United States
Environmental Protection
Agency

Environmental Research
Laboratory
Gulf Breeze FL 32561

Research and Development

EPA/600/S3-87/016 July 1987

EPA

Project Summary

Field Validation of Multi-Species Laboratory Test Systems for Estuarine Benthic Communities

Robert J. Diaz, Mark Luckenbach, Sandra Thornton, Morris H. Roberts, Jr., Robert J. Livingston, Christopher C. Koenig, Gary L. Ray, and Loretta E. Wolfe

The major objective of this project was to determine the validity of using multi-species laboratory systems to evaluate the response of estuarine benthic communities to an introduced stress. Over a 5-year period, experiments in Apalachicola Bay, Florida, and the York River, Virginia, sought to (1) develop criteria for microcosm tests to evaluate the capacity of microcosms to model natural communities in the presence and absence of pollution-induced stress, and (2) assess the validity of extrapolating test results from one location to another. Individual species response patterns in the microcosms were highly variable and seldom showed good agreement with patterns in the field. Species richness in the microcosms and field sites showed good temporal agreement and provided a conservative indicator of community response to a toxic stress. An ecologically based guild approach to grouping species proved to be a powerful and reliable method of extrapolating from microcosm test results to responses of field communities.

This Project Summary was developed by EPA's Environmental Research Laboratory, Gulf Breeze, FL, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

Objectives

The principal goal of environmental toxicology is to understand pollution-induced changes in ecological systems, primarily in an effort to predict the environmental consequences of a toxicant. Recently, emphasis has been placed on using multi-species laboratory systems to evaluate the response of aquatic ecosystems to an introduced stress. The assumption has generally been made that multi-species systems provide more realistic estimates of the effects of toxicants on complex natural ecosystems than do single species tests.

A requisite part in the development of a laboratory multi-species aquatic test system is field verification. The authors define field verification as the testing of the capacity of specific laboratory test systems to predict the responses of ecosystems to toxicants. The process of field verification raises methodological considerations concerning criteria for conducting tests and interpreting data. An essential part of this process is simultaneous investigation of community dynamics in the laboratory test system and in the natural community from which it was derived.

In a 5-year project designed to field verify multi-species laboratory systems (microcosms) for use with estuarine benthic communities, four major issues were addressed: (1) development of criteria for conducting microcosm tests

and for interpreting the data; (2) evaluation of the capacity of a microcosm to track natural field communities in the absence of a toxicant; (3) comparison of community response patterns of laboratory and field communities to a pollution-induced stress; and (4) determination of the validity of extrapolating from microcosm tests conducted in one location to natural communities in another. Focus has been on the infaunal macroinvertebrate communities from unvegetated, soft-sediment sites in Florida (Apalachicola estuary) and Virginia (York River estuary). The approach was to synoptically conduct field and laboratory experiments at both locations over different seasons.

Experimental Approach

The study sites in the Apalachicola Bay system (East Bay and St. George Sound) were located in polyhaline and oligohaline areas, and those in the York River were located in the meso-polyhaline portion of the estuary. All sites were shallow (1-2 m), unvegetated areas. Sediments in the oligohaline sites were silty sand, and in the polyhaline and meso-polyhaline sites sediments were predominantly fine sands.

Microcosms ranging from approximately 0.1 to 1.0 m^2 were used to evaluate the influence of microcosm size on the system response. Microcosms were constructed of a series of cores collected with hand-operated box cores (10 cm x 20 cm x 10 cm deep). Core samples were placed in trays on a seawater table in the same arrangements as the original field orientations. For the duration of the experiments, microcosms were maintained in flow-through seawater tables where conditions of light, temperature and salinity were similar to the field. Macroinvertebrate samples in both the field and laboratory were collected in random sampling designs with coring devices (5 cm diameter, VIMS; 7.5 cm, FSU). Samples were preserved, rinsed onto 500- and 250-μm-mesh sieves, and the organisms identified to species and enumerated.

Over a 5-year period seven field-laboratory experiments were conducted during spring and fall (Table 1). These seasons represented periods of peak biological activity. While the basic protocol for the tests remained similar, several different field and laboratory treatments were employed in the various tests. In all tests, replicated laboratory microcosms maintained in flow-through systems were sampled simultaneously with field treatments located in the same sites from which the microcosm communities were derived. The various treatments included predator exclusion and inclusion cages, and both field and laboratory dosing with unpolluted, hydrocarbon polluted (from the Elizabeth River, VA) and pentachlorophenol (PCP) polluted sediments. Toxicant dosing procedures in the field and laboratory were investigated by testing both the total volume and application method of toxicant-laden sediments. Sampling interval and duration were similar in a field-laboratory tests, but an effort wa made to determine the appropriat sampling schedule to observe response in the dosed experiments.

A major focus of the work was t identify response variables that ade quately reflect important community level responses in both the field an laboratory systems. Community-wid descriptors (total number of individuals number of species, species diversity an evenness), numerical abundances o dominant species, and biomass an productivity estimates were all used a

Table 1. Sampling Schedules for the Combined (FSU-VIMS) Experimental Program (1981 1985)

- *I. Weekly samples*
  - *A. FSU*
    - *1. oligohaline stations (11/24/81-11/17/83)*
    - *2. polyhaline station (11/25/81-3/15/84)*
  - *B. VIMS*
    - *1. polyhaline marine lab station (10/13/79-12/18/83)*
- *II. Microbiological data*
  - *A. FSU*
    - *1. oligohaline stations (fall 1982; spring 1983)*
    - *2. polyhaline stations (spring 1982)*
  - *B. VIMS*
    - *1. marine lab station (spring 1982)*
- *III. Combined (field-laboratory) experiments*
  - *A. Spring 1982*
    - *1. Florida*
    - *2. Virginia*
  - *B. Fall 1982*
    - *1. Florida*
    - *2. Virginia*
  - *C. Spring 1983*
    - *1. Florida*
    - *2. Virginia*
  - *D. Fall 1983*
    - *1. Florida*
    - *2. Virginia*
    - *3. Treatments included:*
      - *a. Field controls*
      - *b. Field predator exclusion cages*
      - *c. Field predator inclusion cages*
      - *d. Microcosm controls*
      - *e. Field and lab treatments dosed with PCP*
  - *E. Spring 1984*
    - *1. Virginia only*
  - *F. Spring 1985*
    - *1. Florida (station ML)*
    - *2. Virginia*
    - *3. Treatments included:*
      - *a. field controls*
      - *b. microcosm controls*
      - *c. replicate lab and field treatments dosed with PCP*
      - *d. azoic sediments*
  - *G. Fall 1985*
    - *1. Florida (station ML)*
    - *2. Virginia*
    - *3. Treatments as in F.3.*

measures of community response. A more fruitful approach was to categorize individual species into guilds based upon ecological similarities and to treat these groups as ecological units in assessing response to toxicant stress.

Project Results

The laboratory-field experiments coupled with the weekly monitoring data permitted us to address each of the objects stated above. Criteria for conducting microcosm toxicity tests are outlined below and detailed in the final report. Definition of ecologically relevant functional groupings and the classification of species into guilds provided the basis for successful comparisons between laboratory and field communities and between the Florida and Virginia test systems.

Comparisons of different microcosm sizes suggested that small benthic microcosms (approximately 0.1 m^2) provide good laboratory systems. These small microcosms had similar community patterns to larger (0.8 - 1.0 m^2) microcosms and are easier to construct and handle. Sampling of small microcosms is a destructive process and each replicate microcosm may be sampled at only one time, therefore requiring a large number of replicate microcosms.

Experimental durations of up to 6 weeks and sampling intervals of 1-2 weeks were used in the various field-laboratory experiments. In one experiment, samples were taken within 24 hours after dosing. Experience suggests that early sampling after dosing (within 24-48 hours) followed by increasing intervals up to 5 weeks provided a good sampling regime. Beyond 5 weeks microcosms may begin to experience changes in sediment geochemistry that cause the laboratory to diverge greatly from the field.

The addition and monitoring of toxic substances are critical steps in any laboratory microcosm test system. The experiments suggest that a good dosing procedure was to apply a 1-cm thick layer of toxicant-laden sediments to the surface. The authors found that this same amount of uncontaminated sediments did not have noticable effects upon the community. Dosing with less sediment (1 mm thick) and the same toxicant load proved to be less effective. In tests using benthic microcosms derived from other habitats, it is important to include a treatment that adds uncontaminated sediments.

The microbiotic component in a laboratory microcosm is highly variable, and its capacity to predict natural trends depends on a combination of habitat characteristics in the area of origin. Microbial communities in microcosms deviate progressively with time from field associations and extended equilibration periods are ill advised. In microcosms of sediments from polyhaline areas, microbes did not follow field conditions as closely as those in microcosms from oligohaline portions of the estuary. Without detailed knowledge of microbial ecology in the source area an interpretation of results from laboratory microcosms could be misleading.

Comparisons of field and laboratory community dynamics of infaunal macroinvertebrates revealed much variation between experiments and locations, but some generalization emerged. The population dynamics of many dominant species in both the Apalachicola Bay and the York River estuary were similar in the microcosms and the field. This generalization was qualified by finding that some species occasionally underwent rapid population blooms in the laboratory microcosms. For instance, in Florida experiments the polychaete *Mediomastus ambiseta* sometimes exhibited large population increases in the laboratory relative to the field. A similar response was observed in Virginia for the oligochaete *Paranais littoralis*. These population increases are related to the ability of these organisms to reproduce in the microcosm where new individuals survive better than in the natural field site. This pattern has been experimentally observed in field cage treatments for *M. ambiseta*. The major population fluctuations in both the field and the laboratory (for undosed treatments) were associated with recruitment events. Since recruitment intensities for most species differ between the laboratory and the field, recruitment events may result in substantial differences between field and laboratory populations. Moreover, the year-to-year variability in the timing and intensity of recruitment for any given species introduces a stochastic element into microcosm testing when single-species fluctuations are emphasized.

Using community-level parameters to describe field and laboratory system responses avoids some of the variability associated with individual species fluctuations. Species richness provided a good descriptor of the macroinvertebrate communities (Figure 1). In undosed treatments, species numbers in the field and the laboratory were often similar and very conservative. In dosing experiments with both hydrocarbon contaminated sediment and PCP-laden sediment, the species richness component showed similar negative responses in both the field and laboratory communities (Figure 1). By contrast total numbers of individuals in the system fluctuated widely, largely as a result of recruitment events, and were not particularly responsive to toxicant treatments. Species diversity and evenness measures reflect combinations of these two components and were variable in their correspondence between the laboratory and field sites. Species richness is an important component of natural systems which is modeled well in aquatic microcosms and proved to be a sensitive community-level response to stress by a toxicant.

Grouping species into guilds based upon functional groups according to trophic, mobility, and dispersal modes proved to be a powerful approach for interpreting community responses. This approach served two purposes. First, it permitted identifying those guilds of organisms for which laboratory microcosm populations do not serve as good analogs of natural populations in the absence of any toxicant (e.g., those brooding or asexually reproducing species which have capability of blooming within the microcosm). These types of organisms may be excluded *a priori* from analyses to assess toxic effects. The second advantage is that identifying types of organisms that act as similar ecological units facilities comparisons between microcosms and field sites from different locations. For instance, while the species composition varies between the Virginia and Florida sites, functionally similar ecological groups are found in both sites and provide a basis for comparison. Figure 2 shows summary examples of this approach for one guild which was modeled well in the laboratory and one which was not.

Comparisons between field-laboratory experiments in Florida and Virginia identified both similarities and differences. The most notable difference is that major recruitment periods at each location occur in different seasons. The major recruitment period in the York River estuary is in the spring with only a minor fall recruitment; the pattern in the Apalachicola Bay is temporally reversed. Since, as noted above, recruitment events play a major role in the population

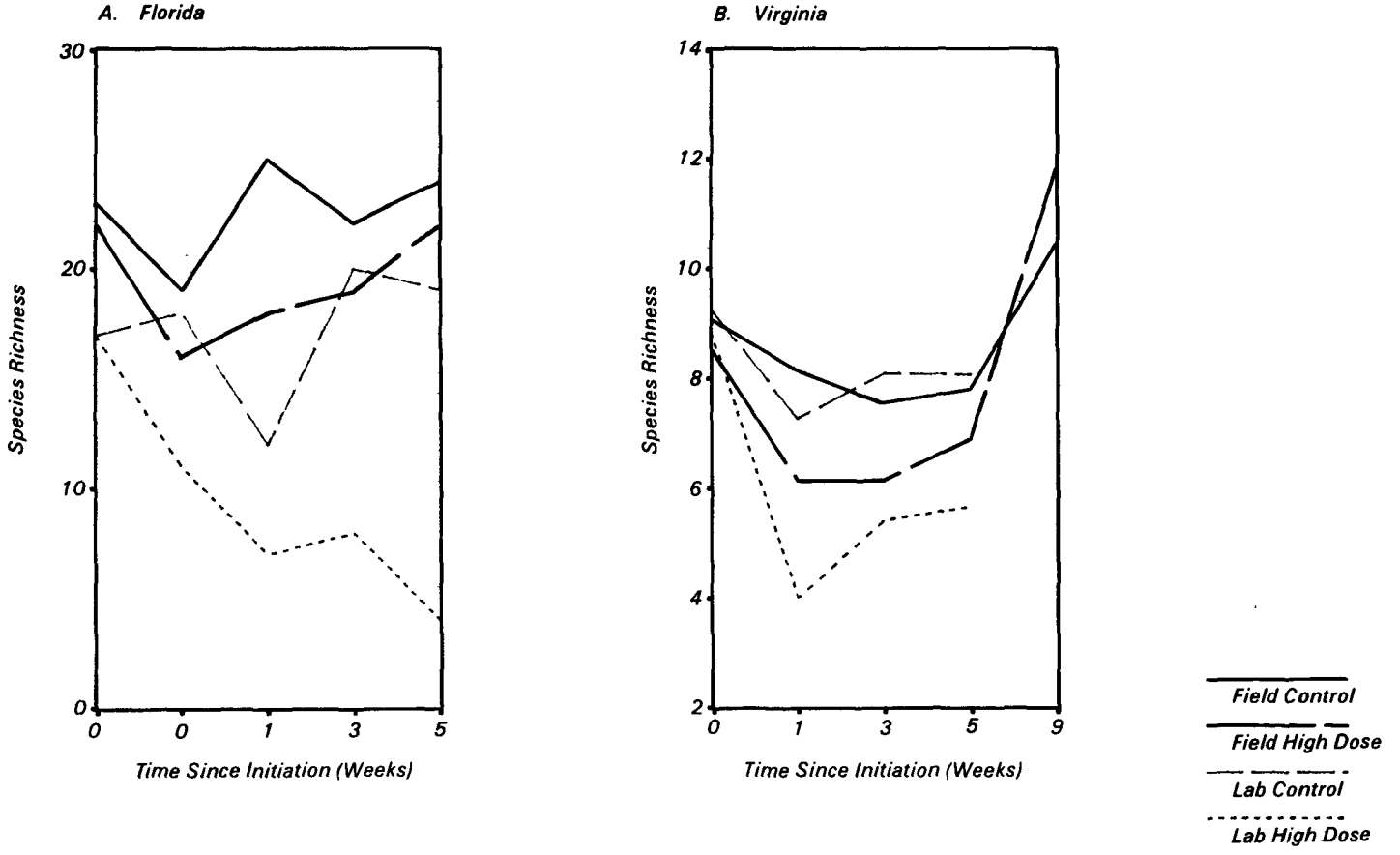

Figure 1. Species richness in field/laboratory PCP-dosed treatments from fall 1985 in Apalachicola Bay, Florida and the York River, Virginia.

fluctuations observed, this becomes an important issue when attempting to infer the responses of natural communities to a toxic stress in one location based upon microcosm experiments located in another.

Conclusions

The authors conclude that properly constructed and replicated multi-species laboratory test systems with estuarine macrobenthic invertebrates can serve as effective tools for predicting responses to toxicant stress. Several caveats apply. Variability in natural estuarine systems is high, necessitating large numbers of experimental replicates and samples to observe even major responses. Species richness measures provide a conservative indicator of community response to pollution-induced stress which is not subject to much of the variation observed for abundance measures. However, this measure may also gloss over much of the ecologically relevant response to the stress. The population dynamics of individual species within laboratory microcosms are too variable to provide adequate models of field populations, but grouping species into ecologically similar guilds alleviates much of this problem. The detailed ecological data required to construct these groups may be difficult to obtain for many species.

The authors emphasize the importance of good ecological characterization of the habitats from which the microcosms are derived and the habitats about which inferences are to be made. When considerations of different recruitment seasons are taken into account, similar community responses to a toxicant are found in both the Florida and Virginia experiments. Experiments with PCP dosing conducted in Florida during the spring of 1985 had a similar response (in species richness) to experiments in the fall of 1985 in Virginia, and the fall experiments in Florida resembled those from the spring in Virginia. At each location experiments conducted during the peak reproductive seasons resulted in blooms of single species in the microcosm. A functional guild approach to analyzing community response patterns enhances the ability to make predictions concerning toxic responses in one site based upon laboratory experiments conducted at another site.

It was concluded that laboratory microcosms can provide a valuable tool for assessing natural benthic community responses to introduced toxicants, provided that the caveats and conditions described above are heeded. The authors recommend using microcosms to provide realistic estimations of field effects.

Recommendations

The results of the field-laboratory comparison experiments indicate that multi-species laboratory aquatic microcosms may yield valuable information regarding the responses of natural communities to pollution-induced stress. However, several very important cautions are offered for conducting and interpreting microcosm toxicity tests and in extending the findings to natural systems:

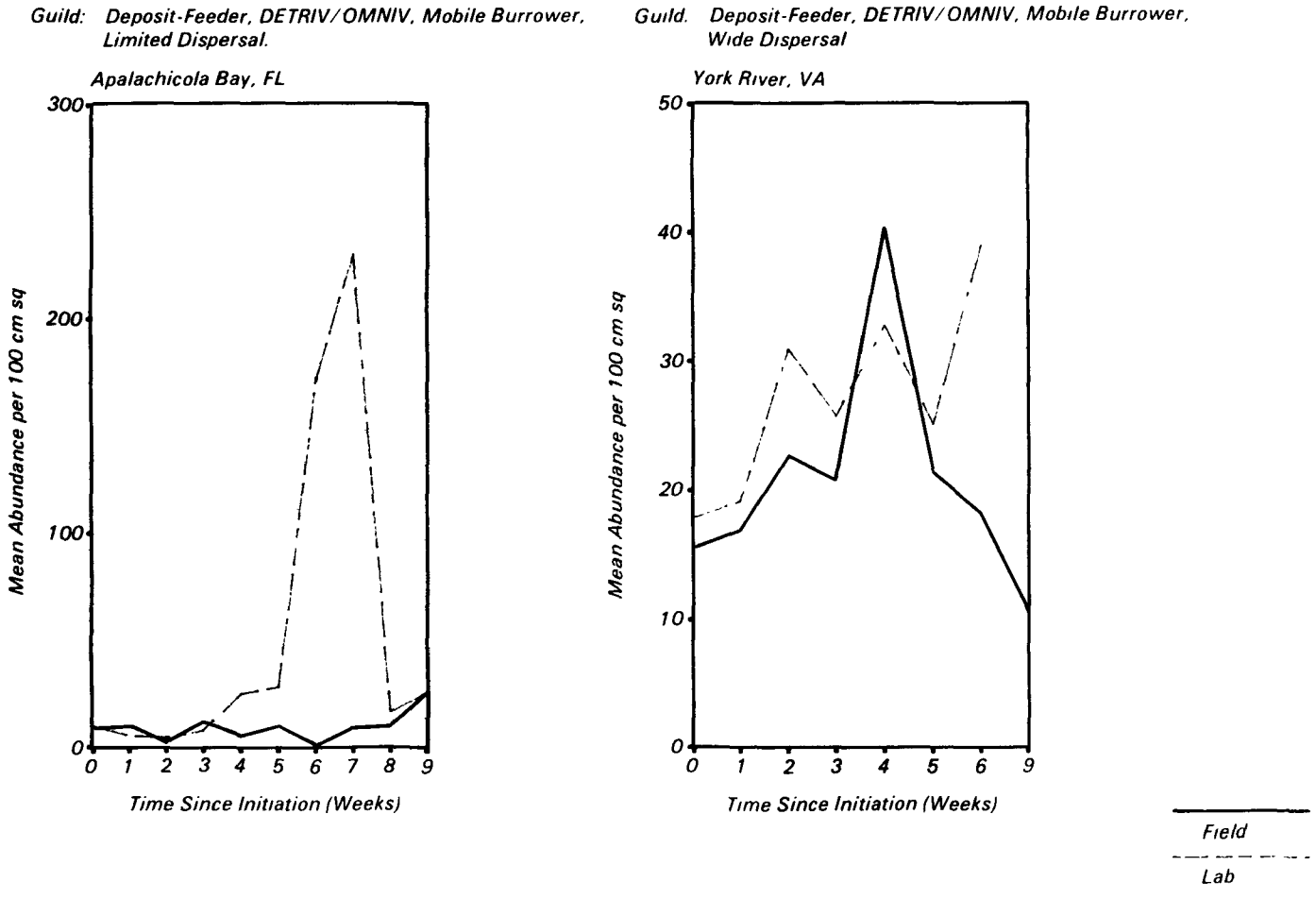

Figure 2. *Comparison of temporal patterns for 2 guilds in the laboratory and the field. Data are composites of the control treatments in all tests. Data for the guild with limited dispersal (for which only Florida data are shown) reveal that although lab and field abundances track one another well initially, individuals in this group may undergo population blooms in the lab The guild with wide dispersal (Virginia data shown) shows a consistent pattern through the first 5 weeks with some divergence between lab and field thereafter.*

1. Close attention must be paid to physio-chemical characteristics of microcosms and it is important that these lie within realistic ranges for field values at the time the experiment is conducted.

2. Monitoring of toxicant levels and distribution within the microcosms throughout the experiment is necessary to evaluate dissipation and breakdown of toxicants.

3. The high spatial variability inherent in benthic communities necessitates that sufficient replicates be employed.

4. The temporal variation in recruitment adds a nearly random component to the community response in microcosm tests from year to year and site to site. To overcome this problem it is mandatory that microcosm tests, while being properly timed to correspond with biologically important reproductive seasons, identify and exclude species with aberrant recruitment patterns in the laboratory test system.

5. Successful extension of toxicity test results from laboratory microcosms to the field sites from which they were derived and beyond to other sites requires detailed knowledge of the systems. Just as important as an understanding of the physical conditions of the habitats is a good knowledge of the reproductive seasons and modes, trophic types, and life history characteristics of the component fauna. Since species composition will vary between sites it is necessary to characterize the response of different "ecological types" (guilds) in microcosms if the results are to be useful