowKey(intRowCount, 0) = intCellCopy
            intRowCount = intRowCount + 1

        Loop

        'Fill row key with sorting position values
        intRowCount = 0
        intColCount = 0
        intPosCount = 0

        Do While intRowCount < intArrRows - 2

            intCellCopy = intPosCount
            arrayRowKey(intRowCount, 1) = intCellCopy + 1

            intPosCount = intPosCount + 1
            intRowCount = intRowCount + 1

        Loop

        ''Sort row key values
```

```
        'Switch values
        intRowCount = 0
        intColCount = 0
        intPosCount = 0

        Do While intPosCount < intArrRows - 2
            intColCount = 0

            Do While intColCount < intArrRows - 2
                intRowCount = 0

                Do While intRowCount < intArrRows - 2

                    intCompare1 = arrayRowKey(intColCount, 0)
                    intCompare2 = arrayRowKey(intRowCount, 0)
                    intSwitch1 = arrayRowKey(intColCount, 1)
                    intSwitch2 = arrayRowKey(intRowCount, 1)

                    If intCompare1 < intCompare2 And intSwitch1 <
intSwitch2 Then

                        arrayRowKey(intColCount, 1) = intSwitch2
                        arrayRowKey(intRowCount, 1) = intSwitch1

                    Else

                        arrayRowKey(intColCount, 1) = intSwitch1
                        arrayRowKey(intRowCount, 1) = intSwitch2

                    End If

                    intRowCount = intRowCount + 1

                Loop

                intColCount = intColCount + 1
            Loop

            intPosCount = intPosCount + 1
        Loop
        ''Fill temporary row array
        'Get row value from row key and use to sort original matrix
into temporary row array
        intRowCount = 0
        intColCount = 0

        Do While intRowCount < intArrRows - 2
            intColCount = 0
            intPosCount = arrayRowKey(intRowCount, 1)
            intPosRef = intPosCount - 1

            Do While intColCount < intArrCols

                intCellCopy = arrayMatrix(intRowCount + 1, intColCount)
```

```
        arrayTempRows(intPosRef, intColCount) = intCellCopy

        intColCount = intColCount + 1
    Loop


    intRowCount = intRowCount + 1

Loop


'' Fill species key array
' Fill key array with first two labels
strKeyValue = arraySpecies(0)
arraySpeciesKey(0) = strKeyValue
strKeyValue = arraySpecies(1)
arraySpeciesKey(1) = strKeyValue

' Fill key array with sorted labels
intRowCount = 0
intColCount = 0

Do While intRowCount < intArrRows - 2

    intPosCount = arrayRowKey(intRowCount, 1)
    intPosRef = intPosCount + 1

    strKeyValue = arraySpecies(intRowCount + 2)
    arraySpeciesKey(intPosRef) = strKeyValue

    intRowCount = intRowCount + 1
Loop


' ' ' Fill temporary matrix ' ' '
'' Fill intial row of the temporary matrix
' Fill row zero with data from matrix array
intRowCount = 0
intColCount = 0
Do While intColCount < intArrCols

    intCellCopy = arrayMatrix(intRowCount, intColCount)
    arrayTempMatrix(intRowCount, intColCount) = intCellCopy

    intColCount = intColCount + 1
Loop

'' Move sorted row data
' Fill temporary matrix with temporary row array
intRowCount = 0
intColCount = 0
Do While intRowCount < intArrRows - 2
    intColCount = 0
    Do While intColCount < intArrCols
```

```
                intCellCopy = arrayTempRows(intRowCount, intColCount)
                arrayTempMatrix(intRowCount + 1, intColCount) =
intCellCopy

                intColCount = intColCount + 1
            Loop

            intRowCount = intRowCount + 1
        Loop

        ''Fill final row of the temporary array
        'Fill final temporary array row with matrix array data
        intRowCount = 0
        intColCount = 0
        Do While intColCount < intArrCols

            intCellCopy = arrayMatrix(intArrRows - 1, intColCount)
            arrayTempMatrix(intArrRows - 1, intColCount) = intCellCopy

            intColCount = intColCount + 1
        Loop

        ' ' ' Sorting array columns ' ' '
        '' Fill column key array
        'Fill column key array with values from temporary matrix
        intRowCount = 0
        intColCount = 0
        Do While intColCount < intArrCols - 2

            intCellCopy = arrayMatrix(intArrRows - 1, intColCount + 1)
            arrayColKey(0, intColCount) = intCellCopy

            intColCount = intColCount + 1
        Loop

        'Fill column key array with sorting position values
        intRowCount = 0
        intColCount = 0
        intPosCount = 0
        Do While intColCount < intArrCols - 2

            intCellCopy = intPosCount
            arrayColKey(1, intColCount) = intCellCopy + 1

            intPosCount = intPosCount + 1
            intColCount = intColCount + 1

        Loop

        ''Sort column key values
        'Switch values
        intRowCount = 0
        intColCount = 0
```

```
        intPosCount = 0

        Do While intPosCount < intArrCols - 2
            intRowCount = 0

            Do While intRowCount < intArrCols - 2
                intColCount = 0

                Do While intColCount < intArrCols - 2

                    intCompare1 = arrayColKey(0, intRowCount)
                    intCompare2 = arrayColKey(0, intColCount)
                    intSwitch1 = arrayColKey(1, intRowCount)
                    intSwitch2 = arrayColKey(1, intColCount)

                    If intCompare1 < intCompare2 And intSwitch1 <
intSwitch2 Then

                        arrayColKey(1, intRowCount) = intSwitch2
                        arrayColKey(1, intColCount) = intSwitch1

                    Else

                        arrayColKey(1, intRowCount) = intSwitch1
                        arrayColKey(1, intColCount) = intSwitch2

                    End If

                    intColCount = intColCount + 1

                Loop

                intRowCount = intRowCount + 1
            Loop

            intPosCount = intPosCount + 1
        Loop

        ''Fill temporary column array
        'Get column value from column key and use to sort temporary row
packed matrix into temporary column array
        intRowCount = 0
        intColCount = 0

        Do While intColCount < intArrCols - 2
            intRowCount = 0
            intPosCount = arrayColKey(1, intColCount)
            intPosRef = intPosCount - 1

            Do While intRowCount < intArrRows

                intCellCopy = arrayTempMatrix(intRowCount, intColCount
+ 1)
                arrayTempCols(intRowCount, intPosRef) = intCellCopy
```

```
        intRowCount = intRowCount + 1
    Loop

    intColCount = intColCount + 1
Loop

'' Fill sites key array
' Fill key array with first two labels
strKeyValue = arraySites(0)
arraySitesKey(0) = strKeyValue
strKeyValue = arraySites(1)
arraySitesKey(1) = strKeyValue

' Fill key array with sorted labels
intRowCount = 0
intColCount = 0

Do While intColCount < intArrCols - 2

    intPosCount = arrayColKey(1, intColCount)
    intPosRef = intPosCount + 1

    strKeyValue = arraySites(intColCount + 2)
    arraySitesKey(intPosRef) = strKeyValue

    intPosCount = intPosCount + 1
    intColCount = intColCount + 1
Loop

' ' ' Fill final packed matrix ' ' '
'' Fill intial column of the packed matrix
' Fill column zero with data from matrix array
intRowCount = 0
intColCount = 0
Do While intRowCount < intArrRows

    intCellCopy = arrayTempMatrix(intRowCount, intColCount)
    arrayMatrix(intRowCount, intColCount) = intCellCopy

    intRowCount = intRowCount + 1
Loop

'' Move sorted column data
' Fill temporary matrix with temporary column array
intRowCount = 0
intColCount = 0
Do While intColCount < intArrCols - 2
    intRowCount = 0

    Do While intRowCount < intArrRows

        intCellCopy = arrayTempCols(intRowCount, intColCount)
        arrayMatrix(intRowCount, intColCount + 1) = intCellCopy
```

```
                intRowCount = intRowCount + 1
            Loop

            intColCount = intColCount + 1
        Loop

        ''Fill final column of the packed array
        'Fill final temporary array column with temporary matrix data
        intRowCount = 0
        intColCount = 0
        Do While intRowCount < intArrRows

            intCellCopy = arrayTempMatrix(intRowCount, intArrCols - 1)
            arrayMatrix(intRowCount, intArrCols - 1) = intCellCopy

            intRowCount = intRowCount + 1
        Loop
        intRowCount = 0
        intColCount = 0

        Do While intRowCount < intArrRows

            arraySpecies(intRowCount) = arraySpeciesKey(intRowCount)

            intRowCount = intRowCount + 1
        Loop

        Do While intColCount < intArrCols

            arraySites(intColCount) = arraySitesKey(intColCount)

            intColCount = intColCount + 1
        Loop

        'MessageBox.Show("End pack.")

        PackMatrix = arrayMatrix
        PackMatrix = arraySpecies
        PackMatrix = arraySites

    End Function

End Module

Module ModloopOutputMatrix

    Public Sub OutputMatrix(ByRef arrayMatrix As Array, ByRef
arraySpecies As Array, ByRef arraySites As Array, ByVal intArrRows As
Integer, ByVal intArrCols As Integer, ByVal strOutputFile As String)

        Dim xlsObject2 As Microsoft.Office.Interop.Excel.Application
        xlsObject2 = New Microsoft.Office.Interop.Excel.Application
```

```
        Dim wbkOutputFile As Microsoft.Office.Interop.Excel.Workbook
        Dim wksOutputSheet As Microsoft.Office.Interop.Excel.Worksheet

        'Open blank output file- Must create blank workbook before
running analysis.
        xlsObject2.Application.Workbooks.Open(strOutputFile)
        wbkOutputFile = GetObject(strOutputFile)
        wksOutputSheet = wbkOutputFile.Worksheets(2) 'matrix is sent to
the second worksheet.

        Dim intRowCounter As Integer
        intRowCounter = 0

        Dim intColCounter As Integer
        intColCounter = 0


        Dim rCell2 As Microsoft.Office.Interop.Excel.Range


        'Load species data
        intRowCounter = 0
        intColCounter = 0

        Do While intRowCounter < intArrRows

            rCell2 = wksOutputSheet.Cells(intRowCounter + 1, 1)
            rCell2.Value = arraySpecies(intRowCounter)

            intRowCounter = intRowCounter + 1

        Loop

        'Load sites data
        intRowCounter = 0
        intColCounter = 0

        Do While intColCounter < intArrCols

            rCell2 = wksOutputSheet.Cells(1, intColCounter + 1)
            rCell2.Value = arraySites(intColCounter)

            intColCounter = intColCounter + 1

        Loop

        'Load matrix data
        intRowCounter = 0
        intColCounter = 0

        Do While intRowCounter < intArrRows
            intColCounter = 0

            Do While intColCounter < intArrCols
```

```
                rCell2 = wksOutputSheet.Cells(intRowCounter + 2,
intColCounter + 2) 'Must have +1 to make the array line up with the
range.  +2 shifts it over 1.

                rCell2.Value = arrayMatrix(intRowCounter,
intColCounter)

                intColCounter = intColCounter + 1

            Loop

            intRowCounter = intRowCounter + 1

        Loop

        'Save output file
        wbkOutputFile.Save()

        'Clean up and close invisible Excel
        wbkOutputFile.Close()
        xlsObject2.Application.Quit()


        'MessageBox.Show("The matrix has been output.") ''''

    End Sub
End Module

Module ModN0
    Public Function No(ByRef arrayMatrix As Array, ByVal intArrRows As
Integer, ByVal intArrCols As Integer, ByRef douMetricVal As Double)

        Dim arrayTempMetric(intArrRows - 2) As Double
        Dim intRowCount As Integer
        Dim intColCount As Integer
        Dim intCheckVal As Integer
        Dim intAbsence As Integer
        Dim boolStart As Boolean

        boolStart = False
        intRowCount = 0
        intColCount = 0
        intAbsence = 0

        Do While intRowCount < intArrRows - 1
            intColCount = 0
            boolStart = False

            Do While intColCount < intArrCols - 2

                intCheckVal = arrayMatrix(intRowCount, (intArrCols - 2)
- intColCount)
```

```
                If intCheckVal > 0 Then boolStart = True

                If boolStart = True And intCheckVal = 0 Then

                    intAbsence = arrayTempMetric(intRowCount)
                    arrayTempMetric(intRowCount) = intAbsence + 1

                End If

                intColCount = intColCount + 1
            Loop

            intRowCount = intRowCount + 1
        Loop

        intRowCount = 0
        douMetricVal = 0

        Do While intRowCount < intArrRows - 1

            intAbsence = arrayTempMetric(intRowCount)
            douMetricVal = intAbsence + douMetricVal

            intRowCount = intRowCount + 1
        Loop

        No = douMetricVal

    End Function
End Module

Module ModpVal
    Public Function subpVal(ByRef arrayMatrixVals As Array, ByVal
intRandomizations As Integer, ByRef douMetricVal As Double)
        Dim intCheckVal As Integer
        Dim intRowCount As Integer
        Dim intAbsence As Integer
        Dim intObsVal As Integer

        intRowCount = 1
        intAbsence = 0
        intObsVal = arrayMatrixVals(0)

        Do While intRowCount < intRandomizations + 1

            intCheckVal = arrayMatrixVals(intRowCount)
            If intCheckVal < intObsVal + 1 Then
                intAbsence = intAbsence + 1
            Else
                intAbsence = intAbsence
            End If


            intRowCount = intRowCount + 1
```

```
        Loop

        douMetricVal = intAbsence / intRandomizations

        subpVal = douMetricVal
        'MessageBox.Show(douMetricVal.ToString and " = p-value.")
    End Function
End Module

Module Fill
    Public Function subFill(ByVal intArrRows As Integer, ByVal
intArrCols As Integer, ByRef arrayMatrix As Array, ByVal intSize As
Integer, ByRef deciFill As Decimal)

        Dim intRowCount As Integer
        Dim intColCount As Integer
        Dim intCount As Integer
        Dim intPlace As Integer

        intRowCount = 0
        intColCount = 0
        intCount = 0

        Do While intRowCount < intArrRows - 2
            intColCount = 0

            Do While intColCount < intArrCols - 2

                intPlace = arrayMatrix(intRowCount + 1, intColCount +
1)

                If intPlace > 0.5 Then

                    intCount = intCount + 1

                Else

                    intCount = intCount + 0

                End If

                intColCount = intColCount + 1

            Loop

            intRowCount = intRowCount + 1

        Loop

        deciFill = (intCount / intSize)

        subFill = deciFill

    End Function
```

```
End Module

Module ModloopDataOutput
    Public Sub DataOutput(ByRef boolNull As Boolean, ByRef
arrayMetricVals As Array, ByVal strOutputFile As String, ByVal
intRandomizations As Integer, ByRef doupVal As Double, ByVal deciFill
As Decimal, ByVal intSize As Integer)

        Dim xlsObject3 As Microsoft.Office.Interop.Excel.Application
        xlsObject3 = New Microsoft.Office.Interop.Excel.Application

        Dim wbkOutputFile As Microsoft.Office.Interop.Excel.Workbook
        Dim wksOutputSheet As Microsoft.Office.Interop.Excel.Worksheet

        Dim rCell3 As Microsoft.Office.Interop.Excel.Range

        Dim intCounter As Integer
        intCounter = 0


        'Open blank output file.
        xlsObject3.Application.Workbooks.Open(strOutputFile)
        wbkOutputFile = GetObject(strOutputFile)
        wksOutputSheet = wbkOutputFile.Worksheets(1) 'data is sent to
the first worksheet.

        ''Output metric values

        'Output observed value
        rCell3 = wksOutputSheet.Cells(1)
        rCell3.Value = "Observed value"
        rCell3 = wksOutputSheet.Cells(2, 1)
        rCell3.Value = arrayMetricVals(intCounter)
        rCell3 = wksOutputSheet.Cells(4, 1)
        rCell3.Value = "Expected values"

        'Output p-value
        rCell3 = wksOutputSheet.Cells(1, 3)
        rCell3.Value = "p-Value"
        rCell3 = wksOutputSheet.Cells(2, 3)
        rCell3.Value = doupVal
        'MessageBox.Show(douObsVal.ToString and " = Observed value.")

        'Output percent fill
        rCell3 = wksOutputSheet.Cells(4, 3)
        rCell3.Value = "Percent fill"
        rCell3 = wksOutputSheet.Cells(5, 3)
        rCell3.Value = deciFill

        'Output matrix size
        rCell3 = wksOutputSheet.Cells(7, 3)
        rCell3.Value = "Matrix size"
        rCell3 = wksOutputSheet.Cells(8, 3)
```

```
        rCell3.Value = intSize


        Do While intCounter < intRandomizations

            wbkOutputFile = GetObject(strOutputFile)
            wksOutputSheet = wbkOutputFile.Worksheets(1) 'data is sent
to the first worksheet.

            rCell3 = wksOutputSheet.Cells(intCounter + 5, 1)
            rCell3.Value = arrayMetricVals(intCounter + 1)
            'MessageBox.Show(rCell3.Value and " = Expected values.")

            intCounter = intCounter + 1

        Loop


        'Save output file
        wbkOutputFile.Save()

        'Cleanup and close Excel application
        wbkOutputFile.Close()
        xlsObject3.Application.Quit()

        'MessageBox.Show("Run complete.")


    End Sub

End Module
```