



## MPowering ecologists: community assembly tools for community assembly rules

Joshua Ladau and Sadie J. Ryan

*J. Ladau (joshua.ladau@gladstone.ucsf.edu), Gladstone Inst. GICD, 1650 Owens Street, San Francisco, CA 94158, USA. – S. J. Ryan, National Center for Ecological Analysis and Synthesis (NCEAS), Univ. of California, 735 State Street, Suite 300, Santa Barbara, CA 93101-5504, USA.*

Null model tests of presence–absence data (‘NMTPAs’) provide important tools for inferring effects of competition, facilitation, habitat filtering, and other ecological processes from observational data. Many NMTPAs have been developed, but they often yield conflicting conclusions when applied to the same data. Type I and II error rates, size, power, robustness and bias provide important criteria for assessing which tests are valid, but these criteria need to be evaluated contingent on the sample size, null hypothesis of interest, and assumptions that are appropriate for the data set that is being analyzed. In this paper, we confirm that this is the case using the software MPower, evaluating the validity of NMTPAs contingent on the null hypothesis being tested, assumptions that can be made, and sample size. Evaluating the validity of NMTPAs contingent on these factors is important towards ensuring that reliable inferences are drawn from observational data about the processes controlling community assembly.

Many fundamental questions about the processes that shape communities cannot be addressed experimentally because experimentation is impractical, unethical, or impossible at the relevant large spatial and temporal scales. Instead, to address these questions, inferences must be made from observational data (Connor and Simberloff 1986). Null model tests of presence–absence data (‘NMTPAs’) are a key tool for making these inferences, allowing community-wide effects of habitat filtering, competition, and facilitation to be inferred from observations of species distribution patterns (Connor and Simberloff 1979). In testimony to their importance, NMTPAs have been used in over 100 recent published studies (Google Scholar search performed on 20 April 2007; Sanders et al. 2003, Gotelli and McCabe 2002, Gotelli and Rhode 2002). Moreover, because an understanding of processes is needed for many conservation-related management decisions, for example, preserving intact assemblages where ecosystem function plays a critical role as the process, NMTPAs can play a role in the development of conservation strategies (Beissinger et al. 1996). In the future, NMTPAs will continue to be important because of the increasing need to make both basic and policy-related inferences from observational data.

In general, null models are constructed to exclude the process of interest (e.g. competition). If the data are found to be inconsistent with the model, then the process is concluded to possibly have affected community assembly (Connor and Simberloff 1979). The data used in NMTPAs

are lists of species present at sites, which are relatively easy to collect and widely available in the literature. The data are usually summarized in a presence–absence matrix, in which rows represent species and columns represent sites. In the matrix, 1 is entered where species are present and 0 where they are absent (Gotelli and Graves 1996).

The application of NMTPAs has two substantial advantages, but also has a serious shortcoming. NMTPAs are often used instead of process-based models because they give primacy to the parsimonious hypothesis – that the process of interest is not occurring – and because they can allow the rate of falsely concluding that a process is occurring to be controlled, as in statistical hypothesis testing (Connor and Simberloff 1983, 1986). However, over 40 NMTPAs have been proposed (Gotelli 2000, Fox 1987), and they often give conflicting conclusions when applied to the same data. Clearly, not all of the conclusions can be correct. Hence, it can be unclear which, if any, NMTPAs should be used.

Given the importance of NMTPAs, their validity should be assessed using appropriate and rigorous criteria. Such criteria are provided by Neyman–Pearson hypothesis testing theory, which is the justification for most common frequentist statistical tests such as ANOVAs and non-parametric methods (Casella and Berger 2002, Lehmann and Romano 2005). The criteria provided by Neyman–Pearson theory include size, which is the maximum rate at which the null hypothesis is rejected when it is true; power, which is the rate at which the null hypothesis is rejected when it is false;

robustness, which is a measure of the dependence of a test's error rates on assumptions; and bias, which is a measure of how much more likely the null hypothesis is to be rejected when it is false than when it is true. These criteria have been used to evaluate many NMTPAs (Wilson 1987, Gotelli 2000, Ladau 2008), and recent applications of NMTPAs have been justified based on closely-related type I and II error rates (rates of incorrect rejection and acceptance of the null hypothesis, respectively) (Gotelli and Rohde 2002, Feeley 2003, Chaves and Anez 2004, Heino and Soininen 2005, Mouillot et al. 2005, Ribichich 2005). If an NMTPA meets these criteria, then it will likely lead to correct ecological conclusions, whereas if it does not, then the conclusions will likely be incorrect.

The size, power, robustness, and bias of NMTPAs depend on the data that are being analyzed. For instance, although a particular test may have low error rates when applied to one data set, that test may have error rates as high as 100% when applied to another. In particular, the validity of NMTPAs depends heavily on the sample size, null hypothesis being tested, and assumptions that can be made about the data (Ladau 2008). It appears difficult to simply generate general rules for the applicability of NMTPAs, so it is necessary to evaluate the validity of NMTPAs on a case-by-case basis.

In this paper, we confirm the results of Ladau (2008) – that the validity of NMTPAs needs to be checked on a case-by-case basis. We consider six published data sets (Connor and Simberloff 1979, Reed 1980, Haefner 1988, Gotelli and Ellison 2002, Feeley 2003), and show that the validity of NMTPAs depends on the sample size, null hypothesis, and assumptions. Moreover, we confirm that in general, strong assumptions are required in order to reach reliable conclusions using NMTPAs in these cases. To facilitate evaluating NMTPAs on a case-by-case basis, we employ MPower, a user-friendly software package detailed in the Appendix 1, for evaluating the size, power, robustness and bias of NMTPAs contingent on the sample size, null hypothesis, and assumptions that are selected by the user.

## Methods

### Formulating null hypotheses

The validity of any hypothesis test depends on the null and alternative hypotheses that it is being used to test. The hypotheses determine what constitutes an error, and hence what it means for a test to be reliable and valid. In previous work, several null and alternative hypotheses have been considered for NMTPAs (Gotelli and McCabe 2002, Ladau 2008). Three canonical null and alternative hypotheses are:

1. *Species interactions.* The null hypothesis ( $H_{01}$ ) posits that at a given site, species occur independently. For instance, if two species can occur on an island, then the probability of species A occurring there is the same regardless of whether species B is already there. Such a situation is inconsistent with effects of species interactions. The alternative hypothesis is consistent with interactions between species, and posits that at a given site, species occur dependently. For instance,

if competition affects the co-occurrence pattern, then species A might be less likely to occur if species B is present.

2. *Heterogeneity of sites.* The null hypothesis ( $H_{02}$ ) posits that each species is equally likely to occur at all sites. Such a situation might arise if sites are uniform and equally accessible. The corresponding alternative hypothesis posits that species have different probabilities of occurring at different sites. Such a situation can arise if sites differ in microclimates, or if some sites are more accessible than others.
3. *Heterogeneity of species.* The null hypothesis ( $H_{03}$ ) posits that at a given site, all species are equally likely to occur. Such a situation could arise if all species have similar dispersal abilities and niche requirements, so that if a site is accessible and hospitable to one species, then it is accessible and hospitable to all species. This null hypothesis differs from the previous null hypothesis in that sites may differ. The corresponding alternative hypothesis posits that species may have different probabilities of occurring at each site. For instance, because a site might be in an arid location, Species A might be more likely to occur than species B.

### Assumptions

The validity of any hypothesis test also depends on which assumptions are reasonable. For example, if assumptions of normality, independence, and equality of variance are met, then a Student's t-test is an appropriate and optimal test of whether two means differ. When the assumptions are invalid but others are met, different, non-parametric tests are appropriate (Zar 1999). Likewise, the reliability of NMTPAs depends on which assumptions are appropriate (Ladau 2008).

Three assumptions are particularly germane to NMTPAs. First, if the null hypothesis is  $H_{02}$  or  $H_{03}$ , then the presence of each species can be assumed to not affect the probability of other species occurring, a scenario consistent with a lack of effects of interactions. Second, if the null hypothesis is  $H_{01}$  or  $H_{03}$ , then species can be assumed to be equally likely to occur at all sites (different species may have different probabilities of occurring, though). Such a situation might arise if sites are uniform and equally accessible. This assumption will hereafter be referred to as site equivalency. Last, if the null hypothesis is  $H_{01}$  or  $H_{02}$ , then all species can be assumed equally likely to occur at each site, an assumption hereafter referred to as species equivalency. However, under this assumption, species may have different probabilities of occurring at different sites. Species equivalency will be reasonable if all species have similar dispersal abilities and niche requirements, so that if a site is accessible and hospitable to one species, then it is accessible and hospitable to all species.

Needless to say, for many data sets, these assumptions are unrealistic – species interact, the sites are heterogenous, and species differ in their dispersal abilities. So why consider these assumptions? If no assumptions are made, then it is very likely that all NMTPAs will be unreliable. That said, some tests – particularly the ones that fix row or column

totals – may require fewer assumptions than other tests (Ladau 2008).

### Presence–absence matrices

The appropriateness of any hypothesis test can depend on attributes of the sample that is being analyzed. For instance, most ANOVAs require a minimal sample size, below which the test will be incapable of giving significant results (Zar 1999). Likewise, the appropriateness and reliability of NMTPAs can depend on attributes of the presence–absence matrix; particularly its dimensions (Ladau 2008). In general, NMTPAs may have low power when applied to small matrices (fewer than seven rows and columns). However, the appropriateness and reliability of NMTPAs can vary unpredictably with matrix size, and should be assessed on a case-by-case basis.

### Analysis of published data sets

To confirm that the validity of NMTPAs varies on a case-by-case basis, we selected six published presence–absence matrices (Connor and Simberloff 1979, Reed 1980, Haefner 1988, Gotelli and Ellison 2002, Feeley 2003) that spanned a range of taxa and geographic locations had widely varying dimensions (Table 1, first column). We used MPower (Appendix 1) to evaluate the validity of four commonly used NMTPAs for analyzing each matrix (Gotelli 2000, Gotelli and McCabe 2002, Gotelli and Ellison 2002, Feeley 2003). The NMTPAs all employed the C-score statistic (Stone and Roberts 1990), and randomization algorithms that fixed the total number of species occurrences, the presence–absence matrix row totals, column totals, or both the row and column totals (Appendix 1; Gotelli 2000). These NMTPAs will hereafter be referred to as  $\delta_{EE}$ ,  $\delta_{FE}$ ,  $\delta_{EF}$  and  $\delta_{FF}$ , respectively. We considered the null hypothesis  $H_{01}$  (interspecific interactions), and we evaluated the validity of each test under four assumptions: both species and sites equivalent, just species equivalent, just sites equivalent, and no assumptions (Ladau 2008).

## Results and discussion

The validity of the tests  $\delta_{EE}$ ,  $\delta_{FE}$ ,  $\delta_{EF}$  and  $\delta_{FF}$  depended on the presence–absence matrix that was being analyzed and the assumptions. If species and/or sites were assumed equivalent,

then in almost all instances, at least one test was robust and powerful and in many cases tests existed that were also unbiased (Table 1). However, if no assumptions were made, then all tests were either non-robust or not powerful (Table 1). Robust tests had type I error rates close to 5%, while non-robust tests frequently had type I error rates approaching 100%. Powerful tests generally correctly rejected the null hypothesis at rates exceeding 30%, while tests lacking power always correctly rejected the null hypothesis at rates below 15%. Overall, for all six data sets, the NMTPAs required assumptions of site or species equivalency to have controlled type I error rates and adequate power. The latter properties are crucial for NMTPAs because otherwise, they will frequently indicate effects of interspecific interactions (non-independent distributions) when they are absent, or not indicate effects when they are present, respectively (Gotelli 2000, Ladau 2008).

Are assumptions of site or species equivalency appropriate for the published data sets? Site equivalency means that unconditional on the occurrence of other species, each species is equally likely to occur at all sites. The data sets tend to cover broad geographic areas, such as the islands of Lake Guri, a reservoir with an area of 4300 km<sup>2</sup> (Feeley 2003), contain sites with differing habitats, and with one exception (Gotelli and Ellison 2002), contain sites with differing areas. Thus, while we are not familiar with all of the details of these systems, it appears that species are likely to differ in their probabilities of occurring at different sites. In the original studies, the authors and investigators may have had this in mind, but it was not explicitly included in the test assumptions. As a case in point, Gotelli and Ellison (2002) compared sites between types (forest, bog) for ant species assemblages. Thus, within each site type, there was the opportunity to include assumptions of site equivalency.

Species equivalency means that unconditional on the occurrence of other species, all species have the same probability of occurring at each site. Gotelli and Ellison (2002) used the similarity in habitat use by ant species as a premise for expecting to find competition; in this way, an assumption of species equivalence may have been appropriate for their data. However, for most assemblages, it is somewhat counter-intuitive to assume species equivalence because species tend to differ in their dispersal abilities and habitat requirements, even at restricted taxonomic scales.

Without detailed information about the sites and specific information about habitat affinities, dispersal abilities, and distributions of the species, it is difficult to assess whether

Table 1. Valid tests. Tests listed without brackets are robust, unbiased and powerful. Those in brackets are robust and powerful, but biased. Tests that are non-robust or not powerful are not listed. See the text for a complete list of the tests considered.

Presence–absence matrix	Assumptions			
	Sites and species equivalent	Sites equivalent	Species equivalent	None
Greater Antillean lizards (6 × 9)	—	$\delta_{FE}$	[ $\delta_{FE}$ ]	—
United Kingdom birds (8 × 26)	$\delta_{EE}$ , $\delta_{FE}$ [ $\delta_{FF}$ ]	$\delta_{FE}$ [ $\delta_{EE}$ , $\delta_{EF}$ , $\delta_{FF}$ ]	$\delta_{FF}$ [ $\delta_{EE}$ , $\delta_{FE}$ ]	—
New England ants (24 × 22)	$\delta_{EE}$ , $\delta_{FE}$	$\delta_{FE}$ [ $\delta_{EE}$ ]	[ $\delta_{EE}$ , $\delta_{FE}$ ]	—
Venezuelan birds (43 × 25)	$\delta_{FE}$ [ $\delta_{EE}$ ]	$\delta_{FE}$ [ $\delta_{EE}$ ]	[ $\delta_{EE}$ , $\delta_{FE}$ ]	—
New Hebrides birds (56 × 28)	$\delta_{FE}$ [ $\delta_{EE}$ ]	$\delta_{FE}$ [ $\delta_{EE}$ ]	[ $\delta_{EE}$ , $\delta_{FE}$ ]	—
West Indian bats (59 × 25)	$\delta_{FE}$ [ $\delta_{EE}$ ]	$\delta_{FE}$ [ $\delta_{EE}$ ]	[ $\delta_{EE}$ , $\delta_{FE}$ ]	—

site or species equivalency can be assumed. Although the NMTPAs examined here require such detailed information to be validated, analyses using different null hypotheses (e.g.  $H_{02}$  or  $H_{03}$ ) may require less information (Ladau 2008). Moreover, an NMTPA has recently been developed that can be proven robust and uniformly most powerful under generally reasonable assumptions, and this can provide a useful alternative in situations in which assumptions cannot be justified (Ladau and Schwager 2008).

The present results, coupled with previous findings, demonstrate the need for assessing the validity of NMTPAs on a case-by-case basis. Generally, in the context of null model testing, size, power, robustness, and bias are closely related to type I and II error rates. Since NMTPAs were introduced in community ecology, type I and II error rates have figured prominently. Indeed, one of the original motivations for introducing null model tests into community ecology was to control the rate of inferring effects of competition when they are absent (Connor and Simberloff 1979, 1986). The current widespread usage of NMTPAs has resulted largely from Gotelli (2000), an evaluation of the type I and II error rates of 36 NMTPAs. Nonetheless, examinations of type I and II error rates have often yielded conflicting conclusions. These conflicts often stem from the fact that the error rates depend strongly on what null hypothesis is being tested, what assumptions are made, and the dimensions of the presence-absence matrix being analyzed (Ladau 2008). The present results corroborate these findings. For instance,  $\delta_{EE}$  was robust, powerful, and unbiased only for the United Kingdom birds and New England ants matrices when all sites and species were assumed equivalent, and  $\delta_{FF}$  only had these properties for the United Kingdom birds matrix when species were assumed equivalent. These findings strongly support the argument for evaluating the applicability of NMTPAs contingent on the data being analyzed, with explicit consideration of assumptions. We provide the user-friendly software, MPower (Appendix 1) to conduct these evaluations.

Although there is a great deal of inconsistency in the appropriateness of tests, some generalizations can be made. If the null hypothesis is independent assembly of species (no effects of interspecific interactions) and if sites or species can be assumed equivalent, then  $\delta_{EF}$  and  $\delta_{FE}$  can be proven robust, respectively (Ladau 2008). Moreover, for all of the matrices examined here,  $\delta_{FE}$  was robust, unbiased, and powerful when sites were assumed equivalent. We hasten to add, however, that we are unaware of any analytical proof of why  $\delta_{FE}$  should have power under this assumption, and although this pattern is suggestive, it is by no means a justification for applying  $\delta_{FE}$  to other matrices. Rather, it suggests that  $\delta_{FE}$  should be one of the first tests evaluated prior to an analysis. For different null hypotheses, different generalizations may apply.

In conclusion, NMTPAs occupy a central position in ecology: they allow theories about the ecological mechanisms of community assembly to be tested with non-experimental data. This capability is important for understanding community assembly when experimentation is impossible, for instance at large spatial and temporal scales. However, the inferences from NMTPAs depend directly on their statistical properties: tests that are non-robust, biased, or without power can provide highly misleading conclusions, which

are less often correct than incorrect. As we confirm here, these properties need to be assessed on a case-by case basis, depending on the sample size, null hypothesis of interest, and assumptions that are appropriate for the data set that is being analyzed. MPower allows ecologists to assess these properties, ensuring that the NMTPAs that they use are reliable. In addition, clearly mapping out, a priori, what mechanisms are thought to be acting, and which assumptions could be considered can give the investigator more power. MPower quite literally empowers the user in these regards.

*Acknowledgements* – This work was supported by a Santa Fe Institute Postdoctoral Fellowship and Gordon and Betty Moore Foundation Grant (to JL), and an NSF Biological Informatics Fellowship (to SJR). This work was conducted while SJR was a Postdoctoral Associate at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant EF-0553768), the Univ. of California, Santa Barbara, and the State of California. We thank Jennifer Dunne, Jon Wilkins, Lauren Buckley and Steven Schwager for suggesting useful improvements to MPower.

## References

- Beissinger, S. R. et al. 1996. Null models for assessing ecosystem conservation priorities: threatened birds as titers of threatened ecosystems in South America. – *Conserv. Biol.* 10: 1343–1352.
- Casella, G. and Berger, R. L. 2002. *Statistical inference* (2nd ed.). – Duxbury Press, Pacific Grove.
- Chaves, L. F. and Anez, N. 2004. Species co-occurrence and feeding behavior in sand fly transmission of American cutaneous leishmaniasis in western Venezuela. – *Acta Trop.* 92: 219–224.
- Connor, E. F. and Simberloff, D. 1979. The assembly of species communities: chance or competition? – *Ecology* 60: 1132–1140.
- Connor, E. F. and Simberloff, D. 1983. Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. – *Oikos* 41: 455–465.
- Connor, E. F. and Simberloff, D. 1986. Competition, scientific method, and null models in ecology. – *Am. Sci.* 74: 155–162.
- Feeley, K. 2003. Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. – *Oecologia* 137: 104–113.
- Fox, B. J. 1987. Species assembly and the evolution of community structure. – *Evol. Ecol.* 1: 210–213.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – *Ecology* 81: 2606–2621.
- Gotelli, N. J. and Graves, G. R. 1996. *Null models in ecology*. – Smithsonian Inst.
- Gotelli, N. J. and Ellison, A. M. 2002. Assembly rules for New England ant assemblages. – *Oikos* 99: 591–599.
- Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. – *Ecology* 83: 2091–2096.
- Gotelli, N. J. and Rohde, K. 2002. Co-occurrence of ectoparasites of marine fishes: a null model analysis. – *Ecol. Lett.* 5: 86–94.
- Gotelli, N. J. and Entsminger, G. L. 2006. *EcoSim: null models software for ecology*. Ver. 7. – Acquired Intelligence Inc. and Kesey-Bear. Jericho, VT 05465. <<http://garyentsminger.com/ecosim.htm>>
- Haefner, J. W. 1988. Assembly rules for greater Antillean *Anolis* lizards: competition and random models compared. – *Oecologia* 74: 551–565.

- Heino, J. and Soininen, J. 2005. Assembly rules and community models for unicellular organisms: patterns in diatoms of boreal streams. – *Freshwater Biol.* 50: 567–577.
- Ladau, J. 2008. Validation of null model tests using Neyman-Pearson hypothesis testing theory. – *Theor. Ecol.* 1: 241–248.
- Ladau, J. and Schwager, S. 2008. Robust hypothesis tests for independence in community assembly. – *J. Math. Biol.* 57: 537–555.
- Lehmann, E. L. and Romano, J. P. 2005. Testing statistical hypotheses (3rd ed.). – Springer.
- Mouillot, D. et al. 2005. Richness, structure and functioning in metazoan parasite communities. – *Oikos* 109: 447–460.

- Reed, T. M. 1980. Turnover frequency in island birds. – *J. Biogeogr.* 7: 329–335.
- Ribichich, A. M. 2005. From null community to non-randomly structured actual plant assemblages: parsimony analysis of species co-occurrences. – *Ecography* 28: 88–98.
- Sanders, N. J. et al. 2003. Community disassembly by an invasive species. – *Proc. Natl Acad. Sci. USA* 100: 2474–2477.
- Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. – *Oecologia* 85: 74–79.
- Wilson, J. B. 1987. Methods for detecting nonrandomness in species co-occurrences: a contribution. – *Oecologia* 73: 579–582.
- Zar, J. H. 1999. Biostatistical analysis. – Prentice Hall.

## Appendix 1

### MPower software package

#### Purpose of MPower

MPower is a user-friendly software package for evaluating the size, power, robustness and bias of NMTPAs contingent on the sample size, null hypothesis, and assumptions that are input by the user.

#### Null model tests

All NMTPAs are statistical hypothesis tests. For every possible set of observed data, they are rules that prescribe an action of either rejecting or not rejecting the null hypothesis (Lehmann and Romano 2005). Many NMTPAs are implemented in EcoSim (Gotelli and Entsminger 2006), a widely used ecological modeling computer program. In EcoSim, an NMTPA is selected by choosing a row and column constraint, co-occurrence index, iterations option,

randomization algorithm, and degenerate matrix option. MPower operates in conjunction with EcoSim (*Program architecture*, below), and can evaluate all of the tests implemented in the Co-occurrence module of EcoSim (Fig. 1).

#### User inputs

MPower allows the null hypotheses and assumptions discussed in the text to be specified. In addition, MPower allows input of the dimensions of the presence–absence matrix (Fig. 1).

#### Program distribution and architecture

A free, compiled version of MPower is available at <www.santafe.edu/~jladau/MPower>. Source code is available upon request from the authors.

To make calculations, MPower begins by simulating data sets contingent on the null hypothesis, assumptions, and data attributes that are selected by the user. For example, if the

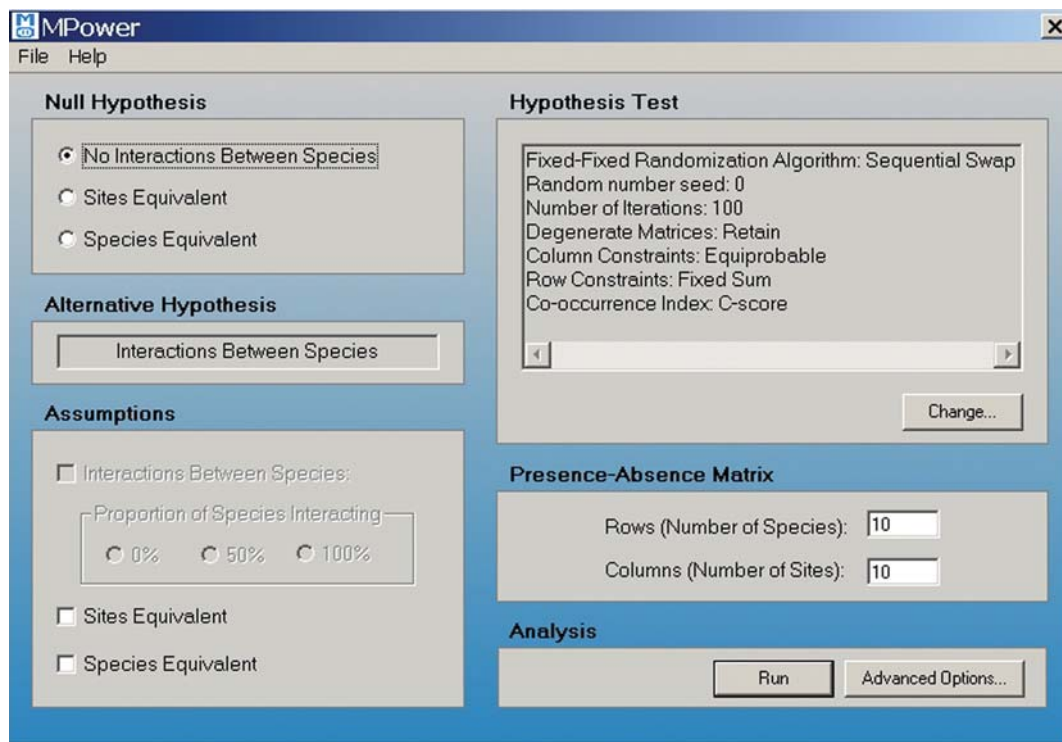


Figure 1. MPower user interface. Contingent on the selected null hypothesis, assumptions, and presence–absence matrix dimensions, MPower assesses the validity of the selected NMTPA (hypothesis test).

null hypothesis of no interactions was selected, sites and species were assumed equivalent, and the size of the matrix was  $10 \times 10$ , MPower would begin by simulating  $10 \times 10$  matrices from nine probability distributions. Because both sites and species were assumed equivalent, in each distribution all occurrence probabilities would be equal; the probability that species  $i$  occurs at site  $j$  would equal the probability that species  $k$  occurs at site  $l$ . (Had other assumptions been selected, distributions having varying probabilities would be used, in addition to these distributions). Moreover, some of the distributions would be elements of the null hypothesis, while others would be elements of the alternative hypothesis. In the example, this means that in some distributions species would occur independently, while in others they would occur non-independently.

MPower next sends each simulated matrix to EcoSim (Gotelli and Entsminger 2006). MPower directs EcoSim to perform the NMTPA that the user has selected, and it

records the result that EcoSim yields – i.e. to reject, or fail to reject the null hypothesis. For each distribution, MPower records the rate with which the null hypothesis is rejected. If a distribution is an element of the null hypothesis, the rate is an estimate of a type I error rate (false positive rate), while if it is an element of the alternative hypothesis, the rate is an estimate of power (true positive rate). These estimates are used to calculate the size, power, robustness, and bias of the NMTPA. In particular, the size is calculated as the maximum observed type I error rate. With regard to power, the maximum and minimum probabilities of rejecting the null hypothesis over the alternative hypothesis are calculated. A test is concluded non-robust if its size exceeds 0.15 (three times the nominal size of 0.05), and biased if its minimum power is below half its size. Additional algorithmic and mathematical details are available in the MPower online help file and in Ladau (2008).