

Null Models for Everyone: A Two-Step Approach to Teaching Null Model Analysis of Biological Community Structure.

Declan J. McCabe* and Evelyn J. Knight

Saint Michael's College, Colchester, VT 05439

*Corresponding Author: dmccabe@smcvt.edu

Abstract: Since being introduced by Connor and Simberloff in response to Diamond's assembly rules, null model analysis has been a controversial tool in community ecology. Despite being commonly used in the primary literature, null model analysis has not featured prominently in general textbooks. Complexity of approaches along with difficulty in interpreting results may reduce classroom application of this technique. Yet, readily available software makes this set of tools accessible to everyone from practitioners to educators. This exercise describes a hands-on approach that can be cheaply implemented with students as a stand-alone exercise or as a prelude to use of full-blown simulations in follow up lab sessions. Along with a detailed hands-on approach, we have provided a brief introduction to a computer-based approach. This paper includes a link to 294 data sets extracted from Patterson's Nestedness software and reformatted for use in EcoSim. These exercises demonstrate how to compare the observed number of checkerboard patterns in real community data to a frequency distribution of checkerboard patterns in randomized communities. The hands-on simulation uses fake data as a starting point and dozens of randomizations to produce a histogram. Computer-based simulations generate thousands of simulations to rapidly analyze real data sets. The results provide a springboard for discussion of the underpinnings of inferential science as applied to ecological communities.

Keywords: Null model analysis; hands-on model; simulation; data mining; community ecology

INTRODUCTION

Diamond's (1975) assembly rules were based upon the assumption that ecologists could recognize the fingerprints of competitive exclusion left on the distributions of species in natural communities. By examining the distributions of species pairs on archipelagos it is possible to identify species that never coexist on a single island. Diamond referred to such exclusive patterns as checkerboard distributions and contended that they were evidence of competitive exclusion. Connor and Simberloff (1983) countered that exclusive patterns were as likely to result from many other mechanisms including random chance. Null model analysis provides a way to statistically compare co-occurrence patterns of observed species distributions with those of randomly placed species.

To run a null model analysis, investigators use the following simple steps (Gotelli and Entsminger, 2004):

- 1). First we calculate an index that presumably tells us something about the structure of an observed set of natural communities summarized in a species by site presence-absence matrix (Table 1). For the purposes of this exercise we will use the number of checkerboard distributions based on Diamond's (1975) contention that checkerboard distributions are consistent with competitive exclusion.
- 2). Subject to some constraints, we generate a set of randomized communities assembled absent the community structuring mechanism (competition in our example).
- 3). We measure the index (number of checkerboards) from this randomized community.

Table 1. Initial placement of the beads in the trays. Any set of 7 unique colors will work to represent the species. Letters A through P represent the islands or sites where these species are found. If available ice cube trays have just 14 wells, columns *O* and *P* can be omitted without drastically affecting the results.

Species	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Purple	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	1
Red	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Yellow	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0
Green	0	0	0	1	1	1	0	1	0	0	0	0	1	0	0	0
Pink	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0
Dark Blue	1	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0
Light Blue	0	0	1	0	0	0	1	1	1	0	0	1	0	1	0	1

4). We repeat steps 2 and 3 many times generating a histogram representing a distribution of communities that are considered *null* with respect to the structuring influence of competition.

5). We ask if the index calculated from the observed natural communities differs significantly from randomly assembled communities.

Rationale

Null model analysis has been an area of active research since its introduction to the field of ecology (Connor and Simberloff, 1979). This approach to data exploration has not been without controversy and yet remains a mainstay in the published literature. Despite prominence in the primary literature, null models may or may not feature prominently in text books aimed at students of general ecology or community ecology. In general ecology textbooks, null models tend either not to be mentioned or mentioned briefly (Molles, 2013; and Ricklefs and Miller, 2000). In community ecology textbooks, a chapter or only a few pages may be dedicated to null model analysis (Morin, 2011; and Mittelbach, 2012). This exercise is intended to help address this mismatch by providing two entry-level null-model exercises that can be

implemented in the lecture setting and/or the laboratory.

The first exercise is an inexpensive hands-on simulation using plastic ice-cube trays to represent an archipelago of islands and colored beads to stand in for the species inhabiting these islands; this activity can run in a 50-minute lecture. The follow-up exercise uses free or inexpensive software to analyze real data from the primary literature and is better suited to a full-length laboratory session. These exercises facilitate discussion of the philosophical underpinnings of inferential science. Can we ever infer mechanism from pattern? What valuable information would be missed by restricting our vision to the narrow lens of experimental science (Diamond, 2001)?

METHODS

Hands-on simulation.

Summary of the approach

An ice cube tray serves to represent an archipelago of 16 (or 14) islands. Beads of different colors represent different species that occur on the islands with the 14 or 16 ice cube tray wells each representing distinct islands. The initial set of beads in the trays as shown in Table 1 and summarized in Table 2, represents a hypothetical observed set of communities. Students work in pairs

Table 2. The number of checkerboards recorded from the observed data set in Table 1.

	L Blue	D Blue	Pink	Green	Yellow	Red	Purple
Purple	0	0	0	0	0	1	
Red	1	1	1	1	0		
Yellow	0	0	0	0			
Green	0	0	0				
Pink	1	0					
Dark Blue	1						
Light Blue							

Total number of checkerboards observed: 7

or small groups to record the observed set of communities in Table 3 and then repeatedly shuffle the beads to generate a distribution of null communities against which to compare the observed data set. Students record the number of species pairs with checkerboard distributions from the original observed and randomized communities. Each student group runs one or more replicate randomizations and the entire class generates a histogram of the number of checkerboards found on the simulated archipelagos against which to compare the

original observed archipelago. The many student replicate randomizations and resulting histogram comprise a null model of presence/absence data that can be compared to the original observed set of communities in the ice cube tray. EcoSim can be used to run computer simulations to repeat the exercise for sets of real communities on real archipelagos from data sets provided here: https://wikieducator.org/Null_Model_Home.

Materials

Each participating group of students needs the following: Labeled ice cube trays

Table 3. Table to be filled by students recording presence or absence of species on islands. Each student or group needs 1 copy to record the initial observed data set and an additional copy for every randomized data set planned for the activity. All entries should be "1" or "0"; no cells should be blank.

Species	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Purple																
Red																
Yellow																
Green																
Pink																
Dark Blue																
Light Blue																

Table 4. This table is used to record checkerboard distributions by inspection of data collected in table 3. Entries should be recorded as “1” or “0”. The number of checkerboards recorded in this table by each student group is to be recorded in a class-wide histogram on the board.

	L Blue	D Blue	Pink	Green	Yellow	Red	Purple
Purple							
Red							
Yellow							
Green							
Pink							
Dark Blue							
Light Blue							

Total number of checkerboards observed: _____

(Figure 1); beads in 7 colors (Perler® fuse beads work well and are small enough to permit storage of nested trays with beads in place); collecting boxes larger than the ice cube trays to dump beads (shoe boxes are ideal); forceps; a handout made from the pre-filled Table 1; three or four copies of the blank Tables 3 and 4; spare beads to replace the inevitably dropped beads. Note that Table 2 is intended for instructor use.

Labeling the ice cube tray

Labeling should be done on the internal vertical walls of the wells in the ice cube trays. Two labels are necessary in each well:

- 1) Each well should be labeled with a unique letter representing an individual island. Letter labels should be placed on a consistent side of each well so that they are all visible from one end of the tray (Fig. 1 A).
- 2) After rotating the tray 180°, the numbers of species occurring on each “island” should be labeled on the wall opposite the island’s letter label corresponding with Table 1 (Fig. 1 B). An arrow marked down the center of the tray helps to orient the tray.

Initial setup

The initial placement of beads represents a hypothetical set of species occurrences on the model islands. A bead of a given color

in a particular well represents a species on that island. This initial placement of species on islands was designed to yield a reasonable number of checkerboards while leaving enough empty cells in the species-by-site matrix to make randomization possible. Matrices with few empty cells are difficult or impossible to randomize because most species occur at most sites and few novel rearrangements of the communities are possible. The arrangement of species described in Table 1 yields an original ‘observed’ dataset with 7 checkerboard distributions.



Figure 1. A. View of an example ice cube tray showing the letters that represent the names of the islands. B. Opposite view of the tray revealing the numbers of species found on each island. These numbers are used to fix the number of species on each island during the simulation runs.

Running the activity

Original observed data

Ice cube trays with beads in place should be distributed to student groups along with forceps used to handle beads, blank copies of Tables 3 and 4, and boxes used to count and collect beads. The ice cube trays should initially be oriented such that the letters representing the island names face the data recorder. Students should record the observed data matrix of species presences and absences in a copy of Table 3. This original data set (Table 1) created for the purposes of this exercise, is analogous to an observed data set from a natural archipelago. Each species pair that never co-occurs in any one of the islands counts as having a checkerboard distribution and students should record these as a “1” in a copy of Table 4. For example, Purple and Red (in Table 1) have a checkerboard distribution because they are never found on the same island and so a “1” will be recorded (see top data row in Table 2). The purple and yellow species co-occur on island F and on island G (Table 1) and therefore do not have a perfect checkerboard distribution; a zero should be recorded (Table 2). Once every species pair has been scored, the total number of checkerboards should be recorded at the bottom of Table 3 (following the worked example in Table 2). This number is the “Observed number of checkerboards” from the original dataset and will subsequently be compared to the number of checkerboards in the null model, IE a histogram from the randomized data sets.

Randomized data sets

The beads should be dumped carefully into the empty box and mixed. It is essential to work only with the starting set of beads; adding beads will alter the constraints in the null model. The ice cube tray should now be rotated 180° such that students can see the number of species that had previously occurred per island in the original “observed” data set. In an ideal world we might use true randomization to place the beads back onto the islands in our

simulation, however rolling dice or using random number generators adds an additional logistical layer to this exercise that would make it impractical to implement during a 50-minute lecture. This difficulty presents an opportunity to discuss the differences between true randomization and haphazard placement of experimental units or, in this case, beads in wells or species on islands. The beads should now be placed haphazardly into the ice cube tray wells subject to 2 constraints: Constraint 1: The number of species per simulated island should equal the number of species that existed on that island in the observed data set. Constraint 2: Beads of identical color should not be placed in the same well. Once the beads have been placed, bead distribution should be assessed and data recorded in Tables 3 & 4. The output of each simulation will be the number of checkerboards recorded from the randomized species distributions. Each student group’s data should be recorded as a single point on a class-wide histogram (Figure 2). The simulation phase should be repeated until 20 or more points have been added to the histogram. Even with a large class it would be worth repeating the

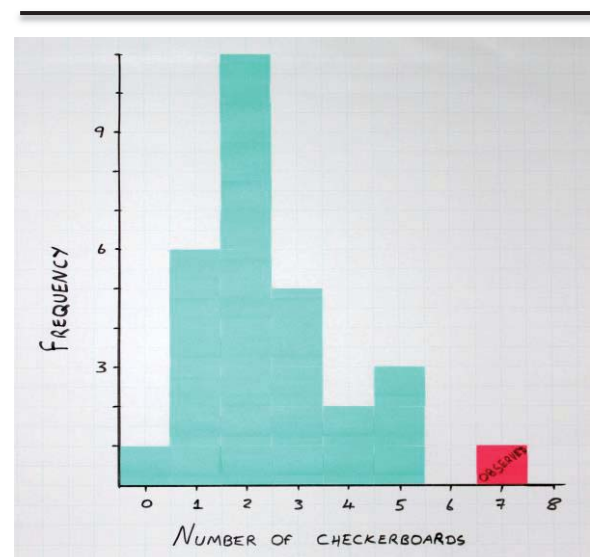


Figure 2. Class-wide histogram for the hands-on null model. Each student or group’s reported number of checkerboards was placed on the histogram using a sticky note to represent the result of their randomization

simulation at least twice per student group to emphasize the point that many replications are needed to generate a null distribution. Figure 2 is the result of four repetitions of the exercise generated by seven student groups in a Saint Michael's College class of 22 students.

Interpretation

The constraints used in this hands-on simulation are the fixed-fixed constraints frequently used in null model analysis. Constraint 1 fixes the number of species per island. Constraint 2 fixes the number of islands upon which a particular species occurs. These constraints provide ample opportunity for student discussion on the nature of null models. The approach to bead placement used to generate the in-class histogram facilitates discussion of true randomization and other alternative approaches.

Computer-based null model.

Summary of the approach

Like the hands-on approach, the computer-based simulation uses the checkerboard index ostensibly to suggest competitive exclusion. Whereas the hands-on method used entirely fictional data, the simulations add realism by utilizing data sets from published literature. We start by measuring the number of observed checkerboards in a published data set. This is followed by measuring this same index from 5,000 data matrices generated by randomly reassigning the original species to islands. Standard statistics are used to compare the observed index to the index measured from the randomized data sets.

Data and software

The EcoSim 2004 software (Gotelli and Entsminger 2004) can be downloaded for free from the following link: <http://www.uvm.edu/~ngotelli/EcoSim/ecosim7.zip>. An updated version of the software is available commercially: <http://www.garyentsminger.com/ecosim/>. After installing the software, data sets extracted from Atmar and Patterson's (1995) Nestedness software can be obtained from

this webpage:

<http://wikieducator.org/NullModelData>. The datasets can be saved as *.txt* files that can be opened directly from EcoSim. The fake data from Table 1 are also hosted on this web site and can be used for comparison with the hands-on simulation.

Initial setup

Once EcoSim opens, click *open* and search for your saved data set. Data should open automatically and you can skip straight to *Running the activity*. If data sets fail to open, the contents of *.txt* files can be copied and pasted into EcoSim software. To do this, click *Edit* and *Edit matrix as text* (this option not available in the commercial version). Delete the existing data and paste in the copied data. Note: Delete any empty lines that may be present before or after the numerical data. Finally, close and click *Yes* to save changes to the main grid.

Running the activity

In EcoSim, you can run the co-occurrence module to check for number of species combinations, number of checkerboards, *Cscore* etc. To do this click *Analyze* and then *1 Co-occurrence* to see the EcoSim Co-occurrence Options screen. Under the *Preferences* tab, there will be a drop-down box directly under the co-occurrence index. As default, it should say *C-score*, click the down arrow and choose *number of checkerboards*. Leave all other settings at their defaults: the fixed sum for both row and column constraints emulates the fixed/fixed constraints discussed above. Under the *General* tab; click *run*. The software will randomly shuffle species onto islands and generate 5,000 random matrices subject to the fixed/fixed constraints. From each of the matrices, Ecosim will count the number of checkerboards and generate a histogram of the results. You will be transferred automatically to the Co-occurrence results screen when the randomizations have been completed.

Data output

Ecosim generates information displayed under a series of tabs:

- *Input matrix* tab: the original observed data set can be seen here.
- *Simulation* tab provides one example of the 5000 randomized data sets.
- *Index* tab contains the histogram of the number of checkerboards recorded from the 5,000 null communities. An average number of checkerboards from the simulated data is calculated and can be compared to the observed number of checkerboards from the raw data. Also presented on this tab are the probabilities that the observed number of checkerboards differs from the randomized number of checkerboards. By convention, we can reject this null hypothesis if the p value is less than 0.05. Two p values are presented representing the left and right tail probabilities. Typically ecologists are interested in p value at the bottom of the display; it refers to the right-hand tail of the distribution, or more checkerboards in the observed data than in the randomized data; the upper value refers to the left-hand tail of the distribution, or fewer checkerboards in the observed data than in the randomized data. The last three columns form the histogram window that summarizes the distribution of checkerboards for the simulated data. The low and high columns are the boundaries of 12 evenly spaced histogram bins representing the number of checkerboards in the randomized data sets. The # simulations column records the number of simulated indices in each bin and could be used to construct a traditional histogram.
- *Summary* tab includes a standardized effect size (SES) of the difference between observed and simulated. SES is expressed in standard deviations and is of value when comparing matrices of very different sizes. For data sets with significantly more checkerboards than in the randomized data, the SES value will exceed 2.

RESULTS AND DISCUSSION

Sample Results

Hands-on ice cube tray simulation

There are seven checkerboard distributions in the hands-on ice cube tray model (Table 2). The 28 randomizations performed by the Fall 2015 Community Ecology Class at Saint Michael's College

yielded an average of 2.36 checkerboards (Figure 2). A one-sample t test can be used to compare the observed value, 7, to the distribution of the randomized values generated in the classroom.

Ecosim analysis of ice cube tray data set

The data from Table 1 can also be downloaded from the Wikieducator site (<https://wikieducator.org/NullModelData>) and opened in Ecosim to run the computer-based null model. The observed number of checkerboards in the starting data set will of course be 7. The average number of checkerboards calculated from the simulated datasets will vary slightly from run to run and in my example (Figure 3) was 2.08 which compares quite well with the value of 2.36 from the average generated from the hands-on simulation (Figure 2). The p value is quite small (tail probability value < 0.0002) indicating that the observed value (7 checkerboards) is significantly larger than the average of the randomized communities (2.08 checkerboards) confirming that the 'species'.... or beads in this case, were indeed non-randomly placed in the trays in the original data set. The authors can

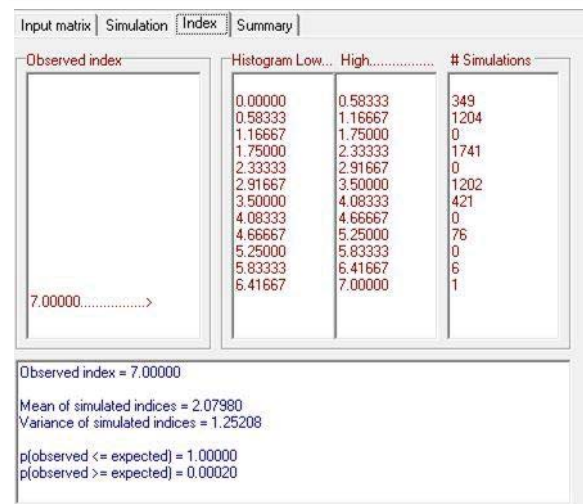


Figure 3. Example Co-occurrence results for computer-based null model using EcoSim 2004 using data from the Ice cube tray simulation. We used Ecosim to run 5000 randomizations and determined the average number of checkerboards (2.08) from the randomized data sets to construct a null model against which to compare the observed number of checkerboards (7).

certainly confirm that we did indeed design the data set in a non-random fashion. Statistically, this p value is defined as the probability that the observed value (7) is greater than or equal to the average of the simulated values (2.08 checkerboards).

Ecosim analysis of Harris' (1973) Galapagos Island bird data set

Because 18 checkerboards were observed in the raw data (Figure 4), we know that there were 18 species pairs among the birds of the Galapagos Islands that never co-occurred on the same island. The mean number of checkerboards from the 5000 simulated data sets that constitute our null model was 6.3 and the observed value (18 checkerboards) was significantly greater than the number in the randomized data or null model (tail probability value < 0.004).

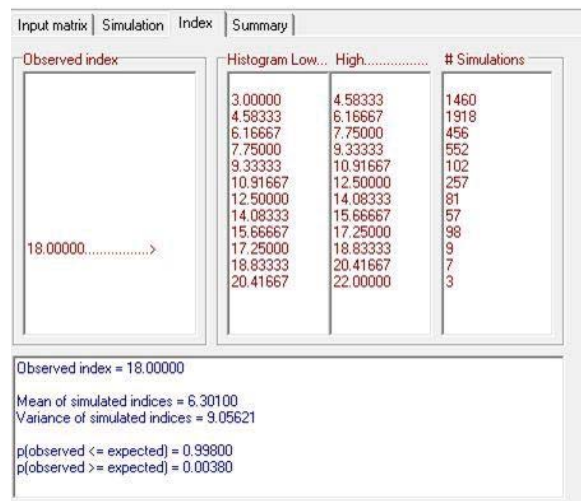


Figure 4. Example Co-occurrence results for computer-based null model using EcoSim 2004. Data set from Harris' (1973) study on 23 bird species on 15 islands in the Galápagos. We ran 5000 iterations to determine the average number of checkerboards in null communities.

Interpretation:

After completing this exercise, we hope you'll agree that running a null model analysis of community data is comparatively easy. However, it is most important that students learn to exercise caution when interpreting the output generated by null model analysis. It is tempting to reach beyond what can be said with certainty and to arrive at illogical conclusions. Many

mechanisms can produce checkerboard distributions, and it can be a valuable exercise to have students come up with a list of such mechanisms. For example, simple differences in habitat requirements could confine one species to rocky locations, and a second species to sandy locations. Such a species pair would have a checkerboard distribution that had absolutely nothing to do with competitive exclusion.

Null models in the context of ecological community development

Historically, some ecologists have considered each ecological community to be a tightly interdependent species group comprising a super organism (Clements, 1916) whereas others saw communities at particular sites to be far more coincidental (Gleason, 1926). Reality lies between these extremes but it is fair to say that biotic, abiotic, and entirely stochastic mechanisms interact and influence the development of biological communities. Null model analysis can reveal patterns that are more or less consistent with each of these mechanisms and suggest experiments that might reveal specific causal relationships. Competition leading to exclusive relationships among species pairs would yield a community with a checkerboard score (Diamond, 1975) in the right tail of the null distribution and would be consistent with an inhibition model of community development (Connell and Slatyer, 1977). The tolerance model of succession (Connell and Slatyer, 1977) might initially yield communities consistent with null communities but with increasing numbers of checkerboards as succession progresses and competitive exclusion intensifies; students could test this using Rey's (1981) defaunation/recolonization data (see "Fumigated Spartina" data sets on the WikiEducator site). Communities with fewer checkerboards than expected by chance appear in the left tail of the null distribution and this pattern would be consistent with the facilitation model of

community development (Connell and Slatyer, 1977).

CONCLUSIONS

The hands-on activity introduces the co-occurrence concept and provides students with a user-friendly path to understanding the null model approach to data exploration. Following the hands-on exercise with the computer-based null model activity will introduce students to powerful null model analysis tools used for research publications (Gotelli and McCabe, 2002). The activities provide entry points into a controversial and often contentious approach to data analysis that continues to be employed by community ecologists. The combined activities facilitate discussion of the value of non-experimental data sets and also the limits inherent in observational data exploration.

ACKNOWLEDGEMENTS

DM's community ecology research is funded by Vermont EPSCoR's OIA Grant #1556770 from the National Science Foundation. We are grateful to Karen Talentino and the Office of the Vice President of Academic Affairs at Saint Michael's College for ongoing support of student research and pedagogical innovations at Saint Michael's College. This project would not have been possible without the help of Luis Miguel Morillas of CPIFP Los Enlaces in Zaragoza, Spain and member of the Wikieducator community, who wrote and implemented the code to transfer Atmar and Patterson's (1995) data sets to Wikieducator.

REFERENCES

ATMAR, J.W., and B.D. PATTERSON. 1995. The nestedness temperature calculator: a visual basic program, including 294 presence-absence matrices. AICS Research, Inc., University Park, NM, and The Field Museum, Chicago.

CLEMENTS, F. E. 1936. Nature and structure of the climax. *The Journal of Ecology* 24:252-284.

CONNELL, J. H., and R. O. SLATYER. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. *The American Naturalist* 111:1119-1144.

CONNOR, E. F., and D. SIMBERLOFF. 1979. The assembly of species communities: chance or competition? *Ecology* 60(6): 1132-1140.

DIAMOND, J. M. 1975. The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. *Biological conservation*. 7(2): 129-146.

DIAMOND, J. M. 2001. Dammed experiments! *Science*. 30 November 2001: Vol. 294 no. 5548 pp. 1847-1848

GLEASON, H.A. 1926. The individualistic concept of plant association. *Bulletin of the Torrey Botanical Club* 53:7-26.

GOTELLI, N. J., and D.J. MCCABE. 2002. Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. *Ecology*, 83(8): 2091-2096.

GOTELLI, N. J., and G.L. ENTSMINGER. 2004. EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465. *Computer software*. <http://garyentsminger.com/ecosim/index.htm>.(accessed on 19-04-2014).

HARRIS, M. P. 1973. The Galapagos avifauna. *Condor*. 75(3):265-278.

MITTELBACH, G.G. 2012. Community ecology. Sinauer Associates, Sunderland, Massachusetts.

MOLLES, M.C. 2013. Ecology concepts and applications. McGraw-Hill, New York, New York.

MORIN, P.J. 2011. Community ecology. Wiley, Chichester, West Sussex.

REY, J.R. 1981. Ecological biogeography of arthropods on *Spartina* islands in Northwest Florida. *Ecological Monographs* 51:237-265.

RICKLEFS R.E., and G.L. MILLER. 2000. *Ecology*. W.H. Freeman and Company, New York, New York.