Design and Analysis of Multispecies Toxicity Tests for Pesticide Registration

Wayne G. Landis  
*Western Washington University*

Robin A. Matthews  
*Western Washington University, robin.matthews@wwu.edu*

Geoffrey B. Matthews  
*Western Washington*

Follow this and additional works at: https://cedar.wwu.edu/esci_facpubs

Part of the Environmental Sciences Commons

Recommended Citation
https://cedar.wwu.edu/esci_facpubs/11
DESIGN AND ANALYSIS OF MULTISPECIES TOXICITY TESTS FOR PESTICIDE REGISTRATION

WAYNE G. LANDIS,1 ROBIN A. MATTHEWS,2 AND GEOFREY B. MATTHEWS3

1Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies, Western Washington University, Bellingham, Washington 98225 USA
2Institute for Watershed Studies, Huxley College of Environmental Studies, Western Washington University, Bellingham, Washington 98225 USA
3Computer Science Department, Western Washington University, Bellingham, Washington 98225 USA

Abstract. The community conditioning hypothesis describes ecological structures as historical, nonequilibrium, and by definition complex. Indeed, the historical nature of ecological structures is seen as the primary difference between single-species toxicity tests and multispecies test systems. Given the complex properties of ecological structures, multispecies toxicity tests need to be designed accordingly with appropriate data analysis tools. Care must be taken to ensure that each replicate shares an identical history, or divergence will rapidly occur. Attempting to realize homogeneity by linear cross inoculation or waiting for an equilibrium state to occur assumes properties that ecological structures do not have. Data analysis must also incorporate the dynamic and hyperdimensional nature of ecological structures. Univariate analysis of individual variables denies the fundamental character of ecological structures as complex systems. A variety of methods, such as correspondence analysis, nonmetric multidimensional scaling, and nonmetric clustering and association analysis, are available to search for patterns and to test their relationships to experimental treatments. Visualization techniques including Space–Time Worms and redundancy analysis are also critical in attempting to understand the dynamic nature of these structures. Reliance upon the traditional analysis methods, such as ANOVA and the estimation of LOECs (lowest observable effects concentrations) or NOECs (no observable effects concentrations), comparable to those of single-species toxicity tests, is to be blind to the unique and complex nature of multispecies toxicity tests. Fundamental design criteria for multispecies toxicity tests, data analysis, and interpretation are presented.

Key words: community conditioning; complexity; mesocosm; microcosm; multispecies toxicity test; multivariate statistics; pesticides.

INTRODUCTION

There has been a renewed interest regarding the use of multispecies toxicity tests and in the impacts of pesticides and other xenobiotics on ecological structures. The decision in the early 1990s to limit the use of field and multispecies tests in pesticide registration (Fisher 1992) has sparked a debate about the appropriateness of these types of evaluations. Although a variety of factors contributed to this action, apparently the field and pond mesocosm tests that were conducted as part of the registration process did not contribute to the evaluation of risk by pesticides in a timely and cost-effective manner. This action was taken despite the number of available methods and analysis techniques.

Over the last 20 yr a variety of multispecies toxicity tests have been developed. These tests, usually referred to as microcosms or mesocosms, range in size from 1 L (the mixed-flask culture) to thousands of liters (in the case of the pond mesocosms). A review by Gearing (1989) listed 11 freshwater artificial stream methods, 22 laboratory freshwater microcosms ranging from 0.1 to 8400 L, and 18 outdoor freshwater microcosms ranging from 8 to 18 000 000 L.

In spite of the variety of methods, multispecies toxicity tests are not commonly used for ecological risk assessment. We suggest two central reasons for this lack of utilization. First, multispecies toxicity tests have fundamentally different properties than single-species acute or chronic tests and must be designed and interpreted accordingly. Second, conventional data analysis as conducted for single-species toxicity tests are inappropriate and misleading for multispecies toxicity tests. We will cover each of these points and provide basic design criteria for multispecies toxicity tests and their analysis.

The fundamental nature of multispecies toxicity tests

Single-species toxicity tests focus on the properties of the organism set apart from its environment and
therefore its ecological and evolutionary context. Organismal structures have distinct properties that are fundamentally distinct (Landis et al. 1995, 1996). First, organisms have a central repository of information, the nuclear and organellar DNA, that does code for the functionality of the structure. The DNA is a central blueprint that is largely protected from alteration (mutation) by repair mechanisms. In asexual organisms this plan is passed to the next generation largely intact. In sexual organisms recombination does occur, but during development the genetic structure is maintained to a large degree. In a fundamental way the genetics of an organism, especially in the gametic cell line, is protected from change. This ahistorical nature and ability for traits to remain intact from generation to generation are the raw material of Darwinian evolution. If the genetic structure did depend upon the historical events of an organism, Lamarck would have correctly described evolution. Indeed, evolution by natural selection is perhaps the most fundamental definition of life.

To say that the inheritance of genetic traits is ahistorical is not to say that certain organismal structures are not historical. Obviously the central nervous system and the immune system are historical in nature and are able to incorporate information about events into the behavioral and immunological responses of an organism. But memory and acquired immunity are not passed to the offspring.

Because of the existence of a genetic blueprint and the ahistorical nature of organisms, a great deal of repeatability is available in a single-species toxicity test. Nutritional states and the quality of culture conditions aside, it is fundamentally possible to repeat experimental results especially with clonal or highly inbred organisms. This feature has been used to advantage to understand modes of action, physiology, pharmacokinetics, and basic toxicological responses. An increase in dose results in an increase in effect until constrained by other toxicological properties of substance. Classical dose–response models do work because the systems are largely linear in nature. However, single-species tests do not include the most fundamental property of ecological structures.

Ecological structures, including multispecies toxicity tests, have a fundamental property of being historical structures. Brooks et al. (1989) in an extensive literature review and detailed derivation concluded that ecological systems are time directed, in other words, irreversible with respect to time. Drake (1991) has experimentally demonstrated the historical aspects of ecological structures. He assembled in varying orders the components of a relatively simple microcosm system. The structure of the system was highly sensitive to the order of the introductions. While stochastic explanations may seem to describe snapshots of ecological systems, knowing the historical dynamics elucidates mechanisms of species interaction and provides deterministic descriptions of the system. Microcosm experiments conducted with a variety of jet fuels (Landis et al. 1993a, b, Matthews et al. 1996) have clearly demonstrated the persistence of information concerning treatment group long after the degradation of the toxicant.

We (Landis et al. 1995, 1996, Matthews et al. 1996) have proposed a framework called community conditioning, based upon the results of numerous multispecies toxicity tests and field research (Matthews and Matthews 1991, G. Matthews et al. 1991, R. Matthews et al. 1991, Landis et al. 1993a, b, 1994). The community conditioning hypothesis is an explicit recognition of the historical and by definition nonequilibrium nature of ecological structures. The basic precept is that ecological communities retain information about events in their history (Matthews et al. 1996). The information can be contained in a variety of formats, from the relative frequencies of alleles or mitochondrial DNA, to the dynamics of predator–prey and competitive interactions. The historical nature of microcosms and other ecological structures classifies them as complex according to the definition of Nicolis and Prigogine (1989).

Complex, nonlinear structures have specific properties, listed by Cambel (1993). A few particularly critical to microcosms are: (1) complex structures are neither completely deterministic nor stochastic, and exhibit both characteristics; (2) the causes and effects of the events the system experiences are not proportional; (3) the different parts of complex systems are linked and affect one another in a synergistic manner; (4) complex systems undergo irreversible processes; and (5) complex systems are dynamic and not in equilibrium; they are constantly moving targets. These properties are especially important in the design, data analysis, and interpretation of multispecies toxicity tests and will be discussed in the appropriate sections.

In addition to multispecies toxicity tests sharing the properties of complex systems as do natural ecological structures, they also have other important characteristics. Multispecies toxicity tests have trophic structure, although simple. The physical aspects of many types of naturally assembled ecological structures can often be mimicked, and there have been many successful attempts at incorporating at least some of the nutrient, sunlight, sediment, soil, and other physical features into toxicity tests. Multispecies toxicity tests have been successful in modeling a variety of ecological structures.

Evolutionary events also occur within multispecies toxicity tests. Species or strains resistant to xenobiotics do arise (Molander and Blanck 1992). Simple microbial microcosms (chemostats) are often used to force the evolution of new metabolic pathways for pesticide and xenobiotic degradation.

Microcosms do not have some of the characteristics
of naturally synthesized ecological structures. Perhaps primary is that multispecies toxicity tests are by nature smaller in scale, thus reducing the number of species that can be successfully integrated into the structure. This feature is very important since after dosing, every experimental design must make each replicate an island to prevent cross contamination and to protect the environment. Therefore the dynamics of extinction and the coupled stochastic and deterministic features of island biogeography produce effects that must be separated from those of the toxicant. Ensuring that each replicate is as similar as possible over the short term minimizes the differential effects of the enforced isolation, but eventually divergence is to be expected.

Coupled with the necessity of making the replicates similar is the elimination of a key ingredient of naturally synthesized ecological structures, the spatial and temporal heterogeneity. Spatial and temporal heterogeneity is crucial to species richness, as in the “Paradox of the Plankton” (Hutchinson 1961). Environmental heterogeneity is key to the establishment of metapopulations, an important factor in the persistence of species. Microcosm systems can be constructed that incorporate spatial heterogeneity as did Drake et al. (1993) by connecting a series of containers by an inoculation scheme, but since each connected container is no longer a replicate, statistical power decreases. Heterogeneity within a replicate container can also be added, but increases the difficulty of beginning the test with similar microcosm replicates.

The design of multispecies toxicity tests runs into a classical dilemma. If the system incorporates all of the heterogeneity of a naturally synthesized ecological structure, then it can become unique, thereby losing the statistical power needed for typical hypothesis testing. If multispecies toxicity tests are complex systems and subject to community conditioning, then the tests are not repeatable in the same sense as a single-species toxicity test or biochemical assay. Since the information about past events can be kept in a variety of forms, from the dynamics of populations to the genetic sequence of mitochondria, it is necessary to be able to incorporate each of these types of data into the design and analysis of the experiment. Assumptions about recovery are invalid, and tend to cloud the now apparent dynamics of multispecies toxicity tests. The ramifications are critical to the analysis and interpretation of multispecies toxicity tests.

DATA ANALYSIS AND INTERPRETATION OF MULTISPECIES TOXICITY TESTS

A large number of data analysis methods have been used to examine the dynamics of these structures. The analysis techniques should be able to detect patterns given the properties of multispecies toxicity tests described in the previous sections. In order to conduct proper statistical analysis the samples should be true replicates and in sufficient number to generate sufficient statistical power. The analysis techniques should be multivariate, able to detect a variety of patterns, and able to perform hypothesis testing on those patterns.

Sample design.—One of the most difficult aspects of designing a multispecies toxicity test is that of having sufficient replication so that the analysis has sufficient power to resolve differences between the reference nondosed replicates and the other treatment groups. This is particularly difficult when examining a broad range of variables with very different distributions and characteristic variances. Obviously logistical considerations are also critical, because of the large size and complexity of multispecies tests. However tempting, it is inappropriate to take several samples from the same microcosm sample and label these samples replicates. This is especially prevalent in artificial streams where individual sampling trays within a stream are considered replicates (Gruessner and Watzin 1996). These samples are not true replicates since each tray is connected by the water to the tray downstream. Such a sampling may underrepresent the true variance and is better used to represent the environmental heterogeneity within a single stream. Such pseudoreplication is best avoided since it invalidates the assumptions of statistics used for hypothesis testing.

Univariate methods.—These techniques, principally ANOVA and nonparametric ANOVA, are the most commonly used analysis methods. However, by definition, these univariate methods of hypothesis testing are inappropriate for multispecies toxicity tests. As such, these methods are an attempt to understand a multivariate system by looking at one univariate projection after another, attempting to find statistically significant differences. Often the powers of the statistical tests are quite low due to the few replicates and the high inherent variance of many of the biotic variables.

A derivative of the ANOVA approach, the Interval of Nonsignificant Difference (IND, Conquest and Taub 1989), is a graphical representation of a series of ANOVAs corrected for Type II error over time. A series of graphs for each variable are then plotted and examined by the investigator for apparently relevant patterns. The IND approach has been used (Landis et al. 1993a, b) to identify patterns that were later confirmed in a blind analysis conducted by Matthews and Matthews (1991) by multivariate techniques. However, the analysis using the IND approach is very laborious and cannot seek interesting relationships among variables.

Perhaps the greatest danger of the use of ANOVA and related univariate tools is the perpetuation of NOELs (no observable effects concentrations), LOECs (lowest observable effects concentrations), and related terms based on univariate hypothesis testing. There is now an energetic discussion of the validity of these measures in evaluating single-species toxicity tests (Chapman et al. 1996, Chapman and Chapman 1997,
Dhaliwal et al. 1997). NOECs and LOECs are so dependent upon statistical power and the concentrations chosen by the experimenter that they are artifacts of the experimental design rather than reflections of the intrinsic hazard of the toxicant. Given the historical nature of microcosm systems such a determination as a NOEC or LOEC is contrary to the properties of complex structures. Instead, measurements such as NOECcommunity are indications of the resolving power of the experimental design and the parameters chosen to be measured rather than a measurement of a real characteristic of ecological structures.

**Multivariate methods.**—There are a variety of multivariate methods that are available for the exploration of patterns within ecological data sets. These have the advantage of examining all of the data and therefore more accurately reflect the nature of ecological structures. Coupled with association analysis, these techniques can also be used to test the hypothesis that the pattern is related to treatment. Although each method described below is multivariate, not all are equal and there is no best method for all cases. Each technique makes different assumptions about the relationships among the variables. Some of the techniques attempt to explain variance, others find clusters based on similarity in a distance measure. In some cases the search for patterns is blind to treatment, in others the treatments are known to the algorithm. Each technique provides the opportunity for a different insight into the patterns that exist within the multispecies toxicity test.

Ludwig and Reynolds (1988) provide an excellent introduction to the assumptions, derivations, and use of several multivariate techniques commonly used for the analysis of ecological communities. Perhaps the most common forms of multivariate analysis are principal components analysis (PCA) and its derivatives. PCA attempts to find orthogonal combinations of variables that account for the variance within a data set. The assumption in PCA is that the relationships are linear, therefore PCA is best used with a relatively narrow range of variables where a linear response can be assumed. Assuming that ecological structures are complex, nonlinear relationships may be the norm. Another drawback of PCA is the emphasis on the explanation of variance, and the corresponding emphasis upon variables that may be highly variable but only contain noise (Matthews and Hearne 1991, Matthews et al. 1995a).

There have been attempts to deal with the issue of nonlinearity in data sets. Detrended principal components (DPC) uses a polynomial expression to remove the nonlinear relationships from the PCA axes. DPC is useful for data sets of moderate nonlinearity. Detrended correspondence analysis uses a more complex algorithm to eliminate the nonlinearity, but requires a more complex computation. Nonmetric multidimensional scaling (NMDS) is a robust method that deals with nonlinearities by using ranks. A recently fashionable technique derived from a principal components approach is the coupling of PCA with redundancy analysis (RDA) (van Wijngaarden et al. 1995, van der Brink et al. 1996). The utility of the technique is that it provides a depiction of the treatment trajectories in an ecological space, and the statistical significance can be examined using a permutation test. One of the proposed benefits of the technique is that it can determine recovery, a dubious distinction in light of our previous discussion. In common with other PCA techniques, the technique does assume a linear response.

One of the noteworthy characteristics of the previously described techniques is that all are based on knowing the treatment groups, which introduces a strong bias into the search for patterns and explanations. Such a bias also makes it difficult to discern new patterns that may be due to other environmental gradients present in the testing facility or outdoor setting. Most of the models assume a linear response. And in common with that assumption is that the variables with the greatest variance are by definition the most important.

Clustering has the advantage of attempting a unbiased search through a data array for patterns. The algorithm has no knowledge of treatment groups and is attempting to detect patterns and conduct a sorting based on a predetermined set of rules. There are a variety of available techniques (Ludwig and Reynolds 1988). We have used clustering based on cosine distance and vector distance on a variety of microcosm data sets (Landis et al. 1995, Matthews et al. 1995b). The sorting or clusters can then be compared to treatment groups or other classification using an association analysis to test for significance.

The typical problems associated with conventional clustering have been extensively discussed (Matthews and Hearne 1991, Matthews et al. 1991, 1995a). Sensitivity to attributes with high variance and the ad hoc nature of combining variables with different metrics are particularly bothersome. Conceptual clustering as implemented by nonmetric clustering and association analysis (NCAA) (Matthews and Hearne 1991) has proven to be a powerful technique in the analysis of data sets with high dimensionality but with replication typical of multispecies toxicity tests (Landis et al. 1993a, b). One of the biggest assets of NCAA is that it is nondimensional, nonmetric, and that it selects the variables important in determining the clusters and rejects those that do not contribute. NCAA does not assume a linear relationship among attributes, in fact it assumes no particular model at all. NCAA is computationally intensive and there is no assurance that a global maxima of clustering has been obtained. Furthermore, NCAA is not available as part of packaged statistical programs.

The characteristic of NCAA to select the important variables in a data set has led to the discovery of the...
phenomenon of community conditioning. NCAA and other pattern recognition tools allow for the discovery of unsuspected patterns, which is useful if the understanding of the nature of ecological structures is to be accomplished.

Visualization.—Methods of visualization that are useful in interpreting the dynamics of ecological structures are also available. An ordination diagram has been used by van der Brink et al. (1996) to plot the path of the various treatment groups using the axes generated by the redundancy analysis.

Landis et al. (1996) have used Space–Time Worms as a method of visualizing the trajectories of the treatment groups. Two variables that NCAA ranks as important in the clustering are plotted along with time. The variability among replicates is represented by the thickness of the cylinder. We have found this technique particularly useful in depicting the changing nature of the ecological structures and in portraying variability as a characteristic of the experiment.

**Basic Design Guidelines for Multispecies Toxicity Tests**

Multispecies toxicity tests come in a wide variety of types (artificial streams, generic freshwater, simulated farm ponds, and even ditches) and they share basic properties. Experimental designs should take into account the advantage of these properties to ensure an interpretable experiment result. We propose the following design parameters for experimental design, analysis, and interpretation.

**Basic principles**

1) Multispecies toxicity tests are complex structures. Complex structures are nonequilibrium, historical, sensitive to initial conditions, and nonlinear. To measure the recovery of such a structure is to measure a property that does not exist for a complex structure (Landis et al. 1996).

2) Multispecies toxicity tests are not repeatable in the strict sense, since each is sensitive to initial conditions. However, common patterns do appear and these should be the focus of the investigation.

3) All impacts can leave lasting effects (Matthews et al. 1996). Therefore determination of a NOEC or LOEC is not warranted.

**Experimental design**

1) Small differences in initial conditions can measurably alter the dynamics of the replicate systems. Attention must be paid to ensure the similarity of the experimental replicates. This includes introducing known components or verifying that the migration pathways from the source are equal for all replicates.

2) Environmental gradients do exist in a laboratory or a field situation. A random block design to take into account such gradients should be used.

3) Since the systems are all sensitive to initial conditions, equal numbers of replicates for each treatment group should be used to give every treatment an equal chance for deviation.

4) Samples taken from the same experimental unit must not be considered as replicates. This is common in experimental stream systems where several samples are taken from the same stream unit. Since these samples are connected by the water flow they are not independent and not replicates.

**Data analysis**

1) Univariate statistical techniques are not appropriate for multivariate structures. Repeated ANOVAs are both unwarranted and misleading. Calculation of NOECs and LOECs is also inappropriate.

2) Multivariate methods are more suitable for the data analysis of multispecies toxicity tests. No one multivariate technique is always best. Given that many responses of multispecies toxicity tests are nonlinear, techniques that do not assume linear relationships may be more robust.

3) Multivariate techniques that account for variability may be misled by noisy variables and miss important relationships.

4) Techniques such as PCA may prevent the detection of novel patterns. Clustering and other exploratory techniques can lead to the discovery of novel patterns and relationships.

5) Do not assume that the combination of variables that are best for determining clusters or treatments on one sampling day will be the most appropriate for every sampling day. As the structure and function of the multispecies toxicity test change over time, so will the important variables.

6) Multivariate visualization techniques do exist and should be used. These techniques can lead to a much better understanding of the dynamic nature of these structures.

**Conclusions**

Microcosm and other multispecies toxicity tests can contribute to the evaluation of risk due to pesticides or other environmental contaminants. Advances in the understanding of model ecological structures and improvements in data analysis make an understanding of the direct and indirect effects attainable. False goals such as NOECs or the determination of recovery should be stricken from analysis, interpretation, and regulation.

Multispecies toxicity tests are particularly useful in predicting the trajectories of ecological structures following a pesticide application. Although specifics may be unique, a search for common patterns can reveal classes of interactions useful for the prediction of risk. Only with these kinds of tests and related field studies can direct experimental evidence for the ecological risks of pesticides be obtained.
ACKNOWLEDGMENTS

We would like to thank our fellow microcosm researchers, especially April Markiewicz, for conducting numerous microcosm and mesocosm studies. Thanks to Frieda Taub and Chris Grue for numerous discussions on this topic. The research has been supported by USAFOSR. Grant Number F49620-94-1-0285.

LITERATURE CITED


