Biogeographical Analysis of Chemical Co-Occurrence Data to Identify Priorities for Mixtures Research

Rogelio Tornero-Velez[1*], Peter Egeghy[1], Elaine A. Cohen-Hubal[2]

[1] National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA tornero-velez.rogelio@epa.gov, egeghy.peter@epa.gov

[2] National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA hubal.elaine@epa.gov

Address correspondence to:

Rogelio Tornero-Velez, PhD
U.S. Environmental Protection Agency
National Exposure Research Laboratory
Mail Drop E205-01, 109 TW Alexander Drive
Research Triangle Park, NC 27709
919-541-9447
919-541-1111 FAX
tornero-velez.rogelio@epa.gov

ABSTRACT

A challenge with multiple chemical risk assessment is the need to consider the joint behavior of chemicals in mixtures. To address this need, pharmacologists and toxicologists have developed methods over the years to evaluate and test chemical interaction. In practice, however, testing of chemical interaction more often comprises *ad hoc* binary combinations and rarely examines higher order combinations. One explanation for this practice is the belief that there are simply too many possible combinations of chemicals to consider. Indeed, under stochastic conditions the possible number of chemical combinations scales geometrically as the pool of chemicals increases. However, the occurrence of chemicals in the environment is determined by factors, economic in part, which favor some chemicals over others. We investigate methods from the field of biogeography, originally developed to study avian species co-occurrence patterns, and adapt these approaches to examine chemical co-occurrence. These methods were applied to a national survey of pesticide residues in 168 child care centers from across the country. Our findings show that pesticide co-occurrence in the child care center was not random but highly structured, leading to the co-occurrence of specific pesticide combinations. Thus, ecological studies of species co-occurrence parallel the issue of chemical co-occurrence at specific

locations. Both are driven by processes that introduce structure in the pattern of co-occurrence. We conclude that the biogeographical tools used to determine when this structure occurs in ecological studies are relevant to evaluations of pesticide mixtures for exposure and risk assessment.

1. INTRODUCTION

The last couple of decades have seen increased legislation concerning chemical mixtures (reviewed by Monosson[1]). The Food Quality Protection Act requires the US EPA to consider the combined effect of pesticides that share a common mechanism of toxicity.[2] Similarly, 1996 amendments to the Safe Drinking Water Act require new approaches for studying the adverse effects of contaminant mixtures in drinking water.[3] Under the Clean Air Act, the EPA is directed to include information on air pollutants that 'may interact with such pollutant to produce an adverse effect on public health or welfare'.[4] Consequently, the EPA and the Agency for Toxic Substances and Disease Registry (ATSDR) have developed guidance documents that outline specific methods for risk assessment of chemical mixtures.[5-8] An important aim of these methods is to assess the nature of the chemical interaction and determine if the effect of mixtures of chemicals is predicted according to additive action.[9-14] From a regulatory standpoint additive action is a simplifying condition because it means that the response of a chemical mixture can be predicted from the individual chemical dose-response curves. On the other hand if additive action does not prevail, the interactions must be accounted for and these are potentially

numerous. The guidance documents suggest that it is beneficial to identify and test ‘mixtures of concern’.

Mixtures may be intentionally formulated, or generated by a chemical process such as combustion or drinking water chlorination, or arise coincidentally in the environment. The question is whether the apparent coincidental mixture is random or structured. Randomness increases the combinatorial possibilities, whereas structuring favors specific combinations and reduces the numbers of realized combinations. At first glance, assessing exposure to mixtures is an intimidating prospect because under stochastic conditions the number of combinations arising by chance increases in a geometric progression, 2^r, as the number of individual species r in the source pool grows. Whereas a source pool of 3 species permits 8 unique combinations, a moderately sized pool of 15 species permits an astounding 32,768 possible combinations (illustrated in Supporting Information).

If chemical co-occurrence is mediated by human actions – with underlying economic, social, and technical drivers – might these actions favor specific combinations of chemicals in the environment? Interestingly, the answer to this question may be informed by the observations of ecologists applying biogeographic approaches. Ecologists have grappled with the question of randomness in the formation of communities of animal

species for at least some forty years.[15] Specifically, how do co-occurrence patterns indicate that an observed community is not simply a chance occurrence? Two important findings came about from the observations of ecologists. First, considering all the combinations that can be formed from a group of related species, only certain ones of these combinations exist in nature .[16] In studying fungus-dwelling insects ecologists Pielou and Pielou[17] observed that many *possible* combinations of species are never realized. Secondly, species interactions exhibit characteristic “checkerboard” patterns. In studying birds in New Guinea and the broader island archipelago in which it resides, Diamond[16] observed a “checkerboard pattern” of “forbidden” species combinations suggestive of competition.

To facilitate the study of species interactions the presence–absence matrix with "null model" randomization arose as the fundamental unit of analysis.[18] Herein, we adapt this approach to study chemical co-occurrence in the environment. The question at hand is whether chemical co-occurrence is random or structured, and if structured, which chemical species associate with one another? Null model analysis addresses this question by comparing the observed matrix with a set of matrices which are obtained through Monte Carlo randomization. Constraints are imposed in the randomization in accordance with the hypothesized form of structure so that the generated matrices embody the

‘null model’. The null model is therefore biased to reflect structuring in the mapping of the species source pool to a realized community. In this study we apply the null model analysis to investigate co-occurrence patterns of chemical species (ie., pyrethroid pesticides) in a nationwide survey of child care centers,[19] and ask whether chemical co-occurrence is due to chance or to structuring processes. In addition, we have developed methods to identify and test specific combinations to help inform risk assessors of potential ‘mixtures of concern’.

2. METHODS

2.1 Pesticide Measurements

The First National Environmental Health Survey of Child Care Centers (CCC) was a collaborative study of the U.S. Department of Housing and Urban Development (HUD), the U.S. Consumer Product Safety Commission (CPSC), and the EPA.[19] The objectives of the pesticide portion of the study were to evaluate pesticide use patterns in child care centers, and to determine pesticide residue concentrations in and around child care centers.

Licensed, institutional child care centers serving children less than 6 years of age within the 48 contiguous United States were randomly selected for participation. Probability-based sampling resulted in the selection of 334 child care centers with 168 eligible centers completing the survey. Up to two classrooms and one multipurpose room where children younger than 6 years of age regularly spent time were randomly selected for sample collection. The community ecology method was applied to the set of averaged floor wipe samples (N=168).

The details of the wipe sample procedure are provided in detail by Tulve et al. 2006.[19] Briefly, floor wipe samples were collected from hard surfaces and in the same room from a fixed area of 929 cm^2. The area was wiped twice, each time with a sterile gauze containing alcohol, and the two wipe samples were

combined and extracted. A multi-residue analysis method was developed and validated for this sampling procedure. The method included 22 organophosphate pesticides, 13 synthetic pyrethroid pesticides, pyrethrins I and II, one synergist (piperonyl butoxide), and one phenyl pyrazole (fipronil). The analysis of deltamethrin by gas chromatography (GC) was unable to rule out determination of tralomethrin which degrades to deltamethrinin in the GC injection port[20] and in the environment.[21]

2.2 Null Model Analysis

Figure 1A shows a presence-absence matrix for an avian community in the Caribbean. This matrix is a subset of the West Indian Finch (WIF) matrix (see Supplemental Information for the full matrix). Figure 1B shows the analogous matrix for study of chemical co-occurrence. In both cases, rows are species and columns are sites, and the entries are the presence (1) or absence (0) of a species at a site. The matrix dimensions consist of the species source pool, specified as R distinct species (rows), and C distinct sites (columns) to be observed. The most basic property of the observed matrix is the overall species occurrence, N. The full WIF archipelago consists of 19 islands (C=19), 17 species of finch (R=17), and 55 total occurrences (N=55). Species were not distributed uniformly throughout the islands. Whereas the *Tiara Canora* finch resided

in just 1 island, the *Tiara Bicolor* populated 17 islands. This finding is not unlike the distribution of pesticides in the environment where one pesticide in a class may dominate the market.

Gotelli[15] has reviewed the origins and evolution of null model analysis and has shown how the various null models relate to one another based on the degree of structuring (Table I). Under Null 1, species are equally likely to occupy all sites and all sites are equally diverse, and therefore Null 1 is the least structured null model, requiring only that the N occurrences in the observed matrix be randomly allocated to an empty R x C matrix. Under Null 8, species are randomized in proportion to their row and column sums. Randomization under Null 8 maintains that some species are dispersed across more sites than others, and some sites exhibit more species diversity than others. Null 9 is the most structured null model, in principle similar to Null 8 yet requiring that row and column marginal totals are maintained fixed in randomization. This is accomplished through the swap procedure, described in method section 2.7. The other null models (2-7) vary in how they impose constraints on randomization with respect to sites and species: equally probable, proportional, or fixed marginal totals.

2.3 Presence-Absence Matrix

Application of the community ecology method was limited to the set of 13 pyrethroids and 2 pyrethrins (hereafter, 'pyrethroids'). Thus, a 15 x 168 matrix of 15 distinct pyrethroids assessed in 168 child care centers constituted the surface residue matrix (ng/cm^2). This matrix was converted into a presence-absence matrix (0,1 matrix) by comparison of the pyrethroid surface residue (ng/cm^2) at each child care center with the appropriate method of detection limit (MDL, ng/cm^2) for each pyrethroid. Figure 2 shows the construction of the 0,1 matrix and its column and row marginal totals. By convention, ecologist use 'sites' as column headers and 'species' as row headers. To maintain consistency with our internal databases we assigned species to columns (pyrethroids) and sites to rows (child care centers). The marginal totals are as follows: let T_j equal the column total for column j (pesticide species j), across all rows 1 to R, where x_{ij} represents the presence (1) or absence (0) of the individual pesticide species. Let S_i equal the row total for row i (child care center i), for all columns 1 to C. Thus,

$$T_j = \sum_{i=1}^{R} x_{ij} \qquad (1)$$

$$S_i = \sum_{j=1}^{C} x_{ij} \qquad (2)$$

$$N = \sum_{i}^{R} S_i = \sum_{j}^{C} T_j = \sum_{i}^{R} \sum_{j}^{C} x_{ij} \ . \qquad (3)$$

The initial probabilities, $P(X_{ij})$, of null models 1-8 in Table I are calculated using R, C, S_i, T_j, and N.

2.4 The Null Model Analysis Algorithm

Details of specific null models are described in subsequent sections. Regardless of the specific null model, the following were the main steps:

(1) The 0,1-matrix was developed as described above.

(2) The identities and frequencies of occurrence of all k-way (1-way, 2-way, 3-way, etc.) combinations in the observed 0,1-matrix M_{obs} were assessed.

(3) M_{obs} was randomized according to null model 1-9 to generate M_{sim}.

(4) Unique k-way combinations observed in M_{obs} were counted in M_{sim}.

(5) The third and fourth steps were repeated 1000 times to generate a frequency histogram for each unique k-way combination. Each histogram represents the expected distribution of occurrence of a particular k-way combination observed in M_{obs} given that the null is true.

(6) An empirical cumulative distribution function (ECDF) was developed for each k-way combination and the corresponding cumulative probability, P_{ECDF}, for M_{obs} with respect to the simulated set M_{sim} was determined. By convention, the null was accepted for values of $0.05<P_{ECDF}<0.95$.

2.5 Randomization by Null Models 1 - 8

The observed matrix was randomized according to Null 1 by first specifying an R X C matrix of zeroes. The empty matrix was filled with 1s such that there was equal probability of converting 0s to 1s for all elements of the matrix. For Null Model 1, the probability of selecting a cell is $p(X_{ij})=1/RC$ (Table I). If the matrix element was already converted to 1, another element was randomly selected. This procedure continued until a 0,1 matrix of N ones was produced, M_{sim}. This constituted 1 randomization of the observed matrix according to Null 1, and was used for Null 1-8 in accordance with the constraints for each null model (Table I). Figure 3 shows a hypothetical 5x6 observed presence-absence matrix and corresponding probabilities, $p(X_{ij})$, for matrix randomization according to Null 1, 2, and 8. For Null 1 and 8 the probabilities of selecting elements are different but sum to 1 for all elements of the matrix. One simulated matrix is

complete when ten 1s have been assigned. Null 2 is an example of a null model with a fixed margin constraint. For Null 2 the probabilities sum to 1 by row. Randomization starts with placements of 1s in the first row until three 1s have been placed (Figure 3, S1=3), and then starts with the second row until all of its placements are complete and so on. The difference between Null 2 and 4, both of which stipulate fixed row sums, is the probability of selecting a column within a given row. Column selection is equally probable for Null 2, while for Null 4 it is proportional to the column marginal totals.

2.6 Null Model 8

Two methods were used to randomize according to Null 8. In the first, the probability of converting a matrix element from 0 to 1, was based on the formula for Null 8 in Table I, $p(X_{ij})=S_iT_j/N^2$. Gotelli [15] is clear to indicate that this probability applies strictly to the first entry, and so this method is designated 'Null8'. The second approach uses a statistical fitting procedure of Navarro-Alberto and Manly[22] to derive p(Xij), designated 'Null8 NM'. These authors note that the underlying species abundance data may be modeled with a log-linear model where the expected value of abundance $E(A_{ij})$ is

given by a constant overall effect γ, a species effect δ_j , and a location effect ϑ_i .

$$\ln(A_{ui}) = \gamma + \delta_j + \vartheta_I \qquad (4)$$

Species j is absent (0) in location if and only if its quasi-abundance is 0 and is present (1) in location i if and only if its quasi-abundance is greater than 0. The A_{ij} and Y_{ij} are linked by

$$\pi_{ij} = P(Y_{ij}=1)=P(A_{ij}> 0) \qquad (5)$$

The generalized linear model for the probability of occurrence of species j on location i is given by

$$\pi_{ij} = 1-\exp(-\exp(\gamma+\delta_j+\vartheta_i)) \qquad (6)$$

Thus, there is set of γ, δ_j and ϑ_i resulting in the optimal probability matrix, $\pi_{ij,}$ for randomizing matrices according to Null 8. Filling a zero matrix with N 1s according to the probability $\pi_{ij,}$ will lead to expected values S_i and T_i which are consistent with the observed matrix. Considering each row total S_i and column total T_j is a random variable, Eqn 6 was

optimized to produce the optimal set γ, δ_j and ϑ_i resulting in π_{ij} which minimized the following objective function:

$$obj = \sum_{1}^{C} \left(T_j - \hat{T}_j\right)^2 + \sum_{1}^{R} \left(S_i - \hat{S}_i\right)^2 \qquad (7)$$

Where $\hat{T}_j$ and $\hat{S}_i$ are the column and row totals, respectively, of the simulated matrix. The matrix π_{ij} specified the correct probability for randomizing matrices under Null 8.

2.7 Randomization by Null 9

Creating a set of random matrices with identical row and column sums is a challenging problem. Filling an empty matrix with randomly placed 1s is challenging because eventually a point is reached in which any further placements violates either a row or column sum. The 'swap' algorithm provides a solution to this problem, beginning with the observed matrices as the starting point, "checkerboard" sub-matrices are identified

$$\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \text{ ('down') or } \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \text{ ('up')}$$

and swapped: a 'down' 2 x 2 sub-matrix is replaced with an 'up' sub-matrix and vice versa. It is important to note that the elements of these sub-matrices do not have to be physically adjacent. The "checkerboard" represents a species pair that does not co-occur on two sites. The first iterations of the swap

algorithm produce matrices that are similar to the observed matrix. Thus, a “burn-in” phase of 10,000 swaps was applied to lose the ‘memory’ of the observed matrix and an additional set of 10,000 swaps was used to compute ECDFs for specific combinations identified in the observed matrix.

2.8 Software

Programs were written in MATLAB (MathWorks, Inc., Natick, MA, USA) to: 1) generate a 0,1-matrix from the surface residue data; 2) search the 0,1-matrix for all k-way combinations and their frequencies of occurrence; 3) randomize matrices according to Null Models 1-9; and, 4) determine P_{ECDF}. In order to randomize according to Null8 NM, an optimization routine employing the Nelder-Mead methods was written to solve for Eqn. (6). The procedure is explained in the Supplemental Information section.

The EcoSim software package[23] was used to compute general measures of species co-occurrence and test for nonrandom patterns. EcoSim applies the same randomization of the observed matrix as used with the MATLAB programs, but with focus on general metrics instead of specific species combinations. We selected a statistic based on Diamond’s checkerboard concept, “CHECKER”[16]. This index quantifies the number of species pairs forming perfect checkerboard distributions in a presence–absence

matrix[15]; this means perfect segregation between species. We also evaluated the “COMBO” statistic, the sum of unique combinations. The raw data for the CCC study and presence-absence matrix is provided in the Supplemental Information.

3. RESULTS

Table II summarizes the general findings of species interaction in the CCC presence-absence matrix. We found 34 pairs of pyrethroids forming perfect checkerboard distributions (CHECKER=34). Randomization by Null 1 produced a mean of 7.21 perfect checkerboards distributions per simulated community (5000 simulations). If the observed matrix has a significantly higher CHECKER score than randomly generated matrices, then a substantial number of species pairs co-occur less often than by chance, suggesting structuring. Randomization by Null 8 and 9 resulted in CHECKER statistics compatible with the observed matrix, 35.9 and 34.9, respectively.

Structured data should exhibit less unique species combinations than chance data. Whereas 39 unique species combinations were observed in the CCC matrix, randomization by Null Model 1 produced a mean of 101 combinations per simulated community (5000 simulations). Randomization by Null 8 and 9 resulted in a COMBO statistic closer to the observed matrix, yet still significantly larger. Is the finding of structuring in the

CCC matrix due to an artifact arising from formulation mixtures? Of the 15 chemical species in the CCC matrix, 3 are isomeric pairs in products: *cis-/trans*-allethrin, *cis-/trans*-permethrin, and pyrethrin I and II. To avoid artifacts from these 'embedded mixtures' we lumped these compounds, resulting in 13 distinct species (CCC Matrix, Lumped), and repeated the analysis with EcoSim. We observed the same essential finding: the CHECKER and COMBO measures could not be reproduced by a chance process (Null 1). We were able to generate non-significant COMBO statistics with NULL Model 9, indicating we had removed some structure by lumping. Interestingly, the same pattern was observed with the West Indian finch (WIF) Matrix. This matrix served as a positive control for our analysis. As expected, the CHECKER statistic was higher than expected for a chance process and the COMBO statistic was lower than expected. Randomization of the WIF matrix by Null 9 rendered a non-significant CHECKER statistic, however the observed COMBO statistic could not be explained by Null 9.

These findings motivated an examination of specific combinations. Figure 4 shows the overall abundance of binary, tertiary, and higher-order combinations as percent of all centers with those combinations. Most pyrethroid mixtures in the 168 child center study were low-order, including many binary combinations (50/168). The majority of these binary

combinations reflected pairing of *cis*-and *trans*-permethrin (46/50). Although we lumped isomer pairs for the EcoSim analysis, we kept them distinct to confirm logical pairings in the identification of specific combinations. Environmental co-occurrence of *cis*- and *trans*-permethrin is expected because of their co-occurrence in the formulation, often as a 40:60 cis:trans mixture. Figure 5 provides a list of the high-order combinations realized in the Child Care Center study. None of the observed combinations conformed to a process where species are equally probable (See Table I; Null 1, 2, 7). Of the 20 observed combinations 10 conformed to processes where some species have greater presence across sites. Interestingly, the choice of site model did not matter. What mattered was whether the randomization process was proportional to species sums (Null 4, 6, 8) or whether it maintained fixed species sums (Null 3, 5, 9).

The most robust high-order combinations captured by these randomizations were the two 5-way combinations: species 4, 6, 8, 9, and 10 and species 5, 6, 8, 9, and 10). Higher order combinations (7-, 8-, and 9-way) were not consistent with any of the null models ($P_{ECDF}>0.994$). The first of the 5-way co-occurrences consists of cyfluthrin (4), cypermethrin (6), esfenvalerate (8), and cis-permethrin (9), and trans-permethrin (10), respectively. The 4-way subset (all except

esfenvalerate), occurred three times and was consistent with the same null models (Null 4, 6, 8 and Null 3, 5, 9). The second 5-way combination consisted of cyhalothrin (5), cypermethrin (6), esfenvalerate (8), and *cis*-permethrin (9), and *trans*-permethrin (10). The 4-way subset (all except cyhalothrin), was the only other 4-way combination to occur three times and was consistent with same null models. Six other 4-way combinations were consistent with same null models (Null 4,6,8 and Null 3,5,9) and shared many of the same pyrethroids.

Null model analysis was less informative when examining low-order combinations (Figure 6). Consistent with the high-order findings, none of the observed 3-way combinations conformed to a process where species are equally probable (Null 1,2,7); however the most frequently observed 3-way combination cypermethrin, *cis*-permethrin, and *trans*-permethrin(6,9,10) was simulated only by two null models specifying fixed species sums (Null 3 and 5). Curiously, the observed high frequency (14) of the 6,9,10 combination was not consistent with Null 9 (P_{ECDF}=0.9736).

Most of the single pyrethroid occurrences and half of the binary occurrences were consistent with null models 1, 7, and 2 (randomization is equally likely across species). The 9,10 paring (*cis/trans*-permethrin) was frequently observed as a

binary combination, but was not explained by any of the null models.

Although not reported in Figure 5 and 6, the method of Navarro-Alberto and Manly ("Null 8 NM") gave a similar P_{ECDF} value as that of the Gotelli ("Null 8", Table I): P_{ECDF} ($Null8_{NM}$= 0.991 Null8 + 0.012, R^2 = 0.906) for all combinations (2-way or greater), and rendered the same decision to reject or accept the null model.

4. Discussion

We draw a parallel with island biogeography in representing the set of child care centers as an ecological niche. Whereas the structuring of avian species result from competition, communalism, and other forms of interaction, the use and occurrence of consumer chemicals at specific locations (daycare centers) is structured in part by social and economic forces. The centers are licensed institutional centers and therefore share properties that may not transfer to home-based day cares, residences, or other settings. Processing the CCC pyrethroid presence-absence matrix through the EcoSim software provided evidence of a "structured community" of pyrethroids. A parallel finding was observed between the CCC pyrethroid matrix and the WIF finch matrix (a positive control): both showed greater numbers of perfect checkerboard distributions and lower numbers

of unique combinations than predicted by chance. Pesticide checkerboards could have resulted because of technical factors. Some pyrethroids are known for their 'knockdown' effectiveness, while others are optimized killing agents.[24] This may provide a technical basis for combining some pyrethroids and not others.

Null model analysis allows for investigation of the degree of structuring and follows classical statistical randomization tests.[25] By classical MacArthur competition theory, species co-occurrence is nonrandom and is less than would be expected if species occurrences were purely stochastic and independent.[26] Here we applied null model analysis to investigate specific combinations with apparent success. The Child Care Center study consisted of 168 centers across the country, thus the realized observations of the CCC can only possess a maximum of 168 unique combinations from a possible set of 32,768. Therefore, it is important to distinguish those combinations which may be spurious from those which are robust and reflective of the underlying structuring process. To achieve this, we conducted null model analysis, comparing observed findings with simulated communities. We calculated tail probabilities from the cumulative frequency of the simulated observations, which were generated according to the rules of the null model in question. By convention we accepted the null hypothesis if the $0.05 < P_{ECDF} < 0.95$. If the particular metric of interest (e.g., a specific

combination, or general measure such as COMBO) was so rare that it was not contained between the 5^{th} and 95^{th} of ranked simulated observations, then we rejected the null model. Initially, this was the aim: to discern whether structure is present and understand it. Once structuring is understood to be operative, the next question is whether particular observations are robust or spurious; that is, to which mixtures should we direct toxicity testing efforts. Of the four observed 5-way combinations, null model analysis differentiated between two 5-way combinations which were structured from two which were apparently spurious (Figure 5). The two 5-way combinations consistent with null models involving chemical species presence (proportional or fixed) contained sub-elements that tended to repeat more often in the 4-way combinations.

Null model analyses did not identify robust combinations higher order than 5-way. Whereas the theoretical distribution of unique numbers of k-way combinations has peak density for 7- and 8-way combinations (see Supplemental Information), the observed density is shifted left with a mode at 4-way combinations. This suggests that structuring forces are operative, thus reducing higher order co-occurrences. Interestingly, null model analyses did not differentiate on the type of randomization of sites, whether equally likely or involving the site occurrence totals. We had anticipated that Null models 8 and/or 9 would prevail and

thus S_i (site sums) and T_j (species sums) would both be important parameters for predicting mixtures. Our findings suggest that T_j is the critical constraint for randomization and predicting mixtures.

In a preliminary analysis, 168 surface wipe samples were sorted by total surface residue (ng/cm^2), and the relative proportions of pyrethroids computed (Figure 5). A pattern emerged at the greatest 10% of loadings indicating a relevant mixture of permethrin (50.1%), cypermethrin (27.7%), cyfluthrin (12.4%), deltamethrin (3.25%), esfenvalerate (2.65%), and cyhalothrin (1.92%). Examining the relative proportions in the top 10% most concentrated samples (Figure 7) helps draw attention to co-occurrences by reducing sample size. Notice the overlap between the most prevalent pyrethroids in the 16 most concentrated samples with the most frequent constituents of the higher order combinations (Figure 5) observed in the full CCC sample (N=168). However, co-occurrence assessment by such means is not as rigorously addressed as is the case with null model approach; that is, all identified pesticides did not co-occur in any one center.

Ecologists favor general measures of species interaction in null model analysis rather than specific species co-occurrences. The single checkerboard represents two species claiming two distinct sites in a mutually exclusive manner.[16] Whether this

pattern arises from competitive interaction or because each site is suited to host just one of the two species remains unknown. Either process represents a form of structuring; therefore, the CHECKER statistic, the total pairs of species forming perfect checkerboard distributions is an overall measure of structure. Through simulations, Gotelli (15) has investigated the Type I error rate for null model analysis using four measures of species co-occurrence, including the COMBO and CHECKER score. Null Models 2 and 9 showed the lowest Type I error rates (< 10%) for both COMBO and CHECKER. Null 8 exhibited a low Type I error rate, but only for the COMBO statistic (6%) (15). Extending these analyses to evaluate the Type I error rate for the study of specific combinations would provide a better understanding of the ability of null model analysis to identify robust mixtures of varying complexity; that is, how the test behaves with low-order and high-order combinations. The inability to adequately simulate the 9,10 binary combination by any of null models, suggest that these approaches may not be suited for identifying low-order combinations. A partial solution to this problem may be to lump isomers. Lumping of isomers for the EcoSim analysis rendered acceptance of the COMBO statistic with null model 9. However, we chose not to lump isomers as we were interested to confirm and show their logical pairing in the observed mixtures. Indeed, the 9/10 *cis-/trans*-permethrin pair co-occur in 19 of

the 20 higher-order mixtures. The 4-way mixture in which they do not co-occur contains neither *cis*- nor *trans*-permethrin (9 nor 10). Acceptance of null models involving equally probable species (Null 1, 7, and 2) for single and binary occurrences suggests a high Type II error rate (accepting the null model when it is false). In short, the confirmation of structuring process in the formation of mixtures requires analysis of higher-order combinations or general measures of species interaction (e.g., COMBO). Even so, the algorithms we have developed can be used to identify the most frequently occurring low-order combinations, even if these are not informative with regards to structuring.

In summary, ecological studies of species co-occurrence patterns parallels the issue of chemical co-occurrence at specific locations. Both are driven by processes that introduce structure in the pattern of co-occurrence. Tools have been developed to determine when this structure occurs in ecological studies and are relevant to the evaluation of pesticide mixtures. Chemical mixtures arise, in part, through non-random processes (economic factors, engineered formulations, differential degradation, etc) such that the observed set of combinations tends to be less diverse than the theoretical random set. The biogeography methods tested here with the CCC case study can be used to identify mixtures of concern to

prioritize risk assessment efforts, to calculate co-occurrence probabilities, and to test if mixtures arise by structuring processes and develop hypotheses. Structuring forces favor specific combinations, and it is those combinations that draw our attention because they have the greatest chance to be realized in the future. This is true provided that the ensemble of structuring forces is a stationary process. In reality, structuring forces shift in time with advances in technology, market factors, and social trends. This speaks to the importance of surveys such as the CCC study of Tulve et al.(19), designed to efficiently survey many sites.

ACKNOWLEDGMENTS

The U.S. Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here under MOBIS contract 23F-8144H to Westat and contract 68-D-99-011 to Battelle Memorial Institute. It has been subjected to Agency administrative review and approved for publication. This does not signify that the contents necessarily reflect the views and policies of the Agency, nor does mention of trade names or commercial products

constitute endorsement or recommendation for use. The authors are grateful to Stephanie Trebes, who participated through the EPA Student Volunteer Program, with help in checking computer algorithms. Earlier versions of this manuscript benefited from the comments of Dr. John Wambaugh and Dr. Michael Breen.

Table I. Nine null models based on the observed presence-absence matrix

	Columns Equally Likely	Column Sums Proportional	Column Sums Fixed
Rows Equally Likely	**Null 1** P(Xij)=1/RC Constraint: N	**Null 6** P(Xij)=Tj/NR Constraint: N	**Null 3** P(Xij)=1/R Constraint: Tj
Row Sums Proportional	**Null 7** P(Xij)=Si/NC Constraint: N	**Null 8** P(Xij)=SiTj/N^2 Constraint: N	**Null 5** P(Xij)=Si/N Constraint: Tj
Row Sum Fixed	**Null 2** P(Xij)=1/C Constraint: Si	**Null 4** P(Xij)=Tj/N Constraint: Si	**Null 9** P(Xij)=[Markov Process] Constraint: Si, Tj

Adapted from Gotelli[15]. Each entry specifies one of nine possible null models which impose structure on the random set of matrices, depending on the observed matrix. In the observed matrix, N is the total species occurrence, R the number of rows, C the number of columns, Si the marginal row sums, and Tj the marginal column sums. Structuring increases left-to-right and top-down, so that Null 1 is the least structured null, and Null 9 the most structured. For null model 1-8, for any given randomized matrix, P(Xij) estimates probability of occupancy of the first cell in the matrix.

Table II. Null Model Analyses on the Child Center Center and West Indian Finch Matrices

	Child Care Center (CCC) Matrix		CCC Matrix, Lumped		West Indian Finch Matrix	
	15 species x 168 sites		12 species X 168 sites		17 species X 19 sites	
	CHECKER	COMBO	CHECKER	COMBO	CHECKER	COMBO
Observed	34	39	20	35	91	10
Null 1	7.21**	101.53**	8.76**	65.61**	70.78**	18.57**
Null 8	35.91	61.91**	21.70	46.33**	51.00**	17.18**
Null 9	34.99	42.78*	19.67	35.41	89.44	15.14**

To remove the effect of structuring due to co-occurring isomers, “CCC Matrix, Lumped” combines 3 species pairs, cis-/trans-allethrin, cis-/trans-permethrin, and pyrethrins I and II into 3 single species (allethrin, permethrin, and pyrethrins). The CHECKER index is the number of species forming checker patterns, and has a theoretical range of 0 to R(R-1)/2, where R = number of rows (species). The COMBO index is the total number of unique combinations, and has a theoretical range, 0 to 2^R, yet is bounded by the number of sites (168). The expected indices are based on 5000 simulations, and ** indicates P-value<0.001, * indicates P-value <0.05.

Figures

Figure 1. A presence-absence matrix for (A)finch species and (B)chemical species. Species are specified in the rows, and sites in the columns, and the entries are the presence (1) or absence (0) of the species.

Figure 2. Creation of a presence-absence matrix from environmental surface wipe data.

Figure 3. Probability calculations for three null models (1,2, and 8) based on the formulas in Table I and the constraints of the observed matrix.

Figure 4. The distribution of mixtures for the CCC study. A k-way mixture= 2 indicates binary mixtures. Thirty percent of centers (50/168) showed binary combinations.

Figure 5. Higher-order pyrethroid combinations observed in the Child Care Center study and tested by null models 1-9. Grey shading indicates that the combination was consistent with the null model, $0.05 < P_{ECDF} < 0.95$.

Figure 6. Lower-order pyrethroid combinations observed in the Child Care Center study and tested by null models 1-9. Grey shading indicates that the combination was consistent the null model, $0.05< P_{ECDF} <0.95$.

Figure 7. Relative proportions of pyrethroids in surface wipe samples of the CCC study for all 168 centers and for those centers with the highest 25% and 10% of total surface residues (ng/cm^2).

Figure 1

A

	Cuba	Hispanolia	Jamaica	Puerto Rico	Guadeloupe	
Loxigilla noxis	0	0	0	0	1	**1**
Melanospiza richardsoni	0	0	0	0	0	**0**
Tiara olivacea	1	1	1	1	0	**4**
Tiara bicolor	0	1	1	1	1	**4**
Tiara canora	1	0	0	0	0	**1**
Loxipasser anoxanthus	0	0	1	0	0	**1**
	2	**2**	**3**	**2**	**2**	**11**

B

	site 1	site 2	site 3	site 4	site 5	
chemical 1	1	0	0	0	1	**2**
chemical 2	1	1	0	1	1	**4**
chemical 3	0	0	0	0	1	**1**
chemical 4	1	0	0	0	0	**1**
chemical 5	0	0	1	0	0	**1**
chemical 6	0	0	1	0	0	**1**
	3	**1**	**2**	**1**	**3**	**10**

Figure 2

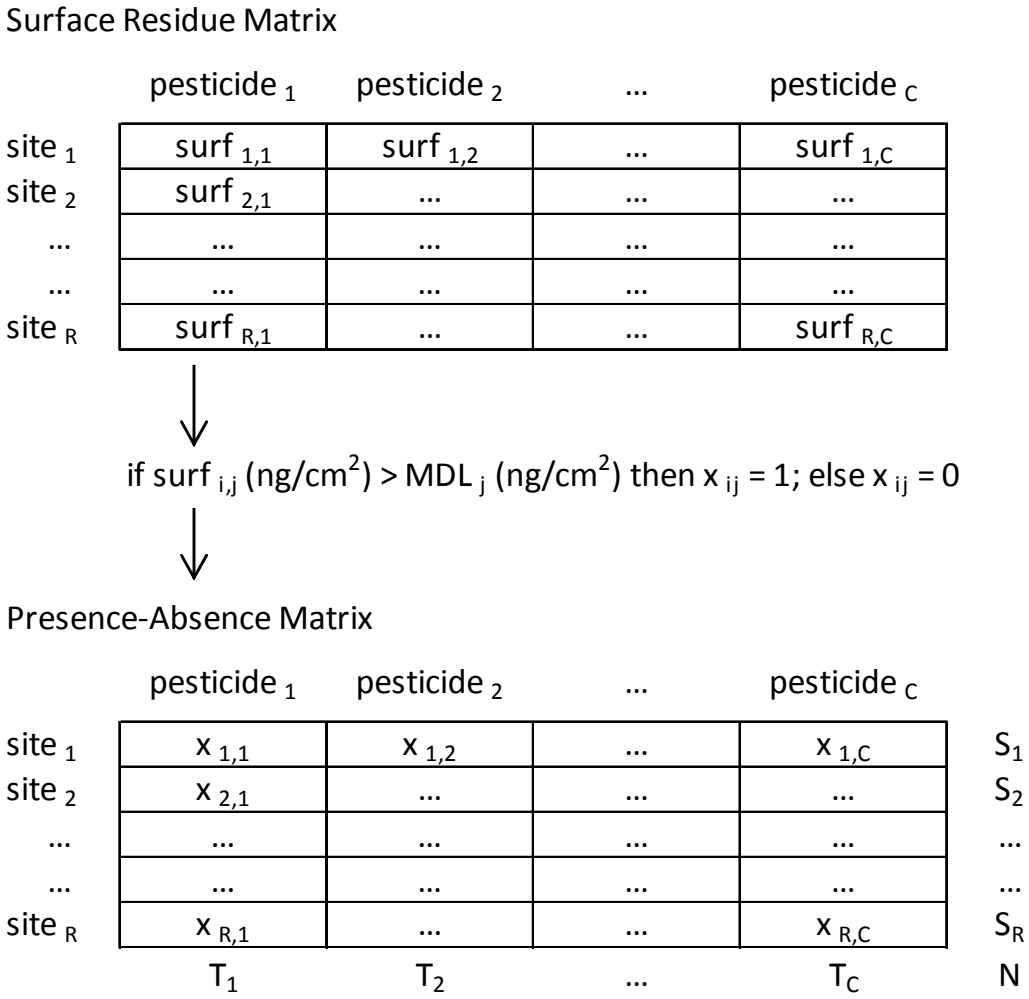

Surface Residue Matrix
pesticide 1
pesticide 2
…
pesticide C
site 1
site 2
…
…
site R
surf 1,1
surf 1,2
surf 1,C
surf 2,1
surf R,1
surf R,C
if surf i,j (ng/cm2) > MDL j (ng/cm2) then x ij = 1; else x ij = 0
Presence-Absence Matrix
pesticide 1
pesticide 2
…
pesticide C
site 1
site 2
…
…
site R
x 1,1
x 1,2
x 1,C
x 2,1
x R,1
x R,C
S1
S2
SR
T1
T2
TC
N

Figure 3

Observed Matrix:

Sites:	Species:						Row Total:
	1	1	0	1	0	0	3
	0	1	0	0	0	0	1
	0	0	0	0	1	1	2
	0	1	0	0	0	0	1
	1	1	1	0	0	0	3
Column Totals	2	4	1	1	1	1	10

Null Model 1

$pr(X_{ij}) = 1/RC =$

0.033	0.033	0.033	0.033	0.033	0.033
0.033	0.033	0.033	0.033	0.033	0.033
0.033	0.033	0.033	0.033	0.033	0.033
0.033	0.033	0.033	0.033	0.033	0.033
0.033	0.033	0.033	0.033	0.033	0.033

Null Model 8

$pr(X_{ij}) = S_i T_j / N^2 =$

0.060	0.120	0.030	0.030	0.030	0.030
0.020	0.040	0.010	0.010	0.010	0.010
0.040	0.080	0.020	0.020	0.020	0.020
0.020	0.040	0.010	0.010	0.010	0.010
0.060	0.120	0.030	0.030	0.030	0.030

$N=10 \rightarrow \sum\sum pr(Xij) = 1$

$N=10 \rightarrow \sum\sum pr(Xij) = 1$

Null Model 2

$pr(X_{ij}) = 1/C =$

0.167	0.167	0.167	0.167	0.167	0.167	$S_1=3 \rightarrow \sum pr(X_{1j}) = 1$
0.167	0.167	0.167	0.167	0.167	0.167	$S_2=1 \rightarrow \sum pr(X_{2j}) = 1$
0.167	0.167	0.167	0.167	0.167	0.167	$S_3=2 \rightarrow \sum pr(X_{3j}) = 1$
0.167	0.167	0.167	0.167	0.167	0.167	$S_4=1 \rightarrow \sum pr(X_{4j}) = 1$
0.167	0.167	0.167	0.167	0.167	0.167	$S_5=3 \rightarrow \sum pr(X_{5j}) = 1$

Figure 4

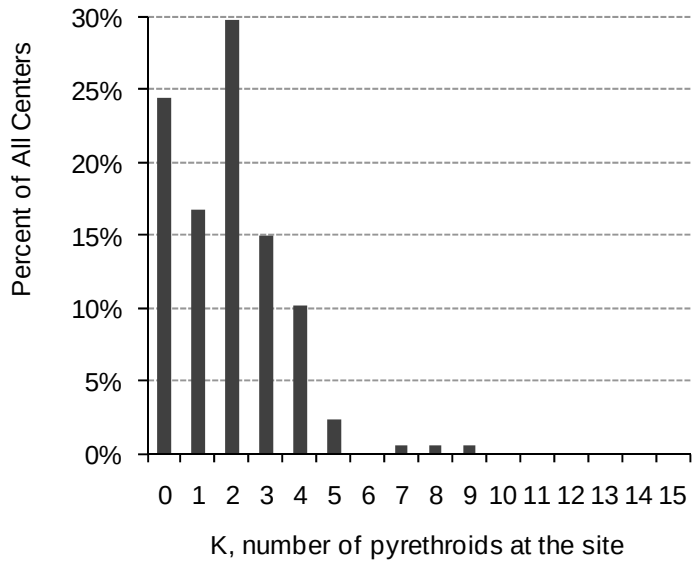

30%
25%
20%
15%
10%
5%
0%
Percent of All Centers
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
K, number of pyrethroids at the site

Figure 5

Pyrethroid Key				
1 *cis* -allethrin	3 bifenthrin	9 *cis* -permethrin	11 pyrethrin I	13 resemethrin
2 *trans* -allethrin	4 cyfluthrin	10 *trans* -permethrin	12 pyrethrin II	14 sumithrin
	5 cyhalothrin			15 tetramethrin
	6 cypermethrin			
	7 deltamethin			
	8 esfenvalerate			

Observed 4-way		Null1	Null7	Null2	Null6	Null8	Null4	Null3	Null5	Null9
		Species Equilikely			Species Sums Proportional			Species Sums Fixed		
Species	Counts	Sites E	Sites P	Sites SF	Sites E	Sites SP	Sites SF	Sites E	Sites P	Sites SF
4, 6, 9, 10	3	0.9995	0.9995	0.9995	0.9125	0.8660	0.8985	0.9085	0.8505	0.79730
6, 8, 9, 10	3	0.9995	0.9995	0.9995	0.9305	0.9150	0.8950	0.9265	0.8820	0.83135
1, 2, 9, 10	1	0.9925	0.9905	0.9915	0.9885	0.9855	0.9870	0.9875	0.9920	0.99605
3, 6, 9, 10	1	0.9940	0.9960	0.9910	0.6100	0.5615	0.5700	0.5685	0.5275	0.44865
4, 5, 6, 7	1	0.9950	0.9935	0.9935	0.9995	0.9975	0.9990	0.9995	0.9995	0.99995
4, 5, 9, 10	1	0.9965	0.9920	0.9915	0.8135	0.7965	0.7950	0.8205	0.8045	0.79175
4, 8, 9, 10	1	0.9900	0.9955	0.9945	0.8410	0.8210	0.8265	0.8230	0.8190	0.84330
5, 6, 9, 10	1	0.9910	0.9945	0.9915	0.4475	0.4075	0.4015	0.4035	0.3340	0.27160
5, 7, 9, 10	1	0.9920	0.9930	0.9955	0.9535	0.9465	0.9570	0.9555	0.9515	0.96110
5, 9, 10, 12	1	0.9950	0.9930	0.9955	0.9125	0.8920	0.8900	0.8850	0.8930	0.92130
6, 9, 10, 14	1	0.9920	0.9940	0.9915	0.8025	0.7685	0.7835	0.7850	0.7690	0.69890
9, 10, 12, 13	1	0.9925	0.9935	0.9945	0.9695	0.9615	0.9625	0.9685	0.9615	0.96645
9, 10, 12, 15	1	0.9950	0.9935	0.9950	0.9710	0.9750	0.9710	0.9730	0.9745	0.98515

Observed 5-way		Null1	Null7	Null2	Null6	Null8	Null4	Null3	Null5	Null9
		Species Equilikely			Species Sums Proportional			Species Sums Fixed		
Species	Counts	Sites E	Sites P	Sites SF	Sites E	Sites SP	Sites SF	Sites E	Sites P	Sites SF
3, 6, 9, 10, 13	1	0.9980	0.9985	0.9995	0.9850	0.9600	0.9790	0.9820	0.9760	0.98820
4, 6, 8, 9, 10	1	0.9990	0.9985	0.9995	0.9075	0.8410	0.9095	0.9320	0.8750	0.91740
5, 6, 8, 9, 10	1	0.9965	0.9990	0.9970	0.9190	0.8545	0.9125	0.9405	0.8895	0.90125
9, 10, 11, 14, 15	1	0.9980	0.9975	0.9980	0.9990	0.9995	0.9990	0.9995	0.9995	0.99995

Observed 7, 8, and 9-way		Null1	Null7	Null2	Null6	Null8	Null4	Null3	Null5	Null9
		Species Equilikely			Species Sums Proportional			Species Sums Fixed		
Species	Counts	Sites E	Sites P	Sites SF	Sites E	Sites SP	Sites SF	Sites E	Sites P	Sites SF
1,2,3,4,6,9,10	1	0.9995	0.9995	0.9990	0.9995	0.9985	0.9995	0.9995	0.9995	0.99995
3,4,5,6,7,8,9,10	1	0.9995	0.9995	0.9995	0.9995	0.9990	0.9945	0.9995	0.9990	0.99565
1,2,5,6,8,9,10,14,15	1	0.9995	0.9995	0.9995	0.9995	0.9995	0.9990	0.9995	0.9995	0.99999

Figure 6

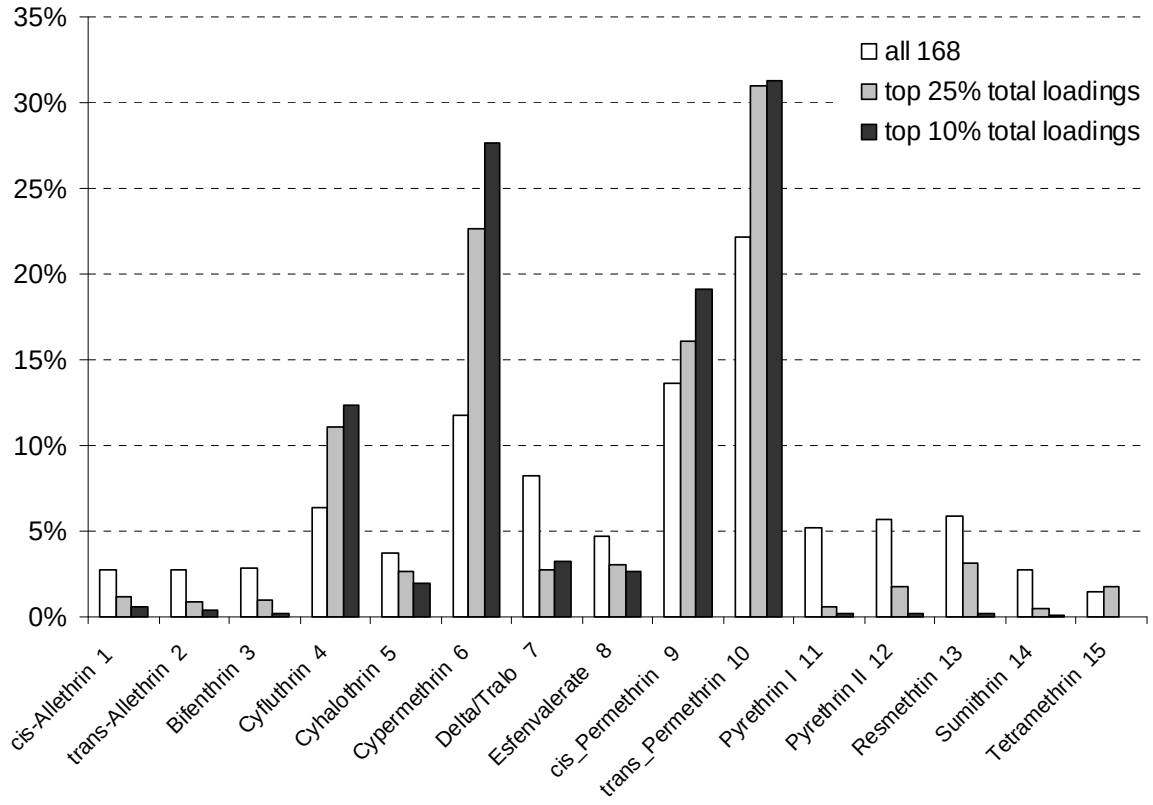
Pyrethroid Key				
1 *cis* -allethrin	3 bifenthrin	9 *cis* -permethrin	11 pyrethrin I	13 resemethrin
2 *trans* -allethrin	4 cyfluthrin	10 *trans* -permethrin	12 pyrethrin II	14 sumithrin
	5 cyhalothrin			15 tetramethrin
	6 cypermethrin			
	7 deltamethin			
	8 esfenvalerate			

		Null1	Null7	Null2	Null6	Null8	Null4	Null3	Null5	Null9
Observed Singles		Species Equilikely			Species Sums Proportional			Species Sums Fixed		
Species	Counts	Sites E	Sites P	Sites SF	Sites E	Sites P	Sites SF	Sites E	Sites P	Sites SF
10	13	0.9995	0.9995	0.9995	0.9490	0.8915	0.9140	0.9650	0.9045	0.1562
9	7	0.9985	0.9965	0.9985	0.4100	0.2610	0.1650	0.5815	0.4020	0.0013
4	3	0.8290	0.7990	0.8045	0.9925	0.9865	0.9490	0.9995	0.9990	0.9998
6	2	0.6510	0.5780	0.5765	0.6110	0.4625	0.2840	0.9245	0.9085	0.8926
5	1	0.3475	0.2845	0.2775	0.7425	0.6545	0.5235	0.9155	0.8910	0.8979
12	1	0.3140	0.3290	0.2685	0.8505	0.8145	0.7020	0.9600	0.9505	0.9516
13	1	0.3535	0.3240	0.2795	0.9055	0.8900	0.8325	0.9725	0.9630	0.9600

		Null1	Null7	Null2	Null6	Null8	Null4	Null3	Null5	Null9
Observed Binary		Species Equilikely			Species Sums Proportional			Species Sums Fixed		
Species	Counts	Sites E	Sites P	Sites SF	Sites E	Sites SP	Sites SF	Sites E	Sites P	Sites SF
9, 10	46	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9980	0.9995	1.0000
3, 6	2	0.9775	0.9850	0.9565	0.9940	0.9945	0.9785	0.9990	0.9990	0.9998
6, 9	1	0.8300	0.8650	0.7830	0.0605	0.1015	0.0230	0.1550	0.3065	0.1688
8, 9	1	0.8210	0.8765	0.7660	0.5250	0.5870	0.3955	0.7090	0.7860	0.8028

		Null1	Null7	Null2	Null6	Null8	Null4	Null3	Null5	Null9
Observed 3-way		Species Equilikely			Species Sums Proportional			Species Sums Fixed		
Species	Counts	Sites E	Sites P	Sites SF	Sites E	Sites SP	Sites SF	Sites E	Sites P	Sites SF
6, 9, 10	14	0.9995	0.9995	0.9995	0.9580	0.9820	0.9995	0.6200	0.6485	0.9736
5, 9, 10	3	0.9995	0.9995	0.9995	0.5815	0.7145	0.8130	0.3520	0.4015	0.7315
9, 10, 12	3	0.9995	0.9995	0.9995	0.8650	0.8990	0.9475	0.7890	0.8285	0.9545
3, 9, 12	2	0.9995	0.9995	0.9990	0.6130	0.7015	0.7685	0.4475	0.4955	0.7072
8, 9, 10	1	0.9640	0.9705	0.9695	0.1780	0.2605	0.3495	0.0575	0.0865	0.1780
8, 10, 14	1	0.9600	0.9730	0.9715	0.9760	0.9795	0.9820	0.9845	0.9885	0.9943
9, 10, 13	1	0.9625	0.9725	0.9720	0.6135	0.6740	0.7225	0.4575	0.5205	0.7356

Figure 7


35%
30%
25%
20%
15%
10%
5%
0%
all 168
top 25% total loadings
top 10% total loadings
cis-Allethrin 1
trans-Allethrin 2
Bifenthrin 3
Cyfluthrin 4
Cyhalothrin 5
Cypermethrin 6
Delta/Tralo 7
Esfenvalerate 8
cis_Permethrin 9
trans_Permethrin 10
Pyrethrin I 11
Pyrethrin II 12
Resmethtin 13
Sumithrin 14
Tetramethrin 15

Literature Cited

1. Monosson E. Chemical mixtures: considering the evolution of toxicology and chemical assessment. Environmental Health Perspectives, 2005; 113(4):383-390.
2. U.S. Congress. Food Quality Protection Act. Public Law 104-170. Washington: U.S. Government Printing Office, 3 August 1996.
3. U.S. Congress. Ammendments to the Safe Drinking Water Act. Public Law 104-182. Washington: U.S. Government Printing Office, 6 August 1996.
4. U.S. Congress. Ammendments to the Clean Air Act. Public Law 101-549. Washington: U.S. Government Printing Office, 15 November 1990.
5. Agency for Toxic Substances and Disease Registry (ATSDR). Guidance for the Preparation of an Interaction Profile for Toxic Substances. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, 2001.
6. Agency for Toxic Substances and Disease Registry (ATSDR). Guidance Manual for the Assessment of Joint Toxic Action of Chemical Mixtures. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, 2004.
7. U.S. EPA. Guidelines for the Health Risk Assessment of Chemical Mixtures. Washington, D.C.1986.

8. U.S. EPA. Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures. Washington, D.C.2000.
9. Bliss CI. The Toxicity oF Poisons Applied Jointly. Annals of Applied Biology, 1939;26(3):585-615.
10. Berenbaum MC. A method for testing for synergy with any number of agents. The Journal of Infectious Diseases. 1978;137(2):122-130.
11. Berenbaum MC. Synergy assessment with growth curves. The Journal of Infectious Diseases. 1984;150(2):304.
12. Gennings C, Carter WH, Jr., Campbell ED, Staniswalis JG, Martin TJ, Martin BR, White KL, Jr. Isobolographic characterization of drug interactions incorporating biological variability. The Journal of Pharmacology and Experimental Therapeutics. 1990;252(1):208-217.
13. Gennings C, Carter WH, Jr., Carchman RA, Teuschler LK, Simmons JE, Carney EW. A unifying concept for assessing toxicological interactions: changes in slope. Toxicological Sciences. 2005;88(2):287-297.
14. Wolansky MJ, Gennings C, DeVito MJ, Crofton KM. Evidence for dose-additive effects of pyrethroids on motor activity in rats. Environmental Health Perspectives. 2009;117(10):1563-1570.
15. Gotelli NJ. Null model analysis of species co-occurrence patterns. Ecology. 2000;81(9):2606-21.

16. Diamond JM. Assembly of Species Communities. In: M.L. Cody and J.M. Diamond E, editor. Ecology and Evolution of Communities. Cambridge, MA: Belkap Press; 1975. p. 342-444.
17. Pielou DP, Pielou EC. Association among species of infrequent occurrence: the insect and spider fauna of Polyporus betulinus (Bulliard) Fries. Journal of Theoretical Biology. 1968;21(2):202-216.
18. Connor EF, Simberloff D. The Assembly of Species Communities - Chance or Competition? Ecology. 1979;60(6):1132-1140.
19. Tulve NS, Jones PA, Nishioka MG, Fortmann RC, Croghan CW, Zhou JY, Fraser A, Cavel C, Friedman W. Pesticide measurements from the first national environmental health survey of child care centers using a multi-residue GC/MS analysis method. Environmental Science and Technology. 2006;40(20):6269-6274.
20. Valverde A, Aguilera A, Rodriguez M, Boulaid M. What are we determining using gas chromatographic multiresidue methods: tralomethrin or deltamethrin? Journal of Chromatography. A. 2002;943(1):101-111.
21. Takatsuki S, Nemoto S, Sasaki K, Toyoda M. [Specific determination of deltamethrin and tralomethrin by preparative HPLC and GC-ECD]. Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku. 1997(115):135-139.

22. Navarro-Alberto JA, Manly BFJ. Null model analyses of presence-absence matrices need a definition of independence. Population Ecology. 2009;51(4):505-12.
23. Gotelli, N.J. and G.L. Entsminger. 2004. EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465. http://garyentsminger.com/ecosim/index.htm, Accessed on March 15, 2011.
24. Nishizawa Y. Development of new synthetic pyrethroids. Bulletin of the World Health Organization. 1971;44(1-3):325-336.
25. Manly BF. Randomization Tests, 4th Edition by Eugene S. Edgington, Patrick Onghena. International Statistical Review. 2007;75(2):269-269.
26. Gotelli NJG, G.R. Null Models in Ecology. Washington, D.C.: Smithsonian Institution Press; 1996.

Supplemental Information

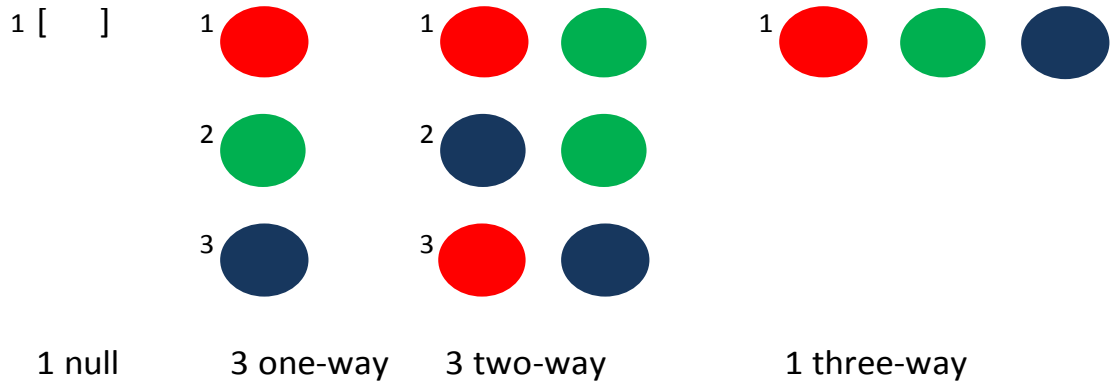


Fig. SI-1. Eight possible combinations from a source pool of three distinct species are enumerated with the binomial coefficient, C(n,k) =n!/(k!(n-k)!. Totals for the null set, 1-way, 2-way, and 3-way possibilities are, respectively, C(3,0)=1, C(3,1)=3, C(3,2)=3, C(3,3)=1. This gives a total of eight (2^r; $2^3 = 8$).

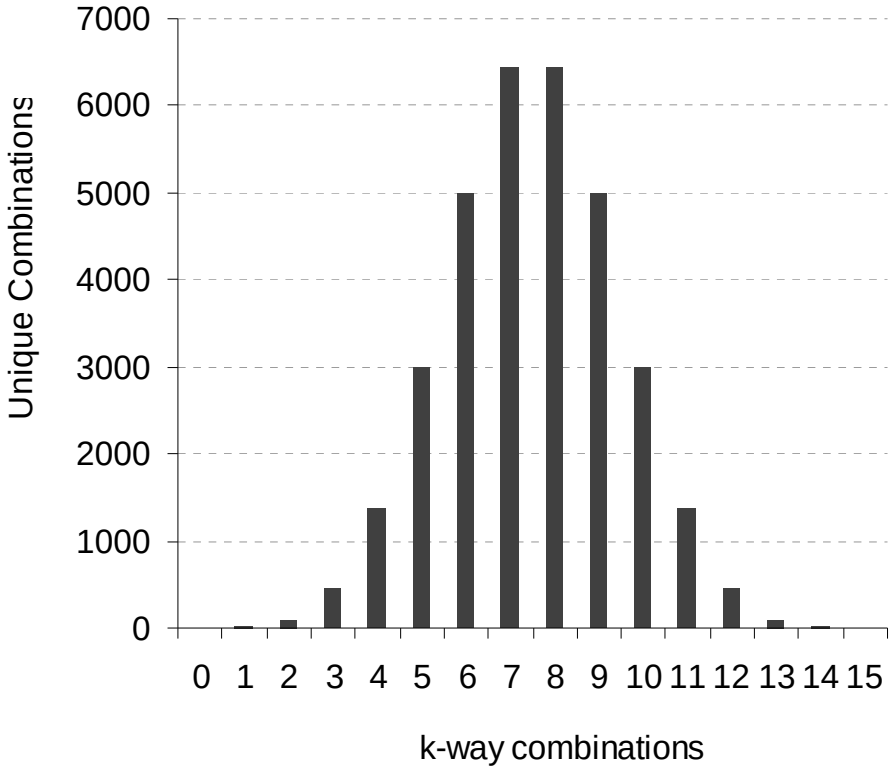


Fig. SI-2. Enumeration of the possible combinations from random sampling of moderately sized source pool of fifteen distinct chemicals is depicted with a histogram. There is a total of 32,768 (2^{15}) unique k-way combinations. The binomial coefficient, C(n,k) =n!/(k!(n-k)!), computes the outcome for the possible k-way combinations; for example, there are 105 binary (two-way) combinations C(15,2)=105.

	Cuba	Hispaniola	Jamaica	Puerto Rico	Guadeloupe	Martinique	Dominica	St Lucia	Barbados	St Vincent	Grenada	Antigua	St Croix	Grand Cayman	St Kits	Barbuda	Montserrat	St Martin	St Thomas	
Carduelis dominicensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxia leucoptera	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Volatina jacarina	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Sprophilia nigrricollis	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Melopyrrha nigra	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
Loxigilla portoricensis	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxigilla violacea	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Loxigilla noxis	0	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	12
Melanospiza richardsoni	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Tiara olivacea	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5
Tiara bicolor	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	17
Tiara canora	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxipasser anoxanthus	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Saltator albicollis	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4
Torreornis inexpectata	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ammodramus savannarum	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Zonotrichia capensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	4	7	5	4	3	3	3	4	2	2	4	2	1	2	2	2	2	2	1	55

Fig. SI-3. The West Indian Finch Matrix (Gotelli and Abele, 1982), an example a species presence-absence matrix used by ecologists.

Table SI-1

Gotteli	iterations:	1000		Navarro-Alberto and Manly	iterations:	1000	
	Total cost	28.3779			Cost	3.5187	
expected	observed	expected	observed	expected	observed	expected	observed
T	col	S	row	T		S	row
1	1.306	1	1.26	1	1.094	1	1.375
1	1.283	2	2.24	1	1.44	2	2.331
1	1.309	2	2.26	1	1.115	2	2.31
1	1.309	2	2.22	1	1.136	2	2.283
2	2.557	2	2.21	2	1.945	2	2.288
1	1.349	2	2.27	1	1.087	2	2.292
2	2.583	1	1.30	2	1.987	1	1.402
12	10.253	2	2.22	12	11.662	2	2.283
1	1.273	4	3.88	1	1.118	4	3.694
5	5.596	2	2.24	5	4.887	2	2.23
17	12.604	2	2.25	17	16.24	2	2.245
1	1.326	4	3.85	1	1.167	4	3.498
1	1.295	3	3.08	1	1.05	3	2.946
4	4.666	3	3.09	4	3.969	3	2.892
1	1.327	3	3.05	1	1.084	3	3.009
3	3.632	4	3.71	3	2.913	4	3.682
1	1.332	5	4.45	1	1.106	5	4.225
		7	5.55			7	6.479
		4	3.89			4	3.536
	25.1972		3.5708		1.0207		2.498

Table SI-1 compares the simulation of the WIF matrix with the two methods for null model 8. Specifically, how accurately does either method estimate the row and column marginal totals? Randomizing under Null 8G1 was accomplished using the formula $p(Xij)=T_jS_i/N^2$. Randomization under Null 8NM required estimation of $P(X_{ij})= 1-\exp(-\exp(\gamma+\delta_j+\vartheta_i))$. Nelder Mead method was applied to find the best set of parameters γ, δ_j and ϑ_I minimizing the objective function:

$$obj = \sum_{1}^{C} \left(T_j - \hat{T}_j \right)^2 + \sum_{1}^{R} \left(S_i - \hat{S}_i \right)^2$$

For each method Null 8G1 and 8M, 1000 matrices were created and the row and column marginal were computed for each. The means of these were compared to the marginal totals of the observed WIF matrix, and the sums of squares were computed, and totaled 28.3779. The same procedure using Null 8 NM gave a sums-of-squares of 3.5187.

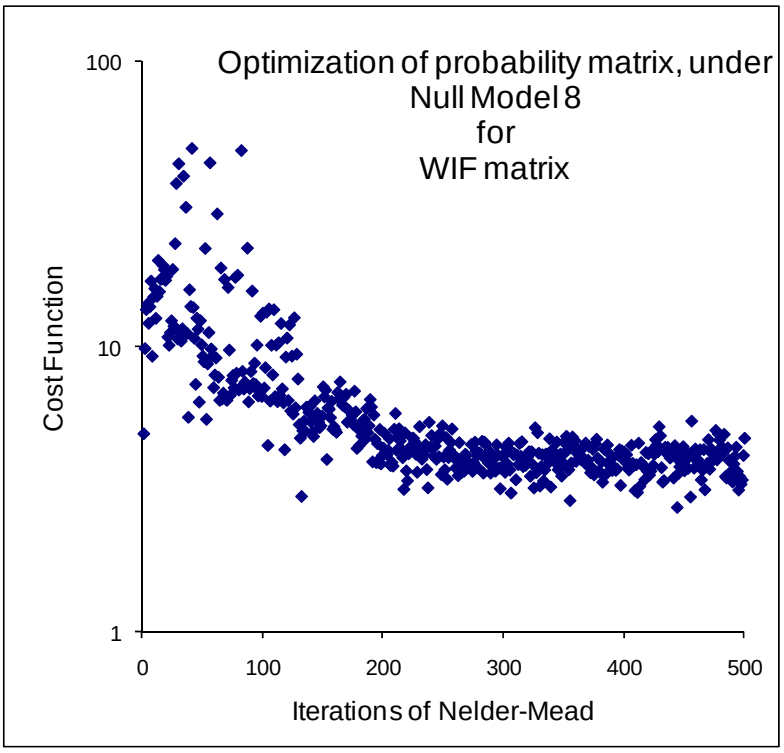


Fig. SI-3 shows that 300 iterations of Nelder-Mead are required to optimize the WIF probability matrix for Null 8 NM. The optimized probability matrix was used to randomized matrices under Null 8NM (reported in Table SI-1).