Draft Guidance Document on Simulated Freshwater Lentic Field Tests  
(Outdoor Microcosms and Mesocosms)

INTRODUCTION

1. Outdoor microcosms and mesocosms are useful in risk assessment when lower-tier and higher-tier laboratory studies (single-species and multi-species) indicate potential risks. They can be an important tool in bridging the gap between these studies and the natural environment. Every microcosm or mesocosm study should be designed to test a specific hypothesis using information gained in previous steps of the risk assessment. This makes every microcosm or mesocosm study unique in at least some aspects of its design. Guidance for conducting microcosm or mesocosm studies is therefore necessarily generic and flexible (Campbell et al. 1999).

2. This document was originally based on consensus methods proposed by experts at a series of meetings convened within Europe and North America (SETAC-Europe, 1991; SETAC/RESOLVE, 1991; EWOFFT, 1992; Hill, et al., 1994). Recent developments and experience have been incorporated from the outcome of international workshops held in France (Campbell et al. 1999) and in Germany (Giddings et al. 2002).

3. One important reason to perform a microcosm or mesocosm study is to determine the ecological relevance of effects identified in laboratory studies. The studies can therefore include a variety of species, functional groups or habitat types. Interpretation of these studies focuses on effects at the community and ecosystem level, potential indirect effects and the recovery potential of sensitive endpoints. A second important reason for conducting a microcosm or mesocosm study is to measure effects of the chemical under more environmentally realistic exposure conditions. Such conditions could include the influence of partitioning to sediments and plants, photolysis, and other processes that may influence the fate of the chemical. Moreover, microcosm or mesocosm studies incorporate natural abiotic conditions (temperature, light, pH, etc.) that may influence the response of certain organisms (Campbell et al. 1999).

INITIAL CONSIDERATIONS

4. Before any microcosm or mesocosm test is conducted, clear objectives should be defined in order to determine which are relevant endpoints and which experimental design (e.g. level of replication, number of treatments) is appropriate. It is the responsibility of the study director to demonstrate that the system is appropriate for achieving the objectives of the study. It may be useful to discuss and agree on the protocol with the relevant authorities evaluating the test results. Any available information should be carefully reviewed, and preliminary laboratory testing should be undertaken when essential information for test design is missing. Factors to be considered include:

   a) Effects properties: the core ecotoxicological data that are always required for registration and other higher tier studies (eg additional single species, population-level studies, indoor multi-species studies) can be used to determine the primary concerns to be investigated. For example, data on the sensitivity of aquatic species can help to focus those populations and communities which should be studied in more detail.

   b) It should be determined what the derived parameters of interest are for each endpoint (e.g. EC$_{50}$ or NOEC).
c) The level of precision that is to be obtained for derived estimates, or the desired power of a relevant hypothesis should be determined as part of determining the objective of the study.

d) The size of effects which is considered of ecological significance should be determined, relative to the endpoints of concern (e.g. differences of 100 – 200% may be appropriate for organisms with a high intrinsic rate of increase; 50% effect may be appropriate to define adverse impact on more slowly increasing populations).

e) Depending on the objectives following on from point (a) to (d), it will be possible to determine:
   - Number of treatments and choice of doses
   - How treatments are to be assigned to the mesocosms (random or block design, etc.)
   - Number of replicated mesocosms per treatment
   - Organisms to be sampled, the size of the sample and how sampling should be carried out.

f) The method of statistical analysis should be determined as part of the setting of the objectives. (Chapman & Maund, 1996).

g) An appropriate exposure regime should be established in order to meet the objectives of the study. Questions that should be addressed include: (i) what are the expected routes of entry of the chemical into aquatic systems (e.g. spray drift or surface run-off); (ii) what is the frequency and timing of the entry of the chemical.

h) Physical-chemical and fate properties (for example, solubility, vapour pressure, octanol/water partition coefficient, adsorption coefficients, hydrolysis and photolysis rates, and biodegradability) should be previously ascertained alongside biological information in order to select sampling times and identify ecological components at greatest risk. A valid analytical method should be available before performing the microcosm or mesocosm test.

i) If it is felt necessary to examine the relationship between loading and resulting exposure concentrations in various compartments or strata then chemical input and fate modelling or field studies specifically designed to determine the fate and behaviour of the chemical can be used. There are a large number of models that can be used to derive initial estimates of persistence and concentration-time functions. Appropriate fate studies may be necessary before starting the microcosm or mesocosm test. Table 1 gives a overview of the models that are currently used in Europe and in the US.

j) Information on the use patterns of the product is necessary.

k) Information on the likely major metabolites should be considered.

**PRINCIPLES OF THE TEST**

5. Microcosms and mesocosms are constructed to simulate parts of natural aquatic ecosystems. These are generally established either by collecting organisms and placing them in tanks or artificial ponds or by enclosing parts of existing ecosystems. The test system thus contains a 'naturally' developed aquatic community, usually containing naturalised sediment and appropriate organisms such as zoo- and phytoplankton, pelagic and benthic macro-invertebrates and macrophytes. It may be appropriate to add
certain organisms (e.g. fish) from external sources. It should be noted that when parts of larger ecosystems are enclosed, the more limited space and boundaries may influence and affect the organisms and processes occurring in the microcosms or mesocosms, thus causing divergence from the larger ecosystem.

6. Application of the test substance can be done by direct addition of the chemical to water or by simulating the actual route of exposure (e.g. by application of spray, to simulate spray drift or by application of a soil-water slurry to simulate erosion runoff). In general it is recommended to have an exposure regime that will allow determination of a dose-response relationship. Exposure is expressed in terms of the concentrations of the chemical in water, as in a laboratory toxicity test. This approach is preferred over a simulation approach where exposure is expressed in terms of the amount of the chemical added per unit surface area or volume. (Giddings et al. 2002).

7. Microcosm or mesocosm studies should generally be designed so that a concentration-response relationship can be developed over a range of concentrations, encompassing those concentrations that reflect exposure in the field. A sufficient number of replicates should be used to allow detection of differences at a sufficiently low level.

8. Structural endpoints and functional endpoints should be considered. Structural endpoints relate to the abundance and biomass of all populations and their spatial, taxonomic and trophic organisation (Brock and Budde, 1994). Functional endpoints are related to all aspects of non-living materials processed by the structure – i.e. nutrient levels, oxygen levels, respiration rate, mineral concentrations, pH, alkalinity, conductivity and organic material content. (Kersting 1994; Brock et al. 2000a). In studies performed with substances that are not herbicides, the functional aspects could also be documented as conditions of the study rather than as endpoints of the study. The microcosm or mesocosm study should focus on taxonomic groups that lower-tier risk assessments have identified as being of concern.

9. Determining rate and extent of recovery of affected taxa can be crucial in the design of microcosm or mesocosm studies. Looking at recovery is one of the key differences between microcosm and mesocosm studies and other higher tier studies and requires substantial ecological knowledge to interpret. If looking at recovery is an objective of the study, the experimental design should be such that recovery can be observed. A sufficiently long post-treatment period of two to three generation times has to be foreseen to allow the detection of repopulation. (Giddings et al. 2002)

10. The statistical method that is to be used to analyse the data should be built into the design of the study.

11. A study can be considered to be valid if the following conditions are met.

- The study should focus on endpoints of organisms that are potentially at risk, which should have been identified in other lower- or higher-tier studies.
- A clear concentration-response should generally be established for the endpoints of concern. In addition, ideally a clear effect level for at least the organisms of concern should be included and at least one level that causes no effects that are considered ecologically significant. (based on expert consideration of ecological function and recovery).
- Variability should be as small as possible. If the variation between replicates is high, then the conclusions drawn from the study are less robust.

VALIDITY OF THE TEST
• The amount of test material applied and the exposure concentration in the water column should be determined analytically at t=0 (start of exposure). Whether or not concentrations are measured in other matrices, depends on the objectives of the study.

• The duration of the study should be appropriate to the life-cycle of the organisms of interest and the time needed for their recovery, if that is an objective of the study.

DESCRIPTION OF THE METHOD

Experimental systems

12. Outdoor meso- or microcosm studies can be performed with artificial tanks or ponds or by enclosing parts of existing ecosystems. When constructing a mesocosm facility it is a good idea to include one or more ‘supply ponds’ which can be used as a source of water, sediment and organisms. A size of 1 to 20 m$^3$ is usually regarded as appropriate for outdoor meso- or microcosm studies. Where planktonic species are the main concern, microcosms of between approximately 100 and 1000 litres can readily be used. It is also possible to use much larger mesocosms, although they may be more logistically demanding than smaller systems (eg Kersting & van Wijngaarden, 1999). The size to be selected for a meso- or microcosm study will depend on the objectives of the study and the type of ecosystem that is to be simulated. In general studies in smaller systems (about 1 - 5 m$^3$) are more suitable for shorter studies of up to three to six months and studies with smaller organisms (e.g. planktonic species). Larger systems are more appropriate for longer studies (e.g. 6 months or longer). The average depth of test systems depends on the objectives of the study, but generally should be between 30 cm and 1 m. In small systems, large daily temperature fluctuations in the test systems can be buffered by partial burial in the ground or immersion in a pond.

13. The microcosm or mesocosms can be constructed from any natural substrate or inert material, such as concrete (sealed appropriately), fiber- or plexiglass or stainless steel. Systems can also be lined with inert PVC to prevent exchange of water with the surroundings. Care should be taken to prevent leaching of plasticisers into the test waters; an epoxy paint may be necessary. In some locations, it may be advisable to cover the test systems with a net to prevent disturbance of the test systems by large birds such as herons or ducks.

Re-use of ponds

14. Re-use of ponds after treatment with toxic chemicals depends on chemical characteristics, particularly the persistence of the chemical, and on biological variability. For non-peristent chemicals there may be no problem if it can be demonstrated that there are no longer any toxic residues present in the water column and in the sediment and the systems can all be readily returned to closely comparable biological systems. Alternatively, the ponds can be drained and left empty for a period of time, or scraped out and relined with new sediments (Hill et al, 1994).

Sediment

15. Sediments should always be included in the test systems because they provide an important buffering element to the systems. Sediment can be collected from supply ponds or natural systems, in which case care should be taken to ensure that organisms which could interfere with the objectives of the study (e.g. fish or highly invasive or exotic macrophytes) are not accidentally introduced. Alternatively, soil can be used as the test system sediment, provided that it has been sufficiently conditioned to have aquatic sediment-like properties. Generally this requires a maturation period of immersion of 3 months or more and addition of a small inoculum of aquatic sediment to encourage the development of a suitable
microflora. Sediment collected from an uncontaminated reference site will contain indigenous flora and fauna which can be used to establish pond communities. If soil is used, then additional organisms may need to be added to develop suitable communities for the study. The sediment should be characterised by analysis of chemical residues, including metals and by determination of the particle size distribution, organic matter content. N and P content, cation exchange capacity, pH and organic carbon content could be quantified to further characterise the sediment. Prior to adding sediment to the individual units, it should be thoroughly mixed to ensure an even distribution of material and benthic organisms and thereby promote inter-replicate similarity. The sediment is added to the microcosm or mesocosms prior to adding water. The depth of sediment in each test system should be > 5 cm.

**Water**

16. If possible, at least part of the water, and preferably all of it, should originate from the zone where the sediment and its organisms were collected. As with the sediment, water used in the microcosm or mesocosm should be characterized for chemical contaminants, nutrients, pH, hardness, dissolved oxygen, and turbidity. When adding water to the test systems, it is advisable to do so gently so that the sediment quickly settles. If large systems are used, or the study director suspects that there may be high levels of variability between the replicates, it is possible to interchange water between microcosm or mesocosms. However, mixing should be discontinued sufficiently before treatment to ensure that the test systems are relatively stable without water exchange between the systems. After dosing, no water exchange should occur among the microcosm or mesocosms.

17. Water level should be maintained at similar levels (within about 20% of the original level) for the duration of the study. This should be done by replacing evaporated water with a well source, ‘conditioned’ mains water, or filtered pond water (preferably water with similar properties to that used to establish the test systems). In very wet weather conditions, test systems can be covered to prevent overflow. In emergencies, water can also be removed, but if this is after treatment, an estimate of the amount of residues removed should be obtained. However, it should be noted that the addition of small volumes of unfiltered water, containing plankton, may simulate natural immigration and water exchange.

**Organisms to include**

18. The microcosm or mesocosm study should focus on taxonomic groups that lower-tier risk assessments have identified as being of concern. Besides naturally colonised sediment, the test system is usually a naturally developed aquatic community with appropriate organisms such as zoo- and phytoplankton, periphyton, bacteria, macrophytes, pelagic and benthic macro-invertebrates. To develop communities suitable to meet the study objectives, it may also be acceptable to add organisms from appropriate external sources. Inclusion of free-living fish is not generally recommended, particularly where effects on zooplankton and macro-invertebrates are key study end-points. If direct effects on fish are of concern, laboratory single species studies are usually more appropriate than microcosm or mesocosm studies. The reason is that, in a confined ecosystem, fish tend to have an unnaturally large influence on the rest of the system. In tests of long duration or high fish biomass density, predation by fish may influence invertebrate responses to the chemical. In most cases, it is recommended that macrophytes should be included in the test systems, even if the objective is only to study phyto- and zooplankton. This is because macrophytes are an important structural and functional part of aquatic ecosystems, providing habitat for organisms and contributing to macro- and micronutrient cycling and influencing physico-chemical conditions.

**Macrophytes**
19. In microcosm or mesocosm studies where macroinvertebrates are the endpoints of major interest, macrophytes should be present, since they provide refugia for the microcosm or mesocosm fauna. Their presence is likely to encourage greater system stability and greater diversity of algae and invertebrates. For auxin-stimulating herbicides it is essential to include relevant macrophyte species, preferably those that develop roots (Brock et al. 2000). Another reason to include macrophytes in these studies is that effects on macrophytes can have indirect effects on the animal communities because of the interdependencies described above.

20. Natural colonisation is acceptable; propagules and seeds will be found within most introduced sediments. Planting macrophytes is also a possibility and this also has benefits because it promotes more uniformity of diversity, abundance and distribution. Planting mature macrophytes (e.g. from a supply pond) also enhances the rate at which the system can ‘mature’. Macrophyte development should be managed to ensure that the objectives of the study are met. For example, some floating (e.g. Azolla or Lemna) or submerged (e.g., Elodea) species can overdominate a system and significantly reduce animal diversity. If the objectives of the study are focused on the plankton, then it may be advisable to maintain areas of open water, limiting macrophyte development to no more than 50% of the bottom area; typically, 25 to 30%. If the study is to focus on macroinvertebrates it may be appropriate to promote the abundance and diversity of submerged macrophytes in order to enhance abundance and diversity of macroinvertebrate species (these are closely ecologically linked). It may also be wise to include emergent species, if aquatic insects are of concern because many aquatic insect species use these plants for emergence or egg-laying.

**Invertebrates**

21. Benthic and planktonic invertebrates will often be added to the microcosm or mesocosms with the sediments and water. Invertebrates typically studied in detail include the zooplanktonic phyla Rotifera and Arthropoda (Branchiopoda: Cladocera, Copepoda); and zoobenthic organisms from the phyla Annelida (Oligochaeta, Hirudinea), Mollusca (Gastropoda, Bivalvia), Arthropoda (Insecta e.g. Coleoptera, Diptera, Ephemeroptera, Hemiptera, Odonata, Trichoptera; Crustacea e.g. Isopoda, Amphipoda, Ostracoda, Decapoda), and Platyhelminthes (Turbellaria). Also, epibenthic invertebrates, and those invertebrates found growing on macrophytes (e.g. Bryozoa) could be studied in microcosm or mesocosm tests. Should it be necessary, any group can be added to the microcosm or mesocosms from field collections or laboratory cultures at the start of the study. Care should be taken to ensure a homogeneous distribution between microcosm or mesocosms prior to dosing. This can be accomplished by compositing samples and by using sample splitters.

**Fish**

22. If one objective of the study is to look at effects on fish populations in a microcosm or mesocosm test then it is recommended to use a larger microcosm or mesocosm. It may be useful to stock the microcosm or mesocosm with a low density of adults and remove adults and larvae after spawning. However, the life stage, number and biomass of fish added also depends on the purpose of the test. For example, should the emphasis be on an insecticide, larval fish may be added to monitor their growth in relation to the invertebrate food base. If applicable, the fish population should have a natural demographic structure and should not exceed the “carrying capacity” of the test system (Brock & Budde, 1994). Biomass densities should generally not exceed 2 g per m$^3$ (Touart, 1988; Fairchild et al., 1992).

23. If fish are included in the test system, they should be added after the test system is reasonably stabilised. Typically, this is from one to four weeks. Some further stabilisation is also recommended to allow the test system to adapt to the presence of the fish. If indirect effects on fish are to be studied, then it may also be possible to introduce caged fish. If free-living fish are included, it may be advisable to include an invertebrate ‘refuge’ where fish are excluded in order to maintain reasonable numbers of invertebrates.
24. Fish used should preferably be representative for fish species living in agricultural areas. Examples of appropriate fish species include bluegill sunfish (*Lepomis macrochirus*), fathead minnow (*Pimephales promelas*), sticklebacks (*Gasterosteus aculeatus*), mosquito fish (*Gambusia affinis*), carp (*Cyprinus carpio*), golden orf (*Leuciscus idus*) or rainbow trout (*Oncorhynchus mykiss*). Which fish species is most appropriate depends on the objective of the test and the size of the test system.

**Maturation time**

25. The period of time the microcosm or mesocosms are adapted prior to chemical dosing will vary with the size of the system and the origin of the introduced sediment/water. A sufficient degree of diversity of the system for the purpose of the study and a certain homogeneity between replicates should have been achieved before the study can begin.

**TEST DESIGN**

26. It is of general value in a generic risk assessment to obtain a dose (concentration) response. This allows wider use of the data under different conditions and different regulatory requirements. In general a microcosm or mesocosm study should include at least three, and preferably five concentrations, with at least two replicates per concentration. More concentrations may be required, depending on the slope of the dose-response curve, for taxa of interest. Statistical techniques may be strongest with more replicates since replication reduces uncertainty in interpretation of results, because test system variability can be better accounted for. However it is also possible to design valid studies with a large number of unreplicated treatment levels. The decision either to favour more replicates of each concentration or to prefer a concentration-response test design with less replicates and an increased number of concentration levels depends on the scientific questions to be answered and on the specific characteristics of the established ponds. (Chapman & Maund, 1996; Campbell *et al.* 1999; Giddings *et al.* 2002).

27. In designing the microcosm or mesocosm study it may be helpful to consult a statistician to help determine which test design is required if an effect in a particular set of measured parameters is to be determined with a specific power. This will be a function of the replicate number and variability of the measurements. A design optimal for one variable will not necessarily be appropriate for another. The importance of focussing on critical endpoints cannot be overemphasized. (Chapman & Maund, 1996).

28. At least one or two of the dose rates should be derived (by modeling, for example) from Good Agricultural Practice, but the experimenter should not be tied only to what is used in agricultural practice. The selection of treatment levels should aim to include at least one concentration that will cause no ecologically significant effects and at least one that will cause clear effects.

29. Data about the natural variability of the biological assemblage of the microcosm or mesocosm system should be available to design the study in such a way that a high number of species can be analysed.

30. Whether treatments should be assigned to experimental units at random or whether a constrained randomization should be employed, such as the arranging of treatments in replicate blocks, depends on the objective of the study (Chapman & Maund, 1996). If the individual microcosm or mesocosms are in a natural pond or aligned in a non-random manner, randomized blocks may be used to allow for a complete parametric analysis with minimum bias. As in any experimental treatment using chemicals, care should be taken not to contaminate other microcosm or mesocosms with test material.
Recovery

31. Determining the rate and extent of recovery of affected taxa can be an important factor in the design of microcosm or mesocosm studies. When considering recovery, it is important to understand the potential influence of life-history and dispersal mechanisms of the organisms involved and possible interactions of these with the exposure regime and tests system. For example, if application is carried out during or after the main reproductive period of univoltine organisms (e.g. certain mayfly species), if affected organisms do not have non-aquatic dispersal mechanisms (e.g. amphipod and isopod crustaceans), or if normal seasonal variations (e.g. phytoplankton) mean that the affected organism becomes absent from controls and treatments, then it may not be possible to demonstrate recovery potential without further experimental manipulation. In such cases, certain experimental techniques may be useful to assist in the evaluation of recovery potential.

32. Periodic reintroduction of organisms, eggs or resting stages into model ecosystems can be used to simulate immigration and reproduction. However this may have practical constraints (e.g. for seasonal reasons) and may also disturb the resident populations. It is also very important to consider whether such immigration would occur naturally for the species concerned. Alternatively, organisms may be placed in cages within the model ecosystem from time to time, to determine when the conditions in the model ecosystem have become suitable to support survival, growth, and reproduction. Another possibility to demonstrate recovery potential is to collect water or sediment from the test system and bioassay it in the laboratory (Campbell et al 1999).

PROCEDURES

Application of the test substance

33. Test substance is added into the test system after the microcosm or mesocosms have stabilized. For pesticides, whether or not to use the active ingredient or the formulation should be determined by earlier laboratory studies. For a generic risk assessment, use of the active ingredient is preferred, unless it is difficult to work with or if the formulation is significantly more toxic. Formulations should generally be used in studies simulating spray drift, but are inappropriate for studies simulating surface run-off. Regulatory objectives will help determine whether the active ingredient or a formulation should be used. (Heger et al. 2000).

34. Two approaches to applying the test substance to the microcosm or mesocosm can be identified.

(i) The first is the ‘toxicological approach’ which consists of direct application of the test substance to the water, usually including mixing, to achieve a uniform distribution. Exposure is then expressed in terms of the concentration of the test substance in the water as in a standard toxicity test.

(ii) The second is the ‘simulation approach’, mimicking the route of entry of the test substance. For pesticides, this amounts to simulating its entry into the water body as under good agricultural practice. For example: drift and direct overspray can be simulated by doing a spray application or erosive run-off can be simulated by making a soil-water slurry. In this case exposure is expressed in terms of the amount of pesticide added per unit surface area or volume (the loading rate).

In both cases concentrations of the test substance established in the water should be analytically determined. The loading (amount of test substance added), the frequency of dosing, and the number of
replicates per treatment necessarily stem from the nature of the chemical, use patterns, routes of entry, variability of the endpoint of concern and objectives of the study.

35. Accounting for multiple exposure events when applying the test substance presents a number of difficulties. Multiple application could be done in the test, however this makes it more difficult to characterise the exposure and interpret the exposure-response relationships.

36. When to apply the test substance depends on how realistic the exposure scenario is to be. It can be argued that it is worst-case to apply the test substance in the spring to midsummer, because the system is assumed to be most sensitive at that time. Applying in the spring also allows a longer time to determine the potential for recovery. Also, if the systems can be assumed to be more sensitive in the spring, the results can be more readily extrapolated to a situation where application is done in the autumn than the other way around. However, there may be a reason to expose the test system in the autumn if it is envisaged that a potential for recovery seen in a spring application has no relevance to the use of the product in the autumn.

**Poorly soluble materials**

37. Water is by far the preferred carrier for the test substance. A range of solvents that are used as carriers in the application of a test substance, can cause effects on the metabolism of model-ecosystems, even at very low concentrations (Brock *et al.* 2000). If a carrier is used, then the concentrations should be equal across all microcosm or mesocosms, including the controls. Controls with and without a carrier can help to determine whether or not the solvent influenced the test systems.

**Sampling**


39. In studies where many endpoints are measured more or less simultaneously, it is recommended to assign measurements and samplings to specific locations in the microcosm or mesocosms to avoid mutual influences or disruption of the individual sampling programmes (van Wijngaarden *et al.*, 1996).

40. It is recommended to take pre-treatment samples, for example on day -14, -7 and 0, in order to assess and demonstrate the suitability of the test system. These pre-treatment samples can also be used to perform covariate analyses in order to reduce residual variance among mesocosms (Chapman & Maund, 1996). Sampling continues after treatment for the duration of the test. The total test duration is dependent on the aim of the study, the fate properties of the chemical, recovery times of the populations of concern. Ideally the study should continue long enough to demonstrate the recovery of the affected species.

41. The sampling regime during the exposure period depends on the objectives of the test, the nature of the chemical(s), and the expected distribution of chemical within the microcosm or mesocosm. Table 2 gives an overview of parameters that should be considered in a typical microcosm or mesocosm test. Depending on the type of compound, certain parameters will need to be emphasised more than others. In tests focusing on invertebrate population responses, zooplankton and emergent insects may be sampled weekly. Minimising sampling stress is of crucial concern.
42. The sampling strategy should ensure that the collection of water samples will not significantly change the microcosm or mesocosm volume. Also the collection of biota specimens should not lower to a significant level the standing stock of the samples species or alter the trophic relationships in the existing food chain.

43. If freeliving fish are added to the system, they may be observed at daily or weekly intervals but generally collected upon test termination. There may be occasions in which marked (e.g., electro-tagged) or caged fish are repeatedly sampled for growth measures. However, mortality and effects can increase dramatically for fish repeatedly sampled. Alternatively, in large systems with many fish, it may be possible to take sub-samples to estimate growth rates.

**Biological measurement**

**Phytoplankton and zooplankton**

44. The planktonic biota should be sampled with a depth-integrating sampler. In small systems a pump or a plankton net could be used to filter plankton. When using a pump for sampling zooplankton, it should be confirmed that (especially the larger) zooplankton is not able to avoid the pumps inlet. In the presence of macrophytes, specific techniques are needed to collect zooplankton in their habitat. Subsamples are used to determine pigment composition, species identification, and cell counts, where appropriate. Population densities are reported as cells (or biomass) per volume. Adult zooplankton are identified to species, where possible and their abundance reported as individuals per volume unit. Several zooplankton taxa undergo microhabitat shifts on a diurnal basis; these may be sampled with an integrating sampler. In general, for planktonic samples, depth integrated samples should be used or at least samples need to be taken at the same time and depth in the different replicates. Epi-benthic species may be sampled with a trap designed to collect them (Lozano, et al., 1992).

**Periphyton**

45. Typically, periphytic biomass and productivity have been estimated using surrogate measures of phytopigments, chiefly chlorophyll-a. Phytopigments (e.g. chlorophyll-a, phaeophytin) are sampled either from "natural" substrates (e.g., macrophyte surfaces), unglazed ceramic tiles or glass microscope slides placed into racks and colonised in the microcosm or mesocosms, for two to four weeks. After exposure, the substrates are scraped and material can be analysed for species presence, pigment biomass, or ash-free-dry-mass. Additional chambers (e.g. light and dark bottles) can be included for estimates of production and respiration, should the emphasis of the test be on factors affecting primary production. The use of colonised substrates allows estimates of colonisation rates and standing crop throughout the test schedule. A series of slides can be pre-exposed within the microcosm or mesocosms and, by collecting slides at frequent intervals, used to monitor changes in periphyton biomass, species composition, and net production. The substrates chosen for use will influence the nature of the periphyton collected. Hence, careful thought should be given to the objectives in estimating primary production or periphytic biomass.

**Primary productivity**

46. If a chemical is expected to cause direct toxicity to algae (e.g. an herbicide), measures of primary productivity should be used. The most practical method is to measure diurnal oxygen fluctuations. It may also be useful to compare primary productivity with the level of light saturation, in order to explain effects seen. At the same time it is also important to analyse algae populations to a sufficiently low taxonomic level to be able to detect changes in species composition of primary producers. If possible total system metabolism should be established (Hill et al., 1994).
Macrophytes

47. If macrophytes are a study endpoint, e.g. in the case of an herbicide study a system could be set up looking at subsamples for biomass estimates. Care should be taken however not to disturb the system too much. It may also be possible to plant macrophyte species of interest in pots placed within the sediment or on the sediment, or suspend at half water depth. In such cases, the entire pot may be removed for measurement of shoot growth, biomass or photosynthesis. During the study estimates of abundance can be done visually (e.g. by mapping or photographically).

Heterotrophic component

48. If effects are expected on primary producers such as macrophytes, it may be informative to look at indirect effects on bacteria and detritivores. Specialised techniques are needed to study structural aspects of microbial populations (Wetzel & Likens, 1991). In practice it may often turn out that impact on microcosmorganisms can only be studied by measuring process rates. (Heger et. al. 2000).

Macroninvertebrates

49. Macroninvertebrates can be sampled using artificial substrates, nets, direct sampling of sediments, or by using emergence traps. Fast-moving invertebrates can best be sampled directly with a net in the water column. In systems smaller than 10m³ it may be advisable not to sample sediments directly. Disturbances to the sediment structure, macrophytes, recycling of nutrients and test material due to sediment sampling may disturb the test system. In these smaller systems, placing sediments in trays prior to application and retrieving the sediment after a period of colonization is one way in which the sediments can be assessed without unduly disturbing the system. To fully characterize the benthic invertebrate fauna, all three methods should be employed in concert. The numbers of collected invertebrates are reported on a per sample basis. Insect emergence rates are reported as numbers emerging per unit of time or per unit area.

Fish

50. Where fish are included, frequent observations should be made at the initiation of the test for dead fish and for abnormal behaviour. Following arrival at the testing facility, fish should be acclimated to the water of the test system for at least a week, before being introduced into the test system. After acclimation, fish should be observed at least weekly and dead fish removed. Fish that die as a result of handling stress or disease may be replaced within the first week of their introduction. At the termination of the test, all fish are collected, counted, measured and weighed. Depending on the test substance and the objectives of the study, fish could be sampled at several time intervals during the study as well, e.g. to measure fish growth. Abnormal growth, external lesions or abnormalities are recorded. Tissue subsamples may be taken for residue analysis, if potential risks of bioaccumulation are indicated in the lower-tier assessments.

51. There is no standard technique presently accepted for determining effects on fish growth and reproduction in microcosm or mesocosms. In designing such a test, it would be necessary to evaluate how best to match biological sampling with the nature of the chemical(s) and also the life history of the organism in question. If the test were to focus on effects on fish early life stage development, the test could be started with eggs or larvae. Test duration would be a function of the fish biomass loading and specific objectives: examples include feeding responses on zooplankton, prey-switching at a critical size, and changes in competitive behaviour as a function of chemical exposure. If the test were to observe effects on reproduction, adult fish at a low stocking density may be added, allowed to spawn, and subsequently collected with their offspring.
In situ bioassays

52. In situ bioassays can provide information on direct effects that can be related to exposure. They may be able to provide information on indirect effects, e.g. in the case of caged fish. In situ bioassays can also be used to compare laboratory and microcosm or mesocosm responses in the same species. They should, however, not dominate the system. The effect of the presence of in situ bioassays on abundance of free-living populations and the sampling programme should be small. (Giddings et al. 2002).

Analytical measurements

Water quality

53. Measurement of water quality (e.g. dissolved oxygen, pH, alkalinity) and nutrient concentrations help to define the ecosystem functioning of the microcosm or mesocosms and help to interpret chemical fate and bioavailability.

Analysis of test chemical

54. The study objectives will determine the appropriate sampling and analysis strategy for the test chemical (Giddings, 1994). There are three different reasons for doing chemical analysis: (i) to confirm that the test substance has been accurately applied to the test system (para 55) (ii) to quantify the chemical exposure and relate it to the ecological responses observed (para 56) and (iii) to look at the chemical’s fate in the aquatic environment under natural or semi-natural conditions (para 57).

55. If the test chemical is soluble in water, and is added directly to the water of the microcosm or mesocosm, then measurement of concentrations in water within a few hours after application can provide confirmation of treatment. Vertically integrated water column samples can be taken from a sufficient number of points in the microcosm or mesocosm to allow calculation of the average chemical concentration in the water. A similar approach can be taken to measure insoluble or highly sorbed chemicals that are added directly to sediment or are mixed with soil before addition. A considerable sampling scheme may be necessary to overcome the spatial variability in concentrations in the sediment.

56. Quantifying exposure may be done by taking three or four samples during one or two half-lives of the test substance, but at least 5 samples before 90% of the substance tested has disappeared. This is done to estimate the temporal pattern of exposure. Subsequently monitoring is continued at a reduced frequency until concentrations fall below the level of biological concern. Analysis of water is most commonly done; sediment analysis may also be important if the chemical has a high partition coefficient, degrades slowly in the sediment or is suspected to be toxic to benthic organisms.

57. Depending on how much information is available before a microcosm or mesocosm study has begun, it may be necessary to obtain information about a) routes and rates of degradation b) chemical partitioning between water and sediment and between dissolved and particulate phases and c) uptake of test chemical by fish and other biota. Sampling should be done dependent on which aspects of fate are to be investigated. If the test substance bioaccumulates it may be necessary to take samples from macro-invertebrates, emerging insects or fish. It is recommended to attempt to reconstruct the mass balance of the test chemical in order to account for all the test material introduced into the test system. This may be done with the aid of a radiotracer (Giddings, 1994), however this is not a requirement.

58. If possible, extraction of residues from water samples should be initiated immediately by the addition of a suitable solvent, when laboratory studies indicate rapid transformation of the test material in natural waters. Otherwise, samples should be refrigerated and extracted as soon as possible. Sediment samples should be frozen immediately. To check stability of test substance during storage, water and
sediment samples from untreated areas should be fortified with analytical standards, stored and analysed in the same manner as samples from treated areas. See also ISO guideline ISO/DIS 10381-6 (1993).

DATA AND REPORTING

Statistical methods

59. Univariate analytical methods, such as ANOVA or regression analysis (or a combination of the two, e.g. William’s test) are best suited to investigate parametric data on effects at the population level of one species or taxon. The power of these methods to detect differences from the control response, should be stated. (Liber et al., 1992).

60. Multivariate analysis is appropriate for describing effects at the community level and can also be employed to indicate which taxa are particularly sensitive to treatment and would warrant specific univariate analysis. One method of multivariate analysis is based on the construction of Principal Response Curves (van den Brink & ter Braak, 1998, 1999) in which canonical coefficients for the identified taxa are plotted against time. This analysis takes into consideration the separate variances between replicates, between time-points and between treatments, thereby allowing clear representation of treatment effects in isolation. This pictorial evaluation of treatment effects can then be converted to an NOECcommunity with statistical significance using Monte Carlo permutation tests.

61. Another method for evaluating microcosm or mesocosm studies is by determining similarity and diversity indices. Whichever method, or combination of methods of analysis is to be adopted, it is essential to build the analytical technique into the design of the study at the outset rather than to search for an appropriate method after the data have been generated.

62. It is recommended to analyse data from microcosm or mesocosm tests both by means of univariate and multivariate techniques to allow an evaluation at the population and the assemblage level. It may then turn out that some NOEC’s from individual species may be lower than the community NOEC. In that case information on the ecological role, legal protection status and the ability to recover of these species should guide a decision on the acceptability of the effect.

63. It is recommended that the performer of a meso/microcosm study spend time to become familiar with the range of methods available for analysis before embarking on any one method. Sparks et al, 1999 provides a comprehensive overview of the range of ordination techniques, methods for the multivariate analysis of grouped data and methods that are available to examine relationships between groups of data. Further recommended literature includes: Chatfield and Collins, 1980; Digby and Kempton, 1987; Krzanowski, 1988 and Manly, 1986; Sokal and Rohlf, 1995.

Appropriate levels of taxonomic resolution

64. Those taxa that are most sensitive should be identified to species (where practicable). Species-level identification may also permit effective use of multivariate statistical approaches for more powerful analysis of community structure. The appropriate level of taxonomic analysis therefore depends on the objectives of the study. Univariate statistics may not be sufficiently powerful to detect differences among groups of rare organisms. In this case, data may be aggregated into larger taxonomic groups before analysis. (Giddings et al. 2002).

65. Other groups of organisms that are identified as less sensitive in lower tiers (e.g. algae in tests with insecticides) may be monitored less intensively (e.g. at a lower level of taxonomic resolution such as
family) or at a community level (e.g. chlorophyll \( a \) concentrations). Identification of organisms should be possible with available taxonomic keys and without breeding larval or nymphal forms through to older stages. Subsamples from abundant species (e.g. chironomids) may be analysed in detail to reduce counting labour and the proportion subsampled should be reported. It may become possible to analyse effects more rapidly in samples taken from microcosms by using flow cytometry for algae and image analysis for invertebrates. Currently such non-taxonomic methods are difficult to use because the significance of any measured change is unknown.

**Structural versus functional endpoints**

66. In general it can be said that conservation of function of an ecosystem is more robust than conservation of structure. However, when testing photosynthesis inhibiting herbicides for example functional endpoints (e.g. oxygen levels) may be more sensitive than structural ones (e.g. densities of algae and biomass of waterplants). In addition functional endpoints are integrative and their measurement can provide indications of the severity of impacts and consequences for the whole ecosystem. It is therefore important to evaluate structural effects in relation to the loss/maintenance of ecosystem function.

**Reporting requirements**

67. The final report should give a full and comprehensive description of the study, including its objectives, design and results. Along with description of the analytic and statistical techniques employed, some or all of the following data should be reported, depending on the objective of the study:

Information on test substance and relevant metabolites:

- identification, including chemical name;
- batch or lot number;
- degree of purity;
- chemical stability under the conditions of the test;
- volatility;
- specific radioactivity and labeling positions (of appropriate);
- method for analysis of test substance and transformation products including limits of analytical detection/quantification;
- physico-chemical properties of the test substance, partition coefficients, rates of hydrolysis, photolysis, etc.

Test systems

- Description of test systems, location, history, dimensions, construction materials, general watershed characteristics, etc.;
- water levels and circulation;
- water quality: description of the chemical/physical parameters of the water used in the test system;
- colonization and introduction of biota;
- sediment characterization;

Experimental design and measured data:

- treatment regime: dosing regime, duration, frequency, loading rates, preparation of application solutions, application of test substance, etc.;
- sampling and analysis, residue monitoring results, analytical method;
- meteorological records;
- Physico-chemical water measurements (temperature, oxygen saturation, pH, etc);
- Sampling methods and taxonomic identification methods used;
- phytoplankton: chlorophyll-a; total cell density; abundance of individual dominant taxa; taxa (preferably species) richness, biomass;
- periphyton: chlorophyll-a; total cell density; density of dominant species; species richness, biomass;
- zooplankton: total density per unit volume; total density of dominant orders (Cladocera, Rotifera and Copepoda); species abundance; taxa richness, biomass;
- macrophytes: biomass, species composition and % surface covering of individual plants;
- emergent insects: total number emerging per unit time; abundance of individual dominant taxa; taxa richness; biomass; density; life stages;
- benthic macroinvertebrates: total density per unit area; species richness, abundance of individual dominant species; life stages;
- fish: total biomass at test termination; individual fish weights and lengths for adults or marked juveniles; condition index; general behaviour; gross pathology; fecundity, if necessary.

Data evaluation:

- Endpoints of the study;
- Statistical methods used;
- Treatment effects with univariate techniques;
- Treatment effect with multivariate techniques;
- Treatment effects with Similarity and Diversity Indices;
- Graphic presentation of results;
- Ecological significance of observed effects;
- Recovery of populations (observed or inferred), with a discussion of relevance to natural recovery processes;
- The highest concentration of test substance that caused no ecologically significant effects.
LITERATURE


ECOFRAM Aquatic report. In prep.

ECOFRAM Terrestrial report In prep.


WILLIAM’S D.A., (1971) A test for differences between treatment means when several dose levels are compared with a zero dose control. (Biometrics, 27, 103-117).

Table 1: Overview of useful models for the purpose of estimating pesticide concentrations in water and soil as discussed by FOCUS in Europe (Forum for the Co-ordination of Pesticide Fate Models and their Use) and ECOFRAM in the US
(The Ecological Committee on FIFRA Risk Assessment Methods)

<table>
<thead>
<tr>
<th>MODEL TYPE</th>
<th>Leaching models</th>
<th>Runoff Models</th>
<th>Spray Drift Models</th>
<th>Surface water receiving models</th>
<th>Surface water watershed models</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECOFRAM</td>
<td>CHEMRRANK, CMLS, PATRIO, SCOGROW, PRZM3, LEACHM/LEACHI, GLEAMS, MACRO</td>
<td>GENEEC, PRZM3, EPICWQ, RICEWQ, GLEAMS</td>
<td>Ag DRIFT, FSCBG, AGDISP</td>
<td>GENEEC, EXAMS, WASP5, RIVWQ</td>
<td>SWAT, EXAMS, SWRRBWQ</td>
</tr>
<tr>
<td>FOCUS</td>
<td>PRZM3, PEARL, PELMO, BAM, PERSIST, FUGACITY, EFATE, TWA</td>
<td>EPIC, GLEAMS, OPUS, PELMO, PRZM3, SWRRBWQ</td>
<td>IDEFICS, MOPED, PEDRIMO, PSMDrift</td>
<td>EXAMS, WASP</td>
<td>ABIWAS, SLOOTBOX, EXAMS, WASP, TOXSWA</td>
</tr>
</tbody>
</table>
Table 2: An example of some typical parameters measured in Microcosm or mesocosm studies and their frequency of measurement ¹

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Suggested frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Water quality</strong></td>
<td></td>
</tr>
<tr>
<td>Water level, pH, temp. DO, turbidity, conductivity, alkalinity, hardness, suspended solids, nutrients (dissolved C)</td>
<td>At least every two week intervals.</td>
</tr>
<tr>
<td>Pesticides, heavy metals</td>
<td>At initiation of the test</td>
</tr>
<tr>
<td><strong>Sediment quality</strong></td>
<td></td>
</tr>
<tr>
<td>Pesticides, metals, particle size, ion exchange capacity, organic content, pH</td>
<td>At initiation of the test</td>
</tr>
<tr>
<td><strong>Phytoplankton</strong></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a /phaeophytin/dry wt., cell counts and additionally species-diversity for long-term tests. Cell density</td>
<td>At least every two weeks</td>
</tr>
<tr>
<td><strong>Periphyton</strong></td>
<td></td>
</tr>
<tr>
<td>Chlorophyl a + phaeophytin + dry wt; Cell counts</td>
<td>At least twice during the study</td>
</tr>
<tr>
<td><strong>Macrophytes</strong></td>
<td></td>
</tr>
<tr>
<td>Identify and estimate abundance visually (+ photographically) Biomass by fresh weight</td>
<td>At infrequent intervals during the peak growth period and at finish</td>
</tr>
<tr>
<td><strong>Macro-invertebrates</strong></td>
<td></td>
</tr>
<tr>
<td>Benthos</td>
<td>Every two weeks; For adult insects; weekly during peak emergence time and at time of test substance application – during other periods less frequent sampling may be sufficient.</td>
</tr>
<tr>
<td>Artificial substrates + emergent insects Optional grabbing Identify to lowest practical taxon</td>
<td></td>
</tr>
<tr>
<td><strong>Zooplankton</strong></td>
<td></td>
</tr>
<tr>
<td>Identify species if possible Density and biomass Record life stages</td>
<td>Weekly</td>
</tr>
<tr>
<td><strong>Fish</strong></td>
<td></td>
</tr>
<tr>
<td>Length/weight</td>
<td>At beginning and end</td>
</tr>
<tr>
<td>Dead fish weighed and measured Gross pathology Sex/fecundity if relevant</td>
<td>At end</td>
</tr>
<tr>
<td><strong>Residues (test substance concentration)</strong></td>
<td>Test material + degradates Frequency dependent on compound. More frequently in the beginning than at the end of the experiment.</td>
</tr>
<tr>
<td><strong>Meteorology</strong></td>
<td></td>
</tr>
<tr>
<td>Air temp, solar radiation, precipitation, windspeed</td>
<td>At appropriate intervals, on site if possible.</td>
</tr>
</tbody>
</table>