Design, analysis, and interpretation of higher tier risk assessment of chemicals in aquatic microcosms

Sanderson, Hans

Publication date:
2002

Design, Analysis, and Interpretation of Higher Tier Risk Assessment of Chemicals in Aquatic Microcosms

Hans Sanderson

2002

Department of Environment, Technology and Social Studies
Institut for Miljø, Teknologi og Samfund
Roskilde University, Denmark

Design, analysis, and interpretation of higher tier risk assessment of chemicals in aquatic microcosms

Doctoral dissertation by Hans Sanderson

Supervisor: Associated Professor Jette Rank

Department of Environment, Technology and Social Studies
Institut for Miljø, Teknologi og Samfund
Roskilde University
Postbox 260, 4000 Roskilde, Denmark
Phone: +45 46 74 24 96, Fax: +45 46 74 30 41, E-mail: hanss@ruc.dk

February 2002

Design, analysis, and interpretation of higher tier risk assessment of chemicals in aquatic microcosms

Hans Sanderson

ABSTRACT

Microcosms have been used in ecological and ecotoxicological science and research for decades. During the 1980s and 1990s, the methodology was gradually evaluated for implementation within the framework of higher tier risk assessment of pesticides both in the United States of America and within the European Union. However, in 1992 the United States Environmental Protection Agency decided to discontinue the requirements for microcosm testing for registration of pesticides and other chemicals primarily due to paramount inherent uncertainties related to the design, conduct, analysis, and interpretation of microcosm experiments. The scientific problem related to microcosm experiments is still the determination of the relative ecological (accuracy) versus statistical (precision) significance of the microcosm design. The aim of the present doctoral dissertation is thus to elucidate the inherent ecological and statistical significance trade-offs in three spatially different microcosm experiments. A further aim is to quantify the uncertainty of the three spatially different designs by using statistical power analysis. Finally, it will analyse a scientific application of precautionary approaches to manage uncertainty from an environmental regulator and policy perspective.

The exordium paper in this dissertation provides a state-of-the-art literature review (1985-2000) on the replicability of microcosm studies for risk assessment of pesticides. The review focuses on the calculation of coefficients of variation (CVs) within the studies as a measure of replication and thus the ability to detect subtle effects. Subsequent to this, a series of microcosm studies were initiated at three independent levels of ecological significance and relevance tentatively set at 25m^2 earthen outdoor ponds > 12 m^3 PVC outdoor ponds > 30L transparent PVC indoor tanks. The statistical power of each level of ecological significance was determined, bearing in mind that a direct comparison of the power is context dependent and impeded due to the unknown magnitude of the extrapolation between the systems and between the compounds used in the tests. Finally, I discuss the scientific application of a reflexive science based on power analysis, analysis of needs and post-normal science in light of precautionary approaches for the environmental risk management of uncertainty and lacking knowledge.

The dissertation presents the following new information. A review illustrates that a majority of microcosm studies suffer from insufficient replication and high CVs, which hampers the detectability of the study. Hence, if the natural background variation amplitude among replicates (s^2) are great, then will the amplitude due to toxic effect size (Δ) will need to be even larger to significantly ($p \leq 0.05$) break-through the noise of natural variation. This could again lead to low statistical power ($1-\beta < 0.8$) and an unacceptable or unknown high risk of a type II error (false negative) (paper I). A similar trend was observed in the three subsequent microcosm experiments I conducted or reviewed, the general pattern in terms of statistical significance is inverted to the ecological significance. The applicability of aquatic microcosm studies and plankton for risk assessment of Roundup is questionable due to rapid removal of the compound from the water column (paper II). The dissertation reports the first published higher tier risk assessments of two dominating perfluoro surfactants (paper III & IV). Paper V is the first published review and ranking of ten different phyto-toxicological endpoints according to ASTM guideline 1913-97 E in microcosm experiments. Finally, the dissertation discusses a scientific application of precautionary approaches via power analysis for handling uncertainty in light of the current EU position, and reflexive modernisation of science (paper VI-VIII). Overall, the dissertation as a whole presents new and quantified aspects of uncertainty associated to higher tier risk assessment of chemicals in aquatic microcosms, plus guidance/recommendations for design, analysis and interpretation of higher tier risk assessment of chemicals in aquatic microcosms.

The dissertation concludes that due to economic and logistical constrains it is generally easier and more feasible to increase power in small scaled microcosms by increased replication. There is a paradox between an increased ecological significance and the need for high effect sizes and thus unrealistic treatment concentrations in microcosms as opposed to realistic exposure under less ecological relevant single species laboratory conditions. Power analysis could be implemented as a short term scientific application of reflexive and precautionary approaches with an extended community peer-review system where the acceptable (Δ) and risk of type I and II errors (α & β) are *a priori* defined based on sound ecological expert judgement for protection of system integrity, societal cost-benefit analysis, and ethics. Hence, after a pilot study a successful microcosm design could be designed with high power. The design of a microcosm experiment is a context dependent evaluation of the relative importance of statistical versus ecological significance and protection aims.

Finally, the thesis reflects upon the future role of microcosm experiments in environmental protection. If, the environmental agencies pay attention to the statistical power of the studies and *a priori* define the entities of the power equation, then microcosm studies are a good higher tier risk assessment test. If they fail to do this environmental protection could be jeopardised due to high risk of type II errors. For research and education they seem invaluable in pursuing better knowledge of manmade chemicals effects on a suite of species, populations and ecosystems, which can only be tested and validated in model ecosystems like microcosms. Moreover, they can be used to develop ecology and ecotoxicology and increase our understanding of extrapolation over time, in space among species, different levels of biological organisation and ecosystems.

Keywords: Microcosm, Higher tier risk assessment, Design, Statistical power analysis, Type II error, Plankton, *Myriophyllum*, ASTM # 1913-97 E, PFOA, PFOS, $Roundup_{2000}$, Precautionary principle.

Table of contents

Introduction

Introduction

The idea of using aquatic model ecosystems, or micro- and mesocosms for higher tier risk assessment of pesticides for registration (and recently also other chemical compounds) originated from a critique against standardised single species laboratory bioassays in the early 1980s (Cairns, 1984) (Kimball & Levin, 1985). The aim was to supplement the single species tests with a higher tier with increased ecological realism of the test and thus reduced extrapolation gap from the test situation to the environment. Microcosm studies could be initiated whenever a pesticide, based on the lower tiers, was thought to impose unacceptable risks for aquatic organisms. Then the higher tier test could confirm or reject the level of acceptability. If the risks from the lower tiers were not confirmed by the microcosm test, the microcosms results would overrule the data from the lower tier (Touart, 1988). However, during the early 1990s the requirements for microcosm testing for registration of pesticides were pulled out of the U.S risk assessment procedure due to lacking interpretability of the results (Fisher, 1992). Meanwhile, smaller scaled microcosm studies were adopted in the EU regulation of pesticides (Dir. 91/414EEC). Ever since, there has been an intense scientific debate of how a reintroduction of the studies in the U.S and a more widely and correct usage of microcosm studies in the EU could be achieved. This is reflected in three recent workshops under Society of Environmental Toxicology And Toxicology (SETAC) and the United States Environmental Protection Agency (USEPA). Often, the uncertainties of microcosm studies are not addressed in quantitative terms but rather in somewhat vague qualitative ditto referring to the need for expert judgements, which is not always feasible from a risk management point of view, due to time and costs constrains (paper I). In addition, this may be legally questionable due to the subjectivity by the expert.

The present thesis will address the usage of microcosms for risk assessment of chemicals in three different scales of ecological realism and attempt to quantify context dependent relative trade-offs between statistical and ecological significance at these scales. The aim is to give empirically based (papers I-V) comments to the latest guidance publications by SETAC-Europe (HARAP & CLASSIC) on design and analysis of microcosm studies. Ecotoxicologists and ecologists, working with ecological risk assessment of chemicals are the principal target group for this thesis.

Statistical power analysis was used to determine the required sample size (number of replicates) and necessary effect size to achieve high power ($1-\beta > 0.8$). The implementation of potentially variable data from microcosm studies is considered in the light of the precautionary principle and the interpretation and implementation of the precautionary principle is also discussed in light of power analysis (paper VI-VIII).

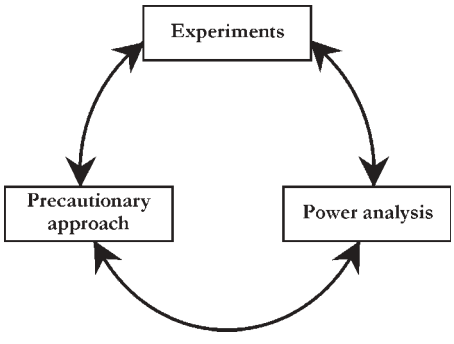

Figure 1: The analytical framework of my thesis can be summarised in the following train of thoughts.

The overriding problemformulation of my thesis could thus be summarised as:

1) To demonstrate experimentally by quantification the trade-offs between statistical and ecological significance of microcosm studies.
2) To comment on recommendations regarding current use of microcosm studies for higher tier risk assessment of chemicals.

In a non-ideal world (constrained by limited resources), I started the experiments with fixed number of replicates (two, three and five, respectively) and worked my way through these and the power analysis and discussed the results and reflected on the methods in light of a precautionary approach. It is important to note that most, or all, higher tier risk assessments are context dependent. This is also the case with the studies of the present thesis. Thus a direct comparison between results is dependent on the effect sizes induced by the compounds being tested. A large effect size results in relatively higher statistical power. Hence an interpretation is still subject to expert judgement due to different effect sizes induced by the compounds tested. Moreover, due to the inherent context dependence, I was not looking for *a* correct design, but to quantify the relative ecological and statistical significance of each experiment, and discuss this in my thesis.

Each paper and experiment in the thesis furthermore have its own individual aims. The first paper (I) was concerned with quantifying the level of Coefficient of Variation (CV) in the scientific literature. I made a review of microcosm studies risk assessing pesticides from 1985-2000. Out of the 129 papers I reviewed 17 were suitable for review, the main threshold were tables with results instead of not precisely readable graphics. Moreover, the review should determine if CV levels could be used to accept or reject studies, which were indicated in the ECOFRAM (1999) report.

Paper (II) aimed at determining the applicability of quasi-natural mesocosm ponds and pelagic plankton for higher tier risk assessment of $Roundup_{2000}$. Despite being one of the most used herbicides in the world the only published aquatic semi-field risk assessment of Roundup was a study conducted more than twenty years ago by Hildebrand et al. (1980). They did not find any effects at a 100 times recommended agricultural dosage, however they neglected to monitor the concentration of the active ingredients in their experiment. Could it be that no effects were expectable due to rapid removal of the compound from the water column and thus lacking exposure of the *Daphnia magna* they monitored, and are quasi-natural and pelagic plankton applicable for higher tier risk assessment in aquatic microcosms?

Papers (III-IV) aimed at an ecological risk assessment of perfluorinated surfactants (PFOS & PFOA) in aquatic microcosms. These very persistent, bioaccumulative and widespread compounds have never previously been risk assessed in microcosms and the research-team I work with are so far the only ones risk assessing these compounds under semi-natural conditions.

Paper (V) aimed at sorting the ten different endpoints of growth inhibition of *Myriophyllum sibiricum* and *Myriophyllum spicatum* according to ASTM guideline 1913 97E, accordingly to relative statistical power and detectability and relative ecological relevance of each endpoint. Furthermore, to determine the relative contribution to variability from both the inherent genetic variability of the plants and the natural variability stemming from the microcosms design.

Paper (VI-VIII) aimed at discussing an scientific and statistical interpretation and implementation of the precautionary principle via statistical power analysis as a mean of contribute a reflexive science, advocated by Ulrich Beck (1992) in his book *„Risk Society"*. Furthermore, integration of science and the precautionary principle was debated, leading up to a SETAC position paper on science and precautionary approaches.

I chose to publish my review of microcosm studies for risk assessment of chemicals (paper I) in the ECOMED journal *Environmental Science and Pollution Research international*, which was chosen because its a broad and science based journal with reference to environmental policy and decision-making and interdisciplinary analysis. This paper was thus also chosen for my analyses on the precautionary principle (papers VI-VII). I published the risk assessment of $Roundup_{2000}$ (paper II) to the Springer-Verlag journal *Archives of Environmental Contamination and Toxicology*, because the comprehensive review paper on environmental risks associated to Roundup was published in a sister journal to *Archives* under Springer-Verlag *Reviews of Environmental Contamination and Toxicology*. The in depth risk assessment of a perfluorinated compound (PFOS) and phyto-toxicology papers is published in the SETAC journal of *Environmental Toxicology and Chemistry* (papers III & V), as a broad and highly qualified target group of scientists from academia, industry and administration. Paper IV containing the risk assessment of PFOA was published in the Elsevier Science journal of *Aquatic Toxicology*. This was chosen to broaden my ecotoxicological target group not only to include SETAC members and readers of *Environmental Toxicology and Chemistry*. The SETAC position paper on Precaution and science is being prepared for the SETAC Globe and later for *Environmental Toxicology and Chemistry*. Finally, I will submit my thesis to SETAC as an empirically based comment of the CLASSIC workshop recommendations.

1.1 Historic and regulative overview of microcosms in risk assessment of chemicals

Hazards associated with chemical substances have challenged mankind for centuries. Over 2000 years ago, Pliny the Elder described a sickness of the lungs that affected slaves who wove asbestos into cloth. In the 16th century Paracelsus expanded medicine and toxicology as the „Luther of medicine" during his difficult and wandering life during the reformation. By 1713, Bernardino Ramazzini had published his work *De Morbis Artificum*, with detailed description of some of the health effects associated with workplace exposure to toxicants. In 1775, Percival Potts described the association between cancer and exposure to soot in London chimneysweepers (Draggan & Reisa, 1980).

The ever widening recognition of chemical hazards has led to contemporaneous theories of endocrine disrupting compounds described by Theo Colborn (1996) in the book „*Our stolen future: Are we threatening our fertility, intelligence and survival? – A scientific detective story*."

The twentieth century has seen a rapid increase in the human population and to satisfy consumption needs, intensive agriculture has been stimulated. The use of agrochemicals (fertilisers & pesticides) was greatly expanded to increase crop productivity in a cost-effective way. The use of pesticides caused environmental problems which were first widely documented by Rachel Carson in her book „*Silent spring*" (1962). The title of her book refers to the pesticide used to combat bark beetles carrying Dutch elm decease on Michigan State University campus. Besides the depletion of the eggshells caused by the spraying of DDT, the niche for birds living of beetles was destroyed and the birds

disappeared and the next spring was silent, eventually all the elms were infected and had to be cut down. So what she actually identified was an unintended adverse indirect ecosystem effect following pesticide usage.

Agricultural pesticides are, as the name indicates, chemicals designed and deliberately released into the environment to control pests that harm crops. This mode of action implies that they may reach non-target areas and organisms - some argue that only 1% of the sprayed pesticide reaches the intended target pest (Levin & Kimball, 1984). Aquatic ecosystems serve as sinks in the agricultural landscape and become contaminated by pesticides due to spray drift, drainage, run-off, atmospheric deposition and/or accidental spills. Since aquatic ecosystems include keystone species related to the target organisms of pesticides, undesirable side effects on aquatic plants and animals may ensue. Consequently, authorities have set criteria to protect aquatic wildlife from pesticide stress in the EU and the US.

The first tier in the EU registration system is the identification of Predicted Environmental Concentration/Predicted No Effect Concentration (PEC/PNEC). If the PEC does not exceed the PNEC, no effects of the pesticide on the aquatic community are expected. On top of these are multiplied uncertainty factors usually 1/100 for acute EC50s or a factor 1/10 for chronic NOEC's. This procedure is generally considered conservative partly because the higher dissipation rate and generally lower bioavailability of pesticides in the field compared to the standardised test conditions in the laboratory and partly due to the worst-case assumptions underlying the determination of PEC. Finally this approach is assumed conservative due to the multiplication of an uncertainty factor. Therefore, if the first tier indicates potential risks, European guidelines for admittance of pesticides on the market offer the possibility to include ecologically more relevant data in an advanced risk assessment procedure, the second tier. Experiments at the ecosystem level are frequently requested and performed to demonstrate that the actual risks of a particular pesticide are acceptable when used under normal agricultural practice (van den Brink, 1999) (Dir. 91/414/EEC).

The lack of ecological realism in the use of single species standardised laboratory tests is often disputed (Cairns, 1988) as are the extrapolations of ecological risks from the laboratory to real world environments. Ideally, the effects of pesticides could be evaluated at ecosystem level under natural conditions in field monitoring programs. However, this approach would be reactive and not proactive and not feasible practically as non-impacted reference sites are rarely available and obscuring variables often makes these field observations difficult to interpret in terms of causal effects, moreover we would risk further potential contamination and ramification of our environment.

Another approach is the use of man-made experimental model ecosystems: mesocosms or microcosms. They are made up of parts of natural ecosystems, which are brought together in a container (earthen ponds, PVC tanks or aquariums etc.) and are left alone to develop into a system that is complex enough to serve as a model for a natural ecosystem in terms of structure and function. Ideally, they serve as a bridge between the laboratory and field across the spatial extrapolation gap. They are manageable, allow replication and hence experimental set up on the one side and on the other side they provide realism in terms of ecological processes and exposure to the chemical (Brock et al., 1995).

The intention is to answer the legitimate question of „*-So What and/or -What if?*" (Cairns, 1984; Kimball & Levin, 1985) addressed to simplistic laboratory

testing, both in terms of whether the initial risk assessment was justifiable or conservatory, and also to focus on secondary ecosystem effects, often posed by risk managers (ECOFRAM, 1999). These two questions can always be posed to an environmental study or result *e.g. -So what,* the *Daphnia* dies under laboratory conditions, but what about the rest of the ecosystem or natural populations? Or -*What if*, the conditions in nature will hamper the effects shown in the laboratory?

The exact definition of, and making the distinction between, microcosm, mesocosm and macrocosm are not easy. Tentative definitions have been attempted declaring microcosms to be experimental ponds or tanks with a water volume less than 15 m^3, and mesocosms being systems larger than 15 m^3 (Crossland, 1994). The macrocosm is the real world. The earliest reference found concerning the use of the term microcosms in a similar fashion to contemporary use defined lakes as naturally occurring replicas of all aquatic environments was Forbes (1887) who described a lake as a little world within itself. „A microcosm within which all the elemental forces are at work and the play of life goes on in full, but on so small a scale as to bring it easily within mental grasp". This statement also helps to explain the substantial number of microcosm studies done for aquatic rather than terrestrial ecosystems (Draggan & Reisa, 1980).

Today, in line with the ECOFRAM aquatic report (1999), all model ecosystems can be referred to as microcosms. There are two major areas where microcosm studies can provide substantial information, first the scientific clarification of ecosystem dynamics and second, to demonstrate the effects of specific stresses on ecosystems. The latter applied mode gained increasing attention during the 1980s in conjunction with the USEPA Toxic Substances Control Act (TSCA). With implementation under the TSCA, it was recognised that these higher tier tests requires a good mix of ecological expertise and a high level of judgement in interpreting abundant and complex microcosm derived data, and that ecologists must find ways of stating the results of their findings more clearly (Draggan & Reisa, 1980).

In 1988 the USEPA produced a technical guidance document for aquatic mesocosm tests to support pesticide registration. Although laboratory testing has been a useful tool for risk managers, ecologists and aquatic toxicologists had recognised the weakness of using single species tests alone for assessing potential ecosystem impacts. The studies should allow the necessary control and replicability to detect ecosystem level effects. In addition, the study must be scientifically credible and performed with appropriate methods. It must also be verifiably accurate with a reasonable confidence of repeatability and applicable to predicting pesticide impacts (Touart, 1988).

On October 29th 1992, the USEPA issued a memorandum stating that although the Agency believes that long-term, indirect effects of pesticide use on aquatic ecosystems may be important, the Agency cannot have a testing scheme in place to accurately measure such effects. Mesocosm testing will not be required for purposes of regulatory decision making (registration or reregistration), as these studies do not generally provide regulatory mangers with information to make better regulatory decisions – decisions can and should be made in the absence of mesocosm studies. Only under unusual circumstances will the USEPA require mesocosm testing. Finally, the Agency will begin to develop a longer-term strategy for obtaining information needed to reduce uncertainty in evaluating ecological risks associated with long-term effects of pesticide use (Fisher, 1992).

The main reason leading to this decision was the lack of accuracy in performing and interpreting mesocosm studies, a problem ecotoxicologists had been warned against as early as in the late 1970s (Draggan & Reisa, 1980). Hence, no microcosm study has ever (1996) been required to register a new chemical for commercial use in the U.S. (Pratt & Cairns, 1996).

Within the EU, requirements for microcosm testing are based on TER (toxicity/exposure rations) triggers from the initial aquatic risk assessment under the EC Plant Protection Product Directive 91/414/EEC (Council Directive, 1991). TER is similar to the PEC/PNEC system - just the other way round. The key trigger values, within 91/414/EEC, which may require microcosm testing are reported in Annex III and VI (concerning the formulated product), and are currently acute TER of ≤100 for fish or aquatic invertebrates (based on LC/EC50), a TER of ≤ 10 for algae (based on EC50) and a chronic TER of ≤ 10 for fish or aquatic invertebrates (based on NOEC). However, the TER trigger values also require expert judgement before deciding whether a microcosm study is required or not (Campbell, 1996).

Microcosm tests should be conducted only if the risk assessment on the basis of standardised toxicity tests and exposure calculations indicates an unacceptable risk for non-target aquatic organisms. The purpose of microcosm testing is to obtain more realistic toxicity data with respect to exposure. Furthermore, indirect effects, more species, and recovery can be tested. The general protection aim and assessment endpoints are; diversity, species levels, function of the system, population and community dynamics, and long-term and indirect effects. The objective of microcosm testing is reduction of extrapolation uncertainty, validation of laboratory test systems and data, realistic exposure regimes, testing of many species, measuring indirect effects, and recovery of populations and communities (Streloke, 2000). To put it very concisely, the objective of a microcosm study is to determine the maximum exposure level of the test chemical that causes ecologically significant changes in population or community structure or ecosystem function in the test system (ECOFRAM, 1999).

The value of freshwater semi-field microcosm testing in ecological risk assessment for pesticide registration has generated considerable debate, not the least after the USEPA 1992 decision to remove all requirements of final-tier micro/mesocosm testing (Fisher, 1992).

But, why were meso-/microcosm studies ineffective in ecological risk assessment? The studies were simply tools that were not well understood and were consequently inappropriately and unfeasible utilised. More studies in smaller scaled microcosms have demonstrated ways in which these systems can be more effectively used in ecological risk assessment. The assumption of risk was difficult to negate in mesocosm studies, because of the inherent variability of the aquatic populations, sampling variability and use of statistical analysis that were inappropriate for ecological data. As a consequence, mesocosm testing was perceived as poor science and a doubtful application in risk assessment. As with all tools, microcosm studies have inadequacies. When a tool is not carefully selected, calibrated, and/or appropriately applied for a required result, and/or modified to suit the specific purpose, these inadequacies appear large. During recent years major scientific meetings, research, and reviews have been used in order to meet and reduce the critique of aquatic microcosm experiments in registration of pesticides AEDG, Wintergreen, Monks Wood, EWOFFT. I'll briefly present the outcome of the latest summit meetings and recommendations of ECOFRAM (1999), HARAP (1999), and

CLASSIC (2002) workshops regarding higher tier risk assessment of chemicals, and reflect on these in the discussion.

The ECOFRAM (Ecological Committee on FIFRA Risk Assessment Methods) aquatic report was published 5/4/1999 by the USEPA, and stressed that microcosm studies need statistical and ecological expertise to analyse and interpret the data. Moreover, it confirmed AEDG recommendation to use smaller-scale microcosm for chemical risk assessment. The USEPA required that the inherent control and replication in the microcosm design should allow an assessment of subtle changes to ecosystem structure and/or function. However, this has not been borne out by experience, even with inherent control and replication there is a large amount of uncontrolled variability in microcosm behaviour, as well as a sensitive dependence on initial conditions that interferes with detection of subtle effects. Moreover, each study must be customised to address the specific concerns for a particular chemical. It is of little value to define general objectives for field aquatic tests because of the many different problems that can arise in the hazard evaluation process. It is necessary to define specific objectives for individual tests in the light of physical and chemical properties, end use of chemicals, quantities manufactured and toxicological profiles. Since the objectives will be different for each chemical it will be impossible to develop rigid protocols for field tests. This has become a guiding principle for the use of model ecosystems. After this ECOFRAM list more than fifty recommendations and research needs. Vital ones of the 50 recommendations were:

1) More species sensitivity data are needed
2) Use of probabilistic risk assessment based on 1) should be increased.
3) Further research into determination of ecologically acceptable effects (ECOFRAM, 1999).

The HARAP (Higher Aquatic Risk Assessment for Pesticides) (1999) workshop (19-22/4-98 in France) built upon the guidance principles for higher-tier aquatic risk assessment. The workshop was funded by the EU-Commission and OECD primarily and SETAC-Europe. Because the first-tier assessment is conservative (and considering the often close taxonomic relationship between target species and non-target aquatic organisms) some pesticides will not pass the first tier and higher-tier assessment will be needed – which may be the case for many pesticides, depending on the level of conservatism of the preliminary risk characterisation. This presents a problem, since at present, regulatory procedures for higher-tier aquatic risk assessments are not well defined. The purpose of the workshop was therefore to examine the different types of methods available, the implications of the data generated from them, and finally to develop guidance on how the methods could be applied to higher-tier aquatic risk assessment in the future (HARAP, 1999).

It was agreed that the most appropriate approach to high-tier risk assessment would be by determining the properties of the compound in question and its use pattern. An important conclusion of the workshop was that higher-tier approaches should be regarded as a new step (or steps) in the risk assessment process. Consequently, the regulatory trigger values used at the first tier need not necessarily be carried over to the higher-tier assessment. This is because data from such higher-tier studies reduce the uncertainty associated with preliminary risk characterisation (*e.g.* by reducing uncertainty relating to species sensitivity and extrapolation of ecological risks). The HARAP workshop defined an Ecologically Acceptable Concentration (EAC) as the concentration at or below which no ecologically adverse effects would be expected (these

can be defined directly from semi-realistic indoor microcosm studies or larger out-door microcosm studies) (HARAP, 1999).

However, the effects could be considered acceptable, if with appropriate expert judgement, they do not pose significant ecological risks to natural aquatic ecosystems, for example if, recovery takes place during the period of the study. Furthermore, the replacement of one species by another with a similar role in the ecosystem may be considered acceptable if the applicant provides clear evidence that the ecological function and community structure in the field situation is unlikely to be significantly affected. However, this does not apply for keystone species (species of high importance for the function of the ecosystem) (HARAP, 1999).

If a valid microcosm study is properly designed, executed, analysed and interpreted, the results may be used in regulatory assessments even without the application of an uncertainty factor. Probabilistic effect assessment offers potential for pesticide regulatory risk assessment but further debate on methods and application is needed (HARAP, 1999).

The CLASSIC (Community Level Aquatic System Studies Interpretation Criteria) the workshop was held between 30 May and 2 June 1999 in Germany (funded by the EU-Commission, OECD and SETAC-Europe). The main issue discussed was the interpretation of data from microcosm tests. Due to the large amount of data from most studies interpretation is very difficult. Furthermore, no harmonised methods for conducting and evaluating these tests are available. The workshop discussed methodological issues like the application of the test compound or the most appropriate composition of the community of test organisms. Analysis and interpretation of results by various statistical methods and the weighing of the relative ecological relevance of the results were also discussed. The potentials and the deficiencies of the test systems were discussed in connection with the recovery potential of populations and communities. It was stated that there is not enough data at present on the abundance of species in waterbodies in agricultural landscapes, which could help to determine the recovery potential of species.

Table 2 *List of the seventeen recommendations of the CLASSIC workshop, which could/ should be followed or considered when performing a microcosm study, presented by Dr. Steve Norman at the 11th, annual meeting of SETAC-Europe 6-19 May in Madrid*

1) All studies are context dependent – no harmonised rules for conduct, analysis and interpretation are available
2) Need for expert judgement in interpreting studies
3) Define treatment concentration based on laboratory test, field use and environmental concentrations
4) Both spray and simple mixture application are acceptable
5) Apply substance during spring due to growth and monitor recovery
6) The test must be practically, technically and economically feasible - focus on most sensitive species
7) Include macrophytes
8) Univariate statistical analysis for single species and Principle Response Curves (multivariate ordination technique) for community data
9) Determine the Ecological Acceptable Concentration (EAC)
10) Preferably include both structural and functional measures
11) Determine recovery time
12) EAC's should protect the system
13) Create database for extrapolation between aquatic systems

14) Include landscape ecology
15) Determine clear protection aims for the study *a priori*
16) Effects with recovery are acceptable
17) No need for uncertainty factors on well executed, analysed and interpreted studies

I will comment on these recommendations in the discussion.

1.2 Detecting an ecological difference

The need for reliable detection of ecological impact on aquatic ecosystems has grown rapidly due to the increased burden placed on aquatic waterbodies by human activity. Without reliable information about changes in ecosystems, and the causes of these changes, environmental management cannot operate efficiently. In theory microcosms are useful tools for the establishment of causal relationships, since they allow controlled experimentation and replication. However, microcosms vary in scale, ecological relevance and complexity. This results in advantages and disadvantages concerning extrapolation to the real environment as well as in the degrees of replicability, repeatability or reproducibility and the associated ability to detect causal relationships for ecotoxicological endpoints (Kraufvelin, 2000).

At this time it must be noted that there is no single *correct* scale on which to describe populations, communities and ecosystems – the choice of scaling must always be a transparent and reasoned context dependent decision. Applied challenges, such as the prediction of the causes and consequences of most environmental problems, require interfacing of phenomena that occur on very different scales of space, time and ecological organisation. A thorough elucidation of the mechanisms underlying observed patterns is the principal key to prediction and understanding. This cannot be done properly without fundamental knowledge of experimental design and statistical analysis (Hurlbert, 1984).

The choice of spatial and temporal scales of an experiment is crucial for its outcome. Every microcosm investigation will suffer from its intermediate position – trapped between the realism of the field studies and the control typically associated with laboratory experiments. Full-scale natural experiments and whole-ecosystem studies operate at higher, more realistic, spatial scales and should thus be better suited for ecosystem management tasks and for understanding and prediction of the ecology of ecosystems, however, this is neither ethically or practically feasible. Any approach has inherent weaknesses and no approach alone is able to give a complete description of how ecosystems function. There is no *a priori* reason to assume that extrapolation of results among microcosms would involve more or less scaling pitfalls than extrapolation from small to large ecosystems (Kraufvelin, 2000) (Petersen et al., 1999). Prior to a context dependent definition of scope and scale of the microcosm study, basic consideration of experimental design and statistical analysis is necessary to among others avoid the pitfalls of pseudoreplication (falsely inflating the number of replicates (n) *e.g.* you can't take three samples from one experimental unit and then claim n=3, because these samples are not independent thus n=1) as described by Hurlbert (1984).

Beside these more qualitative considerations of ecological testing, a more quantitative perspective is also necessary. This concerns the statistical significance of the data resulting from the experiment. The way to determine whether to accept or reject the null hypothesis (of no effects caused by a chemical compound induced to the microcosm) is by statistical analysis of the data.

Since Pierre de Fermat's and Blaise Pascal's early development of probability theory during the renaissance, our prediction of the future has become more manageable in terms of risk assessment and risk management. In 1703 Jacob Bernoulli's defined the law of large numbers that briefly states; the difference between the observed result of a random point-test and its true result will diminish as the number of observations in the random points-test increases (Bernstein, 1996). The law of large numbers is also reflected in the power of a test, which increases with increased numbers of observations (*e.g.* replicates) (Green, 1989). In 1730, Abraham de Moivre drew the structure of the normal distribution, the bell curve, and he also conceived the concept of standard error and standard deviation. These two concepts together constitute the law of averages, which since has become crucial in most techniques for quantifying risk (Bernstein, 1996).

Two years after Thomas Bayes' death in 1764, the Royal Society published Bayes' paper in *Philosophical Transactions* on *a priori* probability also known as the Bayesian theorem. The theorem focuses on the many situations where we have a sound intuitive recognition of the probability for a result but at the same time we want to know how we are able to change the assessment as the situation evolves (this approach is subjective and thus biased). In 1848, Carl Friedrich Gauss published his 16 volume geodetic investigation of the Bayern Mountains. But he also found that the distribution of the values was symmetric around the mean and transposed de Moivre's bell curve into the physical and environmental world. The use of and interest in probability then developed from a mathematical and game/play focus towards using the contemporary techniques in economics (*e.g.* JM Keynes) and hence the political guidance of societal progress (Bernstein, 1996).

Today, probability is an integrated part of our understanding of the world and orientation among various changes and risks. In 1997 the USEPA (EPA/630/R-97/001) sent out guidelines for probabilistic risk assessment by Monte-Carlo simulations and in 2001 SETAC recommended the use of multivariate ordination techniques (Principal Response Curves) to ease the interpretation of multispecies toxicology tests (CLASSIC, 2001). Nobel prize winner Arrow (1992) stated that the mathematically driven apparatus of modern risk management contains the seeds of a dehumanising and self-destructive technology. Our knowledge of the way things work, in society or in nature, *„comes trailing clouds of vagueness. Vast ills have followed a belief in certainty*", he points out.

In the process of breaking free from the past we may have become slaves of a new religion, a creed that is just as implacable, confining, and arbitrary as the old (Bernstein, 1996). This is some of the background for the „regulators dilemma" (having to make a decision with lacking information/knowledge searching for single and magic numbers neglecting uncertainty) (Weinberger, 1985) and subsequently for the precautionary principle.

Today environmental scientists live under the singularity of obsession of statistical significance (α or < 0.05) and Type I error rates (the producer's probability of a false positive) when testing the null hypothesis (H_0). The most common H_0 is the one of „no difference" or „no effect" or to „assume no effect until proven otherwise". In environmental impact assessments, non-rejection of the H_0 typically results in a conclusion of „no impact". In rejecting H_0 by this rule, it is not incumbent upon us to worry about the magnitude of „statistically significant" differences (*e.g.* differences among means), even though that is perhaps the most interesting facet of our data. Statistical significance has come to be treated almost synonymously with biological or

ecological importance and relevance, even though no such relationship exists outcome (Mapstone, 1995).

Neither α nor β (the environments probability of a false negative - Type II error) have any intrinsic meaning in terms of the biological variables we measure or the biological (or economic, political, ethical etc.) importance of an outcome (Mapstone, 1995). Obviously, it is much easier to define statistical significance than to define biological or ecological significance. However, the above mentioned assumptions behind most environmental research and science are part of Arrow's (1992) prophecy of quantitative probability theory.

Green's (1979) ten rules provide a concise (and ideal?) summery of statistical advice for biological research, and will usually be a sound basis for critical assessments of the literature:

Table 1. *Ten statistical principles for ecological research (Green, 1979).*

1. Be able to state concisely to someone else what question you are asking.
2. Take replicate samples within each combination of time, location and any other controlled variable.
3. Take an equal number of randomly allocated replicate samples for each combination of controlled variables.
4. To test whether a condition has an effect, collect samples both where the condition is present and where the condition is absent but all else is the same (*Ceteris paribus*).
5. Carry out some preliminary sampling to provide a basis for evaluation of sampling design and statistical analysis options.
6. Verify that your sampling device or method is sampling the population you think you are sampling, and with equal and adequate efficiency over the entire range of sampling conditions to be encountered.
7. If the area to be sampled has a large-scale environmental pattern, break the area up into relatively homogeneous subareas and allocate samples to each in proportion to the size of the subarea.
8. Verify that your sample unit size is appropriate to the size, densities, and spatial distributions of the organisms you are sampling, then estimate the number of replicate samples required obtaining the precision you want.
9. Test your data to determine whether the error variation is homogeneous, normally distributed, and independent of the mean.
10. Having chosen the best statistical method to test your hypothesis, stick with the result.

Ad 2) Differences among can only be demonstrated by comparison to differences within. Ad 3) Putting samples in „representative" or „typical" places is *not* random sampling. Ad 4) An effect can only be demonstrated by comparison with a control. Ad 5) Those who skip this step because they do not have enough time usually end up losing time. Ad 6) Variation in efficiency of sampling from area/replicate to area/replicate biases among-area/replicate comparisons. Ad 7) If it is an estimate of total abundance over the entire area that is desired, make the allocation proportional to the number of organisms in the subarea. Ad 8) If it is not, as will be the case for most field data, then a) appropriately transform the data, b) use a distribution-free (non-parametric) procedure, c) use an appropriate sequential sampling design, or d) test against simulated null hypothesis (H_0) data. Ad 10) An unexpected or undesired result is *not* a valid reason for rejecting the method and hunting for a „better" one (Green, 1979).

Identifying trade-offs between statistical and ecological significance of environmental science requires careful thought on the experimental design, analytical tools in teasing out the real changes and consideration of the situation and position of the recipient of the conclusions and their possible regulatory and juridical framework. It is a lifework or more just to grasp a lobe of the total picture of the pit-falls and definition of ecological significance, which is also part of the „single large or several small (SLOSS)" (*e.g.* one large microcosm study or many small laboratory bioassays) discourse within ecotoxicology. However, I'll try to convey at least some of the headlines on this interesting and challenging topic that directly or indirectly relates to ecotoxicology and microcosm testing in the following.

1.3 Ecological significance in microcosm tests

Looking for and defining ecological significance is not easy because of the large amplitudes within nature in terms of relevance of season, time, species geographical differences, etc. However, it is important to try and determine the ecological significance of an ecological risk assessment. Pseudoreplication must be avoided. Pseudoreplication may be defined, in analysis of variance terminology, as the testing for treatment effects with an error term inappropriate to the hypothesis being considered. The samples must be independent of each other. This may seem evident and trivial, however, pseudoreplication is probably the single most common fault in the design and analysis of ecological field experiments. It is at least equally common in many other areas of research according to Hurlbert (1984). Pseudoreplication can also occur in the analysis of data, *e.g.* by inflating *n* (number of replicates) in Before-After-Control-Impact-Pairs (BACIP) analysis (Stewart-Oaten et al., 1992) by pseudoreplication in time.

Replication, repeatability and reproducibility are other problematic basic scientific issues for microcosm science (Crane, 1997). With increasing natural realism follows increasing natural variability and thus decreasing ability to get microcosms sufficiently similar to ensure efficient (subtle) effect detection with a reasonable number of replicates.

Another problem is the extrapolation of ecological risks within different scales:

A) Time; acute to chronic toxicity.
B) Space; laboratory to semi-field to field.
C) Complexity; between species, communities, populations and ecosystems.

All these extrapolation gaps are still widely uncertain or unknown (Persoone & Janssen, 1994). Within ecotoxicology today there is a discussion concerning analysing functional (*i.e.* processing of matter and energy) or structural (abundance and diversity species) changes of the ecosystem. Moreover, a precise definition of most of the entities mentioned under A) B) & C) are uncertain within themselves and needs further research.

The assimilative capacity of ecosystems is thought to occur because the systems have redundant parts. The redundancy of functional parameters refers to the principle that changing structure does not alter function, in other words they are buffered by redundant characteristics. If ecosystems have redundant components, then we should determine precisely how much ecological simplification could occur without damage to our life support systems *i.e.* „free" fundamental services provided by ecosystems. If ecosystems lack redundancy, we should redouble our efforts to assess levels of stressors that

produce adverse effects, and we should be more proactive in managing biological resources (Pratt & Cairns, 1996). Determination of the redundancy is also vital for a correct definition of acceptable effect sizes (Δ) in *a priori* power analysis of pre-treatment data prior to determining the test design (#5 in Green's list (1979)).

Structural measures are those with which ecotoxicologists still have the greatest familiarity and compatibility, and which have the longest history of measurement within the field (also chosen for these reasons by the author). However, functional measures may be more sensitive to stressors than structural ditto. The variability in community structure between replicates can be high and possibly problematic in detecting adverse effects at the community and ecosystem levels (Pratt & Cairns, 1996). Ecological functions integrate the collective activities of many species, and stressors can reduce the physiological abilities of individuals without individuals being eliminated from populations. Therefore, a measurement of collective functions might reveal effects that would be missed by enumerating all species („critter counting"). The redundancy argument is a theory based on the assumption that communities rich in species may have an array of taxa in key functional groups or guilds. Therefore, if a species were eliminated, one or more other species could expand to fulfil the functional role (Pratt & Cairns, 1996). Ecological significance for communities or ecosystem levels would be more readily assessable with functional measures if the redundancy theory and the inherent assumptions herein are correct. However, focusing on functional NOEC's could cover an unacceptable change in community structure and loss of biodiversity. Ideally both structural and functional parameters should thus be monitored.

1.4 Ecotoxicology as a science

In June 1969, René Truhaut, a French toxicologist with a pharmacology background, coined the term ecotoxicology – „a new branch of toxicology studying toxic effects to the constituents of ecosystems" (Halfmann, 1995). Ecotoxicology is a science that uses ecological variables to assess the effects of chemicals in the environment. Effects on biotic structure and function must be examined, and these data need to be incorporated (along with knowledge about their uncertainties) into risk assessments (Pratt & Cairns, 1996). The challenge facing those working in the field of ecotoxicology is to settle on a group of ecologically important assessment endpoints (particularly in microcosm tests) that can be reliably measured and, thus, can be used in ecological risk assessments.

Single species acute toxicity tests became the foundation for aquatic toxicology. Why? The tests were simple to conduct, relatively inexpensive, and easily interpreted. However, many began to question whether or not the tests provided adequate information about the effects of chemicals to be truly protective of aquatic life (Dickson, 1995). This was where the microcosm approach came in to play. While ecosystems cannot be created in the laboratory, community-level experimentation is a closer approximation of the ecosystem than independent tests of surrogate species in isolation (Pratt & Cairns, 1996).

Tansley (1935) gives a precise definition of an ecosystem as „... *not only the organism-complex, but also the whole complex of physical factors - the habitat factors in the widest sense. We cannot separate them (the organisms) from their physical environment.*"

It follows that the goal of ecotoxicology as a science is the organisation of knowledge about the fate and effects of toxicants in ecosystems based on

explanatory principles. Inconsistencies arise from the complex interweaving of various scientific, technological and practical goals within this socially obligated endeavour.

Hence, science is concerned with creating an intellectual model of the material world, in other words seeking simplicity in complexity. Technology is concerned with the procedures and tools and their general use to gain or use knowledge to enable scientific discovery. And practice as concerned with how to treat individual cases. Confusing these three in relationship to ecotoxicology and risk assessment can be dangerous (Dickson, 1995).

The technological objective of ecotoxicology is the development and effective application of tools and procedures to acquire a better understanding of toxicant fate and effects in ecosystems. Practical ecotoxicology applies available knowledge, tools, and procedures to specific problems. In ecotoxicology, the necessity for standardisation and the immediate need for action in specific situations, combined with the human fascination with particulars, contribute to our present dearth of innovative science. It encourages a preoccupation with methodology, particulars and *idola quantitatus*. Many of our present practices in ecotoxicology are driven by the history of the field, not their scientific soundness. There is a need for rejection of many of our present paradigms and adherence to a more rigorous falsification process within the sciences of ecology and ecotoxicology (Newman, 1996).

Microcosm testing for higher tier risk assessment of chemicals would fall within the realm of technological and practical ecotoxicology and not ecotoxicology as a science. This distinction is important in terms of what scientific falsification methodology should be recommended and in order to grasps the inertia of implementation of a precautionary approach in risk assessment and risk management. The inertia for changes is considered larger within science, than changing technologies and practices.

1.5 Ecology as a science

Ecology plays a larger role in ecotoxicology and especially in microcosm testing that „mimics" a real ecosystem. Elton (1927) and, later, Andrewartha & Birch (1954) have defined ecology as the science that attempts to predict the abundance and distribution of organisms. Some of the most vital ecological theories in microcosm testing are the ones concerning whole ecosystems, food webs and trophic levels and trophic-cascade hypotheses (Carpenter et al., 1985). I will not pursue these in length here, but just touch upon some criticisms for them.

The trophic-cascade theory was first put forward as a research topic for further investigation by Lindeman (1942). In this paper he concludes that it should be emphasised that the trophic-dynamic principles (*e.g.* the more remote an organism is from the initial source of energy (solar radiation), the less probable that it will be dependent solely upon the preceding trophic level as source of energy) cannot be expected to hold for every single case, according with the known facts of inherent biological variability. *A priori*, however, these principles appear to be valid for the vast majority of cases, and may be expected to possess a statistically significant probability of validity for any case selected at random. Since the available data summarised in his paper are far too meagre to establish a basis, it is highly important that further studies are initiated to test the validity of these and other trophic-dynamic principles, Lindeman (1942) concludes. Lindeman's views have inspired a generation of ecological investigations and still represent a fundamental concept for ecology and is

integral to various treatments of limitations of different trophic levels made popular by the classical work of Hairston, Smith & Slobodkin (1960). These considerations were central in the heated discourse on Top-Down and/or Bottom-Up forces in population community ecology (*Ecology* special feature (1992) 73,3:723-764) and in the subsequent debate of biomanipulation of lakes (Shapiro et al., 1990; McQueen et al., 1986; Persson et al., 1988; DeMelo et al., 1992; Sanderson et al., 1993).

The editors of the special feature in *Ecology* concluded that all of the papers in the special issue agree with the premise that top-down and bottom-up forces act on populations and communities simultaneously. The discussion is no longer about which occurs, but rather about what controls the strength and relative importance of the various forces under varying conditions, and what drives the feedback's and interactions among multiple trophic levels? (Matson & Hunter, 1992).

Hairston, Smith & Slobodkin (1960) and other studies, presume that the division of ecosystems into discrete trophic levels will reveal patterns in ecosystem structure or function, but ignore the serious operational problem presented by such a division. In principle, this division involves only the assignment of the component species of the ecosystem to distinct trophic levels based on their diets, but with the exception of terrestrial plants this is an extremely difficult task. Many organisms are so flexible in their diet that trophic relations and trophic levels change seasonally, ontogenetically and geographically. Omnivores and detritivores are especially difficult to classify yet the former are very common in most ecosystems and the latter often dominate energy exchange (Peters, 1991).

Finally, the recycling of energy among trophic levels, once considered impossible because of an overinterpretation of the laws of thermodynamics, cannot be accommodated into Lindeman's scheme, yet it may be quite common in nature (Peters, 1991).

Some brief examples: Porter (1976) showed that some species of algae that are ingested by *Daphnia* sp. are neither digested nor harmed while passing through the *Daphnia* gut. More important, it appears that they profit from being grazed by exposure to the higher phosphate concentration inside the gut. In this case, the primary interaction in question is the uptake of phosphate by the indigestible alga species, which after passing the gut, increased in abundance.

Another example may be found among cows and other herbivores that depend on intestinal flora to digest their food and in a sense are secondary consumers of the products (wastes) of this flora. Moreover, all mammals grow to 10-20% of their adult size while still totally dependent on their mother's blood supplies or milk for nutrient. The more completely we describe the trophic relations within an ecosystem, the less easily we can divide it into levels and compartments (Peters, 1991). The emergence of the trophic level as a tenet of ecological science might be an accident of history based on an invalid interpretation of thermodynamics. Modifications to the concept have not overcome the basic flaws of theory and, hence there is a need for more research into this, as Lindeman pointed out sixty years ago.

According to Ulanowicz (1988) the trophic level theory has been established as a descriptive and non-predictive model of ecosystems in which individual organisms or species cannot be wholly apportioned to a particular level

(Cousins, 1987), which, fits poorly with the definition of ecology as a science given above. Thereby the existence of a paradigm hierarchy theory of the ecosystem could be questioned (Rigler, 1982). In fact it could contribute to the vagueness of ecology by confusing hypothesis with experiments, and inviting inward-looking research that learns more and more about the artificial, but less and less about natural and managed ecosystems (Carpenter, 1999).

The questions posed to ecotoxicology do however not go away just because the definition of the science of ecology maybe imperfect, and society still need answers. There seems to be two standpoints among ecologists and ecotoxicologists. One state that we know nothing, or next to nothing, and that science is too poor and the other states that this may be true to some extent and that all models are approximations – but some are nevertheless useful. Moreover, who are then to answer all the risk questions posed if not ecologists and ecotoxicologists? The author favours the latter statement, maybe the scientific paradigm is weak but if the techniques and practices are in coherence with public demand of sustainability we must keep on refining these.

1.6 Epistemic approaches in ecotoxicology

Maybe the definition of ecology (and thus of ecotoxicology) given above is not justifiable - compared to for example physics due to random events, multiple causalities, evolution, historicity and self-organising properties in biological systems. However, physics are also submitted to random events and are not expected to predict these, for example the random event of decaying atoms of a radioactive isotope. If the system comprises a large number of atoms of the isotope we can predict empirically how many will decay in a given period of time, but we cannot predict which ones will decay or in which order. This, among other, gave rise to quantum mechanics which are still believed to be random (paper VIII). Another metaphor could be the ball travel in a pinball machine. We know all the laws the ball is subjected to yet we are unable to predict the exact position of the ball during the game 30 seconds ahead.

Environmental toxicologists, like all scientists, are presented with choices over the philosophical frameworks within which they work. However, most scientists do not receive formal training in scientific methods and this may lead to inappropriate choices. Presently there are two dominating philosophical frameworks of science to explain how scientists formulate and justify their theories. The first and most favoured approach is hypothetico-deduction (the testing of *a priori* theories by comparing them with relevant data). The other is classical induction based on observations. Both have strengths and weaknesses. It is important to recognise the strengths of each and use the most appropriate method for achieving the goals for environmental toxicology.

Most modern hypothetico-deductivism follows a version of the falsificationist strategy proposed by Popper (1968). Popper's view was that science progresses through the falsification of theories and their replacement by superior ones. In contrast to inductivists who wish to build up scientific knowledge piece by piece falsificationist's welcome bold conjectures that, if falsified, can be replaced in their entirety by other conjectures. Finally, scientists will arrive at a theory that withstands the most rigorous tests (Crane & Newman, 1996). These theories can then again form the paradigm of the science until replaced by yet more superior ones forming the structure of scientific revolutions (Kuhn, 1962).

There are, as mentioned, several context dependent problems with both approaches. The problem with the inductivist approach has traditionally been

called the „problem of induction“ due to the circularity and possible tautologies in inductive arguments based on observations. For example, what constitutes a sufficiently large number of observations before universal statements can be made? The metaphor of concluding that all Swans are white – as they are in most continents – not knowing that there are black Swans in Australia and thus being wrong.

A problem with the emphasis on falsification in modern hypothetico-deductivism arises when we ask how much of a theory needs to be falsified. Most scientific theories are complex structures that depend upon a collection of universal statements, initial conditions, and auxiliary assumptions. If a prediction made by a theory is false, it may be due to an error in an auxiliary assumption rather than in the main body of the theory itself. Because of this, scientists are usually able to deflect a falsification onto a less important aspect of their theoretical construct and hence protect the main theory (Crane & Newman, 1996).

Lakatos (1974) presents another approach. According to this, theories appear to evolve into complex organised structures over time. These organised structures are described as „scientific research programmes, containing several identifiable features. This approach allows scientists more flexibility than Popper when developing their research programmes. Moreover, Lakatos suggests that immature but growing research should at least initially be sheltered from the full forces of relentless falsification.

The problems faced by environmental toxicologists can be described as either one of retrospective assessment or of predictive assessment. The determination of the fate and effect of single chemicals or mixtures (a retrospective impact study) is usually a historical and spatially well defined problem that may best be solved with emphasis on the hypothetico-deductive approach. The prediction of the fate and effects of (new) chemicals usually has the objective of universal or at least very wide applicability across time, species and habitats, and may best be solved by emphasising inductive approaches (pers. comm. Dr. Peter Calow, 2001) (Crane & Newman, 1996).

In the predictive studies the investigator knows the source (*e.g.* a new chemical or effluent) but needs to estimate the level of exposure experienced by organisms and the toxic effects that result from this exposure. This can be contrasted with an impact study on a known source, in which the task of the investigator is also to discover the level of exposure and effects, but usually by direct measurement rather than estimation. Hence, the explanation of cause and effect is an important aim of impact studies whatever their genesis. Causality within ecology and ecotoxicology must be treated with caution as mentioned above. The criteria of caution by Hume in the mid 1700 was that cause and effect should occur together in space and time, the effect should follow the cause, and an effect should always occur when the cause is present (Crane & Newman, 1996). Peters (1991) suggests abandoning this concept altogether and replace it with the more restricted, but operational search for predictive regularities, for example by adopting multiple working hypothesis.

The hypothetico-deductive method can be a rigorous tool in retrospective impact studies, so long as its potential weaknesses are not ignored. An emphasis on inductive techniques seems to promise the most success in predictive assessments. The systematic combination of both induction and hypothetico-deduction within a pluralistic framework is likely to yield the greatest progress in most areas of environmental toxicology (Crane & Newman, 1996). Again

it is imperative to determine which form of ecotoxicology that is under consideration; science, technology or practice? For the scientific element in both ecotoxicology and ecology there is a strong need for hypothetico-deduction falsification (Peters, 1991). But within technology and even more in practice there are more needs for multiple inductive approaches (Lakatos, 1974). It would therefore be wrong to discard the useful ones due to rigorous falsification, not the least in light of the precautionary principle and the demand for policy- and decision-making on the environmental area, which is very different politically from physics and other core-sciences (Weinberger, 1985).

1.7 Methodology

Scientists like all humans are prone to confirmation bias: the psychological desire to confirm our theories rather that attempt to falsify them. This can lead to a loss of normal levels of scientific objectivity and to what Rousseau (1992) has termed pathological science. Hence, adherence to an immature theory within microcosm testing may represent either a process of positive nurturing until the theory is precise and predictive, or a stubborn adherence to a hopelessly vague and qualitative concept on the part of the scientist.

I was very aware of the confirmation bias starting this PhD. from earlier experiences writing a bachelor-project on biomanipulation and master-thesis involving mesocosm testing. I thereby focused on the methodology and design of microcosm experiments and therefore reversed the objective of my experiments in terms of testing the microcosm design with a compound and not the other way around where we usually test a compound with a test design. This approach meant a higher degree of objectivity for me because then all results were equally relevant and considered necessary. This way I escaped the confirmation bias of having to detect significant effects.

There is nothing wrong with using *a posteriori* hypothesis generated from statistical analysis of data to recommend or design further studies (Green, 1989)(Crane & Newman, 1996). Especially statistical power analysis can aid the design of microcosm testing (Smith, 1995). Moreover, power analysis involve considerations behind setting of acceptable probability of false positives (α) and negatives (β) and acceptable effect sizes (Δ). These can be viewed in light of the precautionary principle in terms of implementation of potentially variable microcosm data in risk management of chemical compounds, which allows me to keep the interdisciplinary perspective throughout my analysis.

Regarding analysis and interpretation, and improving statistical testing by contemporary new analysis, multivariate ordination techniques (*e.g.* Principel Response Curves (PRC)) and Probabilistic Risk Analysis (PRA) were not involved in my work for four primary reasons, regarding PRC:

1) Paul van den Brink has already made a doctoral thesis on multivariate statistical analysis of pesticides in freshwater model ecosystems 1999 where he recommends the use of PRC.
2) Presently we cannot do power analysis on multivariate ordination techniques because of the difficulties to determine acceptable effect sizes at the community and population levels. Hence, the software is not yet developed in CANOCO or related statistical packages.
3) These techniques are primarily analytical tools, which are best suited to lump together data, to give overviews and to generate simpler hypothesis, which then are amenable for further testing.

4) They do not ease the ability among regulators to interpret microcosm data, without additional training for them in multivariate ordination techniques. The multivariate techniques are usually very complex and they involve a lot of assumptions and implicit choices to be made. This is especially a drawback with PRC.

Regarding the PRAs these also involve some problems. According to Forbes & Forbes (1993) PRAs: 1) It is unscientific to ascribe species to a theoretical and unvalidated distribution. 2) To assume that the organisms selected for testing are an unbiased sample (an assumption of the statistical distribution) and 3) In order to scientifically reach 1) & 2), there is a need to generate larger amounts of data (and knowledge of the species biology and ecology). Moreover, species interactions are not accounted for, nor is the relative importance of the species (keystone species) thus the relative ecological relevance is not weighted and the total number of species to fit regression and distribution the curve is often not adequate (see also Petersen, 1999). In other words PRA still needs further scientific testing and validation.

Both PRC and PRA may play an advanced role in risk assessment in the future but full implementation of these techniques will generally require a substantial upgrading of the risk administrators statistical knowledge for in-depth comprehension of these techniques (Streloke, 2000).

The present thesis focuses on the dichotomy between statistical and ecological significance within microcosm testing and its implementation in risk assessment and management. In short, my scientific approach was influenced by a Popperian falsification of the designs by determining the actually required sample size, at the scientific level and a Lakatosian inductive approach in terms of the analysing the technological and practises of microcosm testing by context dependently elucidating trade-offs in design evaluation.

2.0 Design and materials

I used three different microcosms along an ecological relevance gradient. The first set-up consisted of four dough-out earthen out-door ponds of approximately 25m^2 situated at University of Roskilde, Denmark, campus area. These ponds were quasi-natural and had matured for five years before experimentation. For the purpose of this thesis, I was concerned with the applicability of quasi-natural mesocosm ponds for higher tier risk assessment of the herbicide $RoundUp_{2000}$ with emphasis on the potential risk for pelagic phytoplankton and zooplankton. Two of the ponds had been used 3 years before, but no residuals of the compounds were found in the waterphase or top sediment, since these compartments had remained unaffected. The second design was thirty outdoor 12 m^3 PVC pond microcosms designed according to SETAC recommendations (paper V) situated at University of Guelph, Ontario, Canada. This study focused on risk assessment of a perfluorinated compound (PFOS) in terms of toxicity also on phyto- and zooplankton. Moreover, Mark L. Hanson, Keith R. Solomon and I, reviewed three phytotoxicity studies in the 12 m^3 microcosms with monochloroacetic acid (MCA), dichloroacetic acid (DCA) and chlorodifluoroacetic acid (CDFA) (all halogenated compounds). The third microcosm type was used in two studies repeating, on a smaller scale, the Canadian Perfluorooctane sulfonate (PFOS) and Perfluorooctanonic acid (PFOA) experiments in twenty and thirty 30L clear PVC aquariums. These studies focused on the pelagic zooplankton community. Power analysis was conducted in all studies in order to determine the required sample size needed to detect the observed effects with high power,

this measure was used as the yard-stock comparing relative statistical power of the three different designs with regard to spatial dimension and complexity. The extrapolation gap and the relative ecological relevance in terms of how well the design mimics the mother model (the pond) was tentatively ranked and set to; 25 m^2 earthen pond >12m^3 out-door PVC pond > 30L indoor PVC aquarium. I focused on the pelagic plankton communities in the experiments, and on phytotoxicity in the Guelph review (paper V). Moreover, I conducted a state-of-the-art review of the replicability of microcosm studies 1985-2000 quantifying coefficients of variation (CVs) from each study.

Picture 1: 25 m^2 earthen pond

Picture 2: 12m^3 out-door PVC pond

Picture 3: 30L indoor aquarium

2.1 Chemicals

The choice of chemicals in this thesis was not the main issue, as they simply serve for evaluation of the test design and not primarily the other way around which is more commonly used - in order to maximise my objectivity. However, I tried to choose interesting and controversial chemicals for my experiments like the herbicide $Roundup_{2000}$ and the surfactants PFOS & PFOA. More thorough chemical descriptions of the compounds are given in papers II-V. Below I just explain briefly why they were chosen.

The two-component herbicide $Roundup_{2000}$ (and the additive Team-Up) was chosen because it is one of the most widely used herbicides in the world. In addition, if genetically modified (GM) Roundup-ready Soya beans and other Roundup-ready GM plants passes the ecological risk assessments, Monsanto (producer of Roundup products) look to significantly increase the global production and marketing of Roundup. However, the only aquatic semi-field study conducted thus far is a study by Hildebrand et al. (1980), despite the fact that Roundup has been deemed to pose a risk to aquatic organisms and a 2 meter buffer zone to waterbodies has been issued (MST, 2000). Hildebrand et al. (1980) did not find any significant effects despite treatment at a 100 times recommended dosage. Therefore we (Dr. Søren Petersen and I) wanted to determine if this could be attributable methodological problems related to testing Roundup under quasi-natural conditions rather than no effects? Hence, we performed a microcosm test in four ponds (see paper II).

Perfluorooctane sulfonate (PFOS) and Perfluorooctanonic acid (PFOA) belong to the family of perfluorinated surfactants which is widely used in firefighting foams, as water and stain repellent for shoes and clothes and as emulsifiers in the polymerisation of Teflon® for coating and lubricants. Both compounds are extremely persistent (no known environmental half-life (years)), they are ubiquitous in the environment and bioaccumulate as they have been found in blood and tissue of top predators as polar bears, seals, cormorants and humans (EPI (USEPA QSAR and database) estimated LogKoW = 6,3 and BCF = 56.23). They are globally distributed and under intense toxicological and biomonitoring survey by the USEPA and the producer of the majority of

PFOS 3M. However, no testing under field or semi-field conditions are included or have been carried out. This is why the Canadian toxic substances research initiative (TSRI) sponsored a semi-field risk assessment of PFOS and PFOA (papers III & IV). I participated in both in the 12 m^3 outdoor microcosm in Canada and in the indoor microcosm studies. The Danish indoor studies repeated the treatments, sampling and zooplankton toxicological endpoints on a smaller scale of the Canadian study and could thereby serve as a link between the laboratory single species bioassays and the outdoor microcosm study. Unfortunately the outdoor study will not be published until summer/fall 2002 and could thus not be a part of this thesis.

However, instead Mark L. Hanson, Keith R. Solomon and I have prepared a review of three studies performed in the microcosms at University of Guelph, ON, Canada, using haloacetic acids monochloroacetic acid (MCA), dichloroacetic acid (DCA) and chlorodifluoroacetic acid (CDFA). Haloacetic acids are toxic to many plants and have been used in the control of aquatic weeds. They are widespread and found in precipitation and in the aquatic environment, which renders studies on their ecotoxicological relevance important, especially with regard to their phytotoxicity on submerged macrophytes. The ASTM guideline 1913 97E test organism *Myriophyllum* sp. was used in the microcosm studies. We wanted to determine whether the uncertainty introduced by the microcosm design or inherent genetic variability was predominating and thereby we sorted out the ten different growth inhibition parameters in terms of ecological sensitivity and statistical power (paper V).

2.2 Power analysis

Although ecologists have become increasingly sophisticated in applying tests for statistical significance, few are aware of the power of these tests. Statistical power is the probability of getting a statistically significant result given that there is a biologically real effect in the population being studied. In particular if a test is not statistically significant, this is either so because there is no effect or because the study design makes it unlikely that a biologically real effect would be detected. Power analysis can distinguish between this, and is therefore a critical component of designing experiments and testing results (Thomas & Krebs, 1997). In statistical terms, the power of a test is defined as 1-β, and β is the probability of falsely accepting the null hypothesis (H_0) (the claim of no effect) when in fact another hypothesis is true (H_1) (Type II error). The main goal of power analysis is to decide, while in the process of designing an experiment:

1) How large a sample is needed to allow statistical judgements that are accurate and reliable
2) How likely the statistical test will be to detect effects of a given size in a particular situation.

I performed *a posteriori* analysis of power in my studies due to relatively fixed sample sizes two out of three experiments due to logistical and economic constrains. Performing power analysis and sample size estimation is an important aspect of experimental design, because without these calculations, sample size may be too high or too low. If sample size is too low, the experiment will lack the precision to provide reliable answers to the questions it is investigating. If sample size is too large, time and resources will be wasted, often for minimal gain.

I used power analysis to assess the required effect size (Δ) and or replication (n) of the three different scales of microcosms I worked with and thus to

address and quantify the dichotomy of statistical versus ecological significance inherent in designing or evaluating a microcosm study. Furthermore, I discuss the implementation and application of a precautionary approach in light of power analysis. So in a sense, the power analysis is my link between the environmental experiments I conducted and the possible implementation of these results via a precautionary approach in a feed-back-loop affecting how to conduct and interpret microcosm studies technically and in practice.

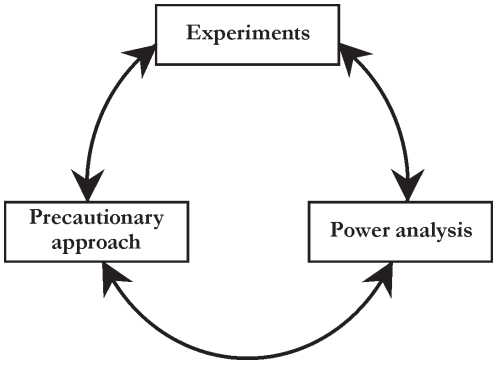

Figure 1: Feed-back-loop and analytical framework of my thesis.

In situations of uncertainty, ecologists following a scientific account of rationality typically minimise Type I rather than Type II statistical error (Shrader-Frechette & McCoy, 1993). There has been a long and sometimes heated debate of relative importance of Type I and II errors within the scientific community (Buhl-Mortensen & Welin, 1998). Type I error is also called „the producers risk" and Type II error is the „consumers risk". Statistically minimising the probability of a Type I error would increase the risk of a Type II error, whereas minimising the probability of a Type II error would not increase the probability of a Type I error (Shrader-Frechette & McCoy, 1993).

Environmental impact assessors and policy-makers typically make the value judgement that Type II errors are preferable to Type I errors in assumed support of an epistemic or scientific concept of rationality under uncertainty. They tend to prefer not to reject the null hypothesis, thus they, according to Shrader-Frechette & McCoy (1993), prefer the risk of not rejecting a dangerous development to the risk of rejecting a harmless development. Consumers and the public generally, however, tend to support an ethical and precautionary concept of rationality under uncertainty. They tend to reject the null hypothesis and prefer Type I error over Type II errors when both cannot be prevented. Preferences for Type II error and minimising Type I error might partly arise from appearing consistent with scientific practice (Shrader-Frechette & McCoy, 1993).

Scientific rationality has traditionally emphasised minimising Type I errors. In order to minimise Type I errors, scientists design studies to guard against the influence of all possible confounding variables, and they demand replication of study results before accepting them as supporting a particular hypothesis. The scientist usually attaches a greater loss to accepting a falsehood than to failing to acknowledge a truth. As a result, there is a certain inertia in the scientific enterprise, often rationalised as the healthy scepticism characteristic for the scientific temper. The preference for Type II errors, for public risks, over Type I errors, is also consistent with the standards of proof required in criminal cases. Our law requires the jury in criminal cases to be sure beyond a reasonable doubt that the defendant is guilty, before deciding against him, a miscarriage of justice must be avoided, thus revealing a preference for Type II

errors. However, in tort cases our law requires the jury to believe no more than that it is more probable than not that the defendant is guilty, thus apparently not preferring neither types of error before the other (Shrader-Frechette & McCoy, 1993).

The rationality of Type I errors before Type II errors within our conduct of science, technology and practice and thus our policy on environmental issues and all other innovation areas are important questions in pursue of a sustainable development for future and present generations. These questions can be treated transparently within a power analysis under the realm of a precautionary principle.

2.3 Precautionary principle

While there may be distinctions to be drawn between precautionary principle and precautionary approach they are used interchangeably here primarily because principle approach is less heavily burdened with years of heated political interpretation debate. The precautionary approach is a distinctive approach to managing threats of serious or irreversible harm where there is scientific uncertainty. This is not new – what is new is the increasing complexity of environmental sciences and the public debate about the ability of governments and administration to respond to such situations. The precautionary approach recognises that the absence of full scientific certainty shall not be used as a reason to postpone decisions where there is an unacceptable risk. Even though scientific information may be inconclusive, decisions have to be made to meet society's expectations that risks are being addressed and living standards may be maintained (Environment Canada, 2001).

It is quite clear that (and an implicit purpose of) using microcosm studies in the risk assessment of chemical compounds include increasing natural variability and thus decreasing detectability in a statistically significant sense. Presently microcosms are not widely used within risk assessment of pesticides and other chemicals due to lack of interpretability of the data and inherently uncertain data (Fisher, 1992) (Pratt & Cairns, 1996). Hence, a potential use and implementation of data from microcosm studies could also front the chemical registration process with considerations concerning how to implement data with a higher ecological significance but lower statistical significance than standardised laboratory bioassays? This is where the precautionary principle comes to my mind as the present framework for implementation of variable microcosm data.

The precautionary principle has been extensively debated and analysed within the juridical and social sciences since the articulation and introduction of the principle in international treaties and conventions in the late 1980s Northsea convention with the London conference in 1987 and early 1990s the Rio declaration in 1992. The lack of consensus regarding interpretation of the principle as been major reason for not implementing it along with the anxiety of it being anti-development and anti-science (papers VII-VIII). On February 2nd 2000 the EU-Commission launched a white-paper on its interpretation of the principle silencing much of the interpretation debate.

Meanwhile, there has been a dawning debate on a scientific interpretation of the precautionary principle, especially within marine environmental sciences. The argument was, and still is, that precaution should not just be restricted to policy-making but should also be reflected in the way environmental science is conducted, analysed and presented. Precautionary decision-making should be founded on precautionary gathered, analysed and interpreted data, it does

References

Andrewartha, HG. & LC Birch. 1954. The distribution and abundance of animals. University of Chicago Press. Chicago, USA.

Arrow, KJ. 1992. „I know a hawk from a handsaw". *In* Eminent economists: Their life and philosophies. *Ed.* M. Szenberg. Cambridge University Press, UK.

Beck, U. 1992. Risk society – towards a new modernity. Sage publications, London UK.

Bernstein, PL. 1996. Against the gods the remarkable story of risk. USA

Brink van den, PJ. 1999. Ecological and statistical evaluation of effects of pesticides in freshwater model ecosystems. Wageningen, Netherlands.

Brock , TMC., AR. Bos, SJH. Crum & R. Gylstra. 1995. The model ecosystem approach in ecotoxicology as illustrated with a study on the fate and effects of an insecticide in stagnant freshwater microcosms, pp 167-186. *In* B. Hock, R. Niessner (*Eds.).* Immunochemical detection of pesticides and their metabolites in the water cycle. Deutsche Forschungsgemeinschaft, Bonn, Germany.

Buhl-Mortensen, L Welin S. 1998. The ethics of doing policy relevant science: The precautionary principle and the significance of nonsignificant results. *Sci Eng Eth* **4**:401-412.

Cairns, J Jr. 1984. Are single species toxicity tests alone adequate for estimating environmental hazards. *Environ Moni Assess* **4**:29-273.

Cairns, J Jr. 1988. Putting the eco in ecotoxicology. *Reg Tox Pharml* **8**:226-238.

Carpenter, SR. JF. Kitchell & JR. Hodgson. 1985. Cascading trophic interactions and lake productivity. *BioSci* **35**:634-639.

Carpenter, SR. 1999. Microcosm experiments have limited relevance for community and ecosystem ecology: Reply. *Ecology* **80**:1085-1088

Carson, R. 1962. Silent spring. Fawcett Crest , New York, USA.

CLASSIC. 2001. Community Level Aquatic System Studies Interpretation Criteria. *In press.*

Colborn, T., D. Dumanoski & JP. Myers. 1996. Our stolen future: Are we threatening our fertility, intelligence and survival? – A scientific detective story. Dutton, New York, USA.

Cousins, S. 1987. The decline of the trophic level concept. *TREE* **2**:312-316.

Crane, M. & MC. Newman. 1996. Scientific method in environmental toxicology. *Environ Rev* **4**:112-122.

Crane, M. 1997. Research needs for predictive multispecies tests in aquatic toxicology. *Hydrobiol* **346**:149-155.

Crossland, NO., F. Heimbach, IR. Hill, A. Boudou, P. Leewangh. P Matthiessen & G. Persoone. 1994. Objectives and design of freshwater field tests. *In* Freshwater field tests for hazard assessment of chemicals. *Eds.* IA. Hill, F. Heimbach, P. Leeuwangh & P. Matthiesen. pp xxv-xxxvii. Lewis Publishers, CRC Press, Boca Raton, USA.

Draggan, S. & JJ. Reisa. 1980. Controlling toxic substances: Historical perspective and future research needs. *In* Microcosms in ecological research. *Ed.* JP. Giesy jr. Technical Information Center, US Dept. of Energy. USA.

DeMelo, R. R. France & DJ. McQueen. 1992. Biomanipulation: Hit or myth. *Limnol Oceano* **37**:197-207.

Dickson, KL. 1995. Progress in toxicity testing – an academics viewpoint. Chapter 12 *In* Ecological toxicity testing: Scale, complexity and relevance. *Ed.* J. Cairns Jr. & BR. Niederlehnder. Lewis Publishers Boca Raton, USA.

Dir. 91/414EEC. EC Plant Protection Product Directive 91/414/EEC (Council Directive, 1991).

ECOFRAM. 1999. Aquatic report. *By* P. Hendley & JM. Giddings. www. Aqex_ecofram_Peer01_may499.doc.

***Ecology* special feature. 1992**. **73**:723-764.

Elton, C. 1927. Animal ecology. Sedgwick and Jackson. London, UK.

Environment Canada, 2001. A Canadian perspective on the precautionary approach/principle – discussion document. Environment Canada homepage.

Fisher, LJ. 1992. Decisions on the ecological and environmental fate data requirements task force. Memorandum to D. Campt. US Environmental Protection Agency, Office of Pesticide Programs, Washington, DC. USA.

Forbes, V. & T. Forbes. 1993. A critique of the use of distribution-based extrapolation models in ecotoxicology. *Func Ecol* **7**:249-254.

Forbes, SA. 1887. The lake as a microcosm. *Bull Ill Nat Hist Surv* **15**:537-550. According to: Draggan, S. & JJ. Reisa. 1980. Controlling toxic substances: Historical perspective and future research needs. *In* Microcosms in ecological research. *Ed*. JP. Giesy jr. Technical Information Center, US Dept. of Energy. USA.

Funtowicz S. & J. Ravetz. 1993. Science for the post-normal age. *Futures* 25:735-755.

Green, RH. 1979. Sampling design and statistical methods for environmental biologists. John Wiley & Sons, New York, USA.

Green, RH. 1989. Power analysis and practical strategies for environmental monitoring. *Environ Res* **50**:195-205.

Halfmann, W. 1995. The boundary between ecology and toxicology: A sociologists perspective. *In* Ecological toxicity testing: Scale, complexity and relevance. *Ed*. J. Cairns Jr. & BR. Niederlehnder. Lewis Publishers Boca Raton, USA.

Hairston, NG. FE. Smith & LB. Slobodkin. 1960. Community structure, population control, and competition. *Am Nat* **XCIV**:421-425.

HARAP. 1999. Guidance document on higher-tier aquatic risk assessment for pesticides. *Eds*. PJ. Campbell, DJS. Arnold, TCM. Brock, NJ. Grandy, W. Heger, F. Heimbach, SJ. Maund, M. Streloke. SETAC-Europe. Brussels, Belgium.

Hildebrand LD, Sullivan DS, Sullivan TP. 1980. Effects of Roundup herbicide on populations of *Daphnia magna* in a forest pond *Bull Environ Contam & Toxicol* **25**:353-357

Hume, D. 1748. An enquiry concerning human understanding. Open Court Publishing Co., Chicago, USA. *In* Crane, M. & MC. Newman. 1996. Scientific method in environmental toxicology. *Environ Rev* **4**:112-122.

Hurlbert, SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol Monograph* **54**:187-211.

Keynes, JM. 1936. The general theory of employment, interest and money. New York. USA.

Lakatos, I. 1974. Falsification and the methodology of scientific research programmes. *In* Criticism and the growth of knowledge. *Eds*. I. Lakatos & A. Musgrave. Cambridge University Press, UK.

Levin S Kimball KD. 1984. New perspectives in ecotoxicology. *Environ Managem* **8**:375-442.

Lindeman, RL. 1942. The trophic-dynamic aspect of ecology. *Ecology* **23**:399-418

Lubchenco, J. 1995. The role of science in formulating a biodiversity strategy. *BioScience* supplement:s7-s9.

Kimball KD & Levin S. 1985. Limitations of laboratory bioassays: The need for ecosystem-level testing. BioScience **35**:1-172.

Kraufvelin, P. 2000. Detecting change in marine zoobenthic communities by large-scale and long-term mesocosm experiments and field studies. Åbo Akademi University Press Finland.

Kuhn, TS. 1962. The structure of scientific revolutions. University of Chicago Press, Chicago USA.

Mapstone, B.D. 1995. Scalable decision rules for environmental impacts studies: Effect size, Type I and Type II errors. *Ecol App* **5**:401-410.

Matson PA & MD. Hunter. 1992. The relative contributions of Top-Down and Bottom-Up forces in population and community ecology. *Ecology* **73**:723.

McQueen, DJ, JR. Post & EL. Mills. Trophic relationships in freshwater pelagic ecosystems. *Can J Fish Aquat Sci* **43**:15711581.

MST. 2000. Risk assessment of Roundup and glyphosate. J.# M 7042-0243 14/9-2000. Danish EPA Miljøstyrelsen, Copenhagen, Denmark.

Newman, MC. 1996. Ecotoxicology as a science. Chapter 1 *In* Ecotoxicology: a hierarchical treatment. *Eds.* MC. Newman & CH. Jagoe. CRC Press Boca Raton, USA.

Persoone G. & Janssen CR. 1994. Field validation of predictions based on laboratory toxicity tests. Chapter 25 *In* Freshwater field tests for hazard assessment of chemicals. *Eds.* IA. Hill, F. Heimbach, P. Leeuwangh & P. Matthiesen. Lewis Publishers, CRC Press, Boca Raton, USA.

Persson, LD., SL. Johansson, G. Andersson & L. Hamrin. 1992. Trophic interactions in temperate lake ecosystems: A test of food chain theory. *Am Nat* **140**:59-84.

Peters, RH. 1991. A critique for ecology. Cambridge University Press, UK.

Petersen, JE., Cornwell, JC., Kemp WM. 1999. Implicit scaling in the design of experimental aquatic ecosystems. *Oikos* **85**:3-18.

Petersen, S. 1999. Effects of toxic compounds on natural phytoplankton – with special emphasis on Tri-Butyl-Tin (TBT), pollution induction community tolerance (PICT) and distribution based extrapolation models. Ph.D.-thesis. Freshwater Biological Laboratory University of Copenhagen and DHI-Institute for the Water Environment. Hørsholm, Denmark.

Popper, K. 1968. The logic of scientific discovery. Hutchinson, London UK.

Porter, KG. 1976. Enhancement of algal growth and productivity by grazing zooplankton. *Science* **192**:1332-1334

Pratt, JR & J Cairns Jr. 1996. Ecotoxicology and the redundancy problem: Understanding effects on community structure and function. Chapter 12 *In* Ecotoxicology: a hierarchical treatment. *Eds.* MC. Newman & CH. Jagoe. Lewis Publishers CRC Press Boca Raton, USA.

Rigler, FH. Recognition of the possible: An advantage of empiricism in ecology. *Can J Fish Aquat Sci* **39**:1323-1331.

Rousseau, DL. 1992. Case studies in pathological science. *Am Sci* **80**:54-63.

Sanderson, H., P. Juul, M. Fotel, R. Pedersen & L. Riemann. 1993. Biomanipulation. University of Roskilde. Denmark.

Sanderson, H. & S. Petersen. 2001. Power analysis as a reflexive scientific tool for interpretation and implementation of precautionary principle in the European Union. *Environ Sci Poll Res Inter* **8**: 2001.

Shapiro J. 1990. Biomanipulation: The next phase - making it stable. *Hydrobiol.* **200/201**:13-27.

Shaw, JL. & JH. Kennedy. 1996. *Editorial.* The use of aquatic field mesocosm studies in risk assessment. *Environ Technol & Chem* **15**: 605-607.

Shrader-Frechette KS, McCoy ED. 1993. Method in ecology: Strategies for conservation. Cambridge University Press. New York, USA.

Smith, EP. 1995. Design and analysis of multispecies experiments. Chapter 5 *In* Ecological toxicity testing: Scale, complexity and relevance. *Ed.* J. Cairns Jr. & BR. Niederlehnder. Lewis Publishers Boca Raton, USA.

Stewart-Oaten A., Bence, JR., Ostenberg, CW. 1992. Assessing effects of unreplicated perturbations: No simple solutions. *Ecology* **73**:1396-1404.

Streloke, M. 2000. Objectives and design of aquatic microcosm studies for pesticide registration. Notes from technical workshop: Regulatory evaluation of aquatic microcosm studies with pesticides. Brighton, UK.

Tansley, AG. 1935. The use and abuse of vegetational concepts and terms. *Ecology* **16**:284-307.

Thomas L, Krebs CJ. 1997. A review of statistical power analysis software. *Bull Ecol Soc Am* **78**:126-139.

Touart, LW. 1988. Aquatic mesocosm test to support pesticide registrations. USEPA, Washington, USA.

Ulanowicz, RE. 1987. Growth and development. Ecosystems phenology. Springer-Verlag, New York, USA.

Underwood, AJ. 1995. Ecological research and (and research into) environmental management. *Ecol App* **5**:232-247.

USEPA 1997. Guidelines for probabilistic risk assessment by Monte-Carlo simulations. EPA/630/R-97/001. Washington DC, USA.

Weinberger, A.M. 1985. Science and its limits: The regulators dilemma. *Issues in Sci Techn* **fall**:59-72.

not make much environmental sense to try and be politically precautionary with scientifically un-precautionary data. Precaution should involve how sound environmental science is conducted. Which, on a short term (technologically and by practical ecotoxicology) could be addressed by transparently defining the entities of a power analysis (acceptable risks of Type I & II errors (α & β) and acceptable effect size (Δ)) of the study (preferably *a priori* but also *a posteriori*).

On the longer term discussions of, epistemic changes of science could include considerations of objectivity (as you can not, and should not, be 100% objective in the light of precaution) and post-normality of science (paper VIII).

2.4 Results

The following papers all represent my context dependent empirical findings and results. Furthermore, the scope of this thesis was not to present *a* correct design of microcosm studies, as this is very context dependent, which has also been stressed by previous publications and work-shops (ECOFRAM, HARAP & CLASSIC), but to illustrate and quantify the trade-offs between statistical and ecological significance associated with choice of microcosm design. Furthermore, the reader should note that there is always a variable extrapolation between different studies, endpoints and chemicals, which hampers the direct quantitative comparison between the studies. The ecological significance (and thus magnitude of extrapolation gap to the mother model (typical temporal small farm pond in the agricultural landscape)) of the three different scales of microcosms covered in this thesis was tentatively set as a gradient quasi-natural 3 year old 25 m^2 ponds > 12 m^3 outdoor PVC ponds constructed according to SETAC recommendations > 30L indoor PVC aquariums. I analyse the different design primarily via statistical power analysis and discuss a short term scientific application of precautionary approaches via statistical power analysis. The following eight papers represent results and the empirical basis for my overall conclusions in the last paper.

Acknowledgements

First of all my supportive little family Karin, Amalie & Regitze. My always very encouraging supervisor Jette Rank (University of Roskilde (RUC)), the department for Environment, Technology and Social studies for the grant and support. Søren Petersen (DHI – water & environment) for fruitful inspiration and cooperation. Keith Solomon for his inspiration and very kind and hospitable reception at University of Guelph. Scott Mabury for helping with the chemical IC analysis of PFOS & A at University of Toronto. Tim Boudreau and Mark Hanson, my very good friends and colleagues at University of Guelph. Patrik Kraufvelin at University of Åbo, for commenting on some of my work and good times in Lecce. Andrew Crabtree (RUC) for English proofreading. Bodil Bruun (RUC) for lab technical assistance. Jan Ole Nielsen (RUC) for helping running the microcosms. Ritta Bitsch (RUC) with layout of my papers. Joel Tickner, University of Lowell. Gregory Biddinger, Exxon Mobile. Jerry Ravetz independent philosopher (UK), Rich Purdy independent ecotoxicologist (USA), Ryan Sullivan and Woo-Jay Cheong University of Toronto. And a warm appreciation to all the anonymous referees on my papers continuously improving my work.

Sincerely yours,

Hans Sanderson
University of Roskilde, 020202

I

Replicability and interpretation of micro/mesocosm studies

by

Sanderson H. 2002

Journal of Environmental Science And Pollution Research International
http://dx.doi.org/10.1065/espr2001.09.085

State-of-the-Art Review

Replication of Micro/Mesocosm Studies

Hans Sanderson

Department for Environment, Technology and Social studies, University of Roskilde, POB 260, DK-4000 Roskilde, Denmark; e-mail: hanss@ruc.dk

DOI: http://dx.doi.org/10.1065/espr2001.09.085

Abstract. The objective of this state of the art review was to quantify the replicability of pesticide studies using micro/mesocosms. Low interpretability of micro/mesocosm studies, and inconclusive and highly variable data, resulted in a discontinuation of the use of these studies for the registration of pesticides. Coefficients of variation, CV%, were calculated on the basis of data tables as a measure of statistical 'effectiveness' taken from the literature. The average CV in the investigated studies was 45%; larger out-door mesocosms averaged 51%, and smaller indoor micro/mesocosms averaged 32%. CVs on variables involving animals were higher than CVs on plant end-points, which in turn were higher than abiotic variables for all experiments. However, to enhance the interpretability and implementation of micro/mesocosm studies for pesticide registration, a number of context-dependent steps could be incorporated; 1) determine the appropriate experimental design and number of replicates by using power analysis, 2) Utilise advanced statistical analysis, such as probabilistic effect distribution and principal response curves, 4) report, preferably in quantitative terms using power analysis, the risk of Type II error. The author's primary conclusion is that the level of CVs is context dependent and, therefore, it is not possible to suggest a generally acceptable level of CVs for all experiments. This has been suggested both directly and indirectly in the literature. Moreover, the number of insignificant ($p > 0.05$) results is high, 88% of all test biotic variables had no statistical significance. The average number of replicates were 3–4, which theoretically should yield significant effects at least at the highest test-concentration, then resulting in 75–66% insignificant results.

Keywords: CVs; coefficients of variation (CVs); interpretation; mesocosm; microcosm; pesticide studies; power analysis; replicability of pesticide studies

Introduction

The replicability of micro/mesocosm studies is of major concern when assessing the statistical quality of data from these studies. If replicability is lacking, then so too will be repeatability, reproducibility, predictability and, thus, the utility of the investigation in the registration of pesticides will be impeded (Crane 1997). This paper presents a literature survey of replicability and performance of micro/mesocosm studies undertaken over the past two decades as a follow-up on the coefficients of variation (CVs) review on microcosm/mesocosm studies (Giesy and Allred 1985). This was performed via calculations and comparisons of test variables' CVs accordingly to Giesy and Allred (1985). Micro/mesocosms have been used for some years for assessing the effects of chemicals in aquatic ecosystems (Touart 1988). However, in 1992, the USEPA decided to discontinue the use of mesocosm studies in the registration process concerning pesticides. This was due to concerns surrounding the objectivity of the studies and their usefulness for registration and risk management decisions. "Although the agency believes that long-term, indirect effects of pesticide use on aquatic ecosystems may be important, the agency does not have a testing scheme in place to *accurately measure* (my italics) such effects", Fisher 1992). The objective of a pesticide mesocosm experiment is best expressed in a FIFRA-SAP report as follows: To determine the maximum exposure level of the test pesticide that causes no ecologically significant changes in population or community structure, or in the ecosystem function of the test system (USEPA 1987). This objective, however, is impeded by the highly variable nature of data derived from micro and mesocosm studies, presumably due to high natural variability.

Crane (1997) identified a number of characteristics for predictive multispecies tests in aquatic toxicology, these being the repeatability, reproducibility and interpretability of mesocosm studies. In relation to these, this paper assesses the relative importance of replicability and statistical power of micro/mesocosm studies in relation to the interpretability of aquatic multispecies tests, as the first target of the critique by Crane (1997). The exact ranges of the CVs in micro and mesocosm studies is still poorly understood, the ECOFRAM report (Hendley and Giddings 1999) estimates that, for taxonomic variables, CVs typically range from 50–100%, a feature which makes the detection of subtle effects difficult. Although micro and mesocosm studies are highly individual and it is difficult to develop rigid protocols for testing in them (Crossland et al. 1994), the present state of the art review addresses the variability between micro and mesocosm replicates and thus tries to quantify CV ranges in papers published over the last two decades to update Giesy and Allred's review from 1985.

The advantages and disadvantages of micro/mesocosm studies are two aspects of the same question, namely the issue of ecological realism versus the repeatability of the design of the study. It has been suggested that reduced repeatability and reproducibility corresponds to an increased scale and ecological realism (Kraufvelin 1999). In addition to the ecological uncertainty principle (Maguire et al. 1980), the interactions of organisms with their biotic and abiotic envi-

ronment are mostly non-linear. In a statistical sense, they are not independent. It has been suggested that ecosystems have emergent properties such that the whole system is greater than the sum of its parts. The many nonadditive interactions observed in ecological studies are evidence of this. For this reason, the prediction of effects on systems other than the whole ecosystem that we are interested in may constitute unreasonable simplifications, and predictions made from any reduced system may thus be inaccurate (Giesy and Allred 1985). The reason why replication is important is that it is needed to guarantee detection of true responses of the treatment relative to the unexplained natural variability and the magnitude of statistical error acceptable and thus the statistical power of the investigation; moreover there is a risk of committing a Type II error (false negative) (Kraufvelin 1998).

Replicability can have several connotations Giesy and Allred (1985) used the following definition: "*Replicability means, in this context, the establishment of more than one individual experimental unit within a particular experimental treatment. Statistically, this is a measure of within-treatment variance. To assess replicability of a system is to determine the similarity of replicate experimental units of an experimental treatment at a given point in time and space that, by definition and design, are meant to be identical*". However, a restricted degree of replicability could also be that replicates never will be and do not have to be identical/duplicates. Replicability simply reflects the variability between whatever is being sampled and the best action to prepare for future decisive statistical analyses is simply to try to increase the number of replicates – perhaps this notion is also implicit in the definition used by Giesy and Allred (1985). If the effects of the studied disturbances (also low level) are large and beyond any doubt in turn, we do not have much of a problem (Kraufvelin 1999).

The level of CVs or replication within micro/mesocosm science is not quantified, moreover, the use of CVs levels and evaluation of the studies on this background is also unsettled. The objective of this paper is to deliver a literature-based quantification of the replicability of aquatic micro/mesocosms, by calculating the coefficient of variation (CV% = standard deviation / mean·100) between experimental units over the past two decades. In addition, it has been performed in order to assess the role of CVs in the evaluation and interpretation of micro/mesocosm studies since indications have been seen in the literature that some levels of CVs were acceptable and some were not. In this respect, it is important to mention that the omission of specific research papers, articles or designs does not imply a lack of conceptual scientific merit or skill.

1 Methods

Biosis, Cambridge Scientific Abstracts, Current Contents, Poltox and SCI were primary databases used in this review. The scan was limited by one or more of the following keywords: Aquatic micro/mesocosm, experimental ponds, replication, pesticide risk assessment, biological or ecological effects. This search yielded 129 papers that used micro/mesocosms. Of these papers, 16 were not replicated, 96 had other deficiencies preventing the use of the results (no tables, but only graphic reporting, fate studies, predation or nutrition studies, marine enclosures, etc.). Thus, only a selection of the studies can be presented, which limits the extent to which generalisations can be made from this literature study. The scan resulted in 17 well-reported and/or replicated papers for analysis, 8 micro and 9 mesocosm studies. However, the original study by Giesy and Allred (1985) did not contain as many studies (total <50), moreover, they did not perform statistical power analysis of the data gathered – this is why this important analysis was also excluded in the present paper. Also in coherence with the original review, this review only analyses ANOVA studies, which is also the most common approach within micro/mesocosm science. The survey obviously does not include all available information on micro/mesocosm studies for the past 15 years, but only a limited number of random and representative articles suitable for a quantifiable evaluation of CV. The relative ecological realism of each study analysed here is also peripherally addressed, this covers how well the model copies the original natural environmental recipient. Realism is, *ceteris paribus*, lower in a small indoor study without sediment than in a larger outdoor study with sediment. Ecological realism is a crude estimate of the model accuracy where CV is a measure of precision in this context.

The use of CV is based on the fact that is has been used as an objective measure of the degree of test system replicability since the birth of micro/mesocosm studies (Abbott 1966). Moreover, Giesy and Allred (1985) used the CV in their review, which this paper seeks to bring up to date. The CV may be used to assess the statistical 'effectiveness' of ecological experiments, which are based on univariate ANOVA rather than regression designs. If the standard deviation is too high in comparison to the mean, the inherent variability may then be nearly as large as the quantity being measured. Therefore, it may be very difficult to demonstrate statistical differences between controls and treatments, unless there are very large effects. On the other hand, a low CV value indicates that the standard deviation is small compared to the mean, and it becomes easier to detect statistical significance (Conquest 1983). Because the CV expresses variability as a fraction of the mean under consideration, it is possible to numerically and statistically compare and add variation between different experiments that otherwise cannot be compared in an unbiased and non-standardised way. The CVs and overall mean CVs determined in this paper were calculated from data tables provided in the studies selected from the literature review.

I acknowledge the dichotomy of ANOVA versus regression design in achieving more powerful and better estimates e.g. of the $NOEC_{community}$ and that this has not been resolved satisfactorily to date (Liber et al. 1992). Regression designs can provide an opportunity to include a broader range of concentrations, because there is only limited need for replication, which can be used to better define thresholds of toxicological response using non-linear techniques. However, due to variation around each point estimate, these can vary substantially, thus making the fitting of a descriptive curve difficult (low R^2 value) and resulting in low power and a high

risk of committing a type II error. Ideally it's not a question of either-or but rather a question of both at the same time, that is well replicated and high power point estimates, and then regression on these for extrapolation.

Although important, an additional power analysis of all the data and the solving of the above dichotomy fall outside the realm of the present paper.

2 Results

All the results presented in **Table 2** (see appendix) are based on the studies in **Table 1** (see appendix).The combined overall average CV of the studies in this survey was 45%, and the average number of replicates was 3.5. The average CV of larger mesocosms with higher natural realism was 51%, for the less realistic and smaller indoor meso/microcosms the average CV was 32%. CVs on variables involving animals ($x \approx 47\%$) were generally higher than CVs on plant end-points ($x \approx 31\%$). There was no relationship between the age of the investigation and the CV. Significant differences were found for 12% of all the variables and 88% were not significantly different from the controls in this survey. Seven papers showed some significant differences, on one or more variables. Ten papers showed absolutely no significant differences. Green (1989) offers a conceptual way of estimating the needed sample size to detect an effect at a certain level of probability of accepting a null hypothesis when it is in fact false (type II error), which is standard on new statistical packages. Notable, however, was that, of the 17 analysed papers, only one explicitly committed a Type II error by concluding that *Btk* was not harmful to benthic steam invertebrates when no statistically significant effects were detectable, without assessing the power of the study (Richardson and Perrin 1994).

3 Discussion and Conclusions

When micro/mesocosms were taken out of the registration process for pesticides, it was because of the uncertainty in measuring effects and interpretation of the investigation (Fisher 1992). The purpose of micro/mesocosm studies is to reduce the uncertainty and variability in extrapolation from single species laboratory bioassays to real environmental effects. What then is a suitable level of CV in micro/mesocosm studies? Suitable test variables for microcosms have been suggested to be those having a CV lower than 20–30% (Isensee 1976). A suitable level for larger mesocosm studies is still rather unsettled. Mesocosm studies should, implicitly, strive to achieve certainty comparable to standard laboratory bioassays. Persoone and Jansson (1994) found that CVs for single species tests reported in the literature usually exceed 25%, and can be as high as 40–50%. In this comparison, the micro/mesocosm studies in this survey are within a normally acceptable level of CVs with an overall mean of 45%. Laboratory bioassays potentially possess a higher degree of statistical precision. Hence, it is relatively easy to achieve higher power of the study in the laboratory by more replicates or higher effect sizes.

The largest contributing factor to mesocosm uncertainty is sample variability due to natural variability. Sample variability is affected by several confounding factors. Rosenzweig and Buikema (1994) found similar successional patterns in 12 new ponds but the community structure between the ponds was not similar at any time after one year, despite statistically similar environmental characteristics in the ponds. In addition to this, one can add the inherent problem of zooplankton sample variability (Gagnon and Lacroix 1981). Another confounding factor is the accuracy and precision of pesticide concentrations following application. Knuth (1986) found an overall CV average of 28% between replicates in a case study. Schindler (1998), points to confounding aspects of sampling only at daytime, thus not taking into account the vertical migration and night activity of the animals, meaning that possible key-stone species or most sensitive species may be excluded from a test regime only operating during the daytime. High concentrations are most likely to produce significant effects, which are needed to break through the noise of natural variation. For zooplankton, an ecologically significant impact should be designated as at least a 1–2-fold difference or 50–80% reduction. Smaller differences, from an ecological point of view, are probably irrelevant in natural ecosystems because of large seasonal variations, rapid generation times, recolonisation and recovery, possibly reducing the chance of detecting effects under realistic design and application conditions (Farmer et al. 1995). However, the weight of these changes differs in each case, and by the eye and opinion of the beholder.

Howick et al. (1992) showed that CV and sample size in mesocosm studies were inversely proportional, with higher sample sizes yielding lower CVs. Table 2 shows that there is no clear relation between CVs and statistical significance. Fairchild *et al.* (1992) showed significant differences on abundance of Gastropoda despite CVs at 112% and 115%. On the other hand, Richardson and Perrin (1994) failed to show significant differences on the total abundance of insects, despite CVs at 9% and 5%. This illustrates the importance of effect size for determination of significant differences. However, as CVs increases, it becomes increasingly difficult to identify subtle effects. Subtle effects, however, are seldom ecologically significant, as is mentioned above in Farmer et al. (1995).

It is not possible to determine a generally acceptable CV level because each design and variable is very diverse. If the effect size is large enough, it is possible, despite a large CV, to obtain significant differences. It is vital to distinguish between duplication and replication, if realism is increased, the exact duplication of the absolute abundances of all species is not necessary for reasonable simulations of processes occurring in an ecosystem, because the well-being of the system as a whole is of concern (Hammons et al. 1981). Here, it is important not to confuse precision with accuracy; while precision describes the range of results encountered in the experiments, the accuracy determines whether these results give valuable insights into the performance of the natural system of interest or not (Lundgren 1985).

Assessment of recovery time is a highly ecologically significant end-point that, moreover, is relatively easily evaluated for regulators. This means that the experiment should be run until there has been a recovery on the functional effect variables analysed by means of univariate ANOVA (Cambell et al. 1999).

Moreover, probabilistic effect distributions could be implemented to ease the interpretation of micro/mesocosm studies (ibid.). The implementation of power analysis and probabilistic effect distributions not only helps the interpretability of the studies, but also increases the possibility for regulators to determine the level of protection. Determining the optimal size of the abiotic and biotic assemblage of micro/mesocosm studies is not a primary concern or possibility. Rather, the design is a context-dependent matter of optimisation to answer questions within the framework of acceptable inference errors and the amount of unexplained variability in a given risk assessment (Giesy and Allred 1985) (Campbell et al. 1999). CVs cannot measure the ability to show significant effects, and thus they do not readily enhance the evaluation and interpretation of micro/mesocosm studies without *a priori* estimates of β and the power of the experiment (Kennedy et al. 1999).

State of the art, in microcosm science is then that it is not possible to determine an acceptable level of CVs, the level of CV is context dependent for each study and the individual scope and purpose of this individual study. To enhance the interpretability and implementation of micro/mesocosm studies in the pesticide registration process, the following context dependent steps could be implemented; 1) Determine the appropriate experimental design and number of replicates by determination of β and the power of the design *a priori*. 2) Utilisation of advanced statistical analysis, i.e. probabilistic effect distribution and principal response curves. 3) Reporting, preferably in quantitative terms, the certainty and uncertainty of the data and the risk of a type II error. The balancing of ecological accuracy and statistical precision continue to be a challenge of micro/mesocosm science before a routinely and standardised implementation of the methodology in the risk assessment of chemicals are feasible according to the USEPA memorandum of 1992 by Fisher (1992).

Acknowledgements. Go to Patrik Kraufvelin University Åbo, Finland, Søren Petersen Danish Hydraulic Institute & Jette Rank University of Roskilde, Denmark.

References

Abbott W (1966): Microcosm studies on estuarine waters. 1 the replicability of microcosms. J Water Poll Contl **38** (2) 258–270

Bott TL, Kaplan LA (1993): Persistence of a surrogate for a genetically engineered cellulolytic microorganism and effects on aquatic community and ecosystem properties: Mesocosm and stream comparisons. Can J Microbiol **39**, 686–700

Boyle TP, Fairchild JF, Robinson-Wilson EF, Haverland PS, Lebo JA (1996): Ecological restructuring in experimental aquatic mesocosms due to application of diflubenzuron. Environ Toxicol Chem **15**, 1806–1814

Campbell PJ, Arnold DJS, Brock TCM, Grandy NJ, Heger W, Heimbach F, Maund SJ, Streloke M (1999): Guidance document on higher tier aquatic risk assessment for pesticides (HARAP). SETAC-Europe Publication

Caquet T, Lagadic L, Jonot O, Baturo W, Kilanda M, Simon P, Le Bras S, Echaubard M, Ramade F (1996): Outdoor experimental ponds (mesocosms) designed for long-term ecotoxicological studies in aquatic environment. Ecotoxcol Environ Safety **34**, 125–133

Conquest LL (1983): Assessing the statistical effectiveness of ecological experiments: Utility of the coefficient of variation. Intern J Environ Stud **20**, 209–221

Crane M (1997): Research needs for predictive multispecies tests in aquatic toxicology. Hydrobiologia **346**, 149–155

Crossland NO, Mitchell GC, Dorn PB (1992): Use of outdoor artificial streams to determine threshold toxicity concentrations for a petrochemical effluent. Environ Toxicol Chem **11**, 49–59

Crossland NO, Heimbach F, Hill IR, Boudou A, Leeuwangh P, Matthiessen P, Persoone G (1994): Summery and recommendations of the (EWOFFT) pp xxv–xxxvii. In: Hill IR, Heimbach F, Leeuwangh P, Mathiessen P (eds.) Freshwater field-tests for hazard assessment of chemicals. CRC Press USA.

Cuppen JGM, Gylstra R, van Beusekom S, Budde BJ, Brock TCM (1995): Effects of a nutrient loading and insecticide application on the ecology of Elodea-dominated freshwater microcosms. III. Responses of macroinvertebrate detritivores, breakdown of plant litter and final conclusions. Arch Hydrobiol **134**, 157–177

DeNoyelles F, Kettle WD, Sinn DE (1982): The responses of plankton communities in experimental ponds to atrazine, the most heavily used pesticide in the United States. Ecology **63**, 1285–1293

Fairchild JF, La Point TW, Schwartz TR (1994): Effects of a herbicide and insecticide mixture in aquatic mesocosms. Arch Environ Contam Toxicol **27**, 527–533

Fairchild JF, La Point TW, Zajicek JL, Nelson K, Dwyer FJ, Lovely PA (1992): Population-, community- and ecosystem-level responses of aquatic mesocosms to pulsed doses of a pyrethroid insecticide. Environ Toxicol & Chem **11**, 115–129

Farmer D, Hill IR, Maund SJ (1995): A comparison of the fate and effects of two pyrethroid insecticides (lambda-cyhalothrin and cypermethrin) in pond mesocosms. Ecotoxicol **4**, 219–244

Fisher L (1992): Memorandum. Decision on the ecological, fate, and effects task force. 29/10-1992, USEPA

Gagnon M, Lacroix G (1981): Zooplankton sample variability in a tidal estuary: An interpretative model. Limnol & Oceanogr **26**, 401–413

Giesy JP, Allred PM (1985): Replicability of aquatic multispecies test systems. Chapter 14. In: Multispecies toxicity testing. Cairns J (ed) Pergamon Press Inc. UK.

Green RH (1989): Power analysis and practical strategies for environmental monitoring. Environ Res **50**, 195–205

Hammons AS, Giddings JM, Suter II GW, Barnthouse LW (1981): Methods for ecological toxicology A critical review of laboratory multispecies tests. USEPA, pp 58–74

Hendley P, Giddings J (1999): ECOFRAM. Aquatic report, chapter 5/4-99, USEPA

Howick GL, Giddings JM, DeNoyelles F, Ferrington LC Jr, Kettle WD, Baker D (1992): Rapid establishment of test conditions and trophic-level interactions in 0.04 ha earthen pond mesocosms. Environ Toxicol Chem **11**, 107–114

Isensee AR (1976): Variability of aquatic model ecosystem-derived data. Intern J Environ Stud **10**, 35–41

Jak RG, Maas JL, Scholten MCTh (1998): Ecotoxicity of 3,4-dichloroaniline in enclosed freshwater plankton communities at different nutrient levels. Ecotoxicol **7**, 49–60

Johnson BT (1986): Potential impact of selected agricultural chemical contaminants on a northern prairie wetland: A microcosms evaluation. Environ Toxicol Chem **5**, 473–485

Kennedy JH, Ammann LP, Waller WT, Warren JE, Hosmer AJ, Cairns SH, Johnson PC, Graney RL (1999): Using statistical power to optimise sensitivity of analysis of variance designs for microcosms and mesocosms. Environ Toxicol Chem **18**, 113–117

Knuth ML (1992): The accuracy and precision of pesticide concentrations following application to littoral enclosure. In: A Canadian technical report of fisheries and aquatic sciences, pp 137–138

Kraufvelin P. (1998): Model ecosystem replicability challenged by the 'soft' reality of a hard bottom mesocosm. J Exp Mar Biol Ecol **222**, 247–267

Kraufvelin P (1999): Baltic hard bottom mesocosms unplugged: Replicability, repeatability and ecological realism examined by non-parametric multivariate techniques. J Exp Mar Biol Ecol **240**, 229–258

Liber K, Kaushik NK, Solomon KR, Carey JH (1992): Experimental designs for a comparison of the 'ANOVA' and 'regression' design for assessing the impact of tetrachlorophenol on zooplankton populations in limnocorrals. Environ Toxicol & Chem **11**, 61–77

Lundgren A (1985): Model ecosystems as a tool in freshwater and marine research. Arch Hydrobiol/suppl 70 (Monographische Beiträge) **2**, 157–196

Maguire B, Slobodkin LB, Morowitz HJ, More B, Botkin DB (1980): A new paradigm for the examination of closed ecosystems. In: Giesy JP Jr (ed) Microcosms in Ecological Research. Technical Information Centre, Springfield/VA, USA, pp 1–13

Persoone G, Janssen CR (1994): Field validation of predictions based on laboratory toxicity tests. Chapter 25 in: Aquatic mesocosm studies in ecological risk assessment. Graney RL, Kennedy JH, Rodgers JH (eds.) CRC Press, Inc

Pratt JR, Melendez AE, Barreiro R, Bowers NJ (1997): Predicting the ecological effects of herbicides. Ecol App **7**, 1117–1124

Richardson JS, Perrin CJ (1994): Effects of the bacterial insecticide Bacillus thuringiensis var. kurstaki (Btk) on a stream benthic community. Can J Fish Aquat Sci **51**, 1037–1045

Rosenzeig M, Buikema A Jr (1994): Phytoplankton colonisation and seasonal succession in new experimental ponds. Environ Toxicol & Chem **13**, 599–605

Sanderson H, Petersen S (2001): Statistical power analysis of data from quasi-natural mesocosm ponds perturbed with Roundup. Environ Toxicol & Chem 2001 (in press)

Schindler DW (1998): Replication versus realism: The need for ecosystem-scale experiments. Ecosys **1**, 323–334

Shaw JL, Copeland JR Johnson JK (1995): Use of coded wire tags to identify fathead minnow (Pimephales promelas rafinesque) adults in an outdoor microcosm study designed to evaluate consistency in reproduction. Environ Toxicol Chem **14**, 1773–1780

Touart, L (1988): Hazard evaluation division: Technical guidance document. Aquatic mesocosm tests to support Pesticide Registration. EPA/540/09-88-035 OPP, USEPA

USEPA (1987): FIFRA-SAP (1987): A set of scientific issues being considered by the agency on aquatic mesocosm testing to support pesticide registration. 25/6-1987, USEPA

van Wijngaarden RPA, van den Brink PJ, Voshaar JHO, Leeuwangh P (1995): Ordination techniques for analysing response of biological communities to toxic stress in experimental ecosystems. Ecotoxicol **4**, 61–77

Webber EC, Deutsch WG, Bayne DR, Seesock WC (1992): Ecosystem-level testing of a synthetic pyrethroid insecticide in aquatic mesocosms. Environ Toxicol Chem **11**, 87–105

Received: May 17th, 2001
Accepted: August 21st, 2001
OnlineFirst: September 3rd, 2001

Appendix: Table 1 and 2

Table 1: Overview of the experiments reviewed in the article. The table indicates types of pesticide and freshwater system studied, study locations, size and references

Pesticides				
Active Ingredient	**System Studied**	**Size**	**Location**	**Author**
Atrazine	Nw; Ns; Le; out; n= 2	0.045 ha	USA (Kansas)	deNoyelles et al. 1982
Atrazine + Esfenvalerate	Ne; Ns; Le; out; n= 3	0.1 ha, 600-850 m^3	USA (Columbia)	Fairchild et al. 1994
Btk **(a)**	Ne; Ns; Lo; out; n= 5	l.w.d.1.52*0.2*0.2 m	Canada	Richardson & Perrin 1994
Chlorpyrifos	Aw; As; Lo; in; n= 4	l.w.d.110*110*50 cm	Netherlands	Cuppen et al. 1995]
Chlorpyrifos	Aw; As; Lo; in; n= 4	l.w.d.110*110*50 cm	Netherlands	van Wijngaarden et al. 1995
Diflubenzuron	Ne; Ns; Le; out; n= 3-4	700m^3	USA (Columbia)	Boyle et al. 1996
3,4-dichloroaniline	Ne; Ns; Le; out; n= 2	1.2 m^3	Netherlands	Jak et al. 1998
Esfenvalerate	Ne; Ns; Le; out; n= 3	0.1 ha, 700 m^3	USA (Columbia)	Fairchild et al. 1992
Esfenvalerate	Aw; Ns; Le; out; n= 3	0.1 ha	USA (Alabama)	Webber et al. 1992
Glyphosate	Nw; Ns; Le; out; n= 2	25 m^3	Denmark	Sanderson and Petersen 2001
λ-cyhalothrin + Cypermethrin Atrazine, Carbofuran, Fonofos,	Aw; Ns; Le; out; n= 2	25 m^3	UK	Farmer et al. 1995
Phorate, Triallate & Treflan Azinphos-methyl, Chlorpyrifos,	Aw; Ns; Le; in; n= 4	4L	USA (Columbia)	Johnson 1986
Controls	AW; Ns; Le; out; n= 12	5*4*1.5m	France	Caquet et al. 1996
Controls	Aw; As; Lo; in; n= 2	l.w.d. 2.23*0.2*0.13 m	USA (Avondale)	Bott et al. 1993
Controls	Nw; Ns; Le; out; n= 4	17 m^3	UK	Shaw et al. 1995
Controls	AW; As; Lo; in; n= 2	l.w.d. 5*0.35*0.25 m	UK	Crossland & Dorn 1992
Controls	Ne; out; Le; n= 3	4-7L	USA (Pennsylvania)	Pratt et al. 1997

Nw = natural whole system; Aw = artificial whole system; Ne = enclosure in natural system; Ns = natural sediment; As = artificial sediment; Lo = lotic; Le = lentic; out = outdoor; in = indoor; n = number of replicates

(a) *Btk* is a bacterial insecticide *Bacillus thuringiensis kurstaki*

Table 2: Average parameter Coefficient of Variation results and design/complexity of study

Author and variables	CV con	CV low	CV high	CV total	Comments
de Noyelles et al. (1982): Fluorescence increase of control and Atrazine pond phytoplankton communities:	17% (9-23%)	23% (6-60%)	33% (8-60%)	24% (17-33%)	Pond water tested in the lab. Decreasing significance with increasing pesticide concentration and decreasing flourescence. Conclusion: Significant effects were found. Overall CV: 25%
Fairchild et al. (1992)					Though reductions in numbers, these were confounded by high within treatment variability. Low power study with too high inter-variability to robustly display significant effects. Lack of replicability in design impeding ability to show significant effects. Approx.10% significant samples. Too few samples 1 before and 4 after. High realism in design and concentrations. Overall CV: 94%
Diptera/L:	84%	85%	69%	76%	
Odonata:	86%	85%	119%	100%	
Ephemeroptera:	90%	82%	119%	110 %	
Gastropoda:	53%	112%	115%	89%	
Coleoptera:	109%	60%	154%	115%	
Total:	68%	61%	67%	91%	
Richardson and Perrin (1994)					A natural stream divided in 15 flow-through mesocosms. More than a 100-fold the recommended concentration did not result in statistically significant effects. Conclusion: *Btk* is not harmful under normal field application. No test of risk of type II error by estimation of power. High realism in design, but high concentration. Overall CV: 18%
Ephemeroptera/L:	24%	23%	16%	21%	
Plecoptera:	25%	20%	33%	26%	
Coleoptera:	23%	15%	8%	15%	
Trichoptera:	16%	24%	22%	21%	
Diptera:	20%	21%	18%	20%	
Oligochaeta:	7%	13%	10%	10%	
Total:	7%	9%	5%	7%	
Cuppen et al. (1995) Relative amounts of residual dry eight of decomposing shoots of *Elodea nuttalli* and *Populus* leaves.					Approx. 1/3 of the control microcosms showed significant differences within replicates. Relatively low realism in design. Overall CV: 9%
Elodea nuttalli:	12% (4-22%)				
Populus:	6% (2-10%)				
van Wijngaarden et al. (1995)					Significant effects were found on G. pulex. Others were, by large not significantly affected. Low realism in design, relatively. High realism in concentrations. Overall CV: 53%
G. pulex/L:	57% (33-80%)	41% (41%)			
Tubificidae:	39% (26-52%)	45% (22-37%)			
P. antipodarum:	64% (44-85%)	69% (52-109%)			
Boyle et al. (1996)					Only significant fish effects on Bluegill were weight, condition and biomass. Assumed effects were confounded by seasonal changes and natural variations. High realism in design and concentrations. Overall CV: 25%
Zooplankton #/L	25%	41%	67%	44%	
Zooplankton species no.	11%	21%	9%	13%	
Zooplankton dominance	18%	18%	14%	17%	
Insects no/0.25 m^2	24%	90%	36%	50%	
Insects species no.	12%	23%	14%	16%	
Insect dominance	31%	16%	20%	22%	
Chlorophyll *a*	53%	22%	28%	34%	
Gross primary production	14%	11%	13%	13%	
Fish total: no, cm, g, kg/ha & condition	18%	47%	39%	35%	
Jak et al. (1998)					P-values were not reported. High realism in design. Overall CV: 18%
1. Experiment:					
Cladocera population densities 21d	17%	20%	14%	17%	
Copepods do.	21%	16%	10%	16%	
Rotifers do.	9%	0,01%	10%	6%	
2. Experiment:					
Cladocera population densities 21d	17%	15%	49%	27%	
Copepods do.	13%	2%	0.4%	5%	
Rotifers do.	14%	9%	0,1%	8%	
Fairchild et al. (1994)					P-values were not reported on crustacean data. High realism in design and concentrations. Overall CV: 55%
1. Experiment:					
Crustacean #./L	50% (6-100%)	32% (14-53%)	75% (10-100%)	57%	
2. Experiment:					
Crustacean no./L	40% (4-100%)	78% (20-200%)	80% (17-113%)	52%	
Webber et al. (1992)					Some significant differences were found in the highest dose on phytoplankton after application. High realism in design and concentrations. Overall CV: 50%
Phytoplankton abundance (Org/ml)	40% (24-82%)	40% (22-87%)	56% (23-122)	44%	
Macroinvertebrate (Org/sample)	28% (18-32%)	34% (20-52%)	54% (19-91%)	39%	
Emerging adult insects	19% (10-29%)	28% (18-47%)	45% (25-106%)	33%	
Bluegill mean no.	38%	45%	39%	44%	
Bluegill mean wt.	61%	58%	43%	54%	
Bluegill length classes 2-10 cm	72%	91%	103%	96%	

Author and variables	CV con	CV low	CV high	CV total	Comments
Sanderson and Petersen (2001)					Too complex quasi-natural design,
Abundance before and after perturbation:					high natural variability, low power and lack of exposure. Significance
Chlorophyll *a*/L	70%		70%	70%	would in a BACI context require
Cyclops sp.	79%		31%	55%	>90% reduction in abundance. High
Daphnia sp.	299%		313%	306%	realism in design and
Total zooplankton	137%		95%	116%	concentrations. Overall CV: 137%
Farmer et al. (1995)					Normal use of either pyrethoroid
Primary producers			52% (44-60%)		would only transiently impact the aquatic ecosystem. No significantly
Zooplankton/L			62% (51-71%)		adverse effects occurred. Risk of type II error and power was not
Macroinvertebrates			36% (20-57%)		estimated. High realism in design and concentrations. Overall CV: 50%
Johnson (1986)					Caution should be taken when
Growth of green alga *Selenastrium Capricornutum*					applying atrazine, fonofos and triallate near wetland habitats.
Carbofuran	18%	4%	16%	13%	Relatively low realism in treatment
Fonofos	14%	12%	14%	20%	design, high top concentrations.
Phorate	6%	7%	10%	7%	Overall CV: 23%
Atrazine	14%	7%	14%	17%	
Treflan	14%	22%	10%	14%	
Triallate	14%	20%	25%	25%	
Submerged macrophyte growth:					
Carbofuran	18%	4%	16%	13%	
Fonofos	14%	12%	14%	20%	
Phorate	6%	7%	10%	7%	
Atrazine	14%	7%	14%	17%	
Treflan	14%	22%	10%	14%	
Triallate	14%	20%	25%	25%	
Caquet et al. (1996)					Overall CV was 42% at the end of
Abundance of:					the stabilisation period. High realism
Mayflies	40% (19-68%)				in design
Chironomids	36% (9-82%)				
Total insects	31% (12-63%)				
Total arthropods	31% (13-61%)				
Total mullusca	85% (19-120%)				
Total invertebrates	31% (13-61%)				
Bott and Kaplan (1993)					No significant effects were found.
Total bacterial densities:					Relatively low realism in design.
Sediments	45%	75%			Overall CV: 47%
C. glomerata	41%	47%			
Leaf packs	96%	60%			
Alga biomass:					
Sediments	21%	23%			
Leaf packs	49%	78%			
Total viable biomass					
Sediments	55%	254%			
Leaf packs	57%	62%			
Alga primary productivity:					
Sediments	50%	88%			
Shaw et al. (1995)					Significance was not reported on
Survival of fathead minnow:	54%				survival of fathead minnow. High
Total biomass	6%				realism in design and
Total No.	21%				concentrations. Overall CV: 54%
Crossland et al. (1992)					Significant effects were found on
#/L					zooplankton feeding rate, but not on
G. pulex	20%	16%	42%		abundances. Relatively high realism
Beatis sp.	53%	74%	120%		in design and concentrations.
Agapetus sp.	51%	59%	74%		Overall CV: 16%
Elminthidae	43%	58%	37%		
Total invertebrates	21%	15%	36%		
Pratt et al. (1997)					Low levels of atrazine significantly
Effects of atrazine on microbial species richness and chlorophyll *a*:					increased chlorophyll *a* concentration. Relatively low realism
Species	7%	4%	5%	9%	in design. Overall CV: 29%
Chl.a	37%	32%	133%	48%	
Effects of diquat on Chl.a:	10%	19%		15%	

S = Statistical significance, P <05; significant results are underlined; CV con = control; CV low = lowest dose and CV high = CV highest dose

II

Applicability of quasi-natural mesocosm ponds and pelagic plankton for higher tier risk assessment of $Roundup_{2000}$

by

Sanderson H. & S. Petersen. 2002

Archives of Environmental Contamination and Toxicology, in review

Applicability of quasi-natural mesocosm ponds and pelagic plankton for higher tier risk assessment of Roundup$_{2000}$

Hans Sanderson*. *Department for Environment, Technology and Social Studies,* University of Roskilde, PO-Box 260, DK4000 Roskilde, Denmark. Tel: +45 46 75 24 96. Fax *+45 46 74 30 41. E-mail:* hanss@ruc.dk Søren Petersen, ‡. *DHI, Water and Environment, Agern Alle 5, DK2970 Hørsholm, Denmark. E-mail:* sop@dhi.dk

Abstract

The fate and effect of Roundup$_{2000}$ on a suite of autotrophic and heterotrophic planktonic organisms were assessed by an experiment with artificial ponds conducted during spring 1999. Roundup was sprayed as a single event corresponding to a worse case scenario as a direct spraying of the recommended field dosage to the surface of the pond. The half life for disappearance of glyphosate from the water column was < 1 day and as a consequence of the rapid removal from the water column no effects on the pelagic plankton was detectable, although the initial nominal concentration of Roundup was 17.85 mg L^{-1} and glyphosate was 8.56 mg L^{-1} which is higher than the standard (LC_{50} (24h) *Daphnia magna* = 2.0-3.5 mg L^{-1}). In addition a power analysis (the probability of detecting a possible effect) revealed that a relative change of 93 % or even higher was required for the present experimental design (n=2). Or, 14-45 replicates at each treatment level was required to detect a relative change of 50 % as statistical significant at the 5 % level. Thus although artificial ponds have been acknowledge for their proposed high ecological relevance the above results might challenge this assumption, at least for zooplankton.

Keywords: Mesocosms; Design; Roundup$_{2000}$; Power.

Introduction

Despite the wide use of the herbicide Roundup and the toxicological impacts documented through several laboratory assays, studies (Giesy *et al.*, 2000), examining the effects and fate of the product in quasi-natural field test systems are scarce. Moreover the official WHO review (1994) refers to Hildebrand et al., (1980) who found that 100 times recommended agricultural dosage resulted in no effects under field conditions, despite the fact that these treatments were well above the NOEC's from laboratory assays (Giesy *et al.*, 2000).

The Giesy *et al.*, (2000) review moreover concludes that more research is needed in lab-to-field extrapolation of Roundup ecotoxicity. The aim of this paper was thus to investigate whether the dichotomy between the laboratory assays and field studies is attributable to methodological problems associated with testing Roundup in aquatic quasi-natural mesocosm ponds?

* To whom correspondence may be addressed: hanss@ruc.dk

Mesocosm studies represent a realistic exposure of the test organisms to the toxicant and, moreover, they reduce the uncertainty caused by extrapolation of adverse effects from standard laboratory bioassays to the environment by including several different species and ecological interactions such as predation and competition, etc. (Campbell *et al*., 1999).

Although the fate of pesticides in quasi-natural field-test systems might be considered as ecological realistic, the exposure might also vary tremendously from the exposure in laboratory assays. Hence, several protocols for laboratory bioassays recommend a constant exposure of the organisms for days or even weeks. However, following a single pulse addition glyphosate and several other pesticides might only remain in the water column for a few hours or even less in quasi-natural systems. Thus to achieve a proper understanding of an effect study conducted under semi-field conditions considerations of the fate of the pesticide are also needed (Styczen, *et al.,* 2001).

Due to the plethora of abiotic and biotic condition species and ecological interactions an exact replication of a quasi-natural field systems is a difficult task (Crane, 1997) (Sanderson, 2001). Hence, although the plethora might reduce the extrapolation uncertainty it might inherently give rise to a high background variation, which might obstruct the detection of adverse effects. Hence a recent overview of field studies and pesticides revealed that a relative reduction of 50% or even more were needed to obtain statistical significant differences ($p<0.05$) when compared to control systems (Farmer *et al*., 1995). Thus for a proper planning, execution and interpretation of a quasi-natural field experiment the power of the applied statistical tests need to be considered carefully (Møhlenberg *et al.,* 2001).

Or in statistical terms, the problem of high variability may lead to higher type II error rates, or the risk of failure to detect an effect of the pesticide in the treated system, even though an effect was present (Peterman & M'Gonigle, 1992) (Green, 1989).

Traditionally, the data analysis of measures of impact in aquatic systems is a twostep process. First, a statistical analysis is conducted to examine the statistical significance of the data; then, an ecological assessment of these results is conducted to examine causality or ecological significance. Several types of errors influence the interpretation of the results, including the rejection of H_0 when it is in fact true (Type I error), or the non-rejection of H_0 when it is in fact false (Type II error) as mentioned above. Statistical power is inversely related to the probability of making a type II error, i.e. to conclude that a significant effect has not occurred, even though it has. Power increases with increasing numbers of replicates and decreasing variability (s^2) and also depends on the value of α (the risk of a type I error) and the effect size to be detected (Δ), i.e. increases with increasing (relaxed) α and increased effect size (Δ). When the variability of the endpoints is high, there is an increased chance of failing to reject the H_0 when there is an effect (type II error) (Green, 1989) (Ammann *et al*., 1997).

The aim of the study was thus to investigate the applicability of mesocosm ponds and pelagic plankton in higher tier risk assessment of Roundup. We focused on effects on the pelagic plankton communities and on the fate of glyphosate and thus organism exposure. The effect size (Δ) was set to 0.5 (or 50%) as least „ecologically significant" reduction of plankton abundance due to rapid recovery time, the ecological significant change (reduction) is assumed to be 50% for plankton. Below this recovery for species with rapid generation

times will return within a few days or weeks (Farmer *et al.*, 1995) (Ammann *et al.,* 1997).

Materials and methods

Mesocosms

The aim of the mesocosm design was to simulate as far as possible the natural conditions of Danish ponds in the agricultural landscape. The facility consists of four round whole, earthen dugout ponds. The mesocosms were filled in the spring 1994 with water from the local municipal water supply. Pond levels were maintained with water from a larger reservoir pond, which was filled, with municipal pipeline water. The sizes of the ponds are app. $25m^2$ with sloping banks, average maximum depth of approximately 1.0 m. They are located at Roskilde University in Denmark and were investigated spring 1999 (6/4-15/6-99). The ponds became colonised over a period of five years. The waterbodies have been spontaneously colonised by phytoplankton, *Copepods, Cladocerans,* and *Gastropoda* sp. Macroinvertebrates such as *Hydrocores* sp., *Phrygane* sp. and *Ephemeroptera* sp.; *Assellus aquaticus* are abundant in the benthic community. Macrophytes consist primarily of *Elodea Canadensis, Potamogeton* sp., *Myriophyllum spicatum* and *Sparganium* sp. No fish were found in the ponds prior to the investigation by electro fishing, nor were any amphibians present. The ponds lay in series thus windspeed and solar exposures are identical. The sediment is a compacted layer of clay and gravel, topped with loose soil and organic matter from the excavation. The sediment and the macrophytes had prior (in March) to the investigation gently been made uniform by hand macrophytes were regulated so that they covered approximately 25-30% of the area of the bottom (Crossland & Wolff, 1988) in all four ponds, in accordance to OECD guideline (1996) on freshwater lentic field-test. The ponds were randomly chosen for controls or treatment among the four existing ponds. The ponds are independent of each other as they consist of four independent ponds, moreover leaks between the ponds transporting Roundup did not occur and no Roundup or metabolites were found in the control ponds.

Treatment

During the first five weeks of the investigation, the relative structural and functional similarities between the four ponds were investigated on a weekly basis before treatment with the formulated two-component herbicide $Roundup_{2000}$, consisting of containing 48% of the active ingredient glyphosate ($C_3H_8NO_5P$) CAS-No. [1071-83-6] and the additive Teamup.

Roundup was chosen for a number of reasons. Roundup is a broad-spectrum and highly used herbicide both in agricultural use and domestic use (Cox, 1998). Acute lethal concentrations (48-96h) for *Daphnia* of Roundup obtained from single species laboratory tests varies between 2.0 and 3.5 mg/L (Hartman & Martin, 1984) (Servizi *et al.*, 1987) (Giesy *et al.*, 2000). Aged semi-field tube-investigations indicate no toxic effect of Roundup on zooplankton at 100 times applied agricultural concentration (2.2 kg a.i./ha) for that formulation (Hildebrand *et al.,* 1980).

No mesocosm test using Roundup has been reported in the scientific literature (Giesy *et al.*, 2000). Roundup was dosed as a single treatment worst-case. It was applied in an aqueous solution simulating recommended agricultural use and drifts, applied by hand using a manual sprayer, uniformly covering the

whole surface of the ponds. Recommended application rates in this study do not exceed 5.8 kg a.i./ha. Roundup/ha and were applied to each of the two treaded pond by hand using a back-pack spray simulating a worst-case with full over spray of the ponds. The concentration of gylphosate in the pond water was measured 24 hours after treatment and again 30 days after treatment.

It is notable that Roundup is more toxic for invertebrates than the active ingredient gylphosate, this is mainly due to the presence of a polyethoxylated tallowamin (POEA) surfactant (CAS number 61791-26-2), which is a mixture of polyethoxylated long-chain alkylamines synthesised from animal-derived fatty acids, which is the predominant surfactant used in glyphosate-based products (Giesy *et al*, 2000).

Table 1. Mesocosm water quality characterisation. Mean and range across the test period sampled weekly and biweekly (n=2)

Variable	Mean	Range
pH	6.05	5-7
Temperature C^o	10.8	5.1-16.2
O_2 ppm	9.9	7.8-12.9
K ppm	0.05	0-0.5
Ca ppm	60	18.8-97.9
Si ppm	0.4	0-2
Na ppm	8	4.3-13.7
NO_3^- ppm	1.4	<1-5.3
Total P µg/L	< 20	0 - <20

Sampling

The two treated ponds received the same dosage and nominal concentration of Roundup per litre and the same test regime before and after treatment was implemented. We sampled five times before and five times after treatment. Samples were collected once a week.

The organisms were collected with a 2 L horizontal depth-integrating watercollector, from April 1999 to July 1999. Five samples of 2 L water were collected randomly in each pond and pooled (Rosenzweig & Buikema, 1994). Samples for assessment of zooplankton were obtained by filtering ten L of pond water from each pond through a 90 µm mesh sieve, followed by fixation in Lugol´s solution. The physio-chemical background characteristics of the water such as maximum and minimum temperature, dissolved oxygen, pH, Ca, Si, Na, total P and NO_3, were measured weekly, following OECD guidelines (DS 291). Total chlorophyll *a* (ug/L) was determined spectrophotometrically following guidelines (DS 2126) immediately after each sampling occasion. All zooplankton abundance numbers and macroinvertebrates were counted and determined to species immediately after sampling. *Cyclops canthocamptus staphylinus* were counted with a counting cell.

Measurements

Zooplankton abundance numbers were counted using a microscope (type: Leica Wild M3Z). Total chlorophyll *a* was filtered for 30 minutes in GF/C 47mm filter and hereafter extractrated in 10 ml ethanol for 24 hours. They were centrifuged for 10 min. at 12.000 rpm in a Centrivon T-42K centrifuge (Milan, Italy), before spectrophotometric analysis in Perkin-Elmer Lambda 11 uv/vis spectrometer λ = 665 & 750 nm (Ivyland, PA, USA). Phosphate was also determined by spectrophotometric analysis. Nitrate was measured by Nitrate-selectrod and pH by a pH-selectrod on a Radiometer PHM 95. Oxygen and temperature were estimated with electrode OXY197 by wtw Moberg (Weilheim, Germany). Metals were analysed according to OECD guidelines (DS 2214) on flame atom absorption spectrometer, Varian Spectra AA 250 Plus. The determination of glyphosate in the pond water was performed at national accredited laboratories: Technological Institute of Denmark, and the HPLC-fluorescenseanalysis was done after liquid/liquid stirring (DTI, 1999).

Statistics

To conduct an unambiguous power analysis a statistical model, a significance level (α), a power (1-β) and an effect size (Δ) need to be defined and an estimate of the background variance (s^2) is needed using a Student *t*-test approach (Green, 1989). The two factor BACI analysis of variance approach (Stewart-Oaten *et al.*, 1992) using the interaction between day and treatment for statistical tests was also tested, however, the power of the Student *t*-test was highest and thus favoured determined using relationship (1) by Green (1989) moreover Student *t*-test is more widely used and accepted. We acknowledge that multiple *t*-test comparisons can be problematic but compensation for this would only result in further decreased statistical power.

(1) $n = 2(t_\alpha + t_\beta)^2 (s/\Delta)^2$

n = estimated number of samples
t_α = t-value for α
t_β = t-value for β
s = estimated error standard deviation
Δ= effect size

For the Student *t*-test the error variance was estimated as the average of the variances between the replicated mesocosms at each combination of treatment and day. Prior to the power analysis and the statistical tests all data were log transformed, since the assumption of variance homogeneity and normal distributed residuals were full filled for log-transformed (logx+1) data but not for data on a normal scale, the tests were performed in SAS™.

Results

Aqueous concentrations of glyphosate declined rapidly from the water column over the five week test period. Initial nominal concentrations after worst-case treatment regime would result in app. 17.85 mg/L Roundup and thus 8.56 mg glyphosate L^{-1}. Measured concentrations 24 hours post treatment in the ponds were 0.047 and 0.04 mg glyphosate L^{-1} indicating a very rapid removal of the active ingredient from the water column (DTI, 1999). Post-study concentrations at day 30 were 0.0016 and 0.00017 mg glyphosate L^{-1}, significantly lower than after 24 hours ($p < 0.05$) (Fig. 1). The half-life range for the surfactant POEA has been set conservatively to 21-42 days and half-life for glyphosate is set conservatory to 7-14 days (Giesy, *et al.*, 2000).

Figure 1: Glyphosate environmental fate in water

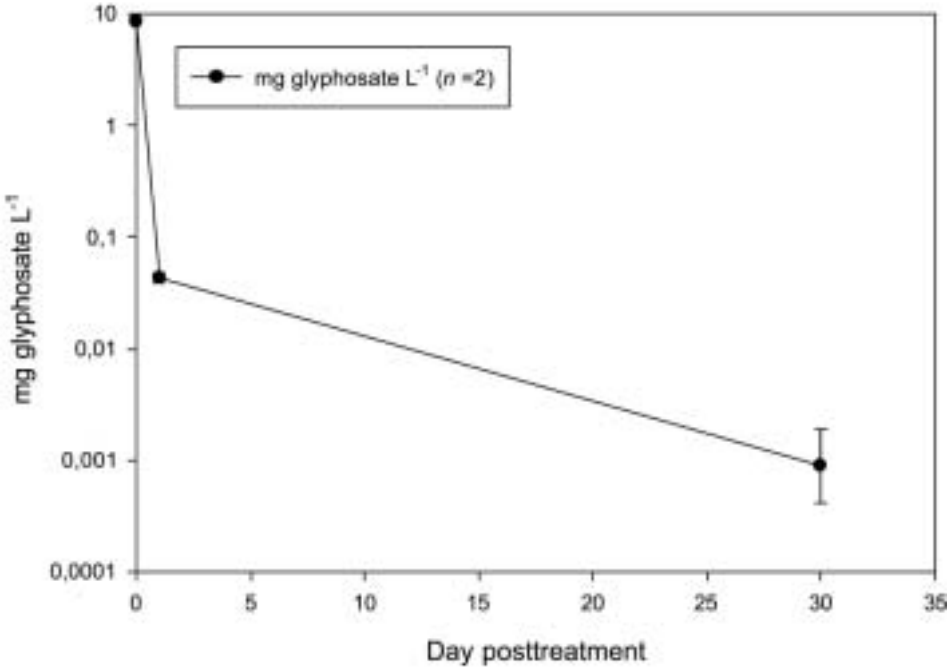

Fig 1: The measured dissipation time of the active ingredient glyphosate in Roundup *in water in the ponds (*n*=2), nominal concentration after treatment (day 0), measured concentration after 24 hours and after 30 days.*

Figure 2: Chlorophyll *a* concentration

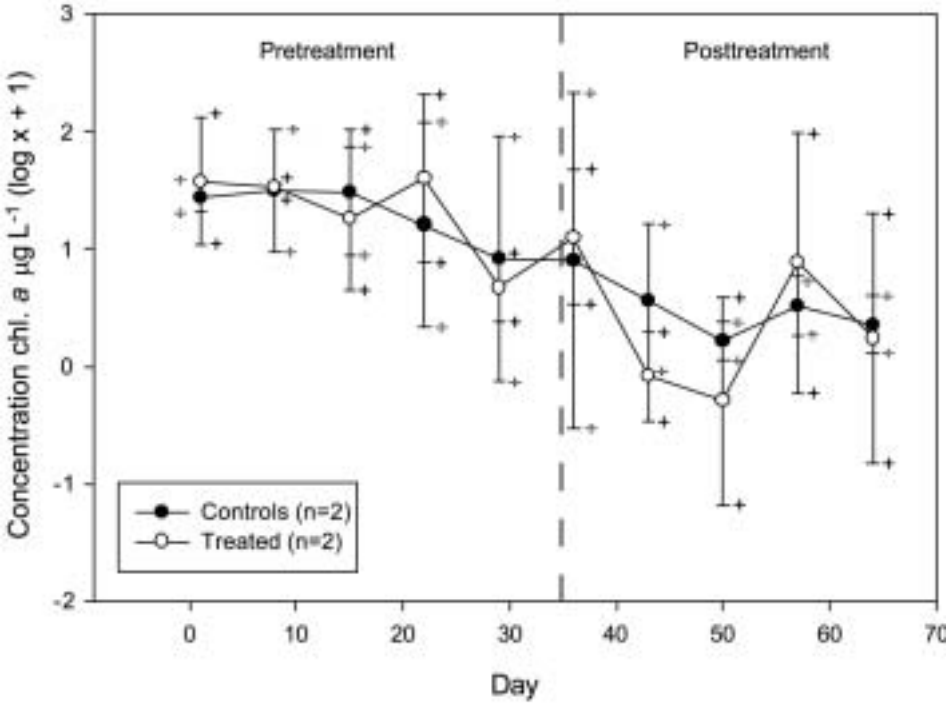

Fig 2: Chlorophyll a *concentration (µg L-1) (log x + 1) in the ponds before and after treatment at day 35. The sign (÷) captures the errorbars (± standard deviation) for the control ponds and ditto (+) captures the errorbars for the treated ponds (*n*=2). No significant effects.*

Figure 3: *Cyclops* sp.

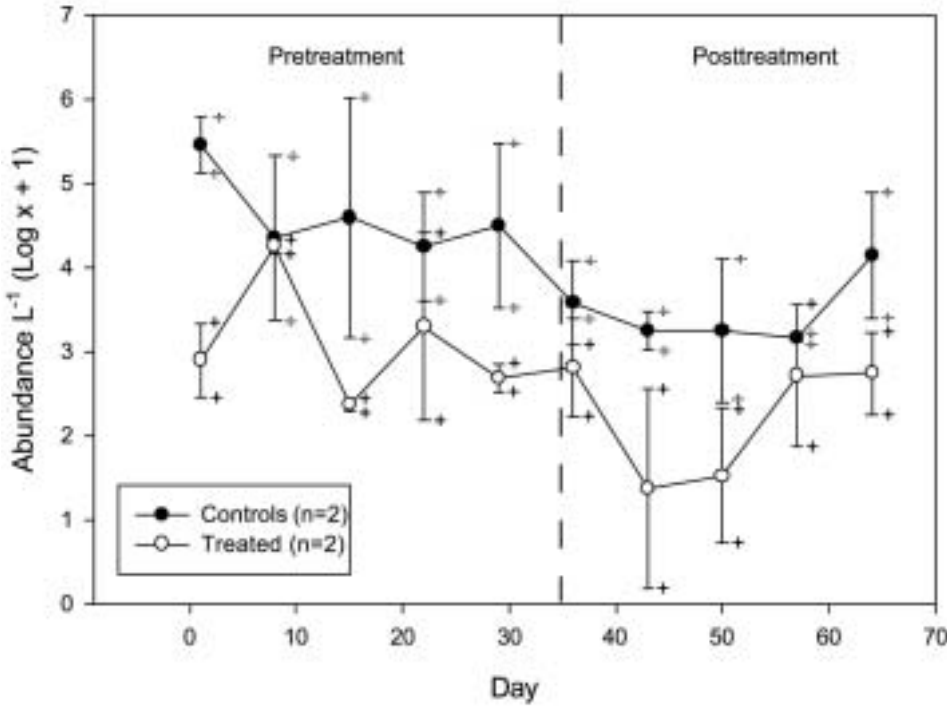

Fig 3: Abundance of Cyclops *sp. L-1 (log x+ 1) in the ponds before and after treatment at day 35. The sign (÷) captures the errorbars (± standard deviation) for the control ponds and ditto (+) captures the errorbars for the treated ponds (*n=*2). No significant effects.*

Figure 4: *Daphnia magna*

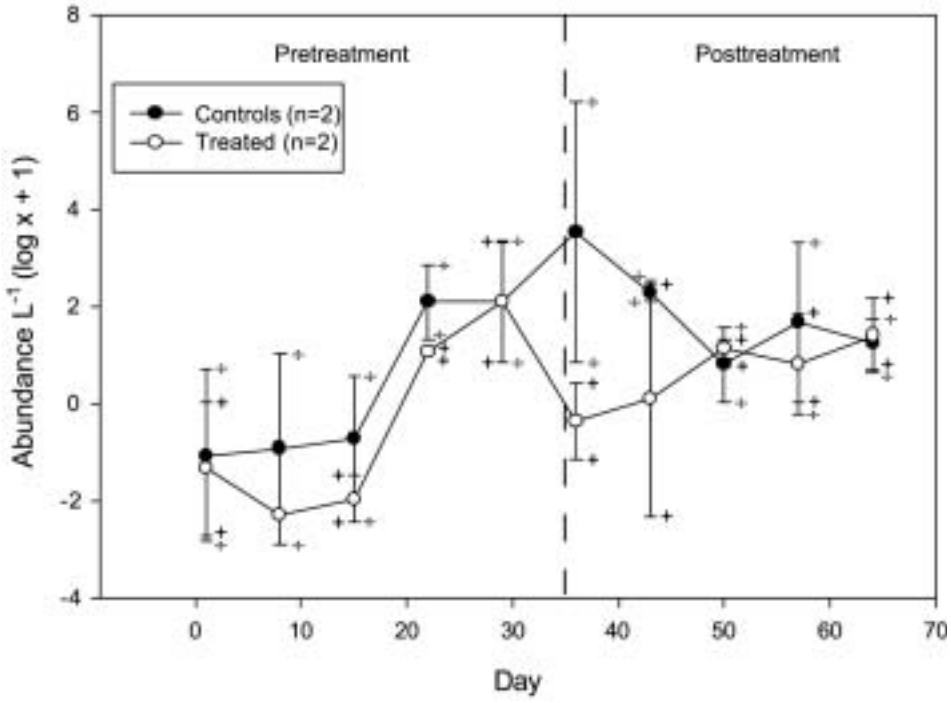

Fig 4: Abundance of Daphnia magna *L-1 (log x + 1) in the ponds before and after reatment at day 35. The sign (÷) captures the errorbars (± standard deviation) for the control ponds and ditto (+) captures the errorbars for the treated ponds (*n=*2). No significant effects.*

Figure 5: Total zooplankton

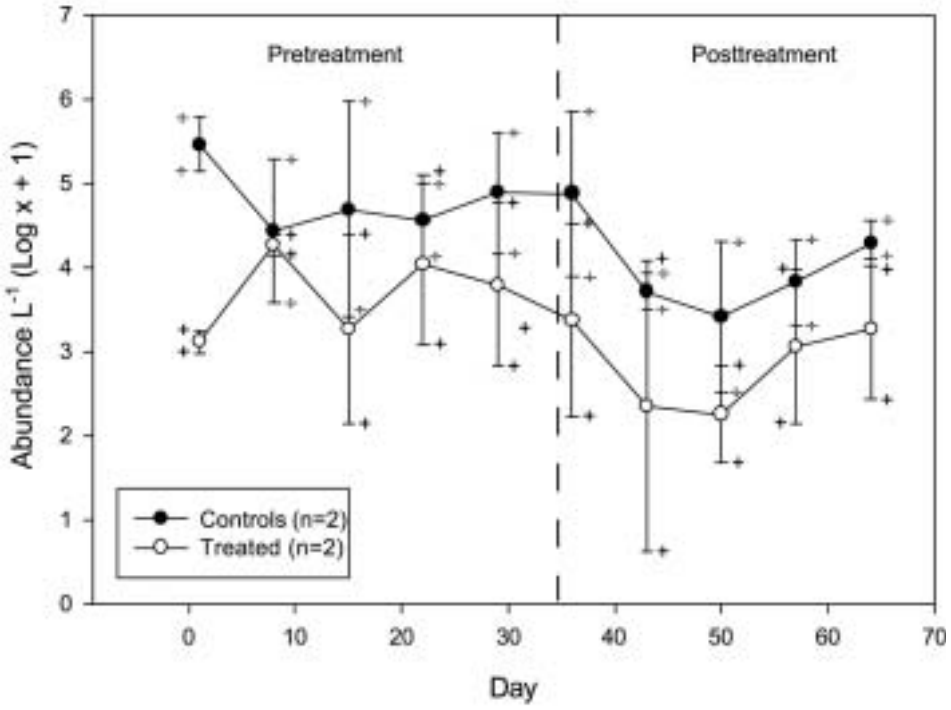

*Fig 5: Total abundance of pelagic zooplankton L^{-1} (log x + 1) in the ponds before and after treatment at day 35. The sign (÷) captures the errorbars (± standard deviation) for the control ponds and ditto (+) captures the errorbars for the treated ponds (*n*=2). No significant effects.*

No changes in zooplankton abundance or chlorophyll *a* content after the application of Roundup were statistically significant at any time point during the study.

Whilst failing significantly to reject the null hypothesis an *a posteriori* analysis of statistical power was implemented. High power ($1-\beta>0.8$ $\alpha=0.05$, $n=2$ and $\Delta=0.5$) not achieved for any parameter, see figure 6. It would moreover, as the graph shows, require more than a 95% reduction in abundance L^{-1} in order to obtain high power (Student *t*-test). Power could also be increased by increasing the number for replicates. The relationship among power and replicates under the following conditions ($\alpha \geq 0.05$, $\beta \geq 0.2$, $\Delta=0.5$) was determined according to relationship (1) by Green (1989) the number of replicates needed for chlorophyll *a* was: >14, for *Cyclops* sp. >15, total zooplankton >20 and for *Daphnia* sp. >45, see figure 7.

Figure 6: t-test power curves (n=2, α=0.05)

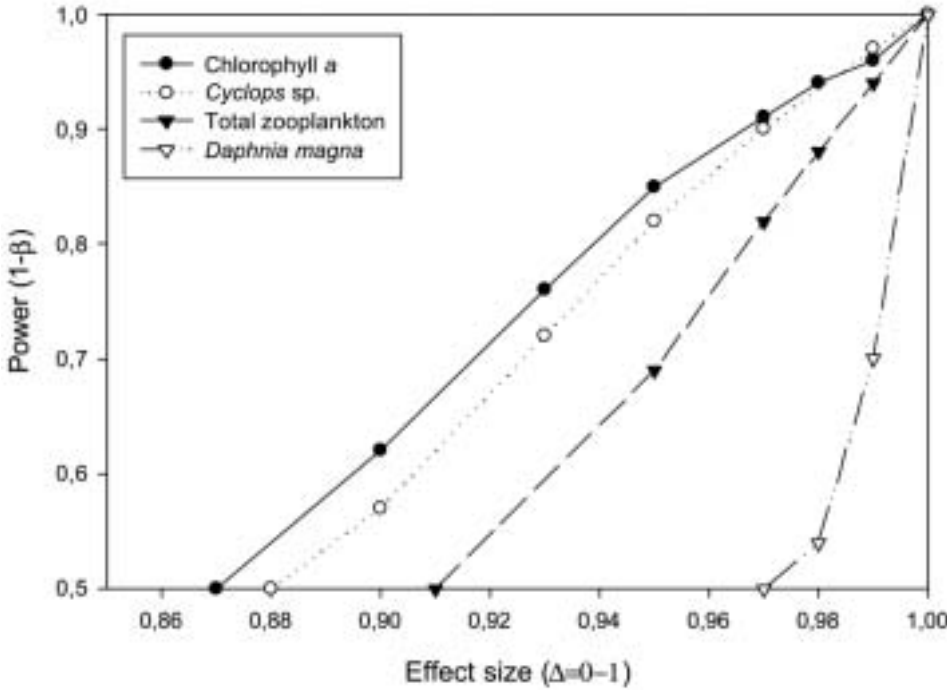

Fig 6: Effect size versus statistical power for the four endpoints. For high power (0.8) the required effect size (Δ) range (0.94-0.99) or 94-99% reduction.

Figure 7: Power versus replicates (Δ=observed, α=0.05)

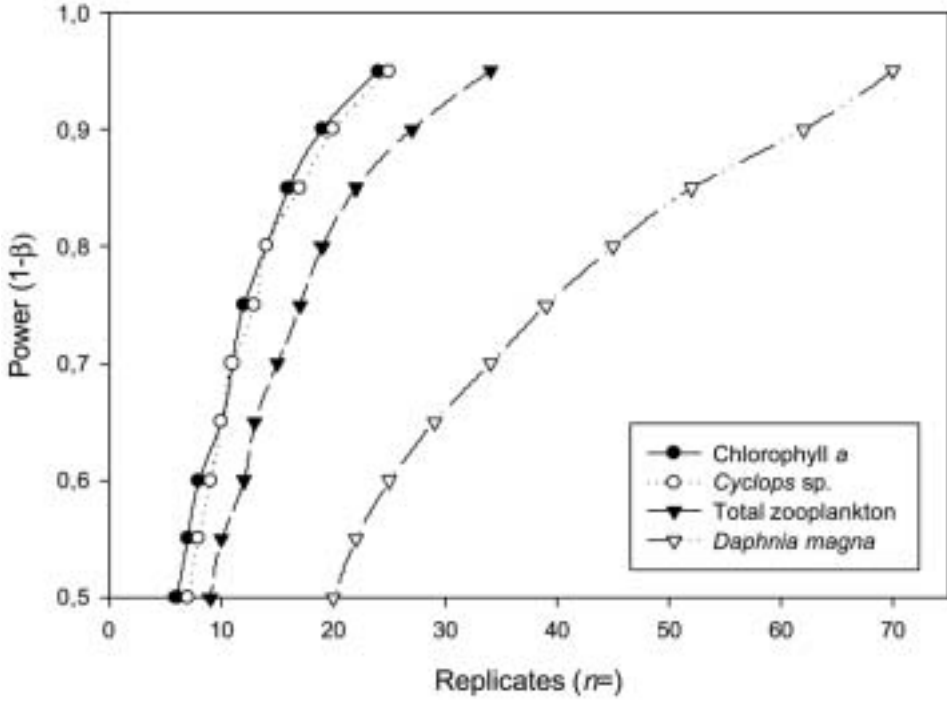

Fig 7: Number of replicates versus statistical power for the four endpoints. For high power (0.8) the required number of replicate ponds (n) range 14-45 ponds.

Discussion and conclusion

No significant effects were found in zooplankton abundance after the application of Roundup. The very rapid removal of the substance from the water column resulted in no significant exposure or detectable adverse effects of the pelagic plankton communities. The concentration of glyphosate was reduced by more than 99% 24 hours post treatment, most likely due the process of reversible adsorption through the phosphonic acid moiety to clay and organic particles (Hildebrand *et al.,* 1980). This means that the pelagic plankton community endpoints where not effective for displaying environmental risks associated with mesocosm risk assessment of this pesticide due to lack of exposure, demonstrating methodological problems of securing calculated exposure in assessing ecological risks of this pesticide in the water column in aquatic mesocosms. The benthic community would probably be more exposed than the pelagic and choosing the benthic community instead would thus probably more effectively reveal environmental risk associated with Roundup and other compounds with similar rapid removal from the water column (Møhlenberg *et al*., 2001). Frank *et al*., (1990) found that in rural areas where glyphosate had been applied, glyphosate was detected in only 2 ponds out of 211, moreover the concentration detected was less than 0.15 mg glyphosate L^{-1}, indicating a rapid removal from the water column.

In this light, the lack of effects at 100 times normal dosage concentration on *Daphnia magna* in a forest pond by Hildebrand *et al*., (1980) is not surprising due to removal from the water column and thus lack of exposure. Fate was not monitored in their study and only nominal concentrations were used.

We did not see any adverse effects on the pelagic plankton communities after treatment with Roundup the most obvious reason for this was the lack of exposure due to rapid dispersal of glyphosate and POEA.

However, from a methodological point of view it is still interesting to examine the detectability of the quasi-natural design and usage of pelagic plankton as toxicological endpoints for higher tier risk assessment of Roundup, recommended by Giesy *et al.,* (2000). Due to the lack of significant differences in the present study, the risk of committing a type II error (accepting a false non-rejection of the null hypothesis) and contributing factors had to be assessed. The reasons for low power were plentiful but originated from the primary problem, the quasi-natural test design and thus high natural variability, which is a common problem within mesocosm science (Kraufvelin, 1998). A determinate factor was the zoo-plankton sample variability, and important sources of variability in estimation plankton abundance in a parcel of 12 water are the patchy distribution of the animals' (Gagnon & Lacroix, 1981), day-and-night variations, migration in the water column, seasonal variations plus possible behaviour i.e. refugeseeking further ad to the sample variability (Schindler, 1998).

This study illustrates the dichotomy between precision and accuracy in quasinatural mesocosm studies, which needs to be considered prior execution of the study. Based on our study, we are able to conclude, that in the process of designing a higher tier aquatic risk assessment in meso/microcosms, according to EU-directive 91/414EEC appendix 2&3, to consideration of the following three items:

1) The environmental fate (*e.g.* by mathematical modelling) and mechanism of action of the product/compound
2) Which organisms are likely to exposed and sensitive based on 1)
3) The statistical power of the study

This study epitomises the fact that, presently, ecological risk assessment in quasinatural mesocosm designs of compounds with similar rapid removal from the water column pose paramount methodological challenges in terms of securing realism, exposure to pelagic organisms and controlling the natural variation to be feasible. Had benthic infauna organisms been used as ecotoxicological endpoint would the risk of low statistical power probably still have been unacceptable (Møhlenberg *et al.*, 2001).

Acknowledgement: Keith Solomon, University of Guelph, Paul van den Brink, University of Wageningen, Jette Rank University of Roskilde.

References

Ammann LP, Waller WT, Kennedy JH, Dickson KL, Mayer FL. 1997. Power, sample size and taxonomic sufficiency for measures of impact in aquatic systems. *Environ Toxicol & Chem* **16**:2521-2431.

Campbell PJ, Arnold DJS, Brock TCM, Grandy NJ, Heger W, Heimbach F, Maund SJ, Streloke M. 1999. Guidance Document on Higher Tier Aquatic Risk Assessment for Pesticides (Harap). SETAC- Europa Publication, Brussels, Belgium.

Council Directive.1991. Council Directive of 15 July 1991 concerning the placing of Plant Protection Products on the market (91/414/EEC). *Off Jour EC* **L230**: 1-32.

Cox C. 1998. Glyphosate (Roundup). Herbicide factsheet. *J Pest Reform/* fall 1998 **18**:3-16.

Crane M. 1997. Research needs for predictive multispecies tests in aquatic toxicology. *Hydrobiol* **346**:149-155.

Crossland NO, Wolff CJM. 1988. Outdoor ponds their construction, management and use in ecotoxicology. In: Hutzinger O. Ed. The Handbook of Environmental Chemistry. Vol. **2**, part D., Reactions and processes, pp. 51-67. Springer-Verlag, Berlin, Germany.

Dansk Standard (DS 291) 1985. Water analysis – Phosphate – Photometric methode. Copenhagen, Denmark.

Dansk Standard (DS 2126)1986. Water quality – Chlorophyll *a* – Spectrophotometric determination in ethanol extract. Copenhagen, Denmark.

Dansk Standard DS (2214)1990. Water analysis – Sampling of natural water for analysis of trace metals and (DS 259). Determination of metals in water, sludge and sediments – general guidelines for determination by atomic absorption spectrophotometry in flame. Copenhagen, Denmark.

DTI-Report 1999. Analyse af fire vandprøver for Glyphosate og AMPA. 1999.06.21. 90705. Århus, Denmark.

Farmer D, Hill IR, Maund SJ. 1995. A comparison of the fate and effects of two pyrethoid insecticides (lambda-cyhalothrin and cypermethrin) in pond mesocosms. *Ecotoxicol* **4**:219-244.

Frank RB, Braun HE, Ripley BD, Clegg BS. 1990. Contamination of rural ponds with pesticide 1971-85, Ontario, Canada. *Bull Environ Contam Toxicol* **44**:401-409.

Gagnon M, Lacroix G. 1981. Zooplankton sample variability in a tidal estuary: An interpretative model. *Limnol & Oceanogr* **26**:401-413.

Giesy JP, Dobson S, Solomon KR. 2000. Ecological Risk Assessment for Roundup Herbicide. *Rev Environ Contam Toxicol* **167**:35-120.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. *Environ Res* **50**:195-205.

Hartman WA, Martin DB. 1984. Effects of suspended bentonite clay on the acute toxicity of Glyphosate to *Daphnia pulex* and *Lemna minor*. *Bull Environ Contam & Toxicol* **33**:355-361.

Hildebrand LD, Sullivan DS, Sullivan TP. 1980. Effects of Roundup herbicide on populations of *Daphnia magna* in a forest pond *Bull Environ Contam & Toxicol* **25**:353-357

Kraufvelin P. 1998. Model ecosystem replicability challenged by the „soft" reality of a hard bottom mesocosm. *Jour of Exp Mar Biol Ecol* **222**:247-267.

Møhlenberg F, Petersen S, Gustavsson K, Lauridsen T, Friberg N. 2001. Mesocosm experiments in the approval procedure for pesticides. A study on effects of mesocosm characteristic and validity of extrapolation methods to protect sensitive species. Pesticide Research No. 56 2001. Bekæmpelsesmiddleforskning fra Miljøstyrelsen, Danish EPA. Copenhagen, Denmark.

OECD guidelines for testing of chemicals. Draft proposal for a guidance document. Freshwater lentic field tests. Draft document (July 1996).

Rosenzweig M, Buikema A Jr. 1994. Phytoplankton colonisation and seasonal succession in new experimental ponds. *Environ Toxicol & Chem* **13**:599-605.

Sanderson H. 2001. Replication of Micro/Mesocosm studies. State-of-the-art review. *Environm Sci & Poll Res Int* **8**. pp xx. DOI:http://dx.doi.org/10.1065/espr2001.09.085.

Schindler DW. 1998. Replication versus realism: The need for ecosystem scale experiments. *Ecosys* **1**:323-334.

Servizi JA, Gordon RW, Martens DW. 1987. Acute toxicity of Garlon and Roundup herbicides to Salmon, *Daphnia* and Trout. *Bull Environ Contam & Toxicol* **39**:15-22.

Stewart-Oaten A, Bence JR, Ostenberg CW. 1992. Assessing effects of unreplicated perturbations: No simple solutions. *Ecol* **73**:1396-1404.

Styczen M, Refsgaard A, Kirkegaard J (eds.). 2001. Model based tool for evaluation of exposure and effects of pesticides in surface waters. 2nd status report – J.nr.M 7041-0120. DHI – Water & Environment, NERI, DIAS. Copenhagen, Denmark.

Peterman RM, M´Gonigle M. 1992. Statistical power analysis and the precautionary principle. *Mar Poll Bull* **24**:231-234.

WHO, 1994. Environmental Health Criteria 159, Glyphosate. World Health Organisation, Geneva, Switzerland.

III

Ecological impact and environmental fate of Perfluorooctane sulfonate on the zooplankton community in indoor microcosms

by

Sanderson H., T. Boudreau, S. Mabury, WC. Cheong, KR. Solomon. 2002

Journal of Environmental Toxicology and Chemistry, in press accepted 15/1-02

Ecological impact and environmental fate of Perfluorooctane sulfonate on the zooplankton community in indoor microcosms

Hans Sanderson*, Timothy M. Boudreau‡, Scott A. Mabury§, Woo-Jay Cheong§ & Keith R. Solomon‡

*) Dept. of Environment, Technology & Social Studies, University of Roskilde, PO-Box 260, DK 4000 Roskilde, Denmark.Tel: +45 46 74 24 96, Fax: +45 46 74 30 41. E-mail: Hanss@ruc.dk. ‡) Centre for Toxicology, University of Guelph, Bovey Blgd. Gordon street, Guelph, Ontario, N1G 2W1. Canada. §) Dept. of Chemistry, University of Toronto, 80 St. George Street, Toronto, Ontario, M5S 3H6. Canada.

Abstract

There is presently a substantial amount of information being gathered concerning the environmental risk associated with the perfluorooctane sulfonate (PFOS) compound. The United States Environmental Protection Agency is requiring more research to be completed before making definitive decisions concerning the regulatory issues covered in the significant new use rule (SNUR 18/10-2000) under the Toxic Substance Control Act (TSCA). However, there are no risk assessment requirements under semi-natural conditions in microcosms. PFOS can enter, and has been found in, the aquatic environment through different pathways including spills associated with use of firefighting foams containing PFOS, leaching from washing Scotchgardä treated clothes with the wastewater, leaching from various coatings, discharges as residual waste from fluorochemical production or by volatilisation and transportation atmospherically. The biota is the sink of PFOS not sediment or soil. The aim of this paper is to determine a 35-day community no observable effect concentration ($NOEC_{community}$) for freshwater zooplankton and the fate of PFOS during the course of study. PFOS persisted in the water phase with only slight reductions over the study, only the decrease from 33.9 mg L^{-1} at day 1 to 29.8 mg L^{-1} at day 35 was significant. A 90-100% reduction (p<0.01) of the total zooplankton population was found after one week of exposure to 30 mg PFOS L^{-1} and a similar reduction after two weeks at 10 mg L^{-1} of PFOS. The *Daphnia magna* 21-day $NOEC_{survival}$ of 12 mg L^{-1} has previously been found in a standard laboratory bioassay by 3M. The rank order of susceptibility for the test community was; Copepoda> Cladocera> Rotifera, assuming all adverse direct effects.

Keywords: Perfluorooctane sulfonate, Plankton, Microcosm, Risk assessment.

* To whom correspondence may be addressed: hanss@ruc.dk

Introduction

Perfluorooctane sulfonate is the dominant and most recalcitrant homologue of a class of chemicals known as perfluoroalkyl sulfonates. It is relatively soluble in water (maximum solubility based on critical micelle concentration is estimated at 460 mg L^{-1}), but also a powerful surfactant, repelling both water and oil. Coupled with the fact that it resists chemical, thermal, and biological degradation, there is concern about risk to the environment with the long-term use of products containing PFOS [1]. The chemical properties of the compounds, mentioned above, and the wide distribution of the chemicals in high trophic levels is strongly suggestive of the potential for bioconcentration/bioaccumulation [2] and for biomagnification with a factor of 50 (liver/food concentration) [1]. The aquatic environment is possibly exposed to PFOS through wastewater from cleaning treated products and leaching from PFOS production. Surface water concentrations ranging from 25-114 ng PFOS L^{-1} has been detected in the proximity of perfluorinated surfactant-related manufacturing facilities in Alabama USA. Because of the proximity to such facility, the PFOS concentrations may not necessarily be indicative of back-ground concentrations [3]

PFOS and related substances have been produced since the 1950s as additives used in synthetic commercial and industrial products. These products include fire fighting foams, shampoo, insecticides, and corrosion inhibitors, among others. Possible sources of human exposure are: Food-wrapping containers that utilise PFOS derivatives, inhalation and dermal exposure from contact with PFOS treated products (detergents, paints, leather impregnation, snackfood wrapping, etc) [4]. Due to the unique hydrophobic and lipophobic properties of PFOS, Scotchgard™ is one of the world's most effective heavy-duty stain repellents for fabrics. The anticipated United States production of PFOS after the significant new use rule (SNUR) takes effect under the Toxic Substances Control Act is a reduction from 6,489,900 pounds in 2000 to 0 pounds by 2003 [2]. On May 16 2000, 3M (St. Paul, MN, USA) announced it would voluntarily phase-out all perfluorinated sulfonates by the end of 2003 based on the persistence and pervasiveness of these anthropogenic compounds [5]. The concern came after reports of PFOS in blood drawn from employees at 3M as well as in the general United States population. The employees were producing the stain-repellent Scotchgard™, in which PFOS is the main active component. The 3M company's agreement to withdraw their best-selling ($3-500 Mill. U.S./year) stain-repellent came as a conscious decision to reduce the potential dispersion of persistent fluorinated compounds in the environment. The presence of organic fluorine compounds in human serum was reported in 1968 [6]. A study with employees involved in the production of Scotchgard™ at 3M in St. Paul MN (U.S.) reported up to 12.83 ppm PFOS in their blood [7], the level in the general US population is 30-44 ppb and the average in children is 54 ppb. The mean human elimination half-life ($t_{1/2}$) has been estimated to 1,428 days (or app. 4 years), the high retention time in the mammalian body is caused by hepatic and bile re-circulation more so than is suggested by the lipophobic properties of PFOS [2].

Despite the widespread use of the compound, relatively little is known about the fate and effects of PFOS, particularly under semi-field conditions. In addition to its occurrence in blood and tissue of humans and wildlife, PFOS is ubiquitous in the environment and has been found globally in water, air and soil samples, however there is a lack of knowledge concerning the typical PFOS levels found in the environment [8]. In view of its widespread occurrence, its extreme persistence and the fact that it bioaccumulates (being found in polar bears, Scandinavian birds, Californian dolphins and Ganges

seals [8]) there is a need to assess the risk of this compound [2]. The purpose of this study was to determine the toxicological effects of PFOS on freshwater zooplankton in 30 L indoor microcosms, to estimate the zooplankton $NOEC_{community}$, and to assess the fate of PFOS during the 35 day long experiment. This investigation constitutes one part of a project under the Canadian Toxic Substances Research Initiative (TSRI) in conjunction with the Canadian Network of Toxicology Centres. The project aims to assess the ecotoxicological risk of PFOS on three different scales: in standard laboratory bioassays, in 30 L indoor microcosms (present study), and in 12m^3 outdoor microcosms.

Materials and Methods

Compound

The potassium salt of PFOS ($CF_3(CF_2)_7S(=0)_2O^-K^+$) (Cas No.: 2795-39-3), received from the Centre for Toxicology, University of Guelph, ON, Canada – donated by 3M St. Paul, MN, USA, was used for the purpose of ecotoxicity testing. All treatments were calculated based on the anion of PFOS ($CF_3(CF_2)_7SO_3^-$) (Cas No.:1763-23-1) which also is the breakdown product of perfluorooctane sulfonyl fluoride and found in Scotchgard™. Three exposure concentrations were used (1 mg L^{-1}, 10 mg L^{-1} and 30 mg L^{-1}) as well as controls, each with five replicates. Aquariums were randomly assigned a treatment concentration. The concentrations were chosen based on the laboratory results from the University of Guelph [9] and laboratory NOEC results from 3M on *Daphnia magna* [10]. The dosing regime mimics that of the outdoor microcosm study conducted at the University of Guelph's Microcosm Research Facility in the summer of 2000 as part of the TSRI project. This study excluded the 0.3 mg L^{-1} of PFOS treatment due to lack of detectable statistically and ecologically significant effects in the other study at this level [9].

Microcosms

The microcosm design consisted of 20, 30 L (46 x 26 x 26 cm) transparent polyvinyl chloride (PVC) aquariums. Polyethylene (PE) and PVC equipment was used in the study to reduce problems associated with adsorption of PFOS to glass. Aquariums were assigned randomly and filled with natural uncontaminated sediment and water from natural ponds situated at the University of Roskilde area in Denmark, on November 1, 2000. Microcosms were allowed to stabilise for 4 weeks in temperature controlled laboratory, under conditions similar to that of the natural ponds from which the water was taken (10°C). The climate lab is situated in an 18 m^2 room with individual light sources 15 cm above each tank and air supply in on three shelves high tank set up. The water and air temperature was gradually increased 2°C per week to 18°C during the four weeks to simulate the conditions of the Guelph study. PFOS was carefully put into solution using microcosm water, added to, and gently hand mixed in the aquariums with a plastic rod. Five replicates for each treatment concentration (1, 10 & 30 mg PFOS L^{-1} nominal concentrations) was used (n=5) including five control aquariums. The dosing was chosen based on laboratory NOEC results (app. 10 mg L^{-1}) [10] [11]. The low concentration (1 mg L^{-1}) was included due to the longer exposure time and possible chronic effects, and the 30 mg L^{-1} was chosen for assumed certainty of adverse effects. Since zooplankton abundance and diversity were the only test endpoints, the addition of phytoplankton *(Scenedesmus acutus*) each week as a supplementary food supply (100 ml of 10^6 cells/ml concentrated algae) to each aquarium during the test period did not compromise endpoints. The end stock (day 35) chlorophyll *a* content was measured to determine if food

shortage could be held accountable for reduced zooplankton abundance in the treated aquariums. The aquariums were under constant aeration and maintained an oxygen concentration of 6 mg L^{-1} over the entire study. Photoperiod consisted of a 12 hour light/dark cycle. The light source was Osram Daylight L 18W/10 lamp lumens of 2852 and colour temperature of 5000K. The zooplankton community consisted of following representative species: *Cyclops diaptomus, Cyclops strenuus, Cyclops canthocamptus staphylinus, Daphnia magna, Keratella quadrata, Phyllopoda* sp.*, Echninorhynchus* sp., *Ostracoda* sp., and total *Rotifera* sp. In addition to zooplankton and pond snails, occasional macrophytes (*Elodea canadansis* and *Myriophyllum spicatum)* and larger invertebrates (*Ephemeroptera* sp., *Assellus aquaticus*) were present. Emergent *Ephemeroptera* sp. trapped under the glass lid were also counted.

Sampling

After the 4 week stabilisation period, the following sampling schedule was implemented: 24 h pre-treatment, 24 h post-treatment, 2 days, 4 days, 7 days, 14 days, 21 days, 28 days and 35 days post-treatment. Zooplankton funnel traps (modified design derived from [12]) were used for pelagic sampling. The traps were suspended 3 cm above the sediment floor in the middle of the aquariums and consisted of a single 10 cm diameter reversed-funnel with a 60 ml confinement jar at the top. Animals were trapped during their cyclic vertical migration. The traps were set 24 h before taking the actual sample to allow for sufficient animal collection. On sampling days, confinement jars were mixed by stirring and 30 ml of the trapped volume was fixed and preserved in Lugol's iodine solution to be used for determination of zooplankton. The remaining 30 ml were returned to the aquarium to reduce the amount of water and zooplankton stock taken out of each aquarium. This might include a possible bias in terms of accuracy and replicability in sampling which is inherent in subsampling, however the procedure and thus uncertainty was consistent throughout the experiment. The samples were identified and enumerated using a microscope (Leica Wild M3Z). During the investigation, pH, total P, and total N were monitored at day 1, day 14 and day 35. pH was determined by selectrod pHM 95, 136R0011N09. Total N was measured on Radiometer Nitrate-selectrod M 27 Ag-9. Total P was determined photometric method according to Danish Standard [10]. After opening in 0.2g $K_2S_2O_8$ at 200 kPa in a CertoClav by Kelomat (Traun, Austria) the absorption was measured on Spetronic 601 by the Milton Roy Company, λ = 880nm (Ivyland, PA, USA). Mean pH for the duration of the investigation was 8.30 (8.28-8.37). Overall mean nitrate concentrations were 5.3 mg L^{-1} (2.2-13.5) and found to be highest in the 30 mg L^{-1} concentration microcosms. Total phosphorus was low during the investigation; total P < 2.0 μ L^{-1} in all microcosms. There were no significant differences between treated and control aquariums in these descriptive background characteristics during the study. The chlorophyll *a* concentration was determined spectrophotometrically on the last sampling day, using 2 L of water from each of the 20 aquariums (n=5) to determine if there was any correlation between food availability and zooplankton abundance confounding the toxic impact of the compound. The water was filtered for 30 minutes (according to guideline [13]) in Whatmanä Glass microfibre Filters Circles (GF/C) 47 mm filter paper and thereafter extracted for 24 h in 10 ml 96% ethanol. The remains of the filters were removed by centrifugation for 10 minutes at 12000 rpm in a Centrivon T-42K centrifuge (Kontron Instruments, Milan, Italy) before analysis was completed in a Perkin-Elmer Lambda 11 UV/vis spectrometer (λ = 665 & 750 nm) (Perkin-Elmer, Nowalk, CT, USA). Dissolved oxygen and temperature were monitored weekly with electrode OXY197 produced by WTW Moberg (Weilheim, Germany). Water samples (one sample (20ml) from three replica due to high r^2 (n=3)) for the fate analysis

were taken at day 1, 8 and 35 and analysed on HPLC Ion Chromatography (IC) method, see below.

Fate analysis methodology

The analysis of the water samples taken from the aquariums was conducted at the University of Toronto, ON, Canada. Instrumentation: The Ion Chromatography (IC) utilised in the PFOS analysis was a Dionex Corporation DX-500 IC system consisting of a GP50 Gradient Pump, AS40 Automated Sampler, LC 25 Chromatography Compartment, and CD 20 Conductivity Detector. The separation was achieved by using Dionex IonPac NG1, 4 X 35 mm, guard column. PFOS was detected by suppressed conductivity using a Dionex ASRS–ULTRA 2mm operated in the chemical suppression mode. Reagents and Procedures: Isopropanol, ACS grade, was used after filtering with Nylon, Sigma, 0.22 Micron, 47 mm. The filtered isopropanol was diluted with 18 MW water to prepare 70 % isopropanol, and 0.10 M Sodium hydroxide was prepared by diluting 50 % w/w aqueous solution obtained from Fisher Scientific (Fair Lawn, NJ, USA). Five mM H_2SO_4 (the chemical regenerant) was prepared from 95–98 % H_2SO_4 obtained from Fisher Scientific (Nepean, ON, Canada). Ion Chromatography (IC) operating condition: The aqueous mesocosm PFOS samples (25 mLand 500 mL injection volumes) were chromatographed with the flow rate of 0.75 mL min^{-1}. The eluent composition was kept at 4 % eluent A and 30 % eluent B for the initial 10 min, where eluent A is 0.10 M NaOH, B is isopropanol (70%), and C is 18 MW deionized water. Then, the gradient was operated from 4 to 2 % eluent A and 30 to 50 % eluent B for 1.5 min, then held at 2 % eluent A and 50 % eluent B for 3 min. The total run time was 14.5 min, with an equilibration time of 4.5 min between the successive runs. IC Quantitation: The quantitation of PFOS was achieved by external calibration. The calibration curves constructed for the mesocosm samples spiked with 1 mg/L PFOS and 10 & 30 mg L^{-1} PFOS, ranged from 0.6 to 6.0 mg L^{-1} and 6.0 to 70 mg L^{-1}, respectively. The calibration curve was linear with r^2 greater than 0.99.

Analysis

To determine the null hypothesis for $NOEC_{community}$, a t-test one-way analysis of variation (ANOVA) regime was used on the most predominant zooplankton species, total zooplankton abundance, species diversity (# of species) and end stock of chlorophyll *a* (not an ecotoxicological endpoint in this study (n=5)) estimates. The data analysis was conducted using Excel 2000 spread-sheets and the power analysis of non-significant results at the 1 mg L^{-1} concentration level was conducted using SigmaStat2 [14]. The data were graphed in SigmaPlot5 [14]. The lowest significance level of $p<0.1$ is signified with one star (*), $p<0.05$ (**), and $p<0.01$ (***). The simplest way to improve statistical power of a test is to relax the related a criterion. This does, however, mean overcoming strict adherence to the $\alpha<0.05$ convention. Public interest could provide the context and justification for this in environmental investigations. This recommended strategy contrasts with the more usual prodevelopment view of '*assume no change until proven otherwise*' [15]. High power in analysis of non-significant results was throughout set to $\alpha=0.05$, $\beta=0.2$, power = $1-\beta= 0.8$.

Results

Environmental fate: There was a slight reduction (12%) in PFOS concentration in the water column over the 35 day exposure period. Initial measured concentration in the 30 mg L^{-1} microcosm was 33.9 (+/-0.8) mg L^{-1} but it decreased to 29.8 (+/-0.2) mg L^{-1} at day 35. However, one unexplainable outlier at day 35 from a 30 mg L^{-1} replicate analysed at 77 mg L^{-1} was excluded from the mean calculation (thus n = 2 for 30 mg L^{-1}). The spiked 10 mg L^{-1} was 12.3 (+/-1.1) mg L^{-1} at day 1 and 11.7 (+/-0.4) mg L^{-1} at day 35 (15% reduction). Finally, for 1 mg PFOS L^{-1} the measured concentration over the study dropped from 1.33 (+/- 0.01) day 1to 1.15 (+/- 0.02) day 8 and to 1.08 (+/-0.07) mg PFOS L^{-1} on day 35. The drop from day 1 to 35 was 19% (Fig.1).

Cyclops: *Cyclops diaptomus* appeared to be the most sensitive species in this zooplankton community. *Cyclops diaptomus* was not found in the samples after one week at 30 mg L^{-1} and after two weeks at the 10 mg L^{-1} concentration, whereas the abundance in the controls was more or less constant after 24 hrs. (Fig. 2). A similar scenario was observed in the population of the smaller *Cyclops canthocamptus staphylinus*. While the population in the controls and 1 mg L^{-1} increased in abundance after one week this did not happen at 10 and 30 mg L^{-1} (Fig 3).

Daphnia magna: Despite the fact that *D. magna* was almost entirely reduced in both the 10 and 30 mg L^{-1} concentration (Fig. 4) only three sample times (at day 7, 14, and 21) yielded significant effects in terms of abundance reduction. The other sample times were not significant due to low power as a result of high within-variability (s^2) among the controls.

Rotifera sp.: After abundance of larger and more sensitive zooplankton species was reduced there was an increase in abundance of smaller and more tolerant species of *Rotifera* (Fig. 5). These organisms were temporally significantly ($p <$ 0.1) less abundant in the controls than in the treated aquariums. There was an obvious shift towards these species in the treated aquariums near the end of the time period.

Total zooplankton: This is the most significant endpoint together with the species diversity of the zooplankton community. There are significant reductions (p<0.01) in the relative total zooplankton populations after one and two weeks at both 10 and 30 mg L^{-1} concentration (Fig. 6). The lowest detected effect concentration on the total zooplankton community was 10 mg L^{-1} after one week.

Species diversity: The number of caught species showed statistically significant ($p < 0.01$) effects on the community's species relative diversity both at the 10 and 30 mg L^{-1} concentration after one week (Fig. 7).

Chlorophyll *a*: This was not an effect parameter in this *investigation*, *Scenedesmus acutus* was added to the aquariums weekly during the investigation but the post-treatment end stock was determined (n=5). Figure 8 shows the mean chlorophyll *a* concentration (μ L^{-1}). There was a significant increasing ($p \leq$ 0.1) tendency in chlorophyll *a* content that correlates with increasing PFOS concentration likely due to a decrease in grazing by the zooplankton community.

Figure 1: Environmental fate of PFOS in the water column

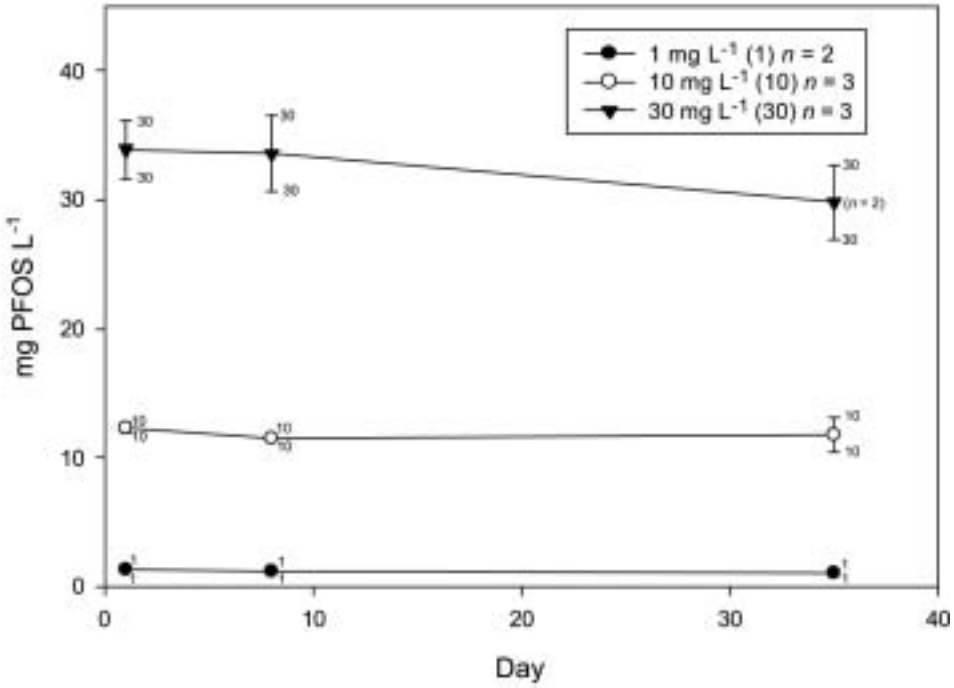

Figure 1. Environmental fate of PFOS in the water column here illustrated as mg PFOS L^{-1} over a 35 day period in 30L PVC aquariums with sediment and macrophytes. Standard deviation (SD) for 1 mg PFOS L^{-1} and the two first sample dates at 10 mg PFOS L^{-1} are too small to display (0.01-0.07 mg L^{-1}). On day 35 in the 30 mg L^{-1} one sample (77mg L^{-1}) had to be discarded altogether most likely due to contamination in the laboratory during Ion Chromatography (IC) analysis resulting in n=2 at this date. n=2-3.

Figure 2: *Cyclops diaptomus* (*n*=5)

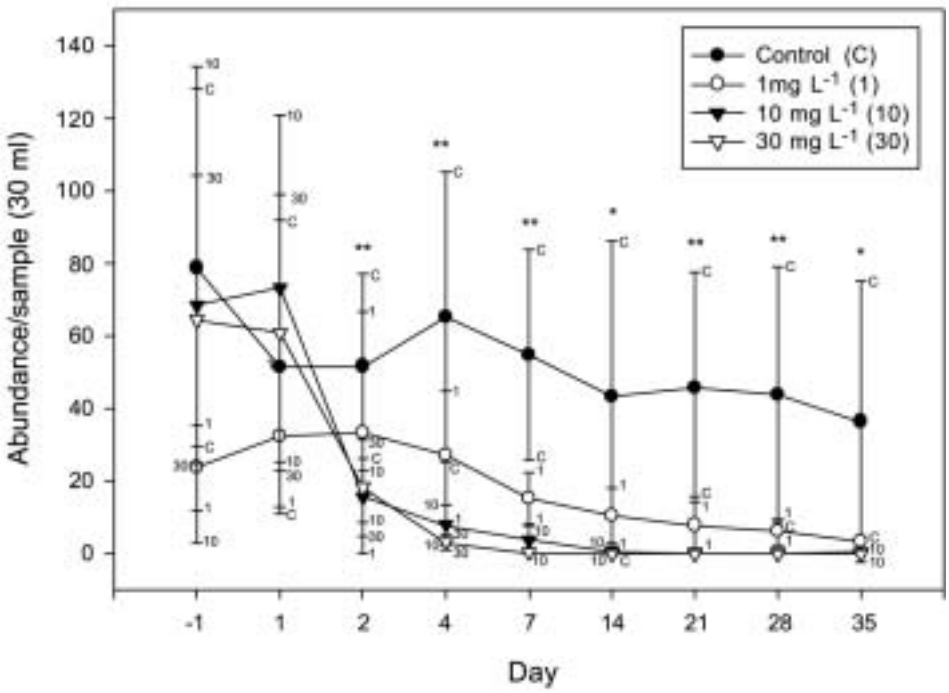

Figure 2. Abundance Cyclops diaptomus *per 30 ml sample. Level of statistical significance illustrated by* p<*0.1 (*),* p<*0.05 (**), and* p<*0.01 (***). Average standard deviation (SD) were for controls = 37; 1 mg L^{-1} = 12; 10 mg L^{-1} = 15; 30 mg L^{-1} = 10 and the total range of SD was 0-49. n=5.*

Figure 3: *Cyclops canthocamptus staphylinus* (*n*=5)

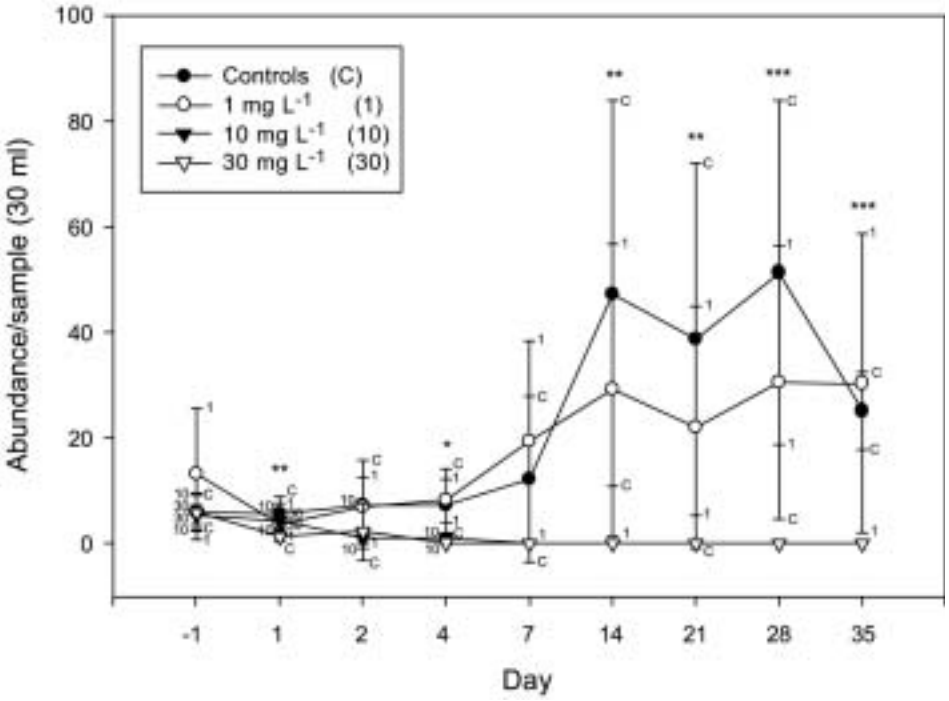

Figure 3. Abundance Cyclops canthocamptus staphylinus *per 30 ml sample. Level of statistical significance illustrated by* p<*0.1 (*), and* p<*0.01 (***). Average standard deviation (SD) were for controls = 16; 1 mg L*$^{-1}$ *= 16; 10 mg L*$^{-1}$ *= 0.8; 30 mg L*$^{-1}$ *= 1.1 and the total range of SD was 0-36. n=5.*

Figure 4: *Daphnia magna* (*n*=5)

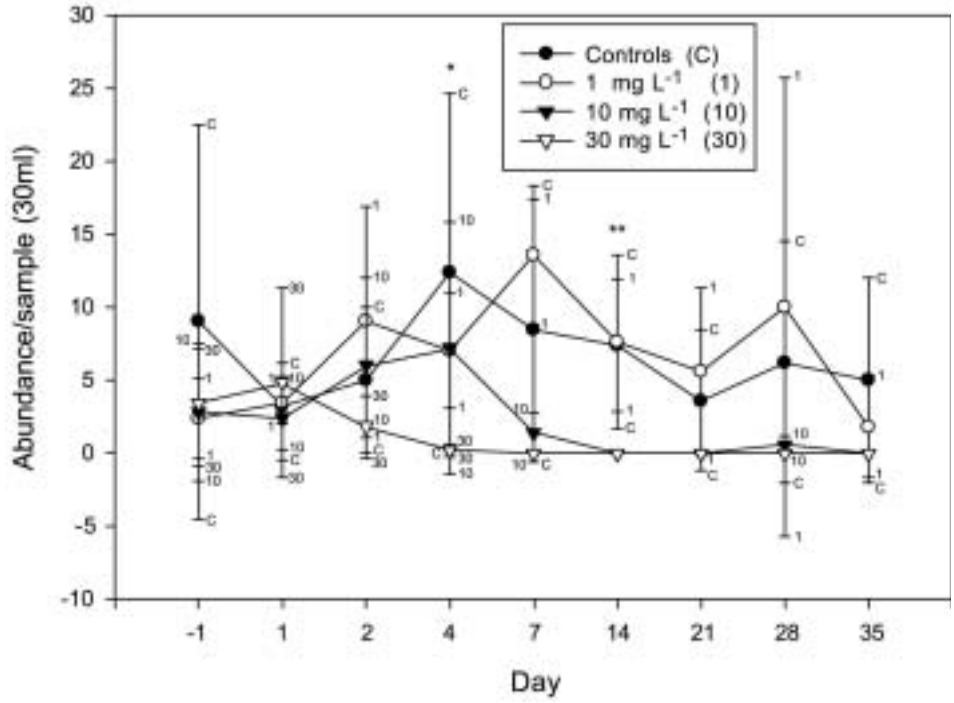

Figure 4. Abundance Daphnia magna *per 30 ml sample. Level of statistical significance illustrated by* p<*0.1 (*),* p<*0.05 (**), and* p<*0.01 (***). Average standard deviation (SD) were for controls = 7.5; 1 mg L*$^{-1}$ *= 5.7; 10 mg L*$^{-1}$ *= 2.7; 30 mg L*$^{-1}$ *= 1.4 and the total range of SD was 0-15. n=5.*

Figure 5: *Rotifera* sp. (*n*=5)

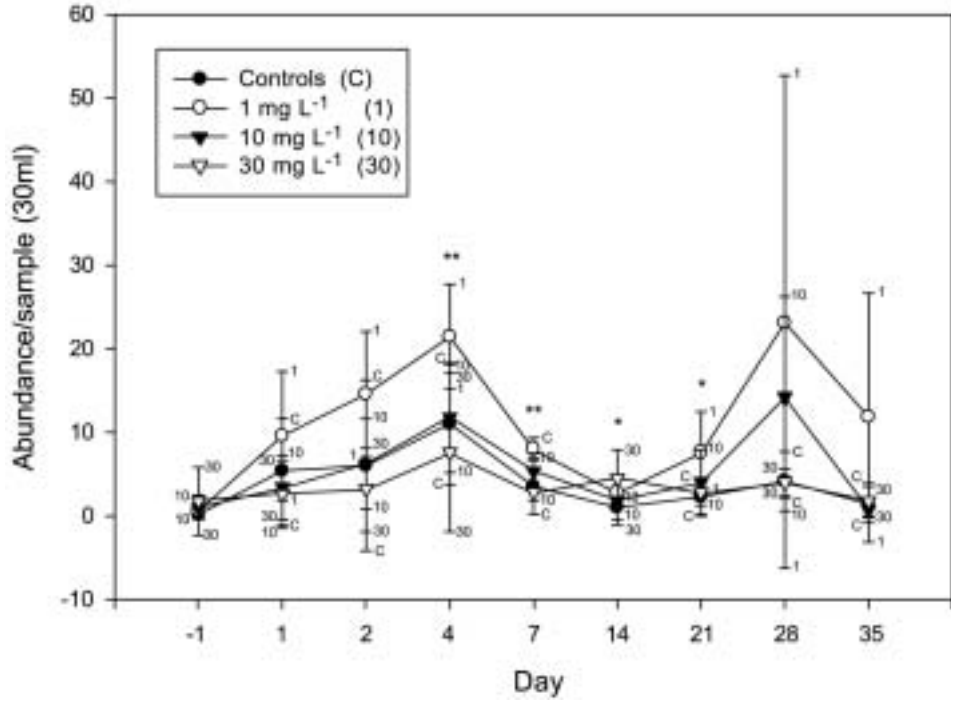

Figure 5. Abundance Rotifera *sp. per 30 ml sample. Level of statistical significance illustrated by* p<*0.1 (*),* p<*0.05 (**), and* p<*0.01 (***). Average standard deviation (SD) were for controls =4.1; 1 mg L*$^{-1}$ *= 8.3; 10 mg L*$^{-1}$ *= 4.3; 30 mg L*$^{-1}$ *= 3.5 and the total range of SD was 0.4-12. n=5*

Figure 6: Total zooplankton (*n*=5)

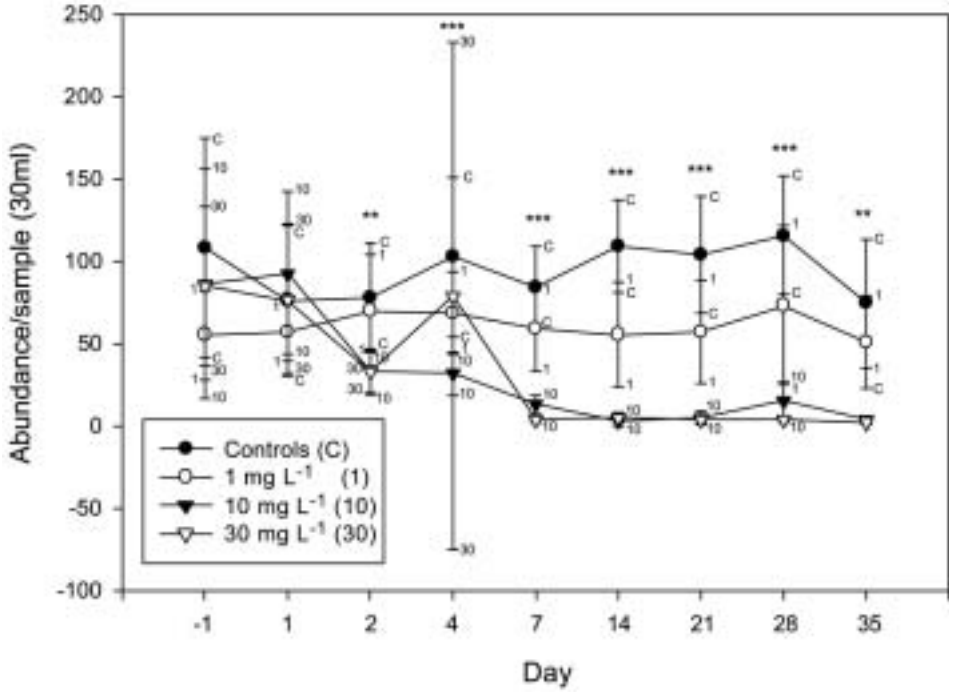

Figure 6. Abundance total zooplankton per 30-ml sample. Level of statistical significance illustrated by p<*0.1 (*),* p<*0.05 (**), and* p<*0.01 (***). Average standard deviation (SD) were for controls = 39.7; 1 mg L*$^{-1}$ *= 30; 10 mg L*$^{-1}$ *= 19; 30 mg L*$^{-1}$ *= 30 and the total range of SD was 1.5-154. High significance on both 10 and 30 mg L*$^{-1}$*. n=5.*

Figure 7: Species diversity (*n*=5)

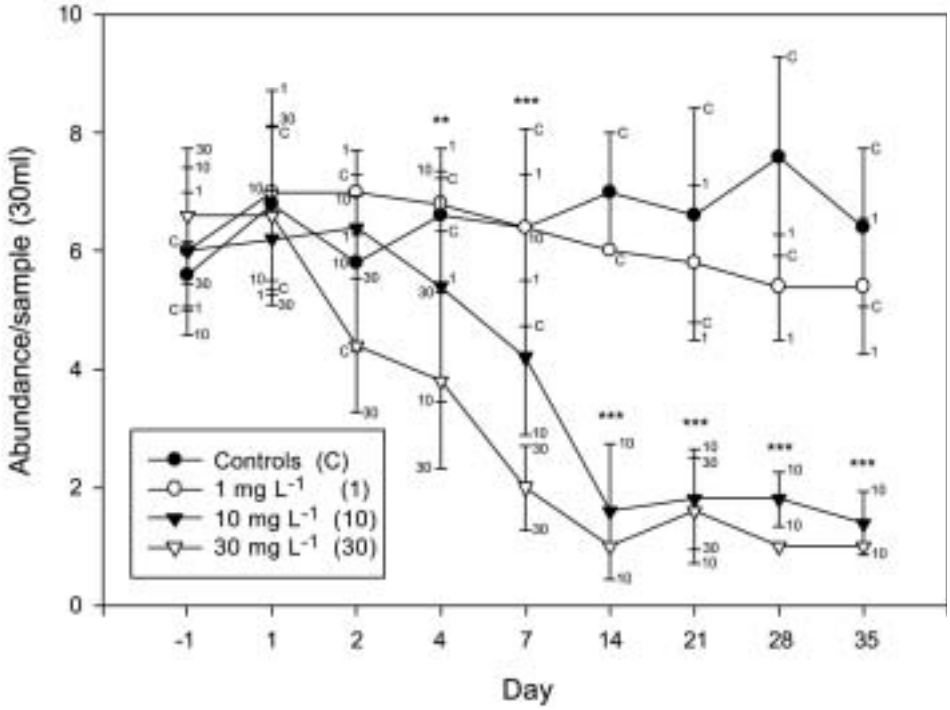

Figure 7. Abundance of species per 30 ml sample. Level of statistical significance illustrated by p<*0.1 (*),* p<*0.05 (**), and* p<*0.01 (***). Average standard deviation (SD) were for controls =1.3; 1 mg L*$^{-1}$ *= 0.9; 10 mg L*$^{-1}$ *= 1; 30 mg L*$^{-1}$ *= 0.7 and the total range of SD was 0-1.8. High significance on both 10 and 30 mg L*$^{-1}$*. n=5.*

Figure 8: End stock (day 35) Chlorophyll *a* μg L^{-1} (*n*=5)

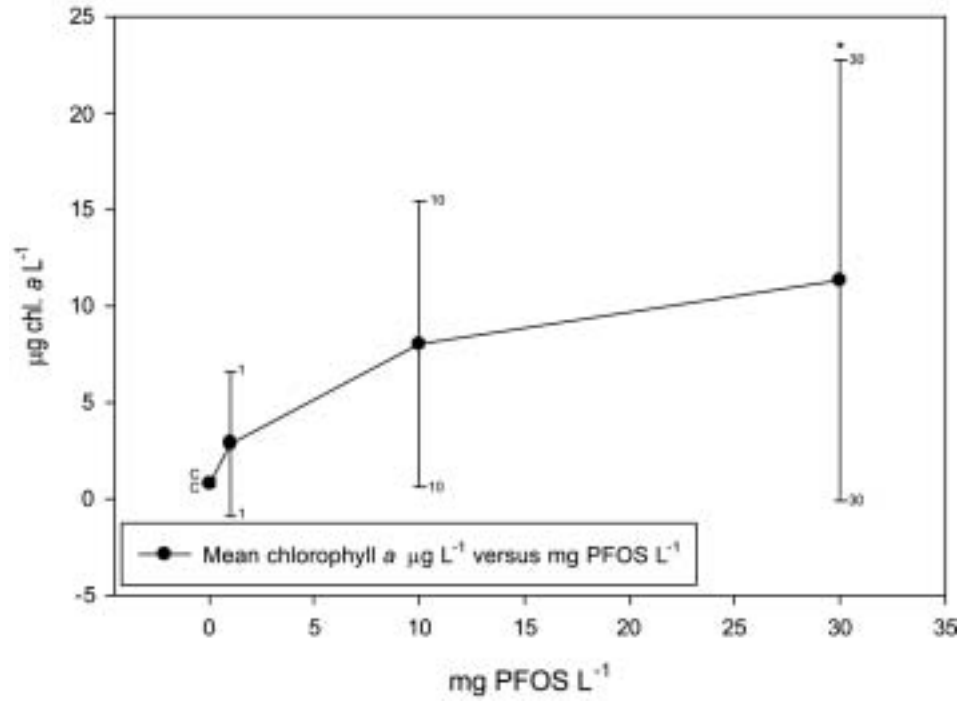

Figure 8. Mean end stock chlorophyll a *concentration (mg L*$^{-1}$*) at day 35. Level of statistical significance illustrated by* p<*0.1 (*),* p<*0.05 (**), and* p<*0.01 (***). Average standard deviation (SD) were for controls = 0.2; 1 mg L*$^{-1}$ *= 3.7; 10 mg L*$^{-1}$ *= 7.4; 30 mg L*$^{-1}$ *= 11.4 and the total range of SD was 0-15. n=5.*

General text Figure 2-7. The unit abundance per 30 ml sample is a relative time, area, depth and volume integrated catchment unit and can therefor not be transformed to a common standard metric unit of (e.g. *abundance L^{-1}) as this would significantly, and wrongly, inflate the actual abundance of animals present in the microcosms.*

There were no statistically significant changes in any endpoints at 1 mg PFOS L^{-1} however the statistical power was too low at this treatment to conclude if there might not have been any adverse effects. At the observed effect sizes (Δ) it would require 1675 replicates for *Daphnia magna*, on average, to yield high statistical power or the acceptable risk of a type II error (accepting a false negative ($\alpha \leq 0.05$, $\beta \leq 0.2$, power = $1-\beta \geq 0.8$)), 973 for *C. canthocamptus staphylinus*, 159 replicates for species diversity, 41 for total zooplankton, 28 for *Rotifera* sp. and 10 replicates for *C. diaptomus*. Had the effect sizes (Δ) been larger or the within treatment variability, in this case the standard deviation (SD), been smaller less replicates would be required to obtain high power. This was the case at 10 and 30 mg PFOS L^{-1} where the effect sizes (Δ) were higher than in 1 mg L^{-1}. Significant changes of diversity and total zooplankton on average was approximately 70% change from control for 10 and 30 mg L^{-1} while only approximately 25% change for 1 mg PFOS L^{-1} were not enough for significance. The power of 1 mg L^{-1} were consistently below 0.8 (ranging 0.1-0.6) lowest for *Daphnia magna* and highest for *Cyclops diaptomus*. The result of the *a posteriori* statistical power analysis is that 1 mg PFOS L^{-1} can not be regarded as a NOEC due to lack of power and detectability at this level.

Discussion and conclusion

Results of the current study confirm other field studies [11], which showed that PFOS is an extremely stable substance, which resists breakdown by chemical or biological processes. The USEPA has stated that it cannot currently conduct a definitive assessment of the environmental transport and partitioning of PFOS because the available data are limited and their accuracy uncertain [2]. Our results show that PFOS persists in the water phase after 35 days in the microcosms. The company 3M determined the 21-d $NOEC_{survival}$ for PFOS on *Daphnia magna* to be 12 mg L^{-1} [10]. The present investigation revealed a significant influence of PFOS on the zooplankton community at the 10 mg L^{-1} concentration after 14 days where several species were markedly reduced or eliminated. The rank order of susceptibility apart from the total relative zooplankton and relative species diversity were: Copepoda> Cladocera> Rotifera, assuming that all adverse effects were direct effects. Moreover, the long term ecological significance of the temporally significant fluctuations at 10 and 20 mg L^{-1} in the structural endpoints should be considered carefully in the context of PFOS's long persistence and its destination in biota as its sink [1]. A zooplankton $NOEC_{community}$ was not detectable with high power in this design. To be able to conclude, with high statistical power ($\beta < 0.2$), that there was no effect at the 1 mg L^{-1} level, forty-one replicate aquariums were estimated to be needed. *Ephemeroptera* sp. (not being a directly analysed toxicological endpoint) were not present in 10 or 30 mg L^{-1} concentrations at any time point. At the 1 mg L^{-1} level there was a total of three emergent *Ephemeroptera* sp. after three weeks - otherwise none. In the controls we observed more animals, the mean was 5 per control after 3 weeks, 4 per control after 4 weeks and 3 per control after 5 weeks. Results of assessment of the algae stock in the microcosms, determined at the end of the study, revealed a significant ($p < 0.1$) increase in chlorophyll *a* concentration with increased PFOS concentration. This is indicative of an indirect effect likely resulting from decreasing zooplankton abundance. This is supported by standard laboratory bioassays on the green algae, *Selenastrum capricornutum* that revealed a 96h EC_{50} of 71 mg L^{-1} and a 96h

NOEC = 44 mg L^{-1} [10], well above the exposure concentrations in our microcosms. A shortage in food supply, therefore, was not responsible for the observed effects. The study, moreover, revealed that the representative rotifer species (*Keratella quadrata*) was relatively insensitive to PFOS and actually increased in abundance. This could be the result of an improved competitive position relative to the other more sensitive zooplankton taxa. The limited observations suggest an adverse effect on *Ephemeroptera* sp. at concentrations near or below 1 mg L^{-1} of PFOS. This warrants further investigation of the susceptibility of emergent insects to PFOS and other fluorinated surfactants at lower concentrations.

The microcosm design was shown to be robust and yielded statistically significant effects at 10 and 30 mg L^{-1} treatments. Moreover, the planktonic ecosystem in the microcosms functioned according to the top-down theory [16] with increased artificial „predation" or poisoning of the zooplankton due to PFOS exposure, reducing grazing on phytoplankton hence, increasing chlorophyll *a* concentration. However, the ecological relevance of the indoor design is not as high as that of the larger outdoor microcosm design at University of Guelph. Thus the full interpretation of the results of this investigation should be made in conjunction with the results of the Guelph investigation. The Guelph investigation may have lower overall statistical significance due to lower number of replicates but higher ecological relevance due to increased size and complexity in the individual enclosures.

In this study the precision of NOEC or LOEC is questionable, moreover, NOEC has generally been shown to be equal to HC5's (hazard concentration for 5% of species) [17]. In this, study the $NOEC_{community}$ could not be generated with confidence due to low power at the 1 mg L^{-1} level. This was, in part due to the chosen ANOVA approach. A regression analysis could possibly have been a more powerful and flexible procedure. The within-treatment variability will primarily affect the confidence intervals, but the parameter estimates will be affected to a lesser extent [18]. Therefore, we feel that an *a priori* power analysis ANOVA approach is most suitable for the testing of explicit null hypotheses. When there is scientific and policy consensus regarding the acceptable effect sizes (D), risk of a Type I error (a), and Type II error (b) and these entities are explicitly put forward the correct experimental design may be set up. Combined with a hybrid of a powerful ANOVA analysis with several treatments allowing regression analysis would probably yield the most robust risk assessment data from a microcosm study [19].

A supplement to the UNEP (United Nations Environmental Programme) POP's (Persistent Organic Pollutants) treaty [20] is currently being prepared by the WWF (World Wildlife Fund). Perfluorinated sulfonates are on WWF's list of unwanted POPs based on their persistence, biomagnification, widespread presence in the biota and human blood, and reports on their toxicity. Outside the US and Canada there is limited environmental regulatory attention paid to Scotchgard™ or PFOS (G. Lyons, personal communication, WWF toxic and policy advisor, May 2001). The EU Commission awaits the outcome of an OECD report, due 2002, before working for voluntary phase out of PFOS according to the USEPA and 3M agreement (U. Sandbaek, personal communication, member EU Parliament [MEP], July 2001). The USEPA's SNUR and 3M's phase-out of PFOS may give a sense of false security as other manufacturers continue to produce fluorinated fatty acids, some of which may be as hazardous as PFOS [1]. PFOS and its precursors are not yet considered hazardous wastes. As a result, PFOS may continue to leak from less stringently controlled waste sites and allowed to accumulate in the environment.

This study demonstrates that PFOS significantly reduced or eliminated the zooplankton community at 10 mg L^{-1}, 2-3 weeks post-treatment. No zooplankton $NOEC_{community}$ was determined due to lack of statistical power at the 1 mg L^{-1} level. In comparison, the 21-day $NOEC_{survival}$ derived from a standard laboratory bioassay was found to be 12 mg L^{-1}. Because the bioassay was performed according to OECD 202 & OPPTS 850.1010 and fulfils all criteria for good laboratory practice (GLP) [10], the use of glass materials must have been ruled out in the experiment as PFOS absorb strongly to glass [21] which thus would reduce exposure of the compound to the organisms. Hence, this can not be attributable to the difference between the laboratory standard bioassay and the present study. This indicates that, in this case, the microcosm approach was more sensitive than the standard bioassay. This is surprising, as the standard laboratory bioassay supposedly would be expected to be the most sensitive due to reduced natural interference of other biotic and abiotic variables. Further investigation of the ramifications on freshwater ecosystems of environment-level concentrations of PFOS, and fluorinated surfactants in general, is needed. Additional chronic toxicity testing should also be addressed since these compounds are so recalcitrant.

Acknowledgements: The authors gratefully acknowledge the people from ANALEST at University of Toronto, Canada, for assisting in performing the IC analysis. And the editor and the two anonymous referees who's comments improved the paper.

References

1. **Purdy R. 2000.** The global pollutant perflourooctane sufonate: Identifying risks by benchmarking against representative POPs, and through food-chain modelling. Poster at SETAC annual meeting in Nashville, TN, (US), November 12-16, 2000. p1-2.

2. **USEPA, 2000.** 40 CFR Part 721 OPPTS-50639; FRL-6745-5 RIN 2070-AD43 Perfluorooctyl Sulfonates; Proposed Significant New Use Rule. In Federal Register Vol. 65, No. 202, pp.62319-62333, October 18th. 2000. USA.

3. **Moody CA, Martin CW, Kwan WC, Muir DCG, Mabury SA. 2002.** Monitoring perfluorinated surfactants in biota and surface water samples following an accidental release of fire-fighting foam into Etobicoke creek. *Environ Sci & Technol* (*In press*, accepted Oct. 22, 2001).

4. **Danish Environmental Protection Agency (MiljøStyrelsen), 2001.** Miljøprojekt #605. Kortlægning af perfluoroktanylsulfonat og lignende stoffer i forbrugerprodukter. 2001MST. Copenhagen, Denmark.

5. **Brown D, Mayer CE. 2000.** 3M to pare Scotchgard products. The Washington Post Washington D.C., May 17, 2000.

6. **Taves DR.1968.** Evidence that there are two forms of fluoride in human serum. *Nature* **217**:1051.

7. **Olsen GW, Burris JM, Mandel JH, Zobel LR. 1999**. Serum Perfluorooctane sulfonate and hepatic and lipid clinical chemistry tests in fluorochemical production employees. *J Occup Environ Med* **41**:799-806.

8. **Giesy J, Kannan K. 2001.** Global distribution of Perfluorooctane sulfonate in wildlife. *Environ Sci & Technol* **35**:1339-1342.

9. **Boudreau TM, Janutka R, Solomon, KR, Sibley PK, Muir DCG, Mabury SA. 2000.** Acute toxicity of perfluorinated surfactants on freshwater primary and secondary producers in laboratory studies. Platform presentation at 20th Annual Meeting of SETAC, Nashville (US); abstract # 348, November 12-16, 2000. p1-2.

10. **UKEPA, 2000.** Review of Environmental effects data of Perfluorooctanyl sulfonate (PFOS) - based substances. Oct. 2000. Chemicals Assessment Section, Wallingford, Oxfordshire UK.

11. **Boudreau TM, Janutka R, Solomon, KR, Sibley PK, Muir DCG, Mabury SA. 2001.** Fluorinated surfactants; aquatic toxicity studies on primary and secondary trophic levels. Poster presentation at 11th Annual Meeting of SETAC Europe, Madrid, Spain; abstract # M7MF144, 6-10 May 2001. p1-2.

12. **Sibley PK. 2000.** Procedures for invertebrate collection in microcosms. Standard Operating Procedure SOP# CTM-010. Centre for Toxicology, University of Guelph, Guelph, ON, Canada. November 30, 2000.

13. **Dansk Standard 2201.** Water quality – Chlorophyll *a* – Spectrophotometric determination in ethanol extract. 1986. Copenhagen, Denmark.

14. **Jandel corporation, 1995.** SigmaStat statistical software version 2.0 and SigmaPlot5. San Raphael, CA. USA.

15. **Fairweather PG. 1991.** Statistical power and design requirements for environmental monitoring. *Aust J Mar Freshw Res* **42**:555-567.

16. **Persson L, Diehl S, Johansson L, Andersson G, Hamrin SF. 1992.** Trophic interactions in temperate lake ecosystems: A test of food chain theory. *Am Nat* **140**:59-84.

17. **Laskowski R. 1995.** Some good reasons to ban the use of NOEC, LOEC and other related concepts in ecotoxicology. *Oikos* **73**: 140-143.

18. **Stephan CE, Rodgers JW. 1985.** Advantages of using regression analysis to calculate results of chronic tests. In: Bahner RC, Hansen DJ (Eds), Aquatic toxicology and hazard assessment: Eighth symposium, ASTM STP 891, pp. 328-338.

19. **Hanson ML, Sanderson H, Solomon KR. 2001.** Power analysis of *Myriophyllum* spp. microcosm toxicity data. *Environ Toxicol Chem*. (*Submitted* Oct. 17, 2001).

20. **UNEP, 2001**. United Nations Environmental Programme Persistent Organic Pollutants (POP's) convention, June 5, 2001. Stockholm, Sweden.

21. **Belisle J, Hagen DF, 1980.** A method for the determination of perfluorooctanoic acid in blood and other biological samples. *Anal Biochem* **101**:369-376.

IV

Ecological impact and environmental fate of Perfluorooctanoic acid on the zooplankton community in indoor microcosms

by

Sanderson H., T. Boudreau, S. Mabury, KR. Solomon. 2002

Journal of Aquatic Toxicology, in review.

Ecological impact and environmental fate of perfluorooctanoic acid on the zooplankton community in indoor microcosms

Hans Sanderson*, Timothy M. Boudreau‡, Scott A. Mabury§ & Keith R. Solomon‡. *) Dept. of Environment, Technology & Social Studies, University of Roskilde, PO-Box 260, DK 4000 Roskilde, Denmark. Tel: +45 46 74 24 96. Fax: +45 46 74 30 41. E-mail: Hanss@ruc.dk. ‡) Centre for Toxicology, University of Guelph, Bovey Blgd. Gordon street, Guelph, Ontario, N1G 2W1. Canada. §) Dept. of Chemistry, University of Toronto, 80 St. George Street, Toronto, Ontario, M5S 3H6. Canada.

Abstract

There is, presently, a substantial amount of information being gathered concerning the environmental risk associated with the perfluorooctanoic acid (PFOA) compound. PFOA may enter the aquatic environment from spills associated with use of fire fighting foams containing PFOA, leaks from various textile and coating processes, washing of treated products, residual waste discharged from fluorochemical production. The aim of this paper was to determine a 35-day community no observable effect concentration ($NOEC_{community}$) or lowest observable effect concentration (LOEC) for freshwater zooplankton exposed to PFOA during a study in 30 L indoor aquatic microcosms. Some significant ($p<0.01$) temporal fluctuations in zooplankton abundance were observed, however, a $NOEC_{community}$ could not be calculated. LOEC for various species varied between 10 and 70 mg L^{-1}. According to LOEC values, the tentative order of organism sensitivity was as follows: *Daphnia magna* > diversity ≥ *Cyclops canthocamptus staphylinus* > *Cyclops diaptomus* > total zooplankton ≥ *Rotifera* sp. The long term ecological significance of these temporal fluctuations could not be determined in this study, however, the overall study cessation analysis showed that the structure of the ecosystem was changed from a more diverse community dominated by larger species towards a less diverse community dominated by smaller more and robust species ($p<0.05$). Additional chronic toxicity testing should also be addressed since these compounds are so persistent and recalcitrant. Moreover, further occupational health and epidemiological risk studies are needed.

Keywords: perfluorooctanoic acid, plankton, microcosm, risk assessment.

* To whom correspondence may be addressed: hanss@ruc.dk

Introduction

Perfluorinated surfactants are used in a large number of industrial applications and consumer products because of their unique surface active properties associated with organic fluorocarbon chemistry. Despite the widespread use of these compounds, relatively little is known about their fate and effects, particularly, perfluorooctanoic acid (PFOA) a known rodent peroxisome proliferator (Gilliland & Mandel, 1993) under semi-field conditions. PFOA may enter the aquatic environment from several sources, including spills associated with use of fire fighting foams containing PFOA, leaks from various textile and coating processes, washing of treated products, residual waste discharged from fluorochemical production, and volatilisation. Evidence of background levels of PFOA in aquatic environments was found during an investigation of an accidental spill involving 22 000 L of fire retardant foam containing perfluorinated surfactants into Etobicoke Creek (Toronto, ON, Canada). PFOA was found upstream of the spill at concentrations ranging between 0.011 - 0.028 mg L^{-1}. In contrast, the downstream surface water concentration of total perfluorinated surfactants ranged from 0–17 000 mg L^{-1} (Moody et al., 2001). The aims of this study were to address the toxicological effects of PFOA on freshwater zooplankton in 30 L indoor microcosms, to estimate the zooplankton $NOEC_{community}$, and to assess the fate of PFOA during a 35-day experiment. This investigation constitutes one part of a project under the Canadian Toxic Substances Research Initiative (TSRI) in conjunction with the Canadian Network of Toxicology Centres. The overall project aims to assess the ecotoxicological risk of PFOA on three different scales: in standard laboratory bioassays, in 30 L indoor microcosms (present study), and in 12m^3 outdoor microcosms.

The production of PFOA via an electrochemical fluorination process at 3M Co. (St. Paul, MN, USA) began in 1947. PFOA is used as an additive in synthetic commercial and industrial products including fire fighting foams, corrosion inhibitors, in the polymerisation of Teflon®, plasticizers, lubricants and wetting agents (Gilliland & Mandel, 1993). The presence of organic fluorine compounds in human serum was first reported in 1968 (Taves, 1968). An occupational exposure study by Olsen et al. (1998) on 3M production employees reported up to 71 mg L^{-1} PFOA in human sera or 10- to 50-fold greater than other non-exposed workers. The greatest likelihood for occupational exposure to PFOA is during the drying process, whereby PFOA is converted from a salt slurry to a salt cake (Olsen, et al, 1998). PFOA has also been found in small concentrations (10-100 μg L^{-1}) in sera of the general population (Gilliland & Mandel, 1993). PFOA resists chemical, thermal, and biological degradation (Welter, 1979), however, the perfluorooctanoate anion has the potential to be generated by either dissociation or metabolism, and can resist degradation and persist in the environment as well. In light of this, its potential for widespread distribution in the environment as the parent compound raises concerns about the risk to the environment as a result of the long term global use of products containing PFOA (Ellis et al., 2001). On May 16 2000, 3M announced it would voluntarily phase-out all perfluorinated sulfonates (PFOS) and also perfluorooctanoic acid (PFOA) by the end of 2003. This decision was based on the persistence and pervasiveness of these anthropogenic compounds, despite vast economical and technological interests (Brown & Mayer, 2000). The decision came after detectable amounts of PFOS and PFOA were found in blood drawn from employees at 3M (Olsen et al., 1998).

The majority of toxicity data comes from laboratory tests in mammals with a focus on rodents (USEPA SNUR, 2000). The 30-d LD_{50} for PFOA in male Fisher rats was 189 (208-175) mg kg^{-1} lethality occurred within the first 5 days.

Sublethal dosage of 100 mg PFOA kg^{-1} resulted in transient decrease in food intake and body weight which were reversed at day 7, liver weights were slightly higher than controls and had a change in oleic and acids composition (Olsen & Andersen, 1983).The unique properties of PFOA are strongly supportive of a potential for persistence and accumulation in surface waters since these appear to be the environmental sink for PFOA. However, compared to PFOS few samples contained PFOA at concentrations greater than the limits of quantification (LOQ) of 2.5-180 ng g^{-1} ww have been reported (AR226-0202). PFOA does not adsorb permanently to either soil or sediment, remaining in the water compartment (Welter, 1979). The water flea (*Daphnia magna*) 48hr EC_{50} is 632 mg L^{-1} and green algae (*Selenastrum capricornutum*) 14 day EC_{50} is 73 mg L^{-1} (3M data sheet, 2000).

Materials and Methods

Compound

Perfluorooctanoic acid ($C_7F_{15}CO_2H$) (CAS No.: 335-95-5). The compound,- The compound, received from the Centre for Toxicology, University of Guelph, ON, Canada (donated by 3M St. Paul, MN, USA) was was as a 19.4% ww solution and used for the purpose of ecotoxicity testing. PFOA has near complete water solubility (>20 g L^{-1}) and an n-octanol/water coefficient = 5 (3M data sheet). The Log Kow = 6.28 was estimated via the KowWin program based on Structural Activity Relationship (SAR) program developed from (Meylan & Howard, 1995) in the EPI suite at www.usepa.gov and a bioconcentration factor (BCF) (using bcfwin, version 2.14) for PFOA was estimated to be 56.23.

Microcosms

The microcosm design consisted of 30, 30 L (46 x 26 x 26 cm) polyvinyl chloride (PVC) aquariums used as previously described in Sanderson et al. (2002). Polyethylene (PE) and PVC equipment was used in the study to reduce problems associated with adsorption of PFOA to glass. Five concentrations were used (1, 10, 20, 30 & 70 mg L^{-1}) plus controls, each with five replicates (*n* = 5). Aquariums were randomly assigned to treatment concentrations. The treatment regime was chosen based on laboratory results (LC_{50} 48h = 313 mg L^{-1} and $NOEC_{48h}$ = 109 mg L^{-1}) from the University of Guelph, ON, Canada (Boudreau et al., 2000), laboratory NOEC from *Daphnia magna* (LC_{50} 48h = 632 mg L^{-1} from 3M data sheet, 2000), and to mimic potential worst-case environmental concentrations without neglecting measured environmental concentrations of up to 22 mg L^{-1} after a spill and 0.011-0.028 mg L^{-1} upstream from the spill (Moody et al., 2001). The treatment regime mimics that of the outdoor microcosm study conducted at the University of Guelph's Microcosm Research Facility in the summer of 2000 as part of the TSRI project (# 0200).

PFOA was carefully added to the microcosm water, and gently hand mixed in the aquariums with a plastic rod. Since zooplankton species abundance and diversity (number of different species) were the only test endpoints, the addition of phytoplankton *(Scenedesmus acutus*) each week as a supplementary food supply (100 ml of 10^6 cells ml^{-1} concentrated algae) to each aquarium during the test period did not compromise endpoints. The zooplankton community composition and sampling procedures were respectively similar and reproduced from to those of Sanderson et al., (2002).

Sampling

After the a 4 week stabilization period, the following sampling schedule was implemented: 24 h pre-treatment, 24 h post-treatment, 2 days, 4 days, 7 days,

14 days, 21 days, 28 days and 35 days post-treatment. Zooplankton funnel traps modified from the design of Sibley et al. (2000) were used for pelagic sampling. The traps were suspended 3 cm above the sediment floor in the middle of the aquariums and consisted of a single 10 cm diameter reversed-funnel with a 60 ml confinement jar at the top. Animals were trapped during their cyclic vertical migration. The traps were set 24 h before taking the actual sample to allow for sufficient animal collection. On sampling days, confinement jars were homogenised by stirring and 30 ml of the trapped volume was fixed and preserved in Lugol's solution to be used for determination of zooplankton. Total phosphorus was low during the investigation; total P < 2.0 μg L^{-1} in all microcosms. Other metals estimated (K, Na and Si) were not significantly different between controls and treated (K_{means} = 0.6-3.8 mg L^{-1}), Na_{means} (7-9.8 mg L^{-1}) and Si_{means} (1.8-3.1 mg L^{-1}). The chlorophyll *a* concentration was determined as previously described (Sanderson et al., 2002).

The residue analysis methodology was in accordance with Sanderson et al. (2002). The analysis of the water samples taken from the aquariums was conducted at the University of Toronto, ON, Canada. Instrumentation: The Ion Chromatography (IC) utilised in the PFOA analysis was a Dionex Corporation DX-500 IC system. The linear calibration curves constructed from control microcosm water samples spiked with 1, 10, 20, 30 & 70 mg L^{-1} PFOA had a r^2 > 0,99. After the initial analysis, and according to the results of Sanderson et al. (2002), we acknowledge that we would not be able to determine, or extrapolate confidently up to, the half-life of PFOA in water. We therefore chose only to illustrate the fate of the three mid-concentrations: 10, 20 and 30 mg L^{-1}.

Analysis

To determine the null hypothesis for $NOEC_{community}$, a Student's *t*-test analysis of variance regime was used on the most predominant zooplankton species, total zooplankton abundance, diversity (number of species) and end stock of chlorophyll *a* estimates (chlorophyll *a* was not an ecotoxicological endpoint). The data analysis was conducted using Excel 2000 spreadsheets (Microsoft 2000) and the power analysis of non-significant results was conducted using SigmaStat2 (Jandel, 1995), and the data were graphed in SigmaPlot5 (Jandel, 1995). Prior to analysis, data were log-transformed (logx+1) to secure normality. Power analysis was conducted in a manner similar to that used in (Sanderson et al., 2002).

Results

Environmental fate: There was a slight reduction in PFOA concentration in the water column over the 35 days in the tanks treated with 20 and 30 mg L^{-1} PFOA over 35 days (± SD). Initial measured concentration in the 10 mg L^{-1} microcosm was 10.6 (± 0.4) mg L^{-1} but it increased to 11.32 (± 0.4) mg L^{-1} at day 35. The spiked 20 mg L^{-1} was 36.8 (± 2.3) mg L^{-1} at day 1 and 27.9 (± 2.5) mg L^{-1} at day 35. The spiked 30 mg L^{-1} was 38.3 (± 1.4) mg L^{-1} at day 1 and 34.8 (± 3.5) mg L^{-1} at day 35. None of the fluctuations at any concentration levels were statistically significant (n=2) (Figure 1). The fluctuations were probably due to evaporation of microcosm water and by the water following the addition of the alga *Scenedesmus acutus*.

Figure 1: PFOA concentration in mid-treatment replicates

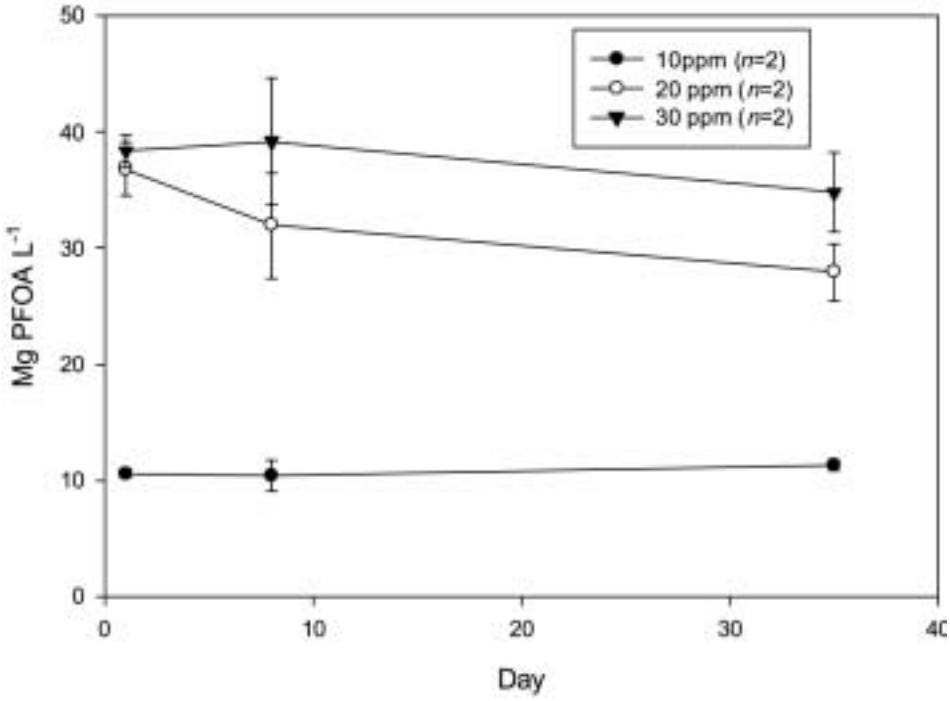

Figure 1. Environmental fate of PFOA in the water column here illustrated as mg PFOA L^{-1} over a 35 day period in 30L PVC aquariums with sediment and macrophytes. n=*2.*

Zooplankton responses: Only temporal structural significant fluctuations were observed. *Daphnia magna* (Figure 2) seemed to be the most sensitive species of the pelagic compartment in the zooplankton community with a LOEC of 20 mg PFOA L^{-1} 24 hours post-treatment. *Cyclops diaptomus* abundance (Figure 3) was significantly reduced at day 7 at 70 mg L^{-1} (p< 0.01); furthermore, 30 & 70 mg L^{-1} had the lowest study cessation readings. A similar response was observed in the population of the smaller *Cyclops canthocamptus staphylinus* (Figure 4). It is notable that the zooplankton abundance trends observed in the 30 and 70 mg L^{-1} microcosms were consistently lower than those found in all other concentrations throughout the study. In contrast, the smaller and more robust species of *Rotifera* sp increased in abundance throughout the study at the two highest treatment concentrations (Figure 5). This could be the result of an improved competitive position relative to other more sensitive zooplankton taxa. Total zooplankton abundance showed no significant trends (Figure 6). Species diversity was highest in low treatment microcosms and lowest in high treatment microcosms (Figure 7).

The overall analysis showed that, compared to controls, total zooplankton abundance (Figure 6) was significantly (p<0.01) increased at 1, 10, 20 and 30 mg PFOA L^{-1}, but not at 70 mg L^{-1} where the log abundance was 3.69 (standard deviation (sd) = ± 0.56) compared to 3.68 (sd = ± 0.16) in the controls. Moreover, the overall species diversity was significantly (p < 0.05) reduced at 10, 30 and 70 mg PFOA L^{-1} indicating a simplification of the community structure. This was seen with a shift from a more diverse community with more total zooplankton species towards less diversity where it was dominated by smaller zooplankton species (*Rotifera* sp.).

Figure 2: *Daphnia magna* (*n* =5)

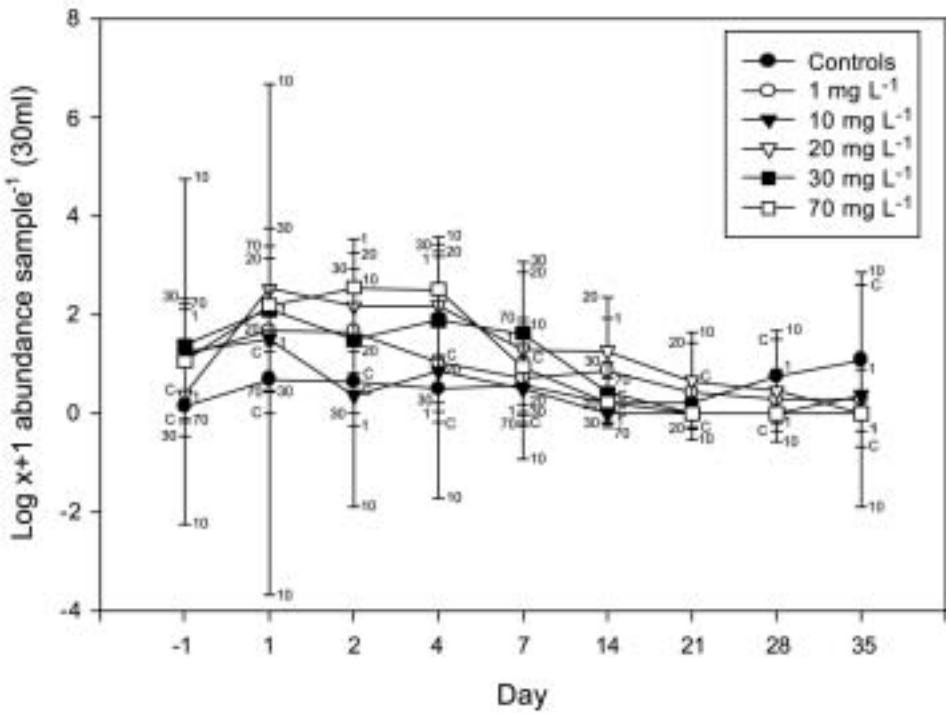

Figure 2. Abundance (logx+1) Daphnia magna *per 30 ml sample. Average standard deviation (SD) were for controls = 1.68; 1 mg L^{-1} = 1.9; 10 mg L^{-1} = 1.9; 20 mg L^{-1} = 1.52; 30 mg L^{-1} = 1.51; 70 mg L^{-1} = 0.96 and the total range of SD was 0-2.8. n=5. Graph legends according to figure 2 or 7.*

Figure 3: *Cyclops diaptomus* (*n*=5)

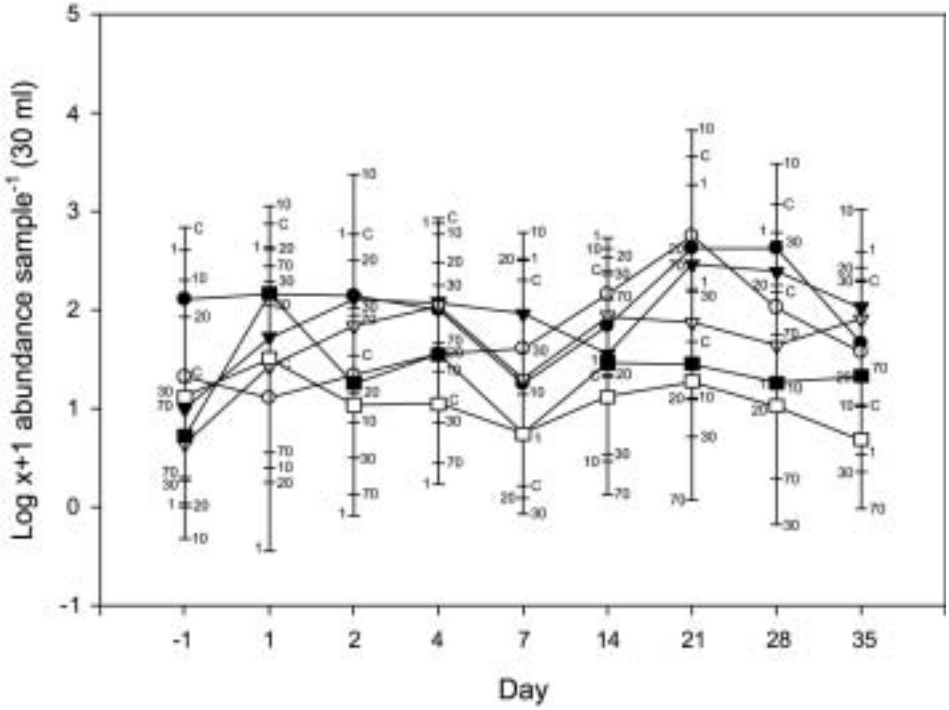

Figure 3. Abundance (logx+1) Cyclops diaptomus *per 30 ml sample. Average standard deviation (SD) were for controls = 0.81; 1 mg L^{-1} = 1.58; 10 mg L^{-1} = 1.55; 20 mg L^{-1} = 1.15; 30 mg L^{-1} = 1.42; 70 mg L^{-1} = 1.7 and the total range of SD was 0.5-2.5. n=5. Graph legends according to figure 2 or 7.*

Figure 4: *Cyclops canthocamptus staphylinus* (*n*=5)

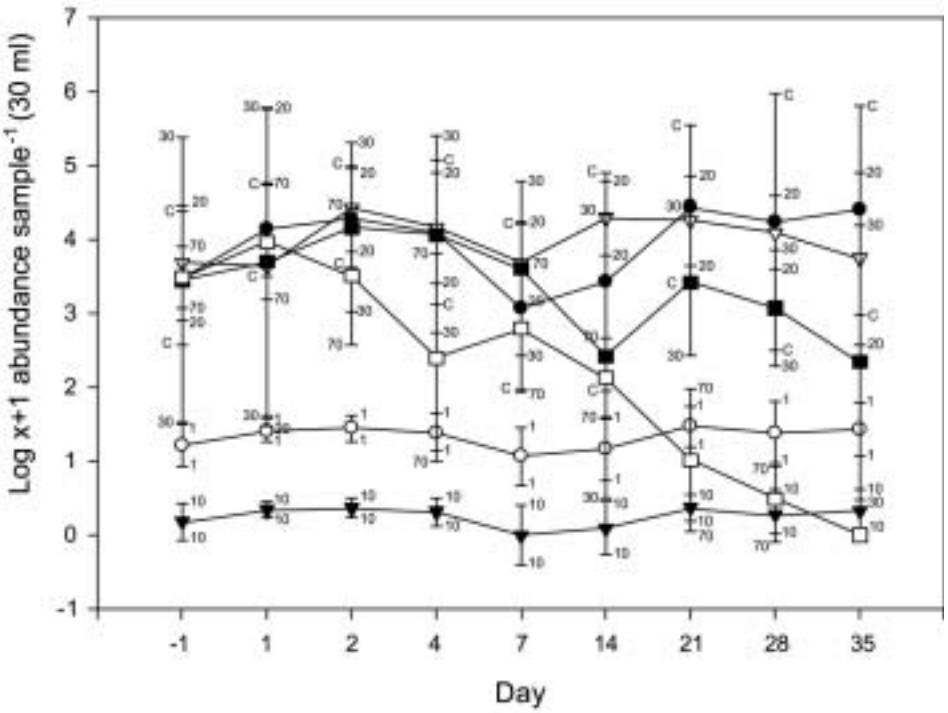

Figure 4. Abundance (logx+1) Cyclops canthocamptus staphylinus *per 30 ml sample Average standard deviation (SD) were for controls = 1.15; 1 mg L^{-1} = 0.96; 10 mg L^{-1} = 0.8; 20 mg L^{-1} = 0.96; 30 mg L^{-1} = 1.92; 70 mg L^{-1} = 1.06 and the total range of SD was 0-2.95. n=5. Graph legends according to figure 2 or 7.*

Figure 5: *Rotifera sp.* (*n*=5)

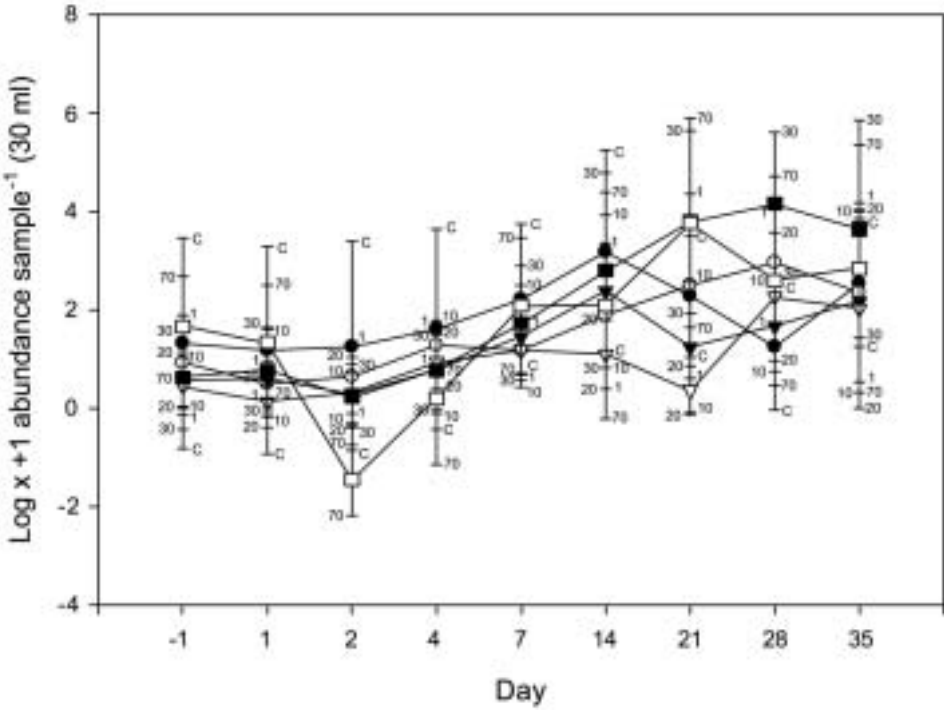

Figure 5. Abundance (logx+1) Rotifera *sp. per 30 ml sample. Average standard deviation (SD) were for controls = 2.49; 1 mg L^{-1} = 1.57; 10 mg L^{-1} = 1.89; 20 mg L^{-1} = 1.43; 30 mg L^{-1} = 2.1; 70 mg L^{-1} = 2.21 and the total range of SD was 0.68-3.39. n=5. Graph legends according to figure 2 or 7.*

Figure 6: Total zooplankton (*n*=5)

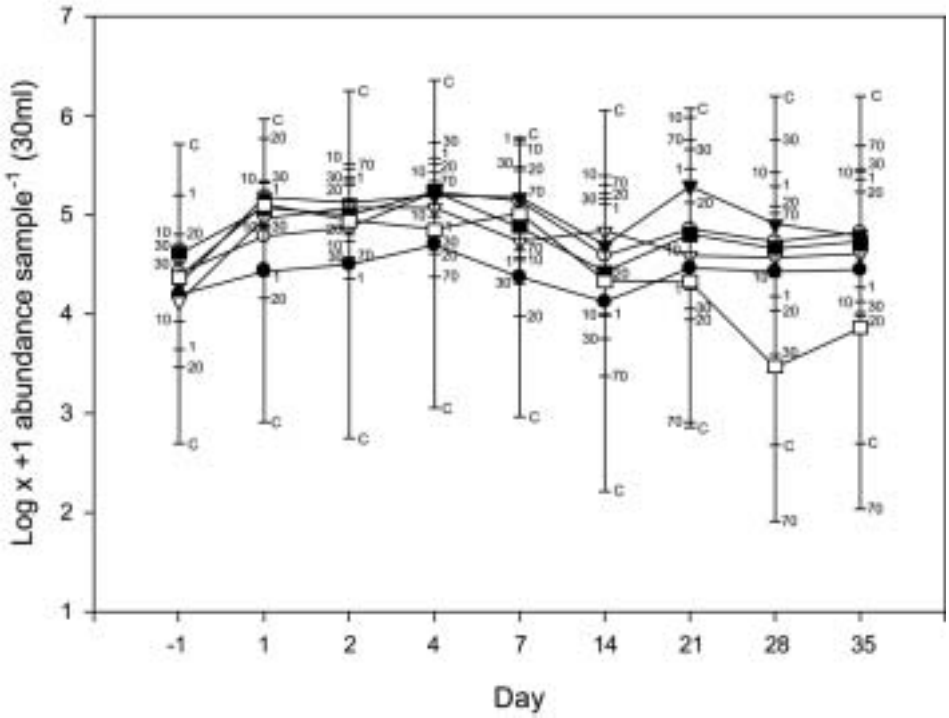

Figure 6. Abundance (logx+1) total zooplankton per 30 ml sample. Average standard deviation (SD) were for controls = 1.65; 1 mg L^{-1} = 0.54; 10 mg L^{-1} = 0.47; 20 mg L^{-1} = 0.55; 30 mg L^{-1} = 0.57; 70 mg L^{-1} = 0.83 and the total range of SD was 0.1-1.91. n=5. Graph legends according to figure 2 or 7.

Figure 7: Species diversity (*n*=5)

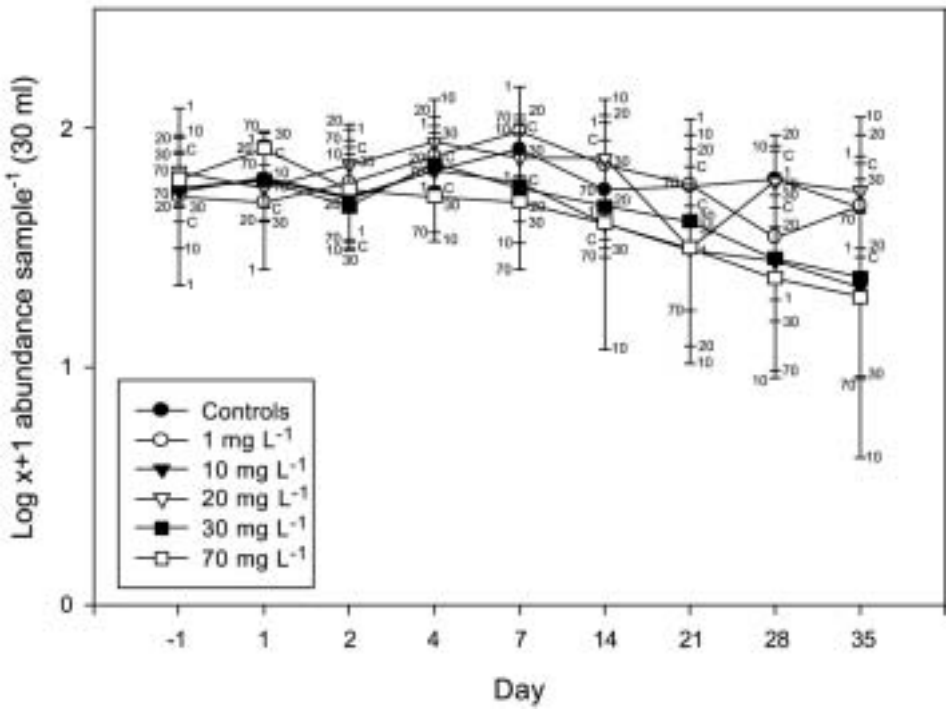

Figure 7. Abundance (log x+1) of species per 30 ml sample. Average standard deviation (SD) were for controls = 0.14; 1 mg L^{-1} = 0.23; 10 mg L^{-1} = 0.35; 20 mg L^{-1} = 0.19; 30 mg L^{-1} = 0.18; 70 mg L^{-1} = 0.22 and the total range of SD was 0-0.71. n=5.

Figrue 2-7. The unit abundance per 30 ml sample is a relative time, area, depth and volume integrated catchment unit and can therefor not be transformed to a common standard metric unit of (e.g. abundance L^{-1}) as this would significantly and wrongly inflate the actual abundance of animals present in the microcosms. Moreover, please note that the y-axis is on (logx+1) scale.

Chlorophyll *a* (not an effect parameter in this investigation) was influenced by the weekly addition of *Scenedesmus acutus*, to the aquariums during the investigation and the total end stock was determined. Mean chlorophyll *a* for controls was 8.69 mg L^{-1} and the overall mean of treated microcosms was 8.75 mg L^{-1}, there were no significant differences in the chlorophyll *a*. No increase in chlorophyll *a* concentration due to less consumption was observed at 70 mg L^{-1} compared to the controls. Highest concentration was 11.73 mg L^{-1} in 20 mg L^{-1}.

We conducted an *a posteriori* statistical power analysis of the data in order to determine the risk of committing a Type II error (false negative) and the number of replicates needed in this study in order to achieve high power ($\alpha = 0.05$, $\beta = 0.2$ and Δ = measured, where α = risk of Type I error or false positive, β = risk of Type II error, and Δ = effect size or the change between control and 70 mg L^{-1}). The changes in *Cyclops canthocamptus staphylinus* had high statistical power seen in the significant differences between controls and 70 mg L^{-1} at day 4 and day 21. Moreover, the fluctuations in *Cyclops diaptomus* (Figure 3) at day 35 had high power indicating no effects at the end of the study.

It is clear from the graphs (Figure 2-7) that within-treatment variability was significant throughout the study even though overlapping error bars (standard deviation) impeding the probability of detecting significant changes and decreasing statistical power of the study. The other parameters suffered from low statistical power, due to high inter-replicate variance and relatively small effect sizes, requiring an unrealistic amount of replicates (mean = 23 aquariums) to compensate. Hence, a total community NOEC was not determined. All significant temporal fluctuations were followed by rapid recovery and therefore, presumably are not ecological significant effects in this study. According to the LOEC's, the tentative order of sensitivity would be as follows: *Daphnia magna* > diversity ≥ *Cyclops canthocamptus staphylinus* > *Cyclops diaptomus* > total zooplankton ≥ *Rotifera* sp.

Discussion and conclusion

The detectable pattern of fate and persistence of PFOA in the microcosms water column was similar to that of similar perfluorinated chemicals (*e.g.* PFOS), where the environmental half-life is high (Boudreau et al., 2001) (Sanderson, et al. 2002). Moreover, it is notable that the compound remained suspended in the waterphase throughout the study, hence the pelagic organisms were also exposed throughout the study.

In this study, there were statistically significant temporal effects on aquatic organisms, however the ecological significance of these trends is difficult to determine. PFOA is persistent and accumulates, thus the long-term ecological significance in aquatic ecosystems is uncertain and mitigation may be difficult. According to Christman et al. (1994) >15% reduction in taxonomic endpoints between control and treated, tentatively represents an ecologically significant change. Therefore, most of the fluctuations in this study may be seen as ecologically significant. However, the ecological significance is often difficult to assess due to:

1) short experiment duration,
2) differences in species lifecycles, and
3) dosing ranges of and power of the experiment.

Furthermore, an assumed insignificant change could be significant under worst-case scenarios, critical seasons, and developmental life stages for certain species in the ecosystem.

The study, moreover, revealed that the representative rotifer species were insensitive to PFOA and actually increased in abundance. This could be the result of an improved competitive position relative to the other more sensitive zooplankton taxa. The small difference in food availability could not explain the trends or the increase seen in *Rotifera* sp. (Snell & Janssen, 1995) (Møhlenberg, et al., 2001).

The use of NOEC's for ecotoxicological investigation has both benefits and disadvantages. The apparent simplicity and interpretability for government agencies and regulators with point estimates - single numbers rather than uncertainty intervals – are useful. However, the precision of NOEC or LOEC is questionable, in fact, NOEC has been shown to equal the hazard concentration for 5% of the species (HC_5) (Laskovski, 1995) (Hoeven, 1997).

When benchmarked against PFOS, PFOA is less toxic to aquatic crustaceans. This is not surprising, since PFOS has been used as an insecticide (Sanderson et al., 2002). Moreover, PFOS is shown to be distributed worldwide with higher concentrations in wildlife tissue and sera (Giesy & Kannan 2001; AR226-0202), and PFOS is more persistent than PFOA. PFOA is a 160 times more toxic in rats when comparing compound levels in the liver with LOEC's. PFOA is a stronger hepatic peroxisome proliferator and inducer of hepatocarcinogenesis in rats, as well as reported to alter reproductive hormones in humans and rodents (Gilliland & Mandel, 1993). In light of this, PFOA constitutes a larger occupational risk than PFOS. On the other hand, PFOS seems to constitute a higher environmental risk, noting that both groups of compounds involve risks at both aspects.

The USEPA has stated that it cannot currently conduct a definitive assessment of the environmental transport and partitioning of PFOA with the limited available data and their uncertain accuracy (USEPA, 2000). Further investigations of the ramifications on freshwater ecosystems of environment-level concentrations of PFOA, and fluorinated surfactants in general, are required to elucidate potential low-level effects. Additional chronic toxicity testing should also be addressed since these compounds are recalcitrant to breakdown (Sanderson et al., 2002). Additional occupational health and epidemiological risk studies are needed to determine the risk to humans in the workplace.

Acknowledgements: The authors gratefully acknowledge Ryan Sullivan and the rest of the staff from ANALEST at University of Toronto, Canada, for assisting in performing the IC analysis. And the editor and the two anonymous referees whose comments improved the paper.

References

AR226-0202. 2000. Composite Analytical Laboratory Report on the Quantitative Analysis of Fluorochemicals in Environmental Samples. Hansen K, Johnson H, Clemen LA, Ellefson M. 3M. Minnesota. USA.

Boudreau, T.M., Janutka, R,. Solomon, K.R., Sibley, P.K., Muir, D.C.G., Mabury, S.A., 2000. Acute toxicity of perfluorinated surfactants on freshwater primary and secondary producers in laboratory studies. Platform presentation at 20th Annual Meeting of SETAC, Nashville (US); abstract # 348, November 12-16, 2000.

Boudreau, T.M., Janutka, R,. Solomon, K.R., Sibley, P.K., Muir, D.C.G., Mabury, S.A., 2001. Fluorinated surfactants; aquatic toxicity studies on primary and secondary trophic levels. Poster presentation at 11th Annual Meeting of SETAC Europe, Madrid, Spain; abstract # M7MF144, 6-10 May 2001.

Brown, D., Mayer, C.E., 2000. 3M to pare Scotchgard products. The Washington Post, Washington D.C., May 17, 2000.

Christman, V.D., Voshell, Jr. J.R., Jenkins, D.G., Rosenzweig, M.S., Layton, R.J., Buikema A.L. Jr., 1994. Ecological development and biometry of untreated pond mesocosms. In: Graney, R.L., Kennedy, J.H., Rodgers, J.H. Jr. (Eds.), aquatic mesocosm studies in ecological risk assessment. CRC Press, Inc. Lewis Publishers Boca Raton, FL, USA. pp 105-129.

Ellis, D.A., Martin, J.W., Mabury, S.A., Muir, D.C.G., 2001. Thermolysis of fluoropolymers as a potential source of halogenated organic acids in the environment. *Nature* **412**: 312-324.

Giesy, J.P., Kannan, K., 2001. Global distribution of perfluorooctane sulfonate in wildlife. *Environ. Sci. Technol.* In press accepted 18/1.

Gilliland, F.D., Mandel, J.S., 1993. Mortality among employees of a perfluorooctanoic acid production plant. *Jour. Occ. Med.* **35**: 950-954.

Hanhijärvi, H., Ylinen, M., Kojo, A., Kosma, V.M., 1987. Elimination and toxicity of perfluorooctanic acid during subcronic administration in the Wistar rat. *Pharm. Tox.* **6**: 66-68.

Van der Hoeven, N., 1997. How to measure no effect, part II: Statistical aspects of NOEC and Ecx estimates. *Environmetrics* **8**: 255-261.

Jandel corporation, 1995. SigmaStat statistical software version 2.0 and SigmaPlot5. San Raphael, CA. USA.

Laskowski, R., 1995. Some good reasons to ban the use of NOEC, LOEC and other related concepts in ecotoxicology. *Oikos* **73**: 140-143.

3M 2000. Material safety, data sheet FC-1090. Document 07-1405-5. C.A.S. NO. 335-95-5. Minnesota. USA.

Meylan, W.M., Howard, P.H., 1995. Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* **84**: 83-92.

Moody, C.A., Kwan, W.C., Martin, J.W., Muir, D.C.G., Mabury, S.A., 2001. Determination of perfluorinated surfactants in surface water samples by two independent analytical techniques: Liquid chromatography/tandem mass spectrometry and ^{19}F NMR. *Ana. Chem.* **73**: 2200-2207

Møhlenberg, F., Petersen, S., Gustavsson, K., Lauridsen, T., Friberg, N., 2001. Mesocosm experiments in the approval procedure for pesticides. A study on effects of mesocosm characteristic and validity of extrapolation methods to protect sensitive species. Pesticide Research No. 56 2001. Bekæmpelsesmiddleforskning fra Miljøstyrelsen, Danish EPA. Copenhagen, Denmark.

Olson, C.T., Andersen, M.E., 1983. The acute toxicity of perfluorooctanoic and perfluorodecanoic acids in male rats and effects on tissue fatty acids. *Toxicol. App. Pharma.* **70**: 362-372.

Olsen, G.W., Burris, J.M., Mandel, J.H., Zobel, L.R., 1998. Serum Perfluorooctane sulfonate and hepatic and lipid clinical chemistry tests in fluorochemical production employees. *J. Occup. Environ. Med.* **41**: 799-806.

Purdy, R. 2000. The global pollutant perfluorooctane sufonate: Identifying risks by benchmarking against representative POPs, and through food-chain modelling. Poster at SETAC annual meeting in Nashville (US), November 12-16, 2000.

Sanderson, H., Boudreau, T.M., Mabury, S.A., Cheong, W.J., Solomon, K.R., 2002. Ecological impact and environmental fate of perfluorooctane sulfonate on the zooplankton community in indoor microcosms. *Environ. Toxicol. Chem*. Accepted 15/1-2002.

Sibley, P., M.L., Harris, R.D., Robinson, R.W., Gensemer, K.T.J., Bestari, M.J., Chappel, J.J.H., Ciborowski, K.E., Day, T.K., George, M.L., Hanson, J.A., Jupp, C.A., Marwood, J.H., McCann, K.A., Munro, E.A., Shaw, T.S., Steele, K.R., Solomon, 2000. Procedures for invertebrate collection in microcosms. Standard Operating Procedure SOP# CTM-010. Centre for Toxicology, University of Guelph, Guelph, ON. November 30, 2000.

Snell, T.W., Janssen, C.R., 1995. Rotifers in ecotoxicology: a review. *Hydrobiologia* **313/314**, 231-247.

Taves, D.R., 1968. Evidence that there are two forms of fluoride in human serum. *Nature* **217**: 1051.

Upham, B.L., Decampo, N.D., Wurl, B., Trosko, J.E., 1998. Inhibition of gap junction intercellular communication by perfluorinated fatty acids is dependent on the chain length of the fluorinated tail. *Int J Cancer* **78**: 491-495.

USEPA, 2000. SNUR in Fed. Reg./Vol. 65, No. 202, pp.62319-62333, 18/10-2000.

Welter, A.N., 1979. Technical Report Summary. 3M. Minnesota. USA.

V

Power analysis of *Myriophyllum* sp. microcosm toxicity data

by

Hanson M., H. Sanderson, KR. Solomon. 2002

Journal of Environmental Toxicology and Chemistry, in review

Power analysis of *Myriophyllum* spp. microcosm toxicity data

Mark L. Hanson†, Hans Sanderson‡*, and Keith R. Solomon†
† Centre for Toxicology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. ‡ Department for Environment, Technology and Social studies, University of Roskilde, PO-Box 260, DK4000 Roskilde, Denmark. Phone: +45 46 74 24 96. Fax: +45 46 74 30 41. E-mail: hanss@ruc.dk.

Abstract
Myriophyllum spp. has been proposed as a new standard laboratory aquatic macrophyte test species for the registration of pesticides. Little is known about the development of this plant under field testing conditions. The main objectives of this investigation were to:

1) determine the power of *Myriophyllum sibiricum* and *M. spicatum* toxicity data derived from an outdoor microcosm bioassay
2) evaluate the variation of ten different aquatic plant effect measures calculate the minimal detectable difference for these effect measures
3) determine the replication required to detect ecologically significant changes from control for these effect measures in field studies
4) make recommendations for future studies with *Myriophyllum* spp.

Control data was used from four different ANOVA studies on *Myriophyllum* spp; each study contained three replicates and was conducted for durations of three to six weeks during the summer of 1999 with five treatment levels, including control. Node number was consistently the most powerful end-point of the ten tested for both plant species. It was possible to detect approximately a 30% change from control after three weeks with high power ($\beta = 0.2$, $\alpha = 0.05$, $n = 3$) for both plant species. The typical range of detectable change with high power was 40-60% for most endpoints. Total root length, pigmentation and wet and dry mass endpoints were generally the least statistically sensitive for both plant species. *M. sibiricum* generally had slightly lower coefficients of variation and thus required fewer replicates than *M. spicatum* to be statistical significant from control values. Initial ecologically significant effect sizes, set tentatively at impacts ≥ 25 % change from control would require 2-19 replicates on average depending on the effect measure and time of assessment. Inherent variability among the plants appears to contribute more to the total variation than did the natural variability of the microcosm. Based on statistical sensitivity, ecological relevance and relative endpoint sensitivity, we recommend using plant length and root effect measures as indicators of toxicity under field conditions.

Keywords: Power analysis, Microcosms, *Myriophyllum spp.*, Bioassay, Replication

* To whom correspondence may be addressed: hanss@ruc.dk

Introduction

Microcosm studies have been the focus of requests for increased repeatability, reproducibility and interpretability (Crane, 1997). Considerations concerning the choice of scale, descriptive endpoints, and number of replicates in order to obtain ecological relevance, statistical significance and high statistical power are vital prior to designing an ecological experiment. Knowledge of statistical power and variation can improve ecological experiments by allowing estimation of sample sizes necessary to detect certain levels of environmental change (Underwood, 1997). This paper aims to quantify aspects of these issues regarding the phytotoxicity testing of the submerged macrophytes, *Myriophyllum sibiricum* and *Myriophyllum spicatum*, in outdoor microcosms.

We analyzed data from a field based bioassay with *Myriophyllum* spp. (Hanson et al., 2001a). The standard laboratory bioassay methodology with the dicot *M. sibiricum* (ASTM, 1999) has been proposed as a required test for registration of pesticides in North America (Davy et al., 2001). Currently, the only macrophyte required for registration is the monocot *Lemna gibba*. The freshwater rooted macrophytes *Myriophyllum spicatum* and *M. sibiricum* are both common aquatic plants in Canadian waters and ecologically significant (Creed, 2000). *M. spicatum,* or Eurasian water milfoil, was introduced into Canadian waters from Eurasia and has produced major changes in aquatic habitats (Aiken *et al.,* 1979; Keast, 1984), while *M. sibiricum* is native to North America (Ceska and Ceska, 1996). These very common macrophytes contribute to primary production, improve water quality, cycle nutrients, generate oxygen, affect flow patterns, provide habitat, and food for other organisms and stabilize the sediment (Chilton, 1990; Duarte and Roff, 1991; Lewis, 1995; Cattaneo et al., 1998). These plants can be adversely affected when pesticides or other phytotoxic chemicals enter the waterway (Forsyth et al. 1997), they are sensitive to auxin-simulating herbicides; current pesticide registration species rely primarily on algae and duckweed, which are not sensitive to these types of herbicides. *Myriophyllum* spp. is thus interesting from an ecological and ecotoxicological point of view.

The field-based assay that has been developed and examined in this paper is meant to be an extension of and complimentary to the standard toxicity bioassay for *Myriophyllum* spp. (ASTM, 1999). Concerns exist about the ability of this assay to be predictive of effects at the field level (ASTM, 1999) and so the bioassay was developed to address some of these concerns. In developing standard test methods for the laboratory a number of criteria have been recommended (Rand et al., 1995). These include:

1) the test should be widely accepted by the scientific community
2) have a sound statistical basis
3) effects should occur over ranges of concentrations within realistic durations of exposure
4) the test should be predictive of effects in the field
5) the data should be useful for risk assessment
6) the test should be economical and easy to conduct
7) it should be sensitive and as realistic possible in design to detect and measure effects.

The current field bioassay meets a number of these criteria easily. A standard laboratory method is already in use and accepted for these plants, a specific desire for more information and assays on the impacts of pesticides on non-target aquatic plants for risk assessment exists (Davy et al., 2001; Lytle and Lytle, 2001) and the toxicity bioassay under consideration is field

based. The test itself, in terms of set-up and expertise, is very simple. Though microcosm studies themselves are very expensive to conduct (Shaw and Kennedy, 1996), the rest of the equipment required is readily accessible at most universities or laboratories making the assay cheap relative to the overall cost of the study. The effect measures, such as root number, plant biomass, chlorophyll content, are easy to evaluate, easily interpretable, sensitive compared to other effect measures. *Myriophyllum* spp. has been found to be very sensitive when compared to other aquatic species (Roshon et al., 1999, Marwood et al., 2001a, 2001b).

The rest of the criteria for a high quality bioassay deal with the statistical sensitivity and organism sensitivity to toxic insults. The sensitivity of a bioassay can be evaluated though the use of power analysis on data generated *a priori* by the bioassay. These types of evaluations have been conducted with some standard laboratory assays (van den Hoeven, 1998), but not for these aquatic macrophytes in the laboratory or the field.

Although natural field or simulated field studies provide a test system that evaluates ecosystem level effects, detection and interpretation of such effects can be difficult. Field systems provide an opportunity to assess the effects of a contaminant under realistic exposure conditions, ecologically important processes such as recovery and effects that are both direct and indirect, as well as species interactions and community responses (Shaw and Kennedy, 1996). The replicability, and thus the design, analysis, and interpretation of micro/mesocosms studies have been an issue for many years (Abbott, 1966). These issues are still important even after the United States Environmental Protection Agency's decision to discontinue the requirement of micro/mesocosm data due to uncertainty in implementation of the data, following the Fisher memorandum in 1992 (Fisher, 1992). Part of the problem is that identification of ecosystem level effects requires the ability to clearly separate treatment-related changes in the system from natural or background variability.

The greater the variability in a particular endpoint, the more difficult it will be to identify or measure stressor-induced alterations. Problems with high variability in microcosm studies have confounded results from many test organisms and endpoints (Sanderson and Petersen, 2001a) and have resulted in adoption of complex statistical models to evaluate some types of data (van den Brink et al., 1997). Statistically, the problem of high variability can be expressed as the likelihood of committing a Type II error, the failure to detect an effect of the stressor in the treated system, even though an effect was present (in this case erroneously accepting a false negative). If this is the case, one could argue that the power of the applied statistical test was too low (Peterman, 1990). An analysis of power and minimal detectable difference should be required if it is not possible to reject the null hypothesis at a specific significance level (*i.e.* $p \leq 0.05$) and when reporting no observed effects concentrations (NOECs). NOECs themselves are controversial with debates surrounding their utility, calculation and relevance (Chapman et al., 1996; Bailer and Oriss, 1997; van den Hoeven, 1997; Crane and Newman, 1999). The legitimate question is when the null hypothesis is not rejected, is this due to no true effect or that the design of the study was insufficient, and thus unable to detect real effects significantly due to low power?

In order to conduct an effective semi-natural field investigation, it is important to understand the limitations of the data that can be generated and how to design a study to maximize the return on time, expense and produce useful

results. This manuscript analyzes and discusses these issues. This was done by:

1) Determining the actual power of the field based toxicity assays for *Myriophyllum* spp. and examining the sensitivity of endpoints to toxicants by ANOVA analysis.
2) Estimating the replication required to determine with high power ecologically significant impacts based on the variation of individual effect measures for *Myriophyllum* spp.
3) Calculating the minimal detectable or observable effect based on the variation of individual effect measures for *Myriophyllum* spp. This will allow for a measure of the statistical sensitivity of the endpoint.
4) Examining the role of subsampling in controlling variability by comparing variation with and without subsamples.
5) Comparing the variation, minimal detectable effect and replication requirements with studies in previous years to characterize the replicability and reproducibility of the field bioassay.
6) Making recommendations about the statistical design of the bioassay, endpoint and species selection, and utility of the test in the risk assessment process.

Materials and methods

The Microcosms

The studies evaluated in this paper were conducted during the summer of 1999 at the University of Guelph Microcosm Facility located at the Guelph Turfgrass Institute, Guelph, ON, Canada. The experimental systems used in this study are described as microcosms after the definition given by Graney et al. (1995) in that they are fabricated tanks large enough to represent a lentic system, holding between 2000 and 15000 L. The facility consists of 30 microcosms that are designed to replicate natural pond systems. The microcosms are approximately 1.2 m deep with a water depth of 1 m, a diameter of 3.9 m and a surface area of 11.95 m^2. Each microcosm has a capacity of approximately 12000 L of water and are constructed according to expert recommendations (SETAC, 1991) to control variability. The microcosms are sunken into the ground with the tops flush with the surface. These systems have been used successfully to examine the fate of anthropogenic compounds in aquatic environments (Bestari et al, 1998a and Bestari et al, 1998b, Ellis et al, 2001) as well as aquatic organisms such as zooplankton, phytoplankton and fish (Sibley et al, 2000, Sibley et al, 2001). A more complete description of the state of the microcosms as they were utilized for the specific studies evaluated in this manuscript are available in Ellis et al, (2001) and Hanson et al, (2001a).

Chemicals

The MCA (99%) and DCA (99%) were obtained from Acros (Acros Organics, Geel Belgium) and the CDFA (98%) was obtained from Aldrich (Aldrich Chemicals, Milwaukee, WI, USA). The compounds were weighted out the day prior to the actual exposure and dissolved in redistilled deinonized water. Each solution was then neutralized to pH 7 to 8.5 with ACS grade sodium hydroxide (Fisher Scientific, Fair Lawn, NJ, USA).

Myriophyllum spp. Experimental Design

This study examined four independent experiments conducted during the summer of 1999 (Table 1).

Table 1. Basic experimental design of the Myriophyllum spp. *studies.*

Study	Compound	Concentrations (mg/L)[a]	Length of Study[b]	Start Date [c]	Finish Date [d]	Sampling days
1	MCA	0, 3, 10, 30, 100	28	June 10, 1999	July 8, 1999	4, 7, 14, 28
2	DCA	0, 30, 10, 30, 100	21	June 23, 1999	July 14, 1999	4, 7, 14, 21
3	MCA/DCA	0, 50 (DCA)/3,6,12 (MCA)	21	Aug. 12, 1999	Sept. 2, 1999	4,7,14, 21
4	CDFA	0, 0.5, 1, 5, 20	42	Aug. 18, 1999	Sept. 29, 1999	7, 14, 28, 42

[a] All concentrations were replicated 3 times.
[b] Duration of *Myriophyllum spp*. Exposure (days).
[c] The date the compound was introduced into the microcosm.
[d] The final date of the *Myriophyllum spp*. exposure.

The compounds examined for their toxicity to *Myriophyllum* spp. were monochloroacetic acid (MCA) (Hanson et al., 2002b), dichloroacetic acid (DCA) (Hanson et al, 2002a), chlorodifluoroacetic acid (CDFA) (Hanson et al., 2001) and a mixture of MCA and DCA (results unpublished). The basic design was an one-way ANOVA with subsamples with each exposure randomly assigned to three separate microcosms. *Myriophyllum spicatum* L. (Haloragaceae) and *Myriophyllum sibiricum* Komarov (Haloragaceae) used in the field studies were obtained from laboratory cultures maintained according to standard methods (ASTM, 1999). Plants were axenically cultured in 50 mL quartz test tubes with Andrews media fortified with 15 g L^{-1} of sucrose in an environmental growth chamber (Model E7H, Controlled Environments, Winnipeg, MB, Canada) for approximately two weeks prior to the initiation of a field study.

Each microcosm was initially stocked with 8 individual 5 cm apical shoots of each plant species to be sampled at four time points. The exception to this was the CDFA study where 4 cm apical shoots of *M. sibiricum* were used due to the small size of the plants in culture at the start of the study. At each sampling event, two plants of each species were removed and assessed for a number of endpoints. The endpoints examined were plant growth (shoot length), root number (primary roots), root length (total and longest of primary roots), wet and dry mass, node number, and chlorophyll *a*, chlorophyll *b* and carotenoid content. The two species in the same genus were tested to determine their relative sensitivities to the contaminants. In each study, an additional 10 of the laboratory-cultured plants that were cultured with the introduced plants were assessed at day -1 for the previously described endpoints. Plants of the same species had their day -1 measurements compared in a one-way ANOVA ($p \leq 0.05$) to test for differences between initial conditions of the studies themselves and the natural variability within the lab cultures (Table 2). If significance was found, Tukey's test ($\alpha = 0.05$) was used as the multiple comparison. At each sampling event, two plants of each species were removed and assessed in the laboratory. The experimental method used to conduct plant toxicity testing in these microcosms has been described previously in more detail (Hanson et al., 2001a).

Statistics

Power analysis was conducted using the statistical package SigmaStat 2.0 (1995) (Jandel San Rafael, CA, USA). To determine the power of the initial ANOVA study designs the data from the studies was analyzed in a one-way ANOVA that reports the power of the comparison with $\alpha = 0.05$. In order to calculate power the quantity Φ must be determined (Zar, 1984), which is

$$(1) \quad \Phi = \sqrt{(((k-1)(\text{group MS} - \text{error MS}))/(k * \text{error MS})}$$

where k is the number of treatments, groups MS is the group or treatment mean square for the ANOVA output and error MS is the error mean square from the ANOVA output. Once ö has been calculated and the degrees of freedom for the group or treatment and error terms are known from the ANOVA output, it is necessary to derive the power from tables such as those contained in Zar (1984).

This analysis was performed for MCA, DCA and CDFA studies, since these studies were designed to be initially analyzed using ANOVA. The mixture study was not included in this analysis as it did not have the same number of concentration levels, but was included in the statistical evaluations described henceforth. Any analysis that did not meet normality or equal variance assumptions was ln, reciprocal, or square root transformed.

The next step was to determine the variation in each of the effect measures for both plant species. The raw control data from the four studies, some of which has been reported (Hanson et al., 2001a), was used as an estimate of the variation for that endpoint as calculated by the sample standard deviation (sd) from

$$(2) \quad sd = \sqrt{s^2}$$

where s^2 is the sample variance as calculated by

$$(3) \quad s^2 = S_i(Y_i - m)^2/(n-1)$$

and n is the number of replicates, Y_i is the value of the endpoint for replicate i and m is the sample mean. Prior to the calculation of the standard deviation, the means were tested for normally distributed residuals using the Kolmogorov Smirnov test ($p \leq 0.05$). All control data met this standard.

The sample standard deviations were transformed to the coefficient of variation (CV)

$$(4) \quad CV = 100sd*Y^{-1}$$

which allows for comparison of the variation between endpoints and studies. This transformation allows for the rapid analysis of power and replicability requirements because the change in each endpoint can be expressed as a function of the control values, which would be 100 percent. The required replication and minimal detectable difference between the control and treated microcosms with high power ($\beta = 0.2$) was done using a Student t-test approach (Green, 1989). To conduct an unambiguous power analysis a statistical model, a significant level (α), a power ($1-\beta$) and an effect size (Δ) need to be defined and an estimate of the background variance (s), in this case the coefficient of variation of the control data, is also required. Estimates of the minimum sample size required to detect significant effects

were determined with the sample size function of SigmaStat2.0 utilizing relationship (5) by Green (1989).

$$(5) \quad n = 2(t_{\alpha, v} + t_{\beta, v})^2 \, (sd / \Delta)^2$$

where, n is the estimated number of samples, $t_{\alpha, v}$ is the t-value for α at v degrees of freedom, $t_{\beta, v}$ is the t-value for β at v degrees of freedom, sd is the estimated error standard deviation of the sample (2) which, in this case, is transformed to the coefficient of variation (3), and Δ is the effect size as a percent change from the control. The degrees of freedom are calculated as 2(n-1). This equation can be rearranged to solve for Δ, which is the minimum detectable change from control. The variable n is set to 3, the number of replicates in the studies. The equation is then

$$(6) \quad \Delta = (\sqrt{2} * (t_{\alpha, v} + t_{\beta, v}) * (sd))/\sqrt{n}$$

Alpha was conservatively set to 0.05 (two-tailed test) and β was set to 0.2 (one-tailed test). Since n=3 the degrees of freedom are 4. Therefore $t_{\alpha, v}$ =2.776 and $t_{\beta, v}$ =0.941. By including the minimal detectable difference with the result of the power of a test a meaningful estimate of the true sensitivity of the test is obtained (van den Hoeven, 1998). As a standard for an effect to be deemed ecologically significant, an ecological threshold value of 25% change from control (Δ) which is equal to the EC_{25} for individual effect measures. Other work has suggested that changes in plant biometrics of >20% be considered a significant ecological impact (Christman et al., 1994) and laboratory and field toxicity results on *Myriophyllum* spp. have reported EC_x values at this level (Roshon et al., 1999; Martin et al., 2000; Hanson et al., 2001d). Still, the selection of this value as the criterion for ecological significance is arbitrary and open to interpretation.

A Student t-test, as opposed to an ANOVA approach, for determining the replication required to detect specific effects was used for two reasons. While using the estimate of standard error of the residuals (SER) from a one-way ANOVA would be ideal, especially since this is the type of test used in the studies that this paper is investigating, this approach is constrained by a number of problems. Firstly, depending on whether or not the exposure regimen was broad enough to induce effects, there can be an influence on the SER. An example of this would be root number. At toxic concentrations, there can be complete reduction in root growth and development, resulting in root number and length values of zero. This would lower the SER and give a biased measure of the pooled variability relative to a study that showed no toxicity at any exposure level. This has been observed with other studies examining replication using an ANOVA approach to calculate power and minimal detectable differences (van den Hoeven, 1998). Both the MCA and the MCA/DCA mixture studies had numerous statistical significant differences detected by ANOVA analysis (Hanson et al., 2001e). By using the t-test approach to determine replication as opposed to an ANOVA where effects have been observed in studies providing an estimate of the variation without this confounding factor. Secondly, the standard protocol (ASTM, 1999) calls for the use of a Student t-test design with single concentrations to evaluate toxicity. Since this design could be adopted for field studies as part of a larger study where macrophytes are not the main focus or expense and time warrant a smaller evaluation, it is important to understand the variation and sensitivity of the test in the context in which it is applied.

The role of subsampling in reducing variability and increasing power was also analyzed. This was done by randomly taking the effect measure values of only one plant sampled from each microcosm at each time-point and determining the coefficient of variation as previously described. This CV was then compared to the CV determined as previously described study by calculating the ratio of the two CVs with one subsample CVs divided by two subsample CVs. Ratio values greater than one indicate a reduction in the variation for that effect measure with increased subsampling.

The consistency of the assay's variation between years was also examined. Studies were conducted in 1998 that used a methodology similar to that of the studies in 1999, including the same plant species, microcosms, experimental set-up and basic ANOVA design (Hanson et al., 2002a, 2002b). The main difference was that total root length was not used as an effect measure. The data from these two studies were evaluated in the same fashion as the studies from 1999 in that their CVs were calculated (3) and a minimal detectable difference was determined (5).

The variation of the standard laboratory assay was compared to the variation observed under field conditions. Laboratory data were based on published studies with *M. sibiricum* (Roshon and Stephenson, 1997, Roshon et al., 1999). The raw data was obtained by contacting the authors of these studies. CVs and minimal detectable differences were calculated as previously described. Since the standard assay is 14 d in duration, the 14 d field study CVs were compared with the laboratory studies with a two-tailed Students t-test ($p<0.05$).

Results

The -1 d samples for both species showed few significant differences between studies, (Table 2).

Table 2. The initial status of the Myriophyllum spp. *laboratory cultures on Day -1 for four field studies. The values are the mean of ten plants with their standard deviation. The results of ANOVA comparison between Day-1 plants within a species are shown. Those endpoints that share the same letter are not significantly different by Tukey's test („<0.05).*

Plant	Compound	Plant Length	Roots	Node Number	Wet Mass	Dry Mass	Chlorophyll-*a*	Chlorophyll-*b*	Carotenoids
M. spicatum	MCA	5	0 ± 0	19 ± 2a	190.9 ± 23.2a	27.9 ± 4.0a	1.017 ± 0.138a	0.389 ± 0.051a	0.352 ± 0.040a
M. spicatum	DCA	5	0 ± 0	17 ± 1b	226.5 ± 38.8a	38.8 ± 6.1b	0.759 ± 0.150b	0.286 ± 0.058b	0.268 ± 0.050b
M. spicatum	MCA/DCA	5	0 ± 0	19 ± 1ab	226.5 ± 43.2a	33.8 ± 3.4ab	1.003 ± 0.201a	0.384 ± 0.074a	0.348 ± 0.065a
M. spicatum	CDFA	5	0 ± 0	19 ± 2a	201.3 ± 27.3a	37.3 ± 2.8b	0.793 ± 0.134b	0.359 ± 0.049ab	0.302 ± 0.035ab
M. sibiricum	MCA	5	0 ± 0	21 ± 2ab	273.6 ± 27.1a	54.5 ± 7.7a	0.716 ± 0.065a	0.283 ± 0.025a	0.269 ± 0.019a
M. sibiricum	DCA	5	0 ± 0	20 ± 1a	241.7 ± 49.9a	51.3 ± 4.1a	0.720 ± 0.121a	0.269 ± 0.051a	0.257 ± 0.060a
M. sibiricum	MCA/DCA	5	0 ± 0	22 ± 2b	191.7 ± 43.0b	34.5 ± 4.1b	0.511 ± 0.055b	0.211 ± 0.030b	0.208 ± 0.022b
M. sibiricum	CDFA	4	0 ± 0	17 ± 2c	126.2 ± 29.1c	18.5 ± 4.5c	0.487 ± 0.094b	0.184 ± 0.041b	0.185 ± 0.033b

The *M. sibiricum* for the CDFA study was likely significantly different due to the fact that a 4 cm apical shoot was used as opposed to a 5 cm shoot. Other differences may be due to the length of time the plants were grown in culture, impacting mainly wet mass and dry mass measurements, or inherent genetic variability within the plants themselves.

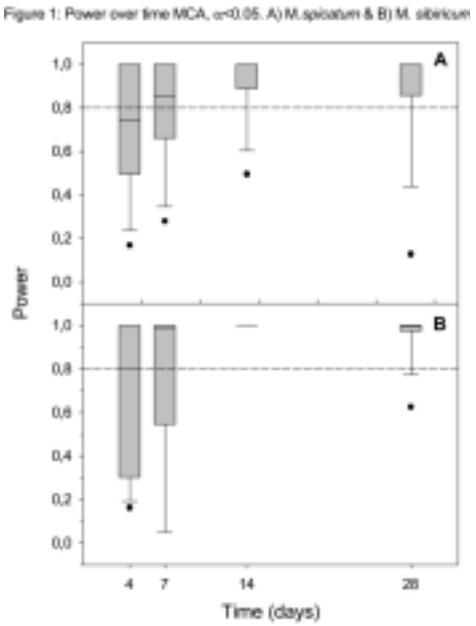

Figure 1. Box plot of the calculated power with alpha < 0.05 of a one way analysis of variance (ANOVA) for Myriophyllum spicatum *(A) and* M. sibiricum *(B) from a study examining the toxicity of monochloroacetic acid (MCA) in aquatic microcosms for ten effect measures. The median is shown as the solid line within the box. The box ends are the 25th and 75th centiles and the whiskers bars are the 10th and 90th centiles. Any points shown outside these areas are considered outliers. Power at 0.8 is considered adequate for an ANOVA.*

Power analysis of the three studies in an one-way ANOVA fashion show that only one of the studies, MCA, had high power for both species of plant (Figure 1). This study reported numerous significant differences from control with well-defined concentration response curves for the majority of endpoints (Hanson et al., 2001e).

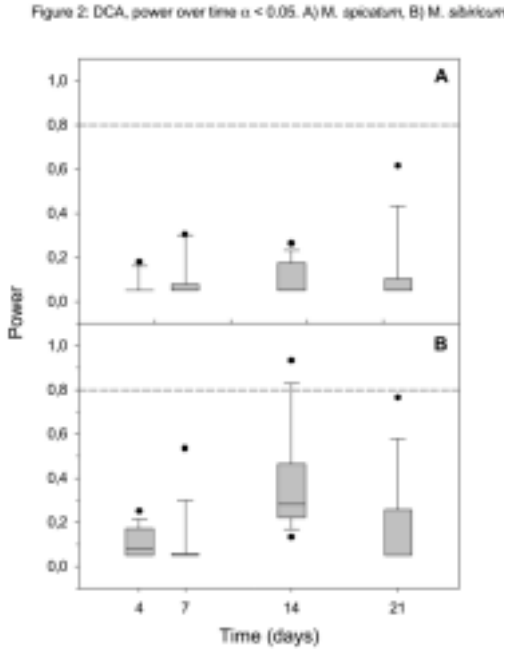

Figure 2. Box plot of the calculated power with alpha < 0.05 of a one way analysis of variance (ANOVA) for Myriophyllum spicatum *(A) and* M. sibiricum *(B) from a study examining the toxicity of dichloroacetic acid (DCA) in aquatic microcosms for ten effect measures. The median is shown as the solid line within the box. The box ends are the 25th and 75th centiles and the whiskers bars are the 10th and 90th centiles. Any points shown outside these areas are considered outliers. Power at 0.8 is considered adequate for an ANOVA.*

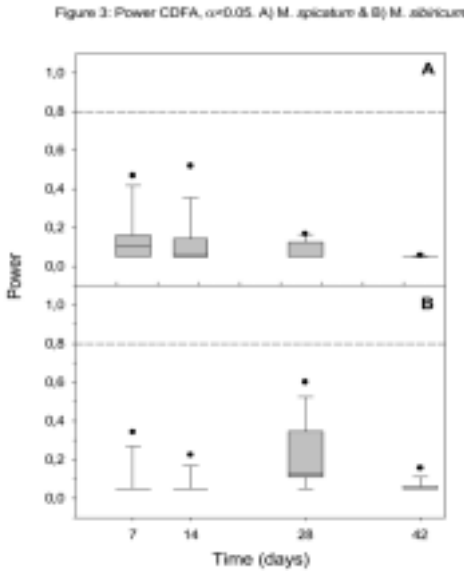

Figure 3. Box plot of the calculated power with alpha< 0.05 of a one way analysis of variance (ANOVA) for Myriophyllum spicatum *(A) and* M. sibiricum *(B) from a study examining the toxicity of chlorodifluoroacetic acid (CDFA) in aquatic microcosms for ten effect measures. The median is shown as the solid line within the box. The box ends are the 25th and 75th centiles and the whiskers bars are the 10th and 90th centiles. Any points shown outside these areas are considered outliers. Power at 0.8 is considered adequate for an ANOVA.*

In the studies with DCA and CDFA, most of the endpoints monitored had low power (Figure 2 and 3) for both plant species. In many of the studies with DCA and CDFA, for both plant species and most effect measures, there

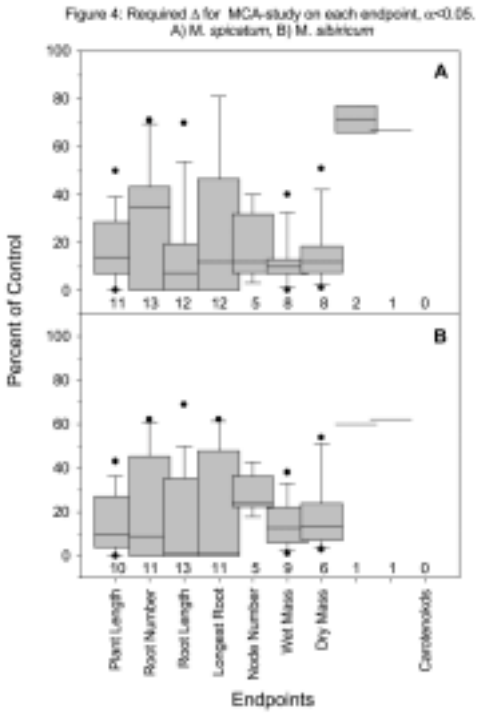

Figure 4. Box plot of the percent of control for effect measures deemed to be significant by a one way analysis of variance (ANOVA) for Myriophyllum spicatum *(A) and* M. sibiricum *(B) from a study examining the toxicity of monochloroacetic acid (MCA) in aquatic microcosms. The median is shown as the solid line within the box. The box ends are the 25th and 75th centiles and the whiskers bars are the 10th and 90th centiles. Any points shown outside these areas are considered outliers. Numbers under each box plot represent the total number of significant differences detected for that effect measure of the specific species.*

was a 95% chance of committing a Type II error if a true difference did exist. Both these studies reported few, if any, significant differences from control and had weak (low r^2 values) concentration-response relationships (Hanson et al., 2001a, 2001d). The significant differences from the MCA

Table 3. The coefficient of variation of Myriophyllum spicatum *endpoints from four separate studies conducted in 1999 and two in 1998. The minimal detectable change from control is provided.*[a]

Compound	Day	Plant Length	Root No.	Root Length	Longest Root	No	Wet Mass	Dry Mass	Chl *a*	Chl *b*	Carot.
MCA	4	11	17	47	47	1	17	24	5	6	6
	7	23	7	17	13	1	25	26	11	15	9
	14	21	5	13	18	4	20	15	13	22	16
	28	8	7	7	11	3	22	26	13	14	14
MCA mean		16	9	21	22	2	21	22	10	14	11
MCA minimum Δ		52	30	68	72	7	68	72	33	46	36
DCA	4	12	33	58	69	11	21	19	23	19	29
	7	11	29	17	18	4	25	26	27	63	15
	14	22	22	22	9	6	30	33	53	59	54
	21	21	17	31	11	6	44	49	41	49	35
DCA mean		16	25	32	27	7	30	32	36	48	33
DCA minimum Δ		52	81	>100	88	23	98	>100	>100	>100	>100
MCA/DCA	4	6	42	80	50	10	9	5	12	18	8
	7	9	15	39	31	7	17	13	36	39	34
	14	8	19	24	16	4	15	12	12	18	13
	21	18	22	32	11	4	25	19	11	13	12
MCA/DCA mean		10	25	44	27	6	17	12	17	22	17
MCA/DCA minimum Δ		33	81	>100	88	13	55	91	55	72	55
CDFA	7	4	0	28	9	7	10	30	19	24	24
	14	5	13	16	2	4	12	14	5	12	20
	28	8	9	10	5	5	2	3	7	7	6
	42	8	13	10	9	16	21	16	9	15	12
CDFA mean		6	9	16	6	8	11	16	10	14	16
CDFA minimum Δ		13	30	52	13	26	36	52	33	46	52
1999 mean[a]		10 ± 8	14 ± 12	24 ± 21	17 ± 19	5 ± 4	16 ± 11	17 ± 13	15 ± 14	21 ± 19	16 ± 14
1999 median		9	15	22	11	4	20	19	12	18	14
1999 max.		23	42	80	69	16	44	49	52	63	54
1999 min.		4	0	7	2	1	2	3	5	6	6
Minimum Δ		33	46	78	55	17	52	55	49	68	52
1998 mean[b]		9 ± 4	21 ± 22	nc	26 ± 27	8 ± 3	15 ± 8	19 ± 15	11 ± 7	17 ± 16	10 ± 5
1998 median		9	15	nc	13	8	16	16	8	9	9
1998 max		18	75	nc	85	12	29	64	26	62	22
1998 min.		3	5	nc	5	3	2	1	4	5	5
Minimum Δ		30	68	nc	85	26	49	62	36	55	33

[a] The values shown are the mean ± the standard deviation of the coefficients of variation shown in the table.
[b] The values shown are the mean ± the standard deviation of the coefficients of variation for *M. spicatum* from two studies conducted in 1998 (Hanson et al., 2001b, 2001c). "nc" stands for not calculated.

study detected by ANOVA ($p \leq 0.05$) with a Dunnett's test ($\alpha = 0.05$) for *M. spicatum* and *M. sibiricum* varied depending on the date of sampling and by species (Figure 4) with most changes greater than 50% from control values. Both species of plant showed similar trends in the relative sensitivity of the

effects measures to detect differences in this study. Pigment and root evaluations being the least sensitive and most sensitive effect measures in regard to these compounds, respectively.

Table 4. The coefficient of variation of Myriophyllum sibiricum *endpoints from four separate studies conducted in 1999 and two in 1998. The minimal detectable change from control is provided.*[a]

Compound	Day	Plant Length	Root No.	Root Length	Longest Root	Node No.	Wet Mass	Dry Mass	Chl *a*	Chl *b*	Carot.
MCA	4	8	21	49	34	6	7	20	11	8	7
	7	8	19	31	18	8	8	23	8	12	8
	14	23	18	9	25	7	14	5	28	29	34
	28	27	24	40	27	7	50	50	4	4	4
MCA mean		16	21	32	26	7	20	24	13	13	13
MCA minimum Δ		52	68	>100	85	23	65	78	43	43	43
DCA	4	4	22	80	69	4	7	6	22	22	50
	7	6	11	15	17	8	9	14	35	80	12
	14	15	9	18	14	4	21	24	27	33	22
	21	19	11	11	12	5	30	28	19	32	19
DCA mean		11	13	31	28	5	17	18	26	42	26
DCA minimum Δ		36	43	>100	91	17	55	59	85	>100	85
MCA/DCA	4	3	13	59	42	7	8	27	1	4	4
	7	2	38	32	11	4	20	24	25	31	25
	14	5	5	17	19	8	3	8	27	34	33
	21	7	13	10	4	7	12	14	9	19	7
MCA/DCA mean		4	17	30	19	7	11	18	16	22	17
MCA/DCA minimum Δ		13	55	98	62	23	36	59	52	72	55
CDFA	7	14	29	13	1	9	11	13	6	8	9
	14	12	17	32	10	11	6	5	18	33	6
	28	7	6	17	6	3	14	11	8	10	6
	42	4	17	14	14	5	18	11	6	3	8
CDFA mean		9	17	19	8	7	12	10	9	14	7
CDFA minimum Δ		30	55	62	26	23	39	33	30	46	23
1999 mean[a]		9 ± 8	14 ± 10	24 ± 21	17 ± 17	5 ± 3	13 ± 12	15 ± 13	13 ± 11	19 ± 19	13 ± 14
1999 median		7	17	17	14	7	11	14	11	19	8
1999 max.		27	38	80	69	11	50	50	28	80	50
1999 min.		2	5	9	1	3	3	5	1	3	4
Minimum Δ		30	46	78	55	17	43	49	43	62	43
1998 mean[b]		8 ± 3	24 ± 17	nc	24 ± 21	7 ± 3	21 ± 9	24 ± 15	17 ± 17	13 ± 8	10 ± 6
1998 median		8	17	nc	20	8	24	23	8	9	8
1998 max		15	69	nc	85	14	36	64	67	29	22
1998 min.		2	5	nc	5	3	4	1	2	6	2
Minimum Δ		26	78	nc	78	23	49	78	55	43	33

[a] The values shown are the mean ± the standard deviation (n=16) of the coefficients of variation shown in the table.
[b] The values shown are the mean ± the standard deviation (n=14) of the coefficients of variation for *M. spicatum* from two studies conducted in 1998 (Hanson et al., 2001b, 2001c). "nc" stands for not calculated.

The CVs were calculated for each study and plant species at each time-point (Table 3 and 4) and were used to determine the number of replicates required to detect significant differences at specific effect levels (Table 5). The CVs

from two studies conducted in a similar fashion in 1998 (Hanson et al., 2001b, 2001c) were calculated and the mean reported (Table 3 and 4). The studies from 1998 had CVs and endpoint ranking similar to that of the studies conducted in 1999. The median values were also reported for these studies as previous work examining power and variation has found the mean to be an unreliable measure of the central tendency due to influences from extreme values (van den Hoeven, 1998). In this study, the median and mean values were not noticeably different (Table 3 and 4), so the mean values were used in subsequent calculations. The distributions of the CVs for each plant species was plotted in order to easily visualize the range of variability within these plants for the examined endpoints (Figure 5). The average of these was taken to derive an overall CV for that endpoint for that specific plant species, regardless of the time of study or compound used.

Figure 5: Distribution of 10 effect measures coefficients of variation and required *n* for four studies. A) M. *spicatum* & B) M. *sibiricum*

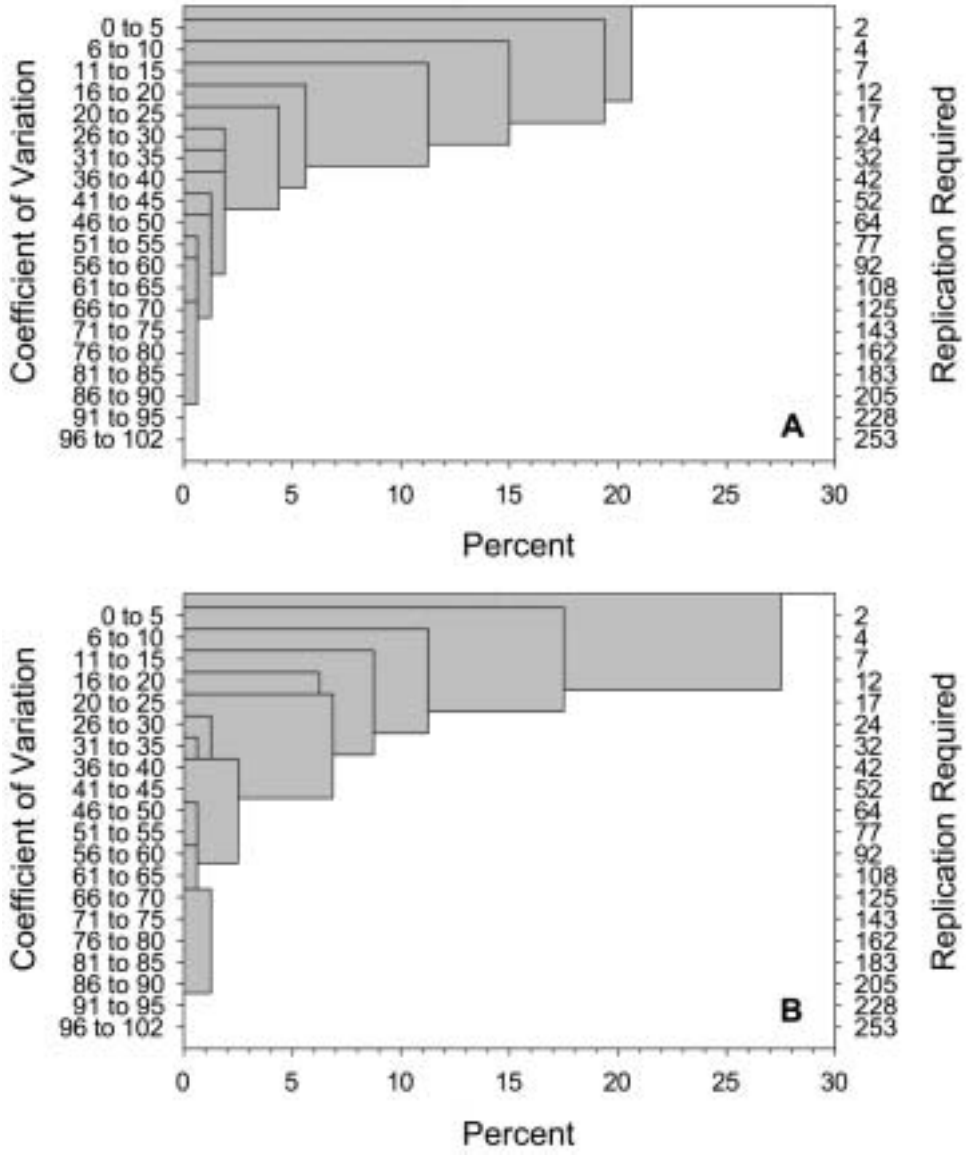

Figure 5. Distribution of the coefficients of variation calculated for Myriophyllum spicatum *(A) and* M. sibiricum *(B) from four studies evaluating 10 effects measures. The right y-axis is the maximum number of replicates required to detect a 25 % change from control with a t-test with* $\alpha < 0.05$ *and* $\beta < 0.2$.

The CVs have a strongly correlated relationship mean r^2=0.88 (range 0.79-0.94) between the two plant species (Figure 6a), implying that the level of variation seen in specific endpoints is similar to both species. The overall levels of variation for *M. spicatum* are summarized as node number < plant length < root number < chlorophyll *a* < wet mass = carotenoids < longest root length = dry mass < chlorophyll *b* < total root length.

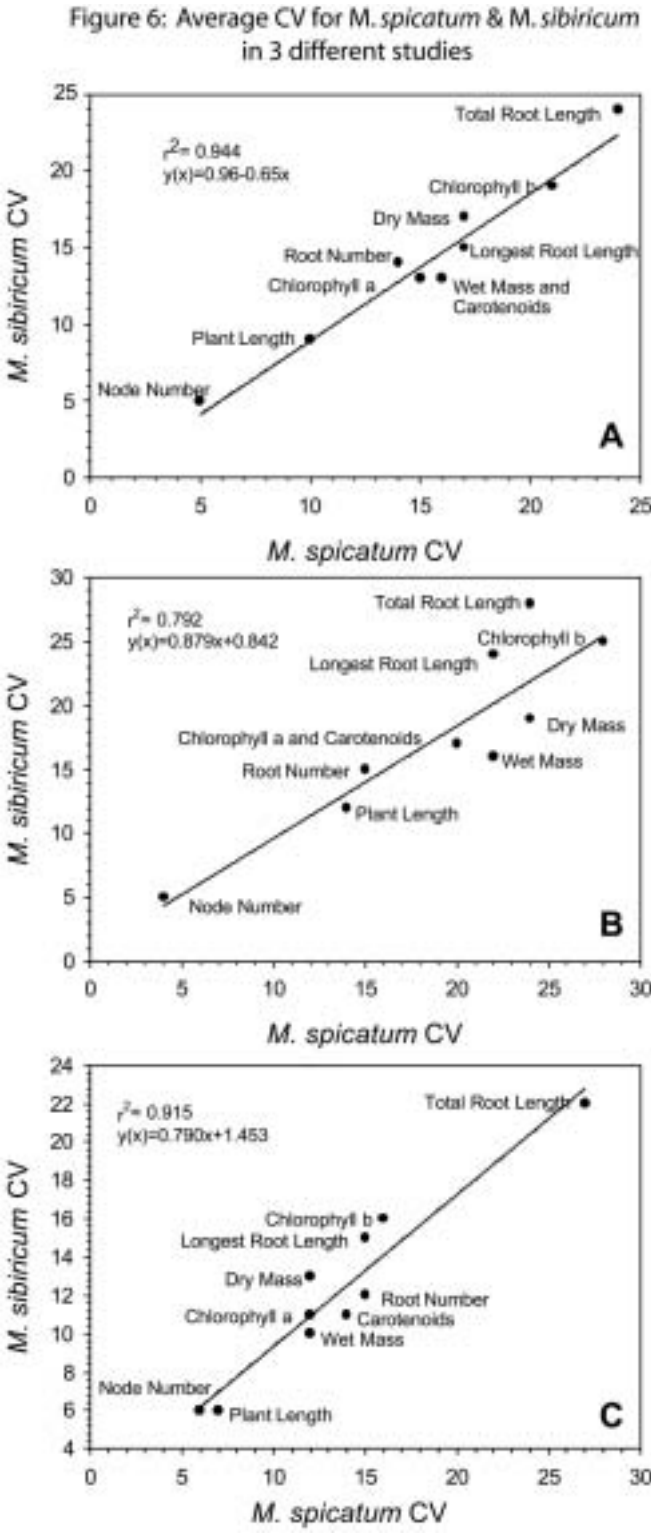

Figure 6. The average coefficient of variation of 10 effects measures from four studies with Myriophyllum spicatum *regressed against those of* M. spicatum *(A), from two studies conducted in the late spring/early summer (B) and two studies conducted in the late summer/early fall (C).* M. sibiricum *and studies conducted in the late summer/early fall had lower variation.*

The overall levels of variation for *M. sibiricum* are summarized as node number < wet mass = carotenoids = chlorophyll *a* < root number < dry mass < chlorophyll *b* < total root length. The regression line of the *M. sibiricum* CVs plotted against the *M. spicatum* CVs had a slope of 0.94 implying that *M. sibiricum* has slightly lower overall variation using these endpoints. The level of effect to be deemed ecologically significant was a $\Delta = 25\%$ change from control. In order to detect such a difference with high power ($\beta = 0.2$, $\alpha = 0.05$), on average, 2 to 19 replicates were required depending on the effect measure and plant species (Table 3, 4 and 5) in a Student t-test design.

Table 5. Number of replicates required for *Myriophyllum spicatum* and *M. sibiricum* to achieve specific differences from control values with a high level of power (β<0.2).

Coefficient of Variation	Percent Change from Control										
	10	15	20	25	30	40	50	60	75	90	99
2	2										
3	3	2									
4	4	3	2								
5	6	4	2								
6	7	4	3	2							
7	9	5	4	3	2						
8	12	6	4	4	3	2					
9	14	7	5	4	3	2					
10	17	9	6	4	3	2					
11	21	15	6	5	4	3	2				
12	24	12	7	5	4	3	2				
13	28	13	8	6	5	4	3	2			
14	32	15	9	7	5	4	3	2			
15	37	17	10	7	6	4	3	2			
16	42	19	12	8	6	4	4	3	2		
17	47	22	13	9	7	5	4	3	2		
18	52	24	14	10	7	5	4	3	2		
19	58	27	16	11	8	5	4	4	3	2	
20	64	29	17	12	9	6	4	4	3	2	
21	71	32	19	13	9	6	4	3	3	2	
22	77	35	21	14	10	6	5	4	3	2	
23	85	38	22	15	11	7	5	4	3	3	2
24	92	42	24	16	12	7	5	4	4	3	2
25	100	45	26	17	12	8	6	4	4	3	3
26	108	49	28	19	13	8	6	5	4	3	3
27	116	52	30	20	14	9	6	5	4	3	3
28	125	56	32	21	15	9	7	5	4	4	3
29	134	60	35	23	16	10	7	5	4	4	3
30	143	64	37	24	17	10	7	6	4	4	3
31	152	69	39	26	18	11	8	6	4	4	4
32	162	75	42	27	19	12	8	6	5	4	4
33	172	77	44	29	21	12	8	6	5	4	4
34	183	82	47	31	23	13	9	7	5	4	4
35	194	87	50	32	23	14	9	7	5	4	4
36	205	92	52	34	24	14	10	7	5	4	4
37	216	97	55	36	25	15	10	8	5	4	4
38	228	102	58	38	27	16	11	8	6	5	4
39	240	108	61	40	28	16	11	8	6	5	4
40	253	113	64	42	29	17	12	9	6	5	4
41	265	119	67	44	31	18	12	9	6	5	4
42	278	125	71	46	32	19	13	9	7	5	5
43	292	131	74	48	34	20	13	10	7	5	5
44	305	137	77	50	35	21	14	10	7	5	5
45	319	143	81	52	37	21	14	10	7	6	5
46	334	149	85	55	38	22	15	11	7	6	5
47	348	156	88	57	40	23	15	11	8	6	5
48	363	162	92	59	42	24	16	12	8	6	5
49	378	169	96	62	43	25	17	12	8	6	5
50	394	176	100	64	45	26	17	12	9	6	6
52	426	190	108	69	39	28	19	13	9	7	6
54	459	205	116	75	52	30	20	14	10	7	6
58	530	236	134	86	60	35	23	16	11	8	7
59	548	244	138	89	62	36	23	17	11	8	7
63	625	278	157	101	71	40	26	19	13	9	8
69	749	334	188	121	85	48	31	22	15	11	9
80	1006	448	253	162	113	64	42	29	19	14	12

The replication required follows the same ranking as mentioned previously for the levels of variation observed, with node number generally requiring the least amount of replication. The same results were observed with calculation of the minimal detectable difference for each effect measure. With three replicates and high power requirements ($\beta = 0.2$ $\alpha = 0.05$) the average minimal detectable difference for both *M. spicatum* and *M. sibiricum* ranged from 17 % to 78% change from control, with node number being the most sensitive and total root length the least.

Timing of the studies had an impact on the levels of variation seen in these studies. When separated into early summer (MCA and DCA) and late summer/fall (MCA/DCA and CDFA) studies, regression analysis showed that the later in the field season the studies were conducted the lower the overall variation for both plant species (Figure 6b). *M. sibiricum* also showed lower levels of variation as compared to *M. spicatum* for the study conducted in the late summer/fall as opposed to the early summer (Figure 6c).

Trends in the variation, replication required and minimal detectable differences were also evident within studies (Table 3 and 4). An increasing trend was noted with plant length and biomass measurements, root measures and node number tended to decrease over the course of the study and pigmentation showed no consistent trends for both plant species.

Table 6. The coefficients of variation for studies conducted by Marwood et al (2001) and Snel et al (1998) for a chlorophyll fluoresence assays and biomass for Myriophyllum spicatum *and* Elodea canadensis.

Effect Measure[a]	Plant Species	Study	Length of Study	Coefficient of Variation	Minimal detectable difference Δ from control (%)
Fv/Fm	*M. spicatum*	Marwood et al., 2001[b]	28	2	7
F/Fm	*M. spicatum*	Marwood et al., 2001	28	4	13
qP	*M. spicatum*	Marwood et al., 2001	28	3	10
1-qN	*M. spicatum*	Marwood et al., 2001	28	9	30
Biomass	*M. spicatum*	Marwood et al., 2001	28	15	49
Photosynthetic Efficiency	*E. canadensis*	Snel et al., 1998[c]	7	2	7
Photosynthetic Efficiency	*E. canadensis*	Snel et al., 1998	28	0	<1
Photosynthetic Efficiency	*E. canadensis*	Snel et al., 1998	35	1	4
Photosynthetic Efficiency	*E. canadensis*	Snel et al., 1998	56	2	7
Biomass	*E. canadensis*	Snel et al., 1998	56	13	43

[a] Fv/Fm the maximum efficiency of electron transport in photosystem II, F/Fm the effective yield of photosystem II photochemistry, qP is the photochemical quenching and qN is the non-photochemical quenching.
[b] Marwood et al., 2001 had an $n = 2$ with 4 subsamples averaged per outdoor micorosm
[c] Snel et al., 1998 had an $n = 2$ using indoor microcosms

Table 7. The coefficients of variation as calculated from Roshon and Stephenson (1997) for a standard laboratory toxicity bioassay with Myriophyllum sibiricum *(ASTM, 1999), the power of the assay ($\beta < 0.2$) and the percent change detectable with the calculated coefficient of variation with $\alpha < 0.05$ and $\beta < 0.2$ for various effect measures.*

Effect measure	Coefficient of variation[a]	Power of test[b]	Minimal detectable change (Δ) from control (%)[c]
Plant length	6 ± 2	1	13
Root number	22 ± 12	0.353	45
Root length	17 ± 7	0.533	25
Wet mass	11 ± 4	0.881	23
Dry mass	13 ± 8	0.759	27

[a] The coefficient of variation reported is the mean ± standard deviation of five separate assay coefficients of variation.
[b] The power of the test was calculated for a t-test with five replicates.
[c] The detectable change (Δ) from control was calculated in for a t-test with five replicates and the mean coefficient of variation for that effect measure.

Subsampling had a distinct impact on the observed variation of the effect measures. Both plant species generally showed increased levels of variation when only one sample was used to calculate the coefficient of variation for that effect measure relative to the coefficient of variation based on the average of two subsamples (Figure 7).

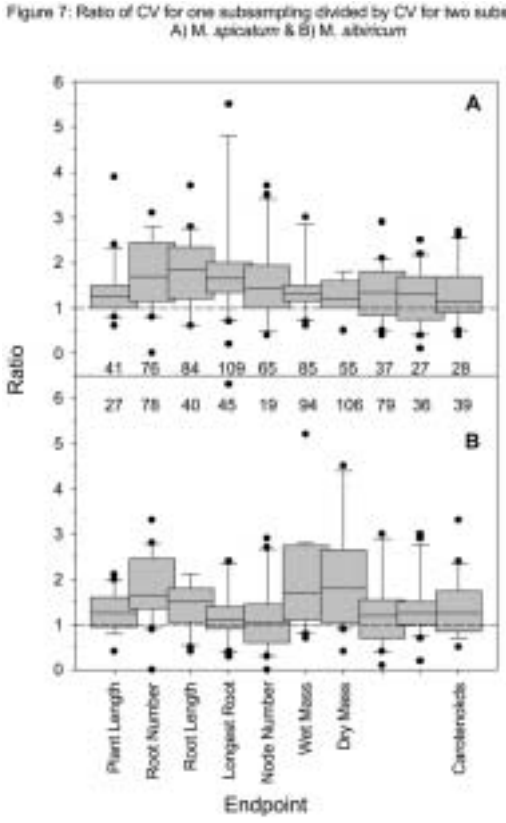

Figure 7. Box plot of the ratio of the coefficient of variation for one subsample by the coefficient of variation for two subsamples for ten effect measure from four studies with Myriophyllum spicatum *(A) and* M. sibiricum *(B). Any value above one shows that subsampling reduces the overall variability. The values were calculated from 4 separate microcosm studies with 4 different evaluation dates and combined for each endpoint. The median is shown as the solid line within the box. The box ends are the 25th and 75th centiles and the whiskers bars are the 10th and 90th centiles. Any points shown outside these areas are considered outliers. The numbers shown in the plots are the average decrease in the coefficient of variation that occurs when using two vs one subsample for that effect measure and plant species.*

Discussion

A major challenge for modern biology is to consider and quantify both probability of Type I and II errors. This will involve more thought about biological and ecological processes operating in nature, so that better models, experiments and more structured predictions can be made. Better quantification of predictive hypotheses will contribute to answering the criticisms on the inadequacies of modern practical ecology and ecotoxicology (Peters, 1991; Underwood, 1997). The use of power analysis raises the issue of how determination of α, β, and ecologically significant effects Δ is conducted and could thus include more transparency in risk management (Sanderson & Petersen, 2001b). An effective power analysis requires that the above mentioned entities be defined *a priori*. The power analysis basically shows us whether the replication design is able to detect a difference, between the treated systems and the controls, which are greater than the environmental criterion. Environmental criteria are not a statistical question as such, but are a benchmark in ecology and management. Ideally, the environmental criteria should be selected such that ecological insignificant effects are less than the environmental criteria, whereas ecological significant effects should be greater. Currently, consensus of how to define scientifically environmental criteria has not been reached by ecologists and environmental managers. Until such a criterion is available, the IC_{25} was used for aquatic macrophytes. It is based on the United States Environmental Protection Agency criteria of >20% change in the biometric of interest to be deemed ecologically significant (Christman et al., 1994).

The actual power of the tests in 1999 was generally low, with the exception of MCA, and therefore the risk of committing a Type II error was high. The most obvious reason for this is that significant impacts were observed which helps to increase the differences between the population means, increasing the power of the test (Zar, 1984). All other parameters in the calculation of Φ (1) would be equal (k, n) or relative similar (error MS), meaning the test would be no more sensitive at detecting smaller changes than the others. The actual coefficient of variation associated with the control data was not any lower than for the other studies and therefore the minimal detectable differences were no less than for the other studies. This demonstrates the need to report both the power of a test and the minimal detectable difference for that study.

There are a number of ways to improve the calculated power or sensitivity (1) of an ANOVA test, including increasing sample size or replication (n), decreasing the number of groups evaluated (k), increasing differences or effect sizes between population means (group MS) and decreasing variability within the population (error MS) (Zar, 1984). With microcosm, or large field studies in general, the ability to increase the number of replicates can be very difficult. At the Guelph Microcosm Facility there are a total of 30 microcosms available. The best replication that can be achieved in a design with even sample size is an n of 15 with only one treatment level and the controls. Increased differences between population means can be achieved through a more conscientious selection of the treatment levels. By evaluating effects observed in the laboratory assay prior to the initiation of a field study and setting treatment levels based on these results, group MS can be reduced. For the compounds in the present study, treatment concentrations could not be increased or set to effect levels seen in the lab for a number of reasons, including cost of the material and environmental realism of the concentrations. When treatment concentrations were based on laboratory studies, such as for DCA, rapid bacterial degradation of the compound in the microcosms

mitigated toxicity (Hanson et al., 2001d) since the laboratory assay is axenic and DCA concentrations remained constant. Efforts to reduce the variation within the population (error MS) can be achieved through subsampling. In this study, subsampling tended to decrease the coefficient of variation observed by 27 to 109% in *M. spicatum* and 19 to 106% in *M. sibiricum*, depending on the effect measure, and is an effective way to increase the power and sensitivity of the test. Regulating such physical parameters such as temperature, light regime and intensity, nutrient availability and pH would also likely reduce this type of variation, but is not feasible or sometimes even desirable with large scale outdoor studies. Smaller microcosm studies conducted indoors such as those of Snel et al (1998) or van den Brink et al (1997) may be more amenable to these types of control. Another way to increase power and sensitivity would be to increase α, the likelihood of making a Type I error. This method has a number of drawbacks, significantly the overwhelming use of α at 0.05 in describing statistical significance in the scientific literature and the potential cost of a false positive in regulating and managing risk.

When selecting effect measures to evaluate toxicity, it is important to understand more than just their potential power and minimal detectable difference. The endpoints should be ecologically relevant and sensitive to toxic insult. Acceptable sensitivity of the test design depends on the criterion for ecological significant effects combined with choice of right ecological endpoint as indicator of plant. These entities can be highly variable within species and over time stress (pers. comm. Dr. John D. Madsen, Minnesota State University & editor of *Jour. Aqua. Plant Mana.* 22/12/00). Node number was consistently the most statistically sensitive effect measure across species and time with an average minimal detectable difference of 17% from control for both species, well within the arbitrary range of 25% change implying an ecologically significant effect. Despite this low variation and high power, few significant differences were found when toxicity was clearly occurring with MCA or DCA exposure (Hanson et al., 2001d, 2001e) as compared with other effect measures (Figure 4). This would imply that node number is not an especially sensitive indicator of toxicity, or at least haloacetic acid toxicity. Its use as an indicator of *Myriophyllum* spp. toxicity should be approached cautiously.

One effect measure that is shared by all plants, and therefore allows for ease of cross species comparison is chlorophyll *a* content. This measure can be a useful indicator of stress in aquatic plants (ASTM, 2000, Marwood et al., 2001). This evaluation found chlorophyll a, chlorophyll b and carotenoids to be, on average, no more variable than the other measures examined. Still, at least in the studies published to date with these plants in the field, few statistically significant impacts have been observed as compared to other effect measures (Hanson et al., 2001d, 2001e) when clear toxicity has been observed with other effect measures (Figure 4). Some studies found chlorophyll's to be a statistically significant and sensitive indicator of stress (Hanson et al, 2001c), but due to variation in other studies (Hanson et al, 2001b) and problems with epiphytic growth (Hanson et al, 2001e), statistical significance was not achieved even though visual observation of pigments would imply dramatic impacts. Studies of natural *M. spicatum* communities in North America found that chlorophyll *a* and *b* concentrations varied greatly between plants with no relationship between pigment concentration and depth of the plant (Marcus, 1980), implying a large, naturally inherent amount of variation in these measures. Problems with pigments as an indicator of toxicity in these plants is not solely due to the variability of the measurements, but

also due to the difficulty of obtaining accurate measurements under field conditions. These effect measures may be more suited to studies where the mechanism of toxicity of the compound of interest is well understood, specifically agents that act on the biosynthesis and functioning of chlorophyll and other pigments themselves, such as some triazoles, isoxazolidinones and isoxazoles, which are bleaching herbicides (Martin, 2000).

Plant length and root measurements appear to combine statistical sensitivity, ecological relevance and sensitivity to toxicants. Next to node number, plant length was consistently the most powerful effect measure. The average minimal detectable difference for *Myriophyllum* spp. plant length with three replicates was 30 to 33%, which is reasonably near our threshold of ecological significance, ≥ 25% change from control. Ecologically, the plant's architecture provides a substrate for fish habitat, epiphytic algae and invertebrates (Chilton, 1990; Duarte and Roff, 1991; Cattaneo et al., 1998). Plant length was also very sensitive to toxicity (Hanson et al., 2001d, 2001e) (Figure 4) making this measure a good candidate for evaluating toxicity. Root measures, while usually the least powerful, were also highly sensitive to toxicity (Hanson et al., 2001d, 2001e) (Figure 4) and hold significant ecological value as the roots stabilize sediments providing a stable substrate for benthic macroinvertebrates and preventing erosion (Lewis, 1995).

The effect measures considered in this paper, with the exception of pigments, are gross morphological endpoints that generally do not react quickly to the introduction of contaminants. Other endpoints in these systems may be more sensitive and have stronger power, such as chlorophyll fluorescence assays, which have been used successfully in these microcosms with *M. spicatum* (Marwood et al 2001a) and in other microcosms (Snel et al., 1998). These endpoints have very low variation in control plants (Table 6). These studies also examined plant biomass and found variation similar to what was observed in the current studies. One point to note is that the variation may be reduced relative to other traditional plant biometrics, but the actual sensitivity of the effect measure may be less than the effect measures of the current studies (Marwood et al., 2001b).

When the CVs and minimal detectable differences of toxicity microcosm bioassays conducted in 1998 (Hanson et al., 2001b, 2001c) were compared with those conducted in 1999, similar levels of variation and ranking of the endpoints were observed, implying that the variation and statistical sensitivity of the bioassay is consistent between years and varying field conditions. This will ease comparison of toxicity of different compounds between various field seasons, a current concern of field level evaluations as most studies tend not to be repeatable or reproducible (Graney et al., 1995). Changes in statistical variation for specific effect measures over smaller time frames (days, weeks) did reveal some general trends. As the data in Tables 3 and 4 demonstrates, variation in plant length, wet mass and dry mass generally increased with length of study, while root number, total root length, longest root and node number tended to decrease with duration of the study. Trends in pigment concentration variation were less evident. Studies conducted later in the field season tended to have lower variation than those conducted earlier in the year. This is due to relatively slower growth in late summer opposed to spring and early summer, likely diminishing the size of the variation observed during that time of year. It may seem reasonable to conduct studies later in the year when variation is lower and the minimal detectable differences are smaller, but this may result in underestimated toxicity. This could occur when comparing *Myriophyllum* spp. exposed and not exposed to

the same compound with the only difference being the growth rate of the plants due to time of the investigation. If the compound suppresses plant growth completely, the study conducted earlier in the season with faster growth rates will have larger effect sizes than the study conducted later in the year with the slower growth rate. The larger the effect size due to the faster growth rate, the more toxic the compound appears to be relative to other toxicants tested with slower growth rates (Huebert and Shay, 1993).

The risk of committing a Type II error when evaluating the data is generally high due to low power, which is due to high variation, too few replicates, and small effect sizes. Low power due to high variation may be due to the choice of *Myriophyllum* spp. as the test organism. Variation observed from results of the standard bioassay with *M. sibiricum* (Table 7 and 8) indicates relatively high background variation among the plants grown under controlled conditions, even with more replicates than these field studies. The inherent variation tends to be larger than the acceptable effect sizes, reducing the detection of adverse effects and increasing risk of Type II errors. Attention towards reducing inherent genetic variability among cultured plants is thus essential to increase power and sensitivity, though this may not be desired in studies, which are meant to replicate the variability that naturally occurs in the field. The general scaling and natural variability of the microcosms do not appear responsible for the majority of variation seen in these studies and thus cannot be held solely accountable for the low power. Variate genotypes of the plants impede the robustness of the indicators and thus lack of a high number replicates (three) resulted in the low power, which again would yield incomplete information to the risk assessor regarding *Myriophyllum* spp. phytotoxicity.

One issue not addressed in this study is the use of regression analysis on the data collected. Previous studies with microcosm data have found that regression analysis has a greater probability of detecting effects not seen by straightforward ANOVA analysis (Liber et al., 1992). Regression analysis also allows for the calculation of effective concentration (EC_x) values and confidence interval about those estimates. Non-linear regression protocols have been developed for plants (Stephenson et al., 2000) and been used successfully on microcosm-derived plant data in conjunction with ANOVA analysis (Hanson et al., 2001d, 2001e), a recommended procedure for microcosm data (Sanderson, 2001). Data analyzed with regression techniques can be used to calculate EC_{10}'s, which can be more conservative surrogates for the NOEC (Crane and Newman, 1999; Moore and Caux, 1997). However, calculation of power of an unreplicated regression design is not possible, hence the risk of a Type II error is unknown (Sanderson, 2001).For regulatory purposes, such as risk assessment of pesticides or other chemicals, it is essential that microcosm studies and bioassays designed to be conducted in such systems be able to confidently detect changes in the biota caused by treatment. Therefore, it is important from a logistical and risk assessment standpoint to know the power and the minimal detectable differences of the study. This will aid in planning of future studies and provide risk assessors with a firmer statistical groundwork upon which to base their recommendations. Currently, the standard method for the laboratory assay with *Myriophyllum* spp. does not require the calculation of power or the minimal detectable difference (ASTM, 1999), though ANOVA and Student t-test analysis are recommended in order to calculate a NOEC. This suggestion has been made in the past for standard bioassays (van den Hoeven, 1997) and should be seriously considered the next time this standard comes under review by ASTM (in 2002) considering the findings of this study.

Another ASTM standard toxicity bioassay uses the aquatic plant *Lemna gibba* (ASTM, 2000), which has also been used in microcosm studies (Hanson et al., 2001a, 2001d), recommends that the power and minimal detectable difference should, as opposed to must, be calculated and reported. This standard (which comes under review in 2003) should also be re-evaluated in this regard.

Power analysis does not solve all problems concerning lack of knowledge and uncertainty in ecological risk assessments. It can test the quality of the statistical testing of the null hypothesis in quantitative terms, but it can never say anything about the quality of the relationship under investigation. Complexity and scale of the ecological risk assessment are thus addressed quantitatively, which could lead to qualitative changes of the design regarding scale, complexity, organisms used, effect measures monitored and duration.

Acknowledgments

The authors would like to acknowledge funding from National Science and Research Council of Canada and the Canadian Network of Toxicology Centres and the helpful comments of P. Sibley in reviewing the manuscript. J. Snel, C. Marwood and R. Roshon kindly provided experimental data to be evaluated.

References

Abbot W 1966. Microcosm studies on estuarine waters. 1 the replicability of microcosms. *J Water Pollut Cont* **38**:258-270.

Aiken SG, Newroth PR, Wile I. 1979. The biology of Canadian weeds. 34. *Myriophyllum spicatum* L. *Can J Plant Sci* **59**: 201-215.

American Society for Testing and Materials. 1999. Standard guide for conducting static, axenic, 14-day phytotoxicity tests in test tubes with the submersed aquatic macrophyte, *Myriophyllum sibiricum* Komarov. E 1913-97. In *Annual Book of ASTM Standards*, Vol 11.05 Philadelphia, PA, pp 1434-1448.

American Society for Testing and Materials. 2000. Guide for conducting static toxicity tests with *Lemna gibba* G3. E 1415 - 91. In *Annual Book of ASTM Standards*, Vol 11.05 Philadelphia, PA, pp 784-793.

Bailer AJ, Oriss JT. 1997. Estimating inhibition concentrations for different response scales using generalized linear models. *Environ Toxicol Chem* **16**: 1554-1559.

Berends AG, Boutonnet JC, de Rooij CG, Thompson RS. 1999. Toxicity of trifluoroacetate to aquatic organisms. *Environ Toxicol Chem* **18**:1053-1059.

Berger TW, Tartowski SL, Likens GE. 1997. Trifluoroacetate retention in a northern hardwood forest soil. *Environ Sci Technol* **31**:1916-1921.

Bestari KTJ, Robinson RD, Solomon KR, Steele TS, Day KE, Sibley PK. 1998. Distribution and composition of polycyclic aromatic hydrocarbons within experimental microcosms treated with liquid creosote. *Environ Toxicol Chem* **17**:2359-2368.

Bestari KTJ, Robinson RD, Solomon KR, Steele TS, Day KE, Sibley PK. 1998. Distribution and composition of polycyclic aromatic hydrocarbons within experimental microcosms treated with creosote-impregnated Douglas Fir pilings. *Environ Toxicol Chem* **17**:2369-2377.

Bowmer, K. H. 1987. Residues of dalapon and TCA in sediments and irrigation water. *Pest Sci* **18**:1-13.

Cattaneo A, Galanti, G, Gentinetta S, Romo S. 1998. Epiphytic algae and macroinvertebrates on submerged and floating-leaved macrophytes in an Italian lake. *Freshwater Biology* **39:** 725-740.

Ceska A, Ceska O. 1986. Notes on *Myriophyllum* (Haloragaceae) in the Far East: The identity of *Myriophyllum sibiricum* Komarov. *Tax* **35**:95-100.

Chapman PM, Caldwell RS. 1996. A warning : NOEC's are inappropriate for regulatory use. *Environ Toxicol Chem* **15**: 77-79.

Chilton, EW. 1990. Macroinvertebrate communities associated with three aquatic macrophytes (*Ceratophyllum demersum*, *Myriophyllum spicatum*, and *Vallisneria americana*) in Lake Onalaska, Wisconsin. *J Freshw Ecol* **5**:455-466.

Christman, VD, Voshell Jr, JR., Jenkins, DG, Rosenzweig, MS., Layton, RJ., and Buikema, AL., Jr. 1994. Ecological Development and Biometry of Untreated Pond Mesocosms. In: Aquatic mesocosm Studies in Ecological Risk Assessment. Eds: Robert L. Graney, James H. Kennedy, and John H. Rodgers, Jr. CRC Press, Inc. Lewis Publishers Boca Raton, FL, USA. pp 105-129.

Crane M. 1997. Research needs for predictive multispecies tests in aquatic toxicology. *Hydrobiol* **346**: 149-155.

Creed RP. 2000. Is there a new keystone species in North American lakes and rivers? *Oikos* **91**:405-408.

Davy M., Petrie, R., Smrchek, J, Kuchnicki, T., Francois, D. 2001. Proposal to update non-tarrget plant toxicity testing under NAFTA. United States Environmental Protection Agency 163 pages http://www.epa.gov/scipoly/sap/2001/june/sap14.pdf

Duarte CM, Roff DA. 1991. Architectural and life history constraints to submersed macrophyte community structure: a simulation study. *Aqua Bot* **42**:15-29.

Ellis D, Hanson M, Sibley P, Shahid T, Fineberg N, Muir D, Solomon K, Mabury S. 2001. The aqueous environmental fate of chloroacetic and trifluoroacetic acids. *Chemosphere* **42**:309-318

Fisher L. 1992. Memorandom. Decision on the ecological, fate and effects task force. 29/10-1992. USEPA.

Forsyth DJ., Martin PA., and Shaw GG. 1997. Effects of herbicides on two submersed aquatic macrophytes, *Potamogeton pectinatus* L. and *Myrionphyllum sibiricum* Komarov, in a prairie wetland. *Environ Poll* **95**:259-268.

Graney, RL., Giesy, JP., and Clark, JR. 1995. Field Studies. In Fundamentals fo Aquatic Toxicology: Effects, environmental fate, and risk assessment. Ed: Gary M. Rand. Taylor and Francis, Washington, DC, USA pp. 257-306.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. *Environ. Res* **50**:195-205.

Hanson ML. 1998. A probabilistic environmental risk assessment of trichloroacetic acid (TCA) in Canadian tap water. In: Canadian Technical Report of Fisheries and Aquatic Sciences No. 2260. Eds: R. Van Coillie, R. Chassé, L. Hare, C. Julien, L. Martel, C. Thellen and A.J. Niimi.

Hanson M., Sibley PK., Solomon KR., Mabury S., Muir, D. 2001a. Chlorodifluoroacetic acid (CDFA) fate and toxicity to the macrophytes *Lemna gibba*, *Myriophyllum spicatum* and *Myriophyllum sibiricum* in aquatic microcosms. *Environ Toxicol Chem* (*in press* 2001).

Hanson M., Sibley PK., Ellis D., Fineberg N., Solomon KR., Mabury S., Muir D. 2001b. Trichloroacetic acid (TCA) fate and toxicity to the macrophytes *Myriophyllum spicatum* and *Myriophyllum sibiricum* in aquatic microcosms. *Aqua Toxicol* (*in press* 2001).

Hanson M., Sibley PK., Solomon KR., Mabury S., Muir D. 2001c. Trichloroacetic acid (TCA) and trifluoroacetic acid mixture toxicity to the macrophytes *Myriophyllum spicatum* and *Myriophyllum sibiricum* in aquatic microcosms. *Tot Sci Environ* (*in press* 2001).

Hanson M., Sibley PK., Solomon KR., Mabury S., Muir D. 2001d. Field level evaluation and probabilistic risk assessment of the toxicity of dichloroacetic acid (DCA) to the aquatic macrophytes *Lemna gibba*, *Myriophyllum spicatum* and *Myriophyllum sibiricum*.. *Ecotoxicol Environ Saf* (*submitted* 2001).

Hanson M., Sibley PK., Solomon KR., Ellis, DA., Shahid., T, Mabury S., Muir D. 2001e. Probabilistic risk assessment of monochloroacetic acid (MCA) toxicity to *Lemna gibba*, *Myriophyllum spicatum* and *Myriophyllum sibiricum* and degradation in aquatic microcosms. *Environ Sci Technol* (*submitted* 2001).

Huebert DB. Shay JM. 1993. Considerations in the assessment of toxicity using duckweeds. *Environ Toxicol Chem* **12**: 481-483.

Kennedy JH., Ammann LP., Waller WT., Warren JE., Hosmer AJ., Cairns SH., Johnson PC. and Graney RL. 1999. Using statistical power to optimize sensitivity of analysis of variance designs for microcosms and mesocosms. *Environ Toxicol Chem* **18**:113-117.

Keast A. 1984. The introduced aquatic macrophyte, *Myriophyllum spicatum*, as habitat for fish and their invertebrate prey. *Can Jour Zool* **62**:1289-1303.

Kühn R., Pattard M.. 1990. Results of the harmful effects of water pollutants to green algae (*Scenedesmus suspicatus*) in the cell multiplication inhibition test. *Water Res* **24**:31-38.

Lewis MA. 1995. Chapter 4 Algae and Vascular Plants. In Rand GM. Fundamentals of Aquatic Toxicology: Effects, Environmental fate, and Risk Assessment, 2nd ed, Taylor and Francis, Washington, DC, USA, pp 135-136.

Liber K., Kaushik NK., Solomon KR., Carey JH. 1992. Experimental designs for aquatic mesocosm studies: A comparison of the „ANOVA" and „Regression" design for assessing the impact of tetrachlorophenol on zooplankton populations in limnocorals. *Environ Toxicol Chem* **11**:61-77.

Lytle, JS, Lytle TF. 2001. Use of plants for toxicity assessment of estuarine ecosystems. *Environ Toxicol Chem* **20**; 68-83.

Marcus BA. 1980. Relationship between light intensity and chlorophyll content in *Myriophyllum spicatum* L. in Canadice Lake (New York, USA). Aquat Bot **9**: 169-172.

Martin H. 2000. Herbicide mode of action categories. Ontario Ministry of Agriculture, Food and Rural Affairs. ISSN 1198-712X.

Martin JW., Franklin J., Hanson ML., Mabury SA., Ellis DA., Scott BF., Muir DCG., Solomon KR. 2000. Detection of chlorodifluoroacetic acid in precipitation: a possible product of fluorocarbon degradation. *Environ Sci Technol* **34**:274-281.

Marwood CA., Dobson E., Smith REH., Solomon KR., Greenberg BM. 2001a. Chlorophyll fluorescence as a bioindicator of photosynthesis inhibition from solar ultraviolet radiation in plants and algae. In Greenberg BM, Hull RN, Roberts Jr MH, Gensemer, RW, eds, *Environmental Toxicology and Risk Assessment: Science, Policy and Standardization: Tenth Volume, ASTM STP 1403*. American Society for Testing and Materials, Philadelphia, PA, USA, pp 27-39.

Marwood CA., Solomon KR., Greenberg, BM. 2001b. Chlorophyll fluorescence as a bioindicator of effects on growth in aquatic macrophytes from mixtures of polycyclic aromatic hydrocarbons. *Environ Toxicol Chem* **20**: 890-898.

Moore DR.J, Caux, PY. 1997. Estimating low toxic effects. *Environ Toxicol Chem* **16**: 794-801.

Peterman RM. 1990. Statistical power analysis can improve fisheries research and management. *Can J Fish Aquat Sci* **47**: 2-15.

Peters RH. 1991. A critique for ecology. Cambridge University Press, Cambridge, England.

Rand, GM, Wells, PG, McCarty, LS. 1995. Introduction to Aquatic Toxicology. In Fundamentals of Aquatic Toxicology: Effects, Environmental Fate and Risk Assessment. Eds: GM Rand. Taylor and Francis, Washington, DC pp.3-66.

Reckhow DA. Singer PC. 1990. Chlorination by-products in drinking waters: From formation potentials to finished water concentrations. *Jour Am Water Works Ass:* 173-180.

Reimann S, Grob K, Frank H. 1996. Chloroacetic acids in rainwater. *Environ Sci Technol* **30**: 2340-2344.

Roshon, R.D. and Stephenson, R.D. 1997. Comparison of two reference toxicants and their effect upon the culture and development of *Myriophyllum sibiricum* in axenic culture. Environmental Toxicology and Hazard Assessment: Seventh Volume. ASTM STP 1333 Eds: EE Little, AJ DeLonay and BM Greenberg. American Society for Testing and Materials. West Conshohocken, PA, pp. 200-214.

Roshon, R.D., McCann, J.H., Thompson, D.G., Stephenson, G.R. 1999. Effects of seven forestry management herbicides on *Myriophyllum sibiricum*, as compared with other non-target aquatic organisms. *Can. J. Forest Res.* **29**, 1158-1169.

Sanderson H. 2001. Replication of Micro/Mesocosm studies. State-of-the-art review. *Environ Sci & Pollut Res* **8**. pp xx. DOI:http:/dx.doi.org/10.1065/espr2001.09.085.

Sanderson H & S. Petersen. 2001a. Applicabillity of quasi-natural ponds ansd pelagic plankton for higher tier risk assessment of $Roundup_{2000}$. *Jour Environ Contam Toxicol* (*submitted* 2001).

Sanderson H & S. Petersen. 2001b. Power analysis as a reflexive scientific tool for interpretation and implementation of the precautionary principle in the European Union. *Environ Sci & Poll Res* DOI:http:/dx.doi.org/10.1065/espr2001.10.095.

SETAC (Society of Environmental Toxicology and Chemistry), 1991. Wintergreen workshop on microcosms report. Pensacola, FL: SETAC Foundation for Education.

Shaw JL., Kennedy JH. 1996. The use of aquatic mesocosm studies in risk assessment. *Environ Toxicol Chem* **15**:605-607.

Sibley PK., Chappel MJ., George TK., Solomon KR., Liber K. 2000. Integrating effects of stressors across levels of biological organization: examples using organophosphorus insecticide mixtures in field-level exposures. *Jour Aqua Ecosys Stress Recovery* **7**:117-130.

Sibley, PK., Harris, ML., Bestari, KT., Steele, TA., Robinson, RD., Gensemer, RW., Day, KE., Solomon, KR. 2001. Response of zooplankton communities to liquid creosote in freshwater microcosms. *Environ Toxicol Chem* **20**:394-405.

Snel, JFH., Vos, J.H., Gylstra, R., Brock, TCM..1998. Inhibition of photosystem II (PSII) electron transport as a convenient endpoint to assess stress of the herbicide linuron on freshwater plants. *Aquat Ecol* **32**:113 123.

Stephenson, GL., Koper, N., Atkinson, GF., Solomon, KR., Scroggins, R.P. 2000. Use of nonlinear regression techniques for describing concentration-response relationships of plant species exposed to contaminated site soils. *Environ Toxicol Chem* **19**:2968-2981.

Underwood, AJ. 1997. Experiments in ecology: Their logical design and interpretation using analysis of variance. Cambridge University Press. UK.

Van den Brink, PJ., Hartgers, EM., Fettweis, U., Crum, SJH., Van Donk, E., Brock, TCM.1997. Sensitivity of macrophyte dominated freshwater microcosms to chronic levels of the herbicide linuron. I. Primary producers. *Ecotoxicol Environ Saf* **38**:13 24.

van der Hoeven N. 1997. How to measure no effect, part II: Statistical aspects of NOEC and ECx estimates. *Envirometrics* **8**: 255-261.

van der Hoeven N. 1998. Power analysis for the NOEC: What is the probability of detecting small toxic effects on three different species using the appropriate standardized test protocols? *Ecotoxiol* **7:** 355-361.

Wilson, RI., Mabury, SA. 2000. Photo degradation of Metolachlor: Isolation, and quantification of monochloroacetic acid. *Jour Agricul Food Chem* **48**:944-950.

Wu WW., Chaddik PA., Davis WM., Powell DH., Delfino JJ. 1998. Disinfection byproduct formation from the preparation of instant tea. *Jour Agric Food Chem* **46**: 3272-3279.

Zar, JH. 1984. Biostatistical Analysis. 2nd Edition. Prentice Hall, INC, Englewood Cliffs, NJ, USA.

VI

Power analysis as a reflexive scientific tool for interpretation and implementation of the precautionary principle in the European Union

by

Sanderson H. & S. Petersen. 2002

Journal of Environmental Science And Pollution Research International.
http://dx.doi.org/10.1065/espr2001.10.095

Commentaries

Power Analysis as a Reflexive Scientific Tool for Interpretation and Implementation of the Precautionary Principle in the European Union

Hans Sanderson[1] and Søren Petersen[2]

[1] Department of Environment, Technology and Social Studies, Building 12.2, University of Roskilde, Postbox 260, DK-4000 Roskilde, Denmark
[2] Danish Hydraulic Institute, Water and Environment, Agern alle' 5, DK-2970 Hørsholm, Denmark

Corresponding author: Hans Sanderson; E-mail: hanss@ruc.dk

DOI: http://dx.doi.org/10.1065/espr2001.10.095

Abstract. The diversity of interpretation, the subsequent lack of implementation, and the enforcement of the precautionary principle have been important issues in the European environmental discourse for the past five years. The European Commission published a communication on the Commission's interpretation of the precautionary principle on February 2nd, 2000. However, the distinction between precaution and prevention is absent in the EU Commission's interpretation, resulting in the communication's lacking relevance for the precautionary principle. The important consequence of the precautionary concept in policy and decision-making is that it should not be based on an assumed certainty of the certainty of environmental knowledge – but rather on a certainty of the uncertainty of environmental knowledge. In other words, the regulation should, to a greater extent, be based on the management of uncertainty, and risk assessments should explicitly present and discuss related uncertainty and lack of knowledge. The management of uncertainty should be based on setting the acceptable level of risk of accepting a failure to reject the null hypothesis of no adverse effects (β). This is done by setting the required power (1-β) according to a socioeconomic cost-benefit analysis. Moreover, the acceptable ecological effect size (Δ) could also be set *a priori* which would have implications for the power of a study. Reversal of the burden of proof could be considered in order to resolve possible legal implications for the risk managers.

Keywords: Precautionary principle; risk assessment; uncertainty; type II error; power analysis

Introduction

Principle of Precaution: Handling of a lack of knowledge and regulating based knowledge beyond reasonable doubt (context dependent and flexible truth claims and early warnings). The handling of epistemic and methodological uncertainty.Traditionally not scientific (§ 15 in the Rio declaration, 1992).

Principle of Prevention: Handling of data uncertainty and the precision of measurable effects and causalities beyond any doubt (95% convention). The scientific handling of test and data uncertainty EU-Communication (1) 2/2 2000.

Highly variable regulatory entities like *assumptions; experiences, common knowledge, opinions or anxieties* would fall in the realm of precaution when untested or un-testable, where-as a Popperian falsification approach of a null-hypothesis would be prevention.The challenge met by power analysis is to merger prevention and precaution and make the acceptable doubt (β) and environmental effect (Δ) transparent, and to integrate science and precaution. We propose that this could be considered within public environmental consensus forums.The difference between prevention and precaution is at present somewhat blurry.

The precautionary principle gained wide acceptance and public awareness following the 1992 Rio Declaration of the United Nations sustainable development meeting, and since then attempts have been made to introduce it into national and international environmental legislation and regulations. Principle 15 of the Rio declaration states: "Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation". So, originally, the idea of the principle was to speed up the rate of sustainable development and overcome the delays in environmental decision-making which were due to scientific uncertainty and the lack of knowledge concerning complex and intrinsic environmental matters (Hey 1992, O'Riordan and Jordan 1995). The history of the precautionary principle has been described in great detail (Freestone and Hey 1996, Harding and Fisher 1999, Raffensberg and Tickner 1999). Furthermore, the political opportunism related to the implementation of principle (Pers. Comm. T. O'Riordan 2001) is not covered in this paper. The explicit management of the lack of knowledge and uncertainty is increasingly important in contemporary society, and has been characterized by the German sociologist Ulrich Beck (1992) as a 'Risk society'. In the risk society the distribution of risks, real and assumed, can be seen as determining, driving forces for societal development in the new millennium which are complementary to the distribution of wealth. Moreover, some argue that the science/policy interface is characterized by a shift from 'hard facts' and 'soft values' to 'soft (scientific) facts' and 'hard (public) values' (e.g. the sinking of oilrig Brent Spar in the North sea and public opposition towards genetically modified organisms and food) pushing decision-makers for a faster and assumed precautionary approach (Pers. Comm. D. Gee 2000).

Hence, the precautionary principle has also been written into the preamble of the Amsterdam treaty as a guiding tool upon which the development of the European Union's (EU) environmental policy should be based. However, the interpretation of the precautionary principle is still unclear and this impedes the implementation and regulatory use of the principle. A communication sent from the EU-Commission in February 2000 addressed the differences in interpretation (EU COM 2000, **1**), and was hence explicitly written into the EU white paper on future new chemicals policy strategy of 28/3-2001, (EU COM 2001, **88**). Where there are lacking, neglected or reluctant data from the companies applying for registration of their products, the regulators are entitled to limit the permit according to the precautionary principle, according to (EU COM 2000, **1**). This power, moreover, was ratified by member states in the Nice treaty (Nice 2001) covering all chemical policy realm.

According to the EU communication, the precautionary principle is intended to be used when there is a reasonable suspicion of unacceptable environmental risk, but the causal relations are scientifically unclear or biased. Thus, further investigation and better documentation are required on the basis of the precautionary principle. On this new basis, a cost-benefit analysis of the pros and cons should be politically weighted and the policy makers should decide on further regulatory action. In this context, precaution is about increased environmental protection through better and more thorough risk analysis, and not about more direct regulatory action based a 'non-scientifically' founded but reasoned suspicion, scares, uncertain or biased scientific documentation. According to the EU Commission's spokesperson Christine Majewski, the communication was also intended to 'open a debate' which might lead to a common understanding of how to assess and manage risks in the face of uncertainty (Santillo 2000). Moreover, the precautionary principle is strictly a political tool only for politicians to use. Claims or considerations of an implementation of precautionary approaches by scientists and researchers is not desirable (EU COM 2000, **1**).

The aim of this paper is to participate in this debate discussing the difference between interpretation of prevention and precaution and, primarily, to discuss possible ways of implementing and managing precaution via statistical power analysis, and to address this as a first step towards increasing the reflexibility of science that Beck (1992) issued as a main precursor for future sustainable development. In power analysis of the null hypothesis, the acceptable effect sizes (Δ), the probability of Type I (α) and also of Type II (β) errors need to be explicitly addressed which could permute a reflexive process between politicians, regulators, stakeholders and scientists. Determination of acceptable anthropogenic change or impact (Δ) of the environment in light of the precautionary principle and reflexive science will also be discussed. Moreover, the aim is to broaden these scientific considerations to the realm of social science and into the regulatory sphere in a relatively accessible form.

1 Interpretation of the Precautionary Principle

The EU Commission's communication concerning the interpretation of the precautionary principle states explicitly, both in writing and verbally, that there is no real difference between the principle of prevention and the precautionary principle (Pers. Comm. Majewski 2000). If there is a lack of knowledge, measures should be taken to gain more scientific knowledge before regulatory action. In this situation delays occur, Weinberg (1985) calls it 'the regulators' dilemma'. How should international lawyers and regulators cope with uncertainties? If an experiment does not confirm or disprove a hypothesis, scientists can continue to gather information, but regulators must choose a course of action. Delaying action in the hope that new information will resolve or, at least, reduce uncertainty, is itself an interim decision (Bodansky 1991). It is difficult to see the legal and regulatory novelty and necessity for the precautionary principle in the EU-Commission's interpretation of a precautionary principle, if it doesn't imply any real difference in the regulatory praxis, except changing the word prevention into precaution in a preamble to the EU treaty.

The problem in interpreting the precautionary principle has been addressed by several authors. Rehbinder (1994) lists nine different possible interpretations of the precautionary principle. The Norwegian sociologist Bratt (1996) defines the difference between the precautionary principle and the principle of prevention as follows: "The prevention of known hazards and prevention of risks of hazards would traditionally fall within the realm of environmental prevention. Whereas the prevention of *possible* risk of environmental hazards would be precaution". In other words, when the risk is known, e.g. that compound X is carcinogenic, persistent and mobile, it is a prevention to regulate the use of X and prevent it from entering the environment and groundwater. However, regulating a compound Y on the basis of possible estrogenic properties would be precaution. Vital in this context is the level of certainty and regulation on the basis of uncertainty and circumstantial evidence (Sanderson 2000). In praxis, the definition of the precautionary principle is most likely a context-dependent interpretation between prevention and precaution. The European court of Justice has consistently defined the precautionary principle as follows: "Where there is uncertainty as to the existence or extent of risk to human health, the institutions may take protective measures without having to wait until the reality and seriousness of those risks become fully apparent". This was used in the Mad Cows (BSE) judgment and it is also consistent in environmental matters (Pers. Comm. David Gee 2000). This interpretation of the original principle of precaution (principle 15 in the Rio Declaration) would imply substantial changes in regulatory praxis. One could be the reversal of the burden of proof, which is often mentioned in relation to the precautionary principle. Traditional tort goals depend heavily on reliable information about causation. If the fact-finder is left in a state of great uncertainty about causation, he is unable to conclude that a litigated compound is either safe or unsafe. Under these circumstances, whichever party bears the burden of proof concerning the question of causation will lose – not because the fact-finder has good reason to conclude that the litigated compound does or does not cause harm, but because of a procedural default rule whose operation is not governed by the truth about causation (Feldman 1995). However, this could jeopardize

the legal status of the risk manager actually using the principle and seriously hamper the implementation of the principle. They could risk personal prosecution, as John Carey of Environment Canada pointed out at the Society of Environmental Toxicology and Chemistry 3rd World Congress in Brighton May 2000, because they could, legally and personally, be held responsible for their decisions – which are clearly unacceptable (Pers. Comm. J. Carey 2000).

Given these problems in interpretation, we will continue by discussing the implementation of the precautionary principle based on a management of uncertainty via statistical power analysis in the context of the EU-Commission's definition of the precautionary principle.

2 Statistical Testing of the Null Hypothesis

The new knowledge produced by the additional scientific work, e.g. ecological toxicity and its adverse effects on the environment or health, is tested by the null hypothesis. The null hypothesis tests whether there is a statistically significant change in the average abundance of the test organism in a comparison between treated and control groups. So the starting point is that there is no difference, and the null hypothesis is then to be falsified by *e.g. t*-test with p or $\alpha \leq 0.05$ as maximal acceptable probability of the rejection being an error (i.e. the null hypothesis is really correct). Thus, the risk of committing a Type I error (accepting a false positive) is statistically tested. Industrial statisticians also call this 'the producer's risk'. If the test shows a statistically significant relationship and the tested relationship also is validated as ecologically relevant, then we need look no further.

But if it is not possible to reject the null hypothesis with more than ($\alpha \geq .05$), the compound or activity is, in theory, often perceived not to have adverse effects. However, this statistical testing is not an example of environmental precaution as it only tests the risk of committing a Type I error that could adversely affect the producer's risk of erroneously rejecting the null hypothesis. Ninety-eight per cent of all marine and aquatic biomonitoring (Peterman 1992) and higher tier aquatic ecotoxicology null hypothesis tests (Sanderson 2001) only calculate the probability of committing a Type I error – or estimations of the producer's risk. Fifty-two per cent of the biomonitoring tests concluded that there was no effect if the change was not significant, $\alpha \geq 0.05$ (Peterman 1990). For intensive information on power analysis see Green (1979).

3 No Significance is not Equal to No Effect

It is relatively easy statistically, using power analysis, to test an ecological risk assessment's null hypothesis in an environmentally precautionary manner. What is and what constitutes a power analysis? In a statistical power analysis, the focus is on the flip side of the coin of the null hypothesis – namely the risk of committing a Type II error, the acceptance of a false negative result. Power analysis (1-β) shows us the probability that our *t*-test could have shown a difference in case there was one in reality. Where α in the *t*-test symbolizes the acceptable risk of committing a Type I error, β symbolizes the risk of committing a Type II error. For any given test, we would like to have the quantity 1-β be as large as possible and the quantity of β as small as possible (Sokal and Rohlf 1995). The power (1-β) of an investigation is related to and influenced by four variables; effect size (Δ), sample size (n), sample variability (σ^2) and α in the following way:

a) If Δ increases ($\uparrow$) then ($\Leftrightarrow$) $\beta\downarrow\Leftrightarrow$ power$\uparrow$

b) If $n\uparrow\Leftrightarrow\beta\downarrow\Leftrightarrow$power$\uparrow$

c) If $\sigma^2\downarrow\Leftrightarrow\beta\downarrow\Leftrightarrow$power$\uparrow$

d) If $\alpha\uparrow\Leftrightarrow\beta\downarrow\Leftrightarrow$power$\uparrow$

If n or Δ is too low or σ^2 is too high, the statistical power of the test is reduced and thus the risk of committing a Type II error is increased. The conclusion of a study that fails to reject the null hypothesis with low power should be that the study should be changed and retired instead of concluding that there is no effect. There is, after all, a substantial difference between accepting an activity or compound and adjusting a test. Sanderson and Petersen (2001) failed to reject the null hypothesis significantly in a *t*-test ($\alpha \leq 0.05$) of no effect in an ecotoxicological risk assessment performed in mesocosms of the herbicide $\text{Roundup}_{2000.}$ The power analysis showed that the mesocosm test design was too variable to show any effects, thus it would on average require an effect size (Δ) at 95% compared with the controls or on average 20 replicates (n) to obtain high power ($\beta \geq 0.2 \Rightarrow$ power 1-β = 0.8).

Gray (1990) provides an example in which a biologist, who has rejected a null hypothesis of no effect for some substance, is confronted by lawyers from industry with the question: "How do you know that the effect you observed is not in fact due to a natural environmental variable that you haven't measured?" In other words, the legitimate question is asked if you have committed a Type I error. However, the equally legitimate one concerning Type II error is almost never asked when the biologist or industry *fails* to reject the null hypothesis of no effect. "How do you know that the absence of a statistically significant effect in an investigation is not just due to a small sample size or sampling variation, which tends to reduce the chances of detecting an effect that is present?" When this question is not asked, and it often is not, incomplete information is being provided to decision-makers (Peterman and M'Gonigle 1992). The power analysis thus tests the statistical power or quality of a statistical relationship and thus also the quality of investigation and methods leading to testing the null hypothesis. The sample variability σ^2 and the required sample size n are submitted to natural laws, whereas α, β and Δ are conventions (Green 1979).

4 The Recipients' Risk β and Acceptable Effects Δ

The 'consumer's or recipient's risk' β as an estimation of the risk of committing a Type II error are calculated in statistical power analysis. For the power of a test to be high, that is an acceptable risk of a false failure to reject the null hypothesis, β is conservatively set to 0.2 – or 4 times higher than the 'producer's risk' α of 0.05. The setting of α and β are purely based on statistics and mathematics and not ecologically or environmental matters (Sokal and Rohlf 1996). Today's precautionary principle using traditional α and β

values in statistical testing of the null hypothesis protects the producer four times more than it does the environment. A way to deal with the precautionary principle following the EU Commission's interpretation, first of all, is through rigorously asking for an estimation of the null hypothesis test's power.

The setting of an acceptable level of β could relay on cost-benefit estimations, where the cost of committing a Type II error should be held against societal benefits of the activity, compound, building, etc. that is being risk assessed. Two examples: When the Danish parliament unanimously decided to initiate a large national consensus study to estimate the costs of reduced usage of pesticides at a national level, the possible adverse public health effects due to pesticide usage were estimated. The following conclusion was reached: "On the basis of the epidemiological studies it is not possible to prove that the amounts of pesticides the public is exposed to pose health hazards. On the other hand, one can neither scientifically prove that a pesticide will not pose a threat" (Bichel 1998). In this case no power analysis was implemented on the epidemiological data and the inherent uncertainty in the data was not quantitatively linked to a precautionary approach. In the subsequent new Danish Environment Ministry's 'pesticide action plan II' for pesticide regulation the conclusion reached was: "On the basis of current knowledge and data it is estimated that public consumption/exposure of pesticides from contaminated food and water presently does not pose any risk to public health" (Danish Environmental Ministry 2000). The risk of a Type II error, in this case the actual human health risks due to exposure of pesticides, and the associated costs of the error have remained un-assessed.

In the next example, estimations of power were made in the Environmental Impact Assessment (EIA) when constructing the Øresund fixed link between Denmark and Sweden. However, when assessing the assumed environmental impacts of the bridge α was set to 0.05 and β was set to 0.25. In other words, the acceptable risk of committing a Type II error (possible effects to the deepwater fauna) was five times as high as the acceptable risk of a Type I error. The socio-economic cost-benefit analysis set the value of the bridge fives times higher than the outcome of Baltic fishery and wildlife. A precautionary setting of α and β, to say 0.05 and 0.1, would have changed the EIA considerably and the decision-making process of the fixed link (Gullett 2000).

The setting of an acceptable effect size Δ is mostly a political question, however it can occasionally be scientifically possible to determine an ecologically acceptable or sustainable effect size. In a power analysis, this means that the property of Δ could at least start an operationalization of otherwise somewhat blurry politically defined goals and criterions, like zero effect, sustainability, ecologically acceptable effects, biodiversity by *a priori* discussing in quantitative terms the acceptable effect sizes, which would then influence the power of a test and thus the design of the investigation testing causal linkage and null hypothesis.

5 Discussion and Conclusions

There must be a difference between precaution and prevention, in regulatory management of uncertainty and the lack of knowledge in order to reduce the 'regulator's dilemma' and the delay in decision and policy-making. This means that the interpretation that the EU-Commission presented in the communication, where there was no significant dissimilarity between prevention and precaution, needs revision for the precautionary principle to have any relevance at all. The most important distinction between precaution and prevention would be that prevention is concerned with the prevention of relatively certain risks and precaution, on the other hand, is concerned with the prevention of relatively uncertain risks. The important change that follows the precautionary concept in policy and decision-making is that they should not be based on the assumed certainty of the certainty of environmental knowledge, but on the certainty of the uncertainty of environmental knowledge. In other words, the regulation should to a greater degree be based on the management of uncertainty, and risk assessments should explicitly present and discuss related uncertainty.

Moreover, the reversal of burden of proof and workload, is also an implicit possibility in power analysis requirements by the regulators to the industry. At the same time, the demands concerning level of acceptable risk of adverse effect and uncertainty are made quantitative and explicit. The setting of an ecologically acceptable effect (Δ) is a political and scientific issue of carrying capacity. Determining the relevant size of effects is not an easy matter. Consensus must be reached about how much impact would be critical for continued functioning of the affected system at the appropriate spatial scale. The systems should be protected from loss of biodiversity, disruption of food webs and loss of integrity in a precautionary context. Science's role in this could be by quantitatively estimating the carrying capacity, recovery time and ecological relevance of the system's inhabitants. This new, explicit and quantitative information on Δ could then be discussed in the consensus forums (industry, stakeholders, regulator, scientists, NGOs, etc.) and submitted to the power analysis before sampling for environmental effects. Moreover, the setting of acceptable risk of a Type II error (β) could be guided by setting α = β + ethics, where the societal internalities and externalities are equal, and could also be decided in the consensus forums. Theoretically, this would go along way in securing Beck's (1992) plea for reflexive environmental science and could be a beginning to avoid an *un*-precautionary hypothesis testing. Mapstone (1996) suggests a new and interesting four-step approach in setting α, β and Δ for new decision rules in environmental impact monitoring programs to negotiate the singularity emphasized by the tyranny of α or $p < .05$ in ecology that everyone uses. The setting of the entities is based on ecological and economic issues and focus on estimating β for the further decision-making (Mapstone 1996).

Power analysis, of course, does not solve all problems concerning lack of knowledge and uncertainty in ecological risk assessments. It can test the quality of the statistical testing of the null hypothesis in quantitative terms, but it can never say anything about the quality of the relationship under investigation. This is the dark realm of probability for Type III errors (wrong question – accurate answer). For example: It is statistically possible *e.g.* in biomonitoring programs to

find significant relationships between two parameters that have no causal linkage whatsoever, like the number of TV sets per capita and abundance of seaweed in the Baltic Sea. The quality of the causal relationships tested is rarely evaluated or questioned because this is not a metric entity but a result of the qualitative scientific development of, in this case, biology and ecology. However, a major challenge for modern biology is to break out of the constraint of obsession with Type I error. This will involve much more thought about biological and ecological processes operating in nature, so that better models and more structured predictions can be made through increased attention to the power of experiments. Finally, better quantification of a predictive hypothesis will go a long way towards answering the criticism raised by Peters (1991) in his accurate reflection on the inadequacies of much of modern, practical ecology and ecotoxicology (Underwood 1997). Bodansky (1991) accuses the precautionary principle for being vague, however law is also full of 'vague' principles (reasonable man, good faith, etc.) needing interpretation in a concrete context, which is the banister's job. There is in principle nothing wrong in this. The communication between scientists and lawyers raises a problem in relation to juridical praxis, science and precaution, because of the misinterpretation of certainty and precision of complex environmental data. This has led Peters (1991) to accuse ecology studies that it suffers from very low normal-scientific (Kuhn 1962) status as it is not able to predict events, where physics with Newton's laws, etc. have a high normal-scientific status. However, we don't feel that this comparison does ecology justice. Since Heisenberg's uncertainty theorem and Einstein's and Bohr's early nuclear physics achievements in the 1920s, physics cannot answer deep complex quantum mechanical questions with certainty, e.g. give a full and certain explanation of an everyday phenomenon such as gravity. Moreover, biology is confounded by self-organizing organisms lead by their historicity imbedded in their DNA which hampers prediction of long term temporal changes due to the simple fact that they seem random or chaotic impeding their computation. So when you ask a complex question, as almost all ecology and environmental ones are, you can't and should not expect precise answers referring to a simplistic, rudimentary and positivistic notion of science dating more that eighty years back in time. In this light, the gap between science and precaution is not necessarily significant. Moreover, the objectivity, public opinion, time constraints and possible outcomes of failures, and of environmental science, are also significantly different from other sciences. Underwood (1995) has analyzed the different types of research related to environmental decision-making. So that their relationships and purposes can become clearer, he defined four major types of research. Some of the primary problems were that ecologists are excessively reactive and not proactive in the use of their findings and defining of problems and solutions, often because researchers are not setting the research agendas (Underwood 1995). In relation with the risk assessment of chemicals or environmental impact assessments, as far as they are well defined with standard guideline techniques, there is not a matter of changing science to policy-making under the influence of precaution, rather it is a matter of changing the technology. Guidelined risk assessments are not so much a science as a technology, which can be changed to suit our needs much more rapidly than epistemic changes of *Science* without profoundly violating our history and culture since the renaissance.

The role of a precautionary science would also be to include non-significant results as early warnings, which are not proven beyond any doubt but beyond reasonable doubt. Science should first and foremost protect humans and the environment, and secondly be guardians of truth, and not visa versa. This means that statistical, non-significant results become highly significant as early warnings for a policy based on the precautionary principle. Waiting to report effects until the risk of committing a Type I error is sufficiently low (5% convention) will often work contrary to a precautionary strategy because it asks too much of environmental data. Indeed keeping silent until proven beyond any doubt and not reporting the inherent uncertainty in data makes the knowledge of little use society paying for the science. The well-established practice of regarding data, not rejecting the null hypothesis, as a support for the null hypothesis will in fact increase environmental risk (Buhl-Mortensen & Welin 1998). But why has there been so little attention to Type II error compared with Type I error within environmental science both qualitatively and quantitatively? Some of the answers would include the lack of knowledge among scientist and the lack of computer-power, although this is not a valid answer for the past decade. Another would be societal rationality and politics, which sustainable development based upon the precautionary principle challenges.

A long-term precautionary principle could also shift the focus away from uncertainty associated with risk assessment of environmental or health risks towards analysis of benefits and necessity of a new chemical or construction before leasing the product, this could be referred to as an analysis of necessity opposed to an analysis of risks. Analysis of societal necessity would more truly reflect a precautionary principle, than risk assessments more or less openly primarily defending the right to market.

The mathematically driven apparatus of modern risk management contains the seeds of a dehumanizing and self-destructive technology. Our knowledge of the way things work, in society or in nature, comes trailing clouds of vagueness. Vast ills have followed a belief in certainty (Arrow 1992). In the process of breaking free from the past we may have become slaves of a new religion, a creed that is just as implacable, confining, and arbitrary as the old? (Bernstein 1996). This is some of the background for the 'regulator's dilemma' (Weinberger 1985) and the possible need for a precautionary principle different from the prevention principle. The implementation of the precautionary principle should rely on the estimation of costs and benefits of the activity being assessed and the socioeconomic cost-benefit involves risk management analysis of BAT (Best Available Technology) and ALARA (As Low As Reasonably Achievable). After this analysis it is possible to determine the level of precaution, and thus risk of committing Type II error (β), and thus the required power of the test ($1-\beta$). The estimation of β, power,

the acceptable effect sizes (Δ) and/or number of replicates, is a fairly simple statistical exercise on an ordinary computer with a standard statistical package.

Acknowledgements. Ulf Kjellerup and Jette Rank (both University of Roskilde) for reading and improving the paper with comments, and to Andrew Crabtree for corrections. And finally to the high quality comments by the reviewers.

References

Arrow KJ (1992): I know a hawk from a handsaw. In: Szenberg M (ed) Eminent Economists: Their life and philosophies. Cambridge & New York: Cambridge University Press, pp 42–50

Beck U (1992): Risk society. Sage Publications, London, UK

Bernstein PL (1996): Against the gods: The remarkable story of risk. John Wiley & Sons, Inc, New York, USA

Bichel S (ed) (1998): Rapport fra Hovedudvalget, Miljøstyrelsen www.mst.dk/udgiv/publikationer/1999/87-7909-296-9/html/helepubl.htm – 101k. Bichel-udvalget,Copenhagen, Denmark

Bodansky D (1991): Scientific uncertainty and the precautionary principle. Environment **33**: 4–5 & 43–44

Bratt C (1996): Når tvivlen skal komme miljøet til gode. Tidsskrift for samfunnsforskning **2**:199–224

Buhl-Mortensen L, Welin S (1998): The ethics of doing policy relevant science: The precautionary principle and the significance of non-significant results. Science and Engineering Ethics **4**: 401–412

Carey J (2000): Personal communication 3rd SETAC World Congress, Brighton, UK, 21–25/5-2000

EU COM (2000): Communication from the European Commission on the precautionary principle (COM 2000, **1**, 02–02-2000). Brussels: EC. Available on the World Wide Web: http://europa.eu.int/comm/off/com/health_consumer/precaution_en.pdf

EU COM (2001): Full text of the White Paper on the Strategy for a future Chemicals' Policy (COM 2001, **88**) (pdf~200K; except el 1, 200K). Stakeholders' Conference. europa.eu.int/comm/environment/chemicals/whitepaper.htm – 11k

Danish Environmental Ministry (2000): Pesticidhandlingsplan II. Copenhagen, Denmark. www.mst.dk/udgiv/Publikationer/2000/87-7944-262-5/html/helepubl.htm – 42k

Feldman HL (1995): Science and uncertainty in mass exposure litigation. Texas Law Review **74**: 24–48

Freestone D, Hey E (1996): Origins and development of the precautionary principle. In: Freestone D, Hey E (eds) The Precautionary Principle and International Law (pp 3–15) The Hague, Kluwer Law International, Holland

Gee D (2000): Personal communications at the chemicals conference 'Chemicals under the spotlight' in Copenhagen 27–28/10-2000

Gray JS (1990): Statistics and the precautionary principle. Marine pollution Bulletin **21**: 174–176

Green RH (1979): Sampling design and statistical methods for environmental scientists. John Wiley & Sons, New York, New York, USA

Gullett W (2000): Environmental decision-making in a transboundary context: Principles and challenges for the Denmark-Sweden Øresund fixed link. Journal of Environmental Assessment and Management 2 (3). In press

Fisher E, Harding R (eds) (1999): Perspectives of the precautionary principle. The Federation Press, Sydney, Australia

Hey E (1992): The precautionary concept in environmental policy and law: Institutionalizing caution. Georgetown International Environmental Law Review **4**: 303–318

Kuhn T (1962): The structure of scientific revolutions. Chicago University Press, USA

Majewski C (2000): Personal communication. 3rd SETAC World Congress in Brighton, UK, 21–25/5-2000

Mapstone BD (1995): Scalable decision rules for environmental impacts studies: Effect size, Type I and Type II errors. Ecological Applications **5**: 401–410

Nice Treaty (2001): Treaty of Nice File Format: PDF/Adobe Acrobat – View as Text Journal of the European Communities C 80/1 'Treaty of Nice Amending the Treaty on European Union, the Treaties Establishing the European Communities and Certain'. Description: Text of the Nice Treaty (PDF format) europa.eu.int/eur-lex/en/treaties/dat/nice_treaty_en.pdf

O'Riordan T, Jordan A (1995): The precautionary principle in contemporary environmental politics. Environmental Values **4**: 1991–2012

Peterman RM (1990): Statistical power can improve fisheries research and management. Canadian Journal of Fishery and Aquatic Science **47**: 2–15

Peterman RM, M'Giongle M (1992): Statistical power analysis and the precautionary principle. Marine Pollution Bulletin **24**: 231–234

Peters RH (1991): A critique for ecology. Cambridge University Press, Cambridge, UK

Raffensberg C, Tickner J (eds) (1999): Protecting public health and the environment: Implementing the precautionary principle. Island Press, Washington, DC, USA

Rehbinder E (1994): The precautionary principle in an international perspective. Pages 91–105. In: Basse EM (ed) Miljørettens grundspørgsmål, bidrag til en Nordisk forskeruddannelse. Gad, København, Denmark

Sanderson H (2000): Håndtering af forsigtighedsprincippet. Vand & Jord 3:88–90

Sanderson H (2001): Replicability and interpretation of micro/mesocosm experiments. ESPR – Environ Sci & Pollut Res–OnlineFirst [DOI: http://dx.doi.org/10.1065/espr2001.09.085]

Sanderson H, Petersen S (2001): Statistical power analysis of data from quasi-natural mesocosm ponds treated with $Roundup_{2000}$. Environmental Toxicology and Chemistry. In press

Santillo D (2000): A precautionary tale: Report from the precautionary principle debate. Third SETAC World Congress, Brighton, May 2000. SETAC Globe **1**: 17–20

Sokal RR, Rohlf FJ (1995): Biometry. 3rd edition, WH Freeman and company, New York

United Nations Environment Programme (UNEP) (1992): Rio declaration on environment and development. Made at the United Nations Conference on Environment and Development, Rio de Janeiro, Brazil. Available on the World Wide Web: http://www.unep.org/Documents/Default.asp?DocumentID=78&ArticleID=1163

Underwood AJ (1995): Ecological research and (and research into) environmental management. Ecological applications **5**: 232–247

Underwood AJ (1997): Experiments in ecology: Their logical design and interpretation using analysis of variance. Cambridge University Press, UK

Weinberger AM (1985): Science and its limits: The regulators dilemma. Issues in Science and Technology Fall:59–72

Received: August 5th, 2001
Accepted: October 18th, 2001
OnlineFirst: October 20th, 2001

VII

Report: International summit on science and the precautionary principle

by

Sanderson H. 2002

Journal of Environmental Science And Pollution Research International

Conference Reports

International Summit on Science and the Precautionary Principle

Lowell Center for Sustainable Production, University of Massachusetts Lowell, USA September 20–22, 2001

Hans Sanderson

Department of Environment, Technology and Social Studies, Building 12.2, University of Roskilde, Postbox 260, DK-4000 Roskilde, Denmark; e-mail: hanss@ruc.dk

Background

The Summit was hosted by the Lowell Center for Sustainable Production, University of Massachusetts Lowell, USA and organized by Dr. Joel Tickner and Sara Wright, and brought together a diverse group of 75 scientists, lawyers, policy analysts and advocates from some 18 countries. The aim was, and still is, to identify ways that environmental science can be conducted to more effectively support precautionary and preventive decision-making, particularly in the face of complex, highly uncertain human and ecosystem health risks. The goals of the Summit included: identifying and illustrating the ways in which scientific methods can either support or limit precautionary decision-making, to build consensus on changes needed in the practice application of science to better support the precautionary principle, and to build a base of support in the scientific community for these changes. Two practical and ongoing outcomes are expected within the near future; a) an edited volume of essays from the Summit and b) a Summit statement providing a vision for science that supports the precautionary principle (see www.uml.edu/centers/lcsp for more information). The Center's staff intends to conduct an outreach with government authorities, academic scientists, and professional organizations throughout the world to discuss the ways in which science can more effectively support precaution.

Science is often mistakenly viewed by policy-makers as an incontrovertible source of knowledge on which to base policy decisions. However, in the context of complex environmental risks, it is much more useful to think of science and policy as dynamically informing each other – science provides critical information on which to base policy and public policy outlines critical societal research and knowledge needs. Lubchenco (1995) suggests a list of characteristics of good scientific communication with policy. In conducting environmental research, scientists should specify:

1) What is known;
2) The certainty with which it is known;
3) What is not known;
4) What is suspected;
5) The limits of the science;
6) Probable outcomes of different policy options;
7) Key areas where new information is needed;
8) Recommended mechanisms for obtaining high-priority information.

Much of the recent debate about the precautionary principle has focused on the questions of whether precaution poses a barrier to trade and of what specific level of evidence is sufficient to act to prevent harm? When the precautionary principle is discussed in the context of its relationship to science, it is often portrayed either as anti-science or as a risk-management principle that is implemented only after objective scientific enquiry takes place. The latter is the case in the European context (EU-COM 2000/1). Both these views are controversial or incorrect. There are ways in which the methods of scientific inquiry often implicitly impede precautionary action, making it more difficult for policy-makers to take action in face of uncertainty (see Kriebel and Tickner 2001 and Kriebel, Tickner and Epstein 2001). Too often scientific research focuses on narrowly defined issues (Sanderson and Petersen 2001); while the problems we face are complex, and require interdisciplinary research methods. Current scientific practice also often attempts to minimize uncertainties, and focus on those aspects of a problem that are quantifiable (Sanderson and Petersen 2001). The Summit group believes that if the precautionary principle is presented to environmental scientists as an opportunity for more and better science, we may find support from researchers who are presently unaware of such developments, or even hostile to a perceived 'attack' on science. Scientists are also needed to respond to critiques of precautionary decisions, particularly when the uncertainties in science are misrepresented. Underwood (1995) accurately addresses the issue of interpretation of uncertainty associated with environmental health risk assessments. Uncertainty is an inevitable conclusion of ecological investigations, and indeed. Physicists have claimed to deal with uncertainty in all their science. This has not caused them to be labeled

incompetent or inadequate. It is important not to mix up uncertainty and quality in science. Ecologists should not be so defensive of the uncertainties that shroud the results of ecological investigations. They (the knowledge of uncertainty) are the best results we are going to get. There is a need for scientists to be more proactive in scientific definition of the research issues and not alone reactive to management and funding-based questions. Development of better methods of ecological investigation into matters of environmental management is long overdue (Underwood 1995).

Conclusions

The Summit statement addresses the points discussed below – and is forthcoming. The world cannot be risk-free, but science and policy can more effectively be used to prevent damage to health and ecosystems, as well as to help reach societal goals and make progress towards a healthier and sustainable future. Applying the precautionary principle can foster innovation in materials, products, and production processes. The goal of precaution is to prevent harm – not progress – and support a sustainable future. Moreover, there a need to find out where science ends and technology start was identified, because technology is much more readily changeable than science end epistemology. Since the question about science and precaution is a very large and broad question, there is a need to divide the role of science in precaution in a short-term (technological) and long-term (epistemic) perspective.

1. **Short-term and epistemic readily technological actions:**

   In this process, Lubchenco's (1995) list of eight questions should be addressed in each environmental investigation. Moreover, Underwood's (1995) reclaim of scientific problem formulation should be noted (reduce risk of Type III error (wrong question – accurate answer)). Research methods and questions should include whole systems, interactions and cumulative causal factors, preferably in interdisciplinary collaborations. Scientists should develop better methods of hazard surveillance, and systems for identifying early warnings, plus expand their focus to preventive opportunities. They should increase transparency and public, stakeholder's and laymen's knowledge of participation in defining the acceptable probabilities of Type I and II error and the acceptable human impact (α, β, & Δ) (Sanderson and Petersen 2001).

2. **Long-term and theoretically not readily scientific actions:**

   First of all, an environmental precaution approach should be the normal state and not something special. We need to shift the focus of research from how much can we pollute with no apparent or detectable effects with a certain amount of uncertainty towards the development of analysis of necessity – who and how much does this product benefit consumers and/or the environment compared to existing similar products? This question collides with two essential rights a) the right to produce and market products and b) the right to good and safe environmental and human health. There is a need to rethink scientific objectivity towards caring. As mentioned above, more holistic analysis, greening of technology and chemistry, better mutual understanding and communication between the public, politician's, administrator's, lawyer's, laymen's knowledge and scientists are needed. Changing of research funding strategies from military research towards sustainable science and research and equity on a global scale. These considerations are appropriate, not the least after the tragic September 11th, and thus, rethinking of defense systems in light of terror. More research into failures to and acceptance of uncertainty, post-normal science (Funtowicz and Ravetz 1993), not seeking a 'magic number' with a neglected uncertainty for regulation.

These were some of the reflections on science's role in a policy under the precautionary principle, in the continuing process of pursuing a sustainable development via sound science and precaution.

Acknowledgements. Lowell Center for Sustainable Production, University of Massachusetts, Lowell, USA, for sponsoring the Summit and Dr. Joel Tickner and Sara Wright and the rest of Center for Sustainable Production (UMass, Lowell) for organizing it.

References

EU COM (2000): Communication from the European Commission on the precautionary principle (COM 2000, **1**, 02-02-2000). Brussels: EC. Available on the World Wide Web: http://europa.eu.int/comm/off/com/health_consumer/precaution_en.pdf

Funtowicz S, Ravetz J (1993): Science for the post-normal age. Futures **25**: 735–755

Kriebel D, Tickner J (2001): Reenergizing public health through precaution. American Journal of Public Health **91**: 1351–1354

Kriebel D, Tickner J, Epstein P (2001): The precautionary principle in environmental science. Environmental Health Perspectives. In Press

Lubchenco J (1995): The role of science in formulating a biodiversity strategy. BioScience supplement: s7–s9

Sanderson H. Petersen S (2001): Power analysis as a reflexive scientific tool for interpretation and implementation of precautionary principle in the European Union. ESPR – Environ Sci & Pollut Res-OnlineFirst [DOI: http://dx.doi.org/10.1065/espr2001.10.095]

Underwood AJ (1995): Ecological research and (and research into) environmental management. Ecological Applications **5**: 232–247

VIII

SETAC position paper on the scientific application of the precautionary principle

by

Sanderson H. G. Biddinger, J. Ravetz, J. Tickner. 2002

Manuscript, summary submitted to the *SETAC Globe*,
whole text to be posted on the SETAC ERAAG webpage

SETAC position paper on scientific application of the Precautionary Principle

Hans Sanderson*1, Jerry Ravetz‡, Joel Tickner§, Gregory R. Biddinger+. *)Department of Environment, Technology and Social studies, building 12.2. University of Roskilde.Postboks 260. 4000 Roskilde. Danmark. Tel: (+45) 46 74 24 96. Fax: (+45) 46 74 30 41. E-mail: hanss@ruc.dk. ‡)106 Defoe House, Barbican. London, UK. §) Department of Work Environment, Lowell Center for Sustainable Production, University of Massachusetts Lowell. One University Ave. Lowell, MA 01854 USA. +) Downstream - Safety, Health & Environment, ExxonMobil Refining & Supply Company, 3225 Gallows Road 8B-914, Fairfax, Va 22037 USA.

Growing awareness of the scale of human impacts on health and environment has led to a recognition of the need to change the ways in which environmental protection decisions are made, and the ways that scientific knowledge informs those decisions. Since the Rio meeting on sustainable development in 1992 the precautionary principle has been written into many different national and international treaties and conventions. The precautionary principle/approach can be seen as a government's tangible commitment to the importance of social values such as health, safety, the environment and natural resources conservation. Principle 15 in of the 1992 Rio Declaration on environment and development states that:"… lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation".

There are two dominating critiques of the precautionary principle, first: Much of the recent debate about the precautionary principle has focused on the questions of whether precaution poses a barrier to trade, and what is sufficient scientific evidence to implement a precautionary principle? And secondly: When the precautionary principle is discussed in the context of its relationship to science, it is often described either as anti-science or as a risk-management principle that is implemented only after objective scientific enquiry takes place. The latter is the case in the European context (EU-COM, 2000 (1)).

The reason for implementation of a precautionary principle is that while scientific information is still inconclusive, decisions will have to be made to meet society's expectations about living standards and to address risks. The scientific process is, and should be, almost always characterised by uncertainty and debate, which is consistent with Sir Karl Popper's scientific falsification theory (1968). The challenge to SETAC members and other environmental scientists and society is to determine what is sufficient scientific certainty to implement a precautionary approach and furthermore, how to achieve a scientific application of precautionary approaches within research?

* To whom correspondence may be addressed: hanss@ruc.dk

We know from experience that the public has a low tolerance for serious or irreversible harm characterised by scientific uncertainty, thus a different approach to public engagement is required – greater transparency and increased public involvement in decision-making (Environment Canada, 2001). Hence, determination of the society's chosen level of protection, acceptable adverse risks and effects should be determined in advance, when ever possible on a case by case basis.

In case of severe uncertainties and risks (the long term implementation of the precautionary principle) post-normal scientific considerations. When the inherent uncertainties, lacking knowledge and multi-dimensionality of risks require that risk management policy should be achieved through a dialogue among concerned interests and stakeholders. The theory of post-normal science has been developed for the comprehension of such situations, which frequently occur when the precautionary principle is invoked. For this theory, the typical case is when facts are uncertain, values in dispute, stakes high and decisions urgent. Then, the tidy, controlled world of the scientific laboratory or the controlled field trial gives only a partial insight into the reality being discussed, and this might actually be misleading. It is for this reason that the quality of the procedures of inquiry depends on an 'extended peer community', including all with a concern for resolving the issue. Also known as 'extended peer review', this conclusion of post-normal science is finding increasing acceptance wherever these contentious issues are being discussed. It could be said that the theory of post-normal science is the essential foundation for the realisation in practice of the precautionary principle in science-related policy issues in light of epistemic uncertainties (Funtowicz and Ravetz, 1994). Moreover, an analysis of societal needs and benefits could be considered as risk management tool in the extended peer review.

Scientists need to respond to critiques of precautionary decisions, particularly when the uncertainties in science are misunderstood. Underwood (1995) accurately addresses the issue of interpretation of uncertainty associated with environmental health risk assessments. Uncertainty is an inevitable conclusion of ecological investigations, and indeed of science. Physicists have claimed to deal with uncertainty in all their science. This has not caused them to be labelled incompetent or inadequate. Hence, it is important not to mix up uncertainty and quality in science. Ecologists should not be so defensive of the uncertainties that shroud the results of ecological investigations. They (the knowledge of uncertainty) are the best results we are going to get, Underwood argues. There is in other words, a need for scientists to be proactive in scientific definition of the research issues and not merely reactive to management and funding based questions. Development of better methods of ecological investigation into matters of environmental management is long overdue (Underwood, 1995).

An effective application of the precautionary principle requires interdisciplinary scientific research, as well as explicitness about the uncertainties involved in this research and its findings. Precautionary decision-making is consistent with sound science because of the large areas of uncertainty and even ignorance that persist in our understanding of complex biological systems, in the interconnectedness of organisms, and in the potential for interactive and cumulative impacts of multiple hazards. Because of these uncertainties, science will sometimes be incapable of providing clear and certain answers to important questions about potential environmental hazards. Waiting for incontrovertible scientific evidence of harm before preventive action is taken can increase the risk of costly mistakes that can cause serious and irreversible harm to ecosystem and human health and well being, and the economy (Lowell statement, 2001).

Some of the ways that scientific information is currently applied in setting policy can work against the ability to take precautionary action, for example by misrepresenting limitations in the state of scientific knowledge. Decision-makers frequently look for high levels of proof of causal links between a technology and a risk before acting, so that their decisions will be protected from accusations of being arbitrary. But often, high levels of proof cannot be achieved, and are not likely to in the foreseeable future. A more complete and open presentation on the part of scientists of the current limitations in understanding of environmental risks will encourage the acceptance on the part of government decision-makers and the public of the idea that precautionary action is a prudent and effective strategy when potential risks are large and uncertainties are large as well. There is a need to improve communication and understanding among scientists and policy makers and lawyers.

Currently some methods of scientific inquiry may occasionally retard precautionary action. For example, research frequently focuses on narrow, quantifiable aspects of problems, thus inadvertently excluding from consideration potential interactions among different components of the complex biologic systems of which humans are a part. The compartmentalisation of scientific knowledge further impedes the ability of science to detect and investigate early warnings and develop options for preventing harm when far-reaching health and environmental risks are involved. Unfortunately, limitations in scientific tools and in the ability to quantify causal relationships are often misinterpreted by government decision-makers, scientists, and proponents of hazardous activities as evidence of safety. However, not knowing whether an action is harmful is not the same thing as knowing that it is safe. An effective implementation of the precautionary principle demands improved scientific methods, and a new interface between science and policy that stresses the continuous updating of knowledge as well as improved communication of risk, certainty, and uncertainty (Lowell statement, 2001).

There is a need for a more effective linkage between research on hazards and expanded research on prevention and restoration. Increased use of interdisciplinary approaches in science and policy, including better and transparent integration of qualitative and quantitative data. Human activities cannot be risk-free. The goal of precaution is to prevent harm, not to prevent progress. Applying precautionary policies can foster innovation in better materials, safer products, and alternative production processes. It can bring benefits beyond the reduction of health and environmental impacts, stimulating both more innovation, via technological diversity and flexibility, and better science. But over-precaution can also be expensive, in terms of lost opportunities for innovation and lost lines of scientific enquiry (EEA, 2002).

Science is often mistakenly viewed by policymakers as an incontrovertible source of knowledge on which to base policy decisions. However, in the context of complex environmental risks it is much more useful to think of science and policy as dynamically informing each other – science provides critical information on which to base policy and public policy outlines critical societal research and knowledge needs. Lubchenco (1995) and the Lowell statement (2001) suggests a list of characteristics of good scientific communication with policy. In conducting environmental research, scientists should specify:

1. What is known.
2. The certainty with which it is known.
3. What is not known.
4. What is suspected.

5. The limits of the science.
6. Probable outcomes of different policy options.
7. Key areas where new information is needed.
8. Recommended mechanisms for obtaining high-priority information.

The European Environmental Agency (EEA) furthermore released twelve late lessons from early warnings considering precautionary approaches and science from 1896-2000. The 12 "late lessons" are:

1. Acknowledge and respond to ignorance, as well as uncertainty and risk, in technology appraisal and public policy-making.
2. Provide adequate long-term environmental and health monitoring and research into early warnings.
3. Identify and work to reduce blind spots and gaps in scientific knowledge.
4. Identify and reduce interdisciplinary obstacles to learning.
5. Ensure that real world conditions are adequately accounted for in regulatory appraisal.
6. Systematically scrutinise the claimed justifications and benefits alongside the potential risks.
7. Evaluate a range of alternative options for meeting needs alongside the option under appraisal, and promote more robust, diverse and adaptable technologies so as to minimise the costs of surprises and maximise the benefits of innovation.
8. Ensure use of "lay" and local knowledge, as well as relevant specialist expertise in the appraisal.
9. Take full account of the assumptions and values of different social groups.
10. Maintain regulatory independence from interested parties while retaining an inclusive approach to information and opinion gathering.
11. Identify and reduce institutional obstacles to learning and action.
12. Avoid "paralysis by analysis" by acting to reduce potential harm when there are reasonable grounds for concern (EEA, 2002).

Moreover, there is need for SETAC members and other environmental scientists, administrators and layers to consider where science ends and scientific technology and practice starts. To shortly epitomise this environmental science is this light is concerned with exploring new knowledge, by using and developing theories and techniques. Technology of science is using established methods and refining of these. And finally, practice is using established scientific methods and guideline approaches. Scientific technology and practice is much more readily changeable than science and epistemology. Since the question about science and precaution is a very large and broad question, there is a need to divide the role of science in precaution in a short term (technological) and long term (epistemic) perspective.

1. Short term methodological actions: In this process Lubchenco's (1995) list of eight questions should be addressed in each environmental investigation. Moreover, Underwood's (1995) reclaim of scientific problem formulation scientific research/should be noted (reduce risk of Type III error (wrong question – accurate answer)). Research methods and questions should include whole systems, interactions and cumulative causal factors, preferably in interdisciplinary collaborations. Scientists should develop better methods of hazard surveillance, and systems for identifying early warnings plus expand their focus to preventive opportunities. Decision-makers should increase transparency and public, stake-holders and laymen knowledge participation in defining the

acceptable probabilities of Type I and II error and the acceptable human impact (α, β, & Δ) (Sanderson & Petersen, 2001). Acceptable Type I and II error levels and effect sizes (Δ) should be determined in advance based on ecologically acceptable simplification of the ecosystem (*i.e.* the recovery time, redundancy and carrying capacity of the system), ethics, cost-effectiveness (internalising the external cost of the product) and public risk acceptability. These entities can be analysed by statistical power analysis producing data based on a common precautionary approach. This will help the scientist designing his or her experiment and testing of hypothesis. Moreover, probabilistic risk assessments and species sensitivity distributions could also guide the prioritisation process, however, there is still need for more research on the species sensitivity, physiology, biology and ecology.

2. Long term epistemic actions: In this process the twelve late lessons from the EEA mentioned above should be carefully be considered. An environmental precautionary approach should be the normal state and not something special. This could involve a shift in focus of research from how much can we pollute with no apparent or detectable effects with a certain amount of uncertainty towards development of analysis of necessity and alternatives – who gets benefits and who incurs costs, and by how much, consumers and/or the environment compared to existing similar products? There is a need to rethink scientific objectivity, how it can be used to enhance caring for people and the environment and still be scientifically objective. Better mutual understanding of constrains on and between the public, politicians, administrators, lawyers, laymen's knowledge and scientists are needed. More research into deficiencies in the management of uncertainty, by scientists and by policy-makers and awareness of post-normal science (Funtowicz and Ravetz, 1994).

The short-term changes of techniques, technology and practices are from a scientific point of view more or less readily applicable, however, they still need some refinement, in terms of defining the methods and context of assessing environmental and public health externalities and public involvement need further research. On the long term perspective there is a need to update the notion of the present scientific paradigm in society and not the least in relation to law and lawsuits. Science is by nature uncertain, otherwise it would cease being science and become dogmas, preventing this was part of the reason behind the reformation and foundation of modern science. Physics is often used as the scientific yardstick of Kuhnian normality (Kuhn, 1962) in this light environmental sciences often comes out weak and insufficient. However, this comparison is, perhaps, not permissible because it ignores the introduction of quantum mechanics by Bohr, Heisenberg and Schrödinger in the 1920s. Heisenberg's uncertainty principle implies that there is a built-in uncertainty in the Universe. It is possible for something to be created out of nothing, given enough time. Just recently, Greiner et al., (2001) found a major unexplainable theoretical logjam studying a massive stellar black hole. Hence, physics is not just a simple science with clear-cut causalities, but complicated, uncertain and at times chaotic just like most other sciences including environmental sciences. Sound science can not always deliver evidence which is particular important to realise also in a legal context.

Finally, we must not forget that we apply precautionary approaches all the time in our daily lives. Thus we deal with uncertianty when we make investments, take out insurance, arrange health care, deciding whether to carry

an umbrella or even choosing a life-companion. Such examples remind us that it is not only uncertainty that is involved; the costs and benefits of the various possible outcomes influence our choice but even our way of setting up the problem. This also happens when decisions need to be made on issues of safety and the environment. Then both the conflicted value-commitments and the conflicted perceptions of the issues are relevant. If this duality is not recognised, then the debate can be mislead.

These approaches would themselves not remove the dilemmas of decision-making under situations of uncertainty and high stakes nor would they eradicate uncertainties or avoid the consequences of ignorance. But they would at least increase the chances of anticipating costly impacts, of achieving a better balance between the societal pros and cons, costs and benefits/needs, of technological innovations and of minimising the costs of unpleasant surprises in the future.

References

EEA, 2002. European Environmental Agency. Late lessons from early warnings: The precautionary principle 1896-2000. Environmental Issue report #22. Copenhagen, Denmark. http://org.eea.eu.int/documents/newsreleases/Newsrelease-10012002-en and http://reports.eea.eu.int/environmental_issue_report_2001_22/en with the report.

EU COM (2000)1: Communication from the EU-Commission, on the precautionary principle. Brussels, 2/2-00. European Commission (EC). (2000). Communication from the Commission on the precautionary principle (COM (2000) 1, 02-02-2000). Brussels: EC. Available on the World Wide Web: http://europa.eu.int/comm/off/com/health_consumer/precaution_en.pdf.

Funtowicz S. & J. Ravetz. 1994. Uncertainty, complexity and post-normal science. Environmental Toxicology and Chemistry 13:1881-1885.

Greiner, J., Cuby, J.G., McCaughrean, M.J., 2001. An unusually massive stellar black hole in the Galaxy. Nature, 414:522-525.

Kuhn, TS. 1962. The structure of scientific revolutions. University of Chicago Press, USA.

Lowell statement, 2001. http://www.uml.edu/centers/LCSP/precaution/.

Lubchenco, J. 1995. The role of science in formulating a biodiversity strategy. BioScience supplement:s7-s9.

Popper, K. 1968. The logic of scientific discovery. Hutchinson, London, England.

Sanderson, H. & S. Petersen. 2001. Power analysis as a reflexive scientific tool for interpretation and implementation of precautionary principle in the European Union. Environmental Science and Pollution Research International. 8: DOI:http://dx.doi.org/10.1065/espr2001.10.1095.

Underwood, A.J. 1995. Ecological research and (and research into) environmental management. Ecological applications 5:232-247.

Discussion and conclusions

Discussion and conclusions

Summary of results

Whilst each paper discusses a specific question, there is a more general question which was raised in the first chapter and which underlines the entire thesis. What are the utility and trade-offs associated with microcosm studies? In order to answer this, a quantitative demonstration of the trade-offs between statistical and ecological significance of experimental microcosm studies, commenting on the current use of microcosms for higher tier risk assessment of chemicals, was necessary.

As mentioned in the introduction, the aim of the thesis was not to determine a single correct, design but to show how the choice of ecological significance (how well the microcosm model mimics the parent macrocosm model) and the statistical power required are context dependent and difficulty compatible. I tentatively defined, and ranked, the ecological significance of the three microcosm systems that I have been working in as: 25 m^2 outdoor earthen ponds, > 12 m^3 outdoor polyvinyl chloride ponds, > 30L indoor aquariums. The 25 m^2 outdoor earthen microcosm ponds had been allowed to stabilise for three years and were quasi-natural or as close to natural as possible. The 12 m^3 outdoor PVC ponds are assumed to be less realistic than the quasi-natural ponds due to the lack of; banks, naturally occurring perennial macrophytes, wintering of the systems. The extrapolation gap from the PVC ponds to the parent model (a farm pond) was anyway smaller than that for the 30L indoor aquariums. This was primarily due to the fact that the indoor aquariums in addition were not influenced by natural changes in weather and that they were smaller. In addition to the differences in the ecological significance of each design there is furthermore always additional uncertainty related to extrapolation of ecological risks from one ecosystem to another and from pond to pond (Persoone & Janssen, 1994). Hence, depending on the context and research question, a smaller and simpler design could generate results of high generality with respect to certain questions and larger more complex systems could also generate less general results on more complex questions with inherent larger extrapolation uncertainty, as a result (paper I).

Regarding the evaluation of the relative statistical power of the three different designs a true comparison would require working with the same compound, at equal exposure times and concentrations and similar sampling techniques and designs. As mentioned in the introduction, this was not possible within the framework of this thesis. Thus, the comparisons in this thesis serve as illustrative demonstrations, which are dependent upon expert judgement due to inherent unknown extrapolation gaps due to scale, sampling techniques and general maintenance of the microcosms. The power of each design is dependent of the effect sizes the compounds produce, including the within treatment variability and initial pretreatment variability. A replicated design in 30L aquariums with PFOS and PFOA produced markedly different statistical power due to differences in effect sizes (see papers III-IV).

Main results

- In the first paper, the average level of Coefficient of Variation (CVs) in the state-of-the-art review (paper I) for larger outdoor microcosms was 51%. The corresponding average CV for smaller indoor microcosms was 32%.

Animal variables expressed higher variability than plant endpoints. The CV values are context dependent and the CV is inadequate as a general quality criterion. Non-significant effects were abundant - 88% of all biotic endpoints did not yield significant effects and < 5% estimated power.

- The effect size is very dependent on the actual concentration of the compound to which the pelagic plankton community is exposed. This places restrains on the physical design and control of the microcosm, which was evident in the risk assessment of Roundup (paper II). The active ingredient glyphosate was rapidly removed from the water column, resulting in very low exposure levels of the pelagic plankton community. The lack control over the exposure resulted in no adverse effects, with the conclusion that the pelagic plankton community is hardly suited for higher tier risk assessment in aquatic microcosm for testing the ecological risks of Roundup. The required sample size (*n*) range was 14-45 ponds depending on the endpoint.

- The 21-day NOECsurvival of Daphnia magna exposed to 12 mg PFOS L^{-1} in single species standardised laboratory test was not conservative to the zooplankton community compared to the results of the 30L indoor microcosm study. We found a 90-100% reduction (p<0.01) of the total zooplankton population after one week posttreatment at 30 mg PFOS L^{-1} and a similar reduction after two weeks at 10 mg L^{-1} of PFOS (paper III). Required sample size range was 2-5, in the PFOS study and 5-31 in the PFOA study, depending on the endpoint.

- It was notable that the inherent genetic variability in the ASTM standardised test organism *Myriophyllum* sp. was greater than the natural variability induced by the microcosm design (paper V). Furthermore, two to nineteen replicates were required to detect a 25% change in treated microcosms from control depending on the character of the endpoint. A 25% change Δ reduction in growth on perennial macrophytes such as *Myriophyllum spicatum* and *M. sibiricum* would normally constitute an ecologically significant effect. Thus, detection of subtle effects (10%) would require unfeasible numbers of replicates (20-30 microcosms). However, this was primarily due to genetic variation as we found the same low power in previous laboratory data.

- Statistical power analysis could be used as a framework in pursuing a reflexive modernisation of science and as an implementation of a precautionary approach within science. This is however, only applicable in environmental areas dominated by methodological uncertainty. In complex areas dominated by epistemic uncertainty and lacking knowledge, considerations of post-normal science and analysis of need and benefits can be recommended (paper VI-VIII).

- Recent, preliminary and unpublished data (Pers. Comm, T. Boudreau, 2002) indicates that the 30L indoor PFOS study had higher statistical power and greater zooplankton sensitivity than the 12m^3 outdoor PFOS study. However, it seems that the opposite might be the case for the PFOA study, in the repeated small-scale studies. This simply confirms the context dependence of microcosm studies and the uncertainties associated with extrapolation of results to other ecosystems. Each design has its inherent strengths and weaknesses and none is superior to others on all occasions. What is important is to be aware of, and to critically selective, of these strengths and weaknesses *a priori*. Moreover, to realise that no amount of statistical analysis and inter-

pretation can afterwards compensate for an initially poorly designed and inadequately replicated experiment (Kraufvelin, 2000).

Microcosm recommendations

As mentioned in the introduction, I will, based on the work in this thesis, comment on the seventeen recommendations made in the Community Level Aquatic System Studies Interpretation Criteria (CLASSIC, 2002) workshop. These recommendations relate to the regulative assessment of microcosm studies but also to the execution of microcosm studies.

I agree with the basic circumstance that all microcosm studies are context dependent, and harmonised rules for conduction, analysis and interpretation are not possible at present. Thus, expert judgement in interpreting the studies is needed. However, this is not good news for the environmental administrators and regulators, who hope and look for harmonised interpretation models to increase administrative and juridical efficiency (Weinberger, 1985). Furthermore, that both spray and simple mixture application are acceptable and that application preferably should be conducted during spring to observe growth and recovery. Moreover, it is true that the study must be feasible and focus on the most sensitive species (if known). Both structural and functional variables should preferably be included and the recovery time should be decisive for the ecological significance of the effect (Pratt & Cairns, 1996). The Ecological Acceptable Concentration (EAC) should be used to protect the system (HARAP, 1999). If the aim is to provide and protect landscape ecology, in terms of possibilities for migration of animals and inclusion of macrophytes, this should be considered initially and made part of the design. Finally, it is important to determine clear protection aims for the study a priori, which processes or characteristics of the environment is in focus

However, I question a few of the recommendations. First of all we have the use of Principle Response Curves (PRC's) for community data (van den Brink, 1999). As mentioned earlier in the introduction, PRC's have some compelling advantages such as a visually good interpretability of data. The major drawback is however, that it would take serious and protracted training of the risk managers in order for them to understand the mechanics behind the PRC ordinations for full utilisation (Streloke, 2000). This will thus, involve more expert judgement (not less) till the techniques are fully incorporated. Moreover, power analysis is not readily possible on multivariate data sets (van den Brink, 1999). I think it is too early to boldly recommend the use of PRC's. I am not sure if it would be generally welcomed among the risk managers presently.

The question concerning the subject of determining the EAC and whether this should protect the system is in addition difficult, since the protection criteria for the definition of EAC are unclear. Recovery time and redundancy (Pratt & Cairns, 1996) of the ecosystem have been mentioned as criteria (HARAP, 1999). If the recovery time is short, the effects might not have been ecologically significant (Møhlenberg et al., 2001). This is however, dependent upon the life cycles of each organism and thus the season in which they are exposed to the stressor. This in turn influences the resilience and redundant capacity of the system (Pratt & Cairns, 1996). However, if the function of the ecosystem is not effected and the ramification of the structure does not comprise keystone species or in other ways important, sensitive or rare species, this could be considered acceptable according to HARAP (1999). Finally, the treatment concentrations should be based on laboratory tests, field use and environmental concentrations. Often, the environmental concentrations,

the field use concentrations, the laboratory effect concentrations and the detectable effect concentrations in the microcosms will range up to several orders of magnitude. This requires several treatment concentrations and thus several microcosms, which impedes the feasibility of the study.

At the 11th annual meeting of SETAC Europe, 6-10 May 2001, Dr. Lisa Thattersfield presented some interesting results about relative ecological realism. She had compared a cylindrical PVC pond, 1 meter in diameter and 1 meter deep, with rectangular sloping ponds, 4m long and of three different depths (0.1, 0.3 & 0.7m) and with banks and macrophytes. The latter design simulates the natural ponds better than the former. The cylindrical design simulates and represents the relatively deep centre of a pond, whereas the rectangular ponds have better refuge possibilities and presents a more diverse habitat for more species. Hence, maybe most microcosm studies over-represents deep ponds characteristics, and neglect the important catchment areas on the banks in the very shallow waters (paper I)?

Finally, Møhlenberg et al. (2001) produced a literature based guidance report to risk managers on how to interpret results from micro- and mesocosm studies. They applied a multivariate statistical method (PLS, partial least squares) to examine relationships between toxic effects of pesticides and system characteristics such as microcosm design, season and location of study to get an overview of the data in their database. These analyses were then supplemented by more detailed traditional statistics to examine differences in sensitivity, potential for recovery etc. within different taxonomic groups. With the aid of the PLS models they were able to evaluate all the studies in the database on a common basis. In short, their conclusions with direct relevance to this thesis are:

1) Overall, the statistical power in the studies was rather low. The average reduction in abundance of zooplankton exposed to insecticides at recorded significant lowest observed effect concentration (LOEC) was 75% (± 21% SD), which is consistent with the findings in the studies of my thesis.
2) The volume of the microcosm units had no influence on the toxicity of pesticides to zooplankton, while depth significantly influenced the toxicity of insecticides to macroinvertebrates with increasing effects (lower LOEC) at decreasing average depth of the microcosm.
3) Cladocerans (*e.g. Daphnia sp.*) are the most sensitive zooplankton to insecticides followed by copepods (*e.g. Cyclops sp.*) and rotifers (*Rotifera sp.*), which also is consistent with the findings in the studies of my thesis.
4) In microcosm experiments where Cladocerans were severely reduced (*i.e.* by 95%) it took more than twelve to fifteen weeks for full recovery. At reductions below 80% of the initial population size recovery was fast, less than 20 days. For copepods an almost identical initial decrease and recovery was obtained. However, the taxonomic level of the species in the studies was crude (*e.g.* Cladocera).

Given the conclusions in this thesis, and the contemporary economical and logistical realities, it seems that only acute and severe adverse effects can be assessed with most experimental designs that are in use today. Many such effects can be assessed by single species tests as well in the laboratory and to very reduced costs. Then the question is; where does this leave microcosm experiments? A context dependent analysis determining the feasibility of a microcosm study is needed. This should include an *a priori* analysis of the, acceptable levels of α, β and Δ based on societal cost-benefit analysis and

ethics (paper VI). This should be done in consensus between the registrant and the applicant and possible stakeholders. The level of acceptable effect sizes (Δ) could be considered according to the recovery time of the species in question, as well as the resilience and functional carrying capacity of the ecosystems at risk. In this process careful attention should be paid to the fate and hereby the actual exposure of the compound and the detectable effect size of the microcosm design and its statistical power (paper II). Based upon this, the applicant may choose whether or not to include a microcosm study for the risk assessment of a compound. The environmental protection agency (and the foreseen proactive and responsible company) on the other hand should consider implementation of precautionary approaches. Furthermore, considerations of the role and epistemology of environmental science, technology and practices and thus the precautionary approaches are needed both from a scientific, administrative and company point of view (paper VII & VIII).

Perspectives for microcosm studies

There seems to be a dichotomy surrounding the conduction, utility, and economy of microcosm studies. On one hand it is compelling from a risk management perspective to try to reduce the extrapolation gap between single species laboratory tests and the environment, answering the "-So what" question (Levin & Kimball, 1984 & 1985), mentioned in the introduction. On the other hand, the microcosm studies generate data, which are not readily interpretable (Fisher, 1992). At the same time there is an increasing scientific interest in more complicated analyses of the impact of manmade chemicals in the environment, involving not least long-term and subtle effects, secondary and ecosystem effects and effects on multiple species plus synergistic effects (van den Brink, 1999). All these questions can or need analysis in model ecosystems like microcosm experiments. Moreover, microcosm experiments can serve as illustrative models for the difficulties and hurdles for conduct, analysis, interpretation and implementation of environmental science for students. As in most applied ecotoxicology most of the research funds for microcosm experiments come from requests by either companies or environmental protection agencies for risk assessment of a chemical. Core scientific research funds for microcosm studies are more meagre and rare. This leaves us in a situation where the costumer (companies or environmental protection agencies) frequently do not get the service they presumably bargained, for, in terms of more readily interpretable data, because the risk assessment goal of the actual microcosm study might be overshadowed by "more interesting" often urgent and necessary core scientific questions, which can be prerequisites to do the risk assessment in the first place. Gradually the interest and implementation of microcosm studies in risk assessment of chemicals (primarily pesticides) decline, along with the funding for costly experiments. Hence, the scientific community replies by issuing various workshops on recommendations regarding how to conduct analysis and how to interpretation of microcosm studies (AEDG, Wintergreen, Monks Wood, EWOFFT, ECOFRAM, HARAP, CLASSIC).

My conclusion is that microcosm studies are very context dependent and do not easily fit into standardised regulatory frameworks, which impedes the utility of the methods for environmental protection agencies. The USEPA has also arrived at the same conclusion (Fisher, 1992). However, there is still, maybe more than ever, a need for microcosm experiments for training of students, and for the pursuit for answering the more complex questions mentioned above on the actual impact of manmade chemicals in the environment. It's moreover important in the process of designing, executing, analysing,

interpreting and implementing microcosm studies to distinguish between science, technology and practice and between theory, model and real environment. Microcosms are models of the real environment by which we are able to test inherent theoretical assumptions. We do in other words not necessarily test the real environment and in that process it is very difficult to achieve high accuracy, precision and generality of the hypothesis tested and the design by which it is tested. Hence, all models are more or less fortunate modifications of a real parent model, and some are useful if this is recognised and put into the right context despite possible flaws - and should thus not be discarded if they are well designed and executed and serves a common good purpose. If these criterions are not considered and followed a microcosm study can either be a waste of time, effort and money. In a worst-case scenario, a badly designed and executed microcosm study may lead to the erroneous acceptance of potentially high environmental risks due to type II errors.

We assume that single species tests are often more sensitive than microcosm tests and thus relatively more conservative (pers. comm. Prof. Valery Forbes, 2001). However, this may also be a misconception confounded by the generally low power in microcosm studies. Moreover, we do not have solid scientific evidence that the most commonly used test organisms in single species tests are conservative. Neither do we have scientific evidence for the conservative capacity of extrapolation of ecological risks among species, over time or between levels of biological organisation. The same applies for uncertainty factors, or any indications of secondary effects (Persoone & Janssen, 1994). Thus, the "SLOSS" (Single Large Or Several Small) testing discourse should not be an exclusive one because both are needed within ecotoxicology. I would, en passant, suggest the term "FLASS" (Few Large And Several Small) in order to avoid pseudoreplication (Hurlbert, 1984).

Further microcosm research[1] is needed to illuminate the conservatism of extrapolation of ecological risks, which are routinely made in hazard and risk assessments in a number of regulatory agencies. In addition, more microcosm research is needed on development of probabilistic risk assessment, where the microcosms might represent a more scientific approach to the problem of extrapolation of ecological risks and thereby better meet the critique by Forbes & Forbes (1993). However, not many of extrapolation techniques are well based in science and they are neither formalised nor are they necessarily harmonized between countries. Many extrapolations of ecological risks are made in the absence of adequate data, simply in order to account for uncertainty Persoone & Janssen, 1994). Within the EU, a factor of 10-100 is generally applied in the first tier risk assessment procedure (Dir. 91/414/EEC). Basically, extrapolations fall into two broad categories.

1) Range extrapolations in which responses in a single species or system are extrapolated from high to low exposures in the same species or system or vice versa and.
2) Ecological data extrapolations in which extrapolations are made from one species to another, from one level of organisation to another or from one type of ecosystem to another.

The key issue in these extrapolations is the development of standardised approaches and procedures that will simplify and speed up the risk assessment

1 The following comprises a short resume from the American Chemistry Council project proposal on the basis of which I have been offered a Post-Doctoral position at University of Guelph (2002-2004).

process by using knowledge from some substances to be applied to others in other regions (database development for comparative risk assessments, e.g. QSAR techniques or the USEPA EPI suite). Questions to be addressed are:

Which method should be used, what data are required, how can uncertainties best be addressed in the methods, etc.? Answering these questions requires risk assessment of compounds in the laboratory, in small scaled indoor microcosms and large scaled outdoor microcosms plus additional literature reviews on the sensitivity and ecology of individual species.

Range extrapolations can be carried out through the use of exposure modelling and the application of assessment factors, however, there is no general agreement on how these methods should be applied or whether assessment factors can vary in relation to toxicological properties and sensitivities.

Ecological data extrapolations are in their simplest form extrapolations between different taxa. Species sensitivity distributions (SSD's) are based on the results of laboratory tests and offer a level of refinement that considers the range of sensitivity across entire groups of organisms or within specific categories determined from knowledge of the mechanism of action and the eco- and toxico-kinetics of the substance. The distributions also more closely approach the issue of assessment of hazards and risks at the ecosystem level where our understanding of redundancy and resiliency play an increasingly important role in the community homeostasis, which should be reflected in the risk assessments (ECOFRAM, 1999).

Extrapolation between levels of biological organisation such as from physiological and biochemical responses in single organisms (bioindicators) to responses at the scale of populations and communities involves consideration of both temporal and spatial issues and is not only a common source of uncertainty, but also a typical for of misinterpretation when responses to exposure are confused with adverse effects. The extrapolation of laboratory single species responses to the population level is another ecologically relevant extrapolation that is routinely conducted without confirmation of its appropriateness or whether extrapolation is inherently protective or not. Population models may be used to assist in these extrapolations of ecological risks but only a few models have been verified under field or semi-field conditions (paper I). Current procedures of higher-tier risk assessment are often based on the extrapolation of responses observed in relatively simple microcosm tests to structurally more complex ecosystems in the field. The predictive value of studies in small microcosm tests, however, depends on factors such as fate and exposure of the stressor and the sensitivity and recovery potential of the populations present. Relatively simple microcosms may be directly used to assess risks of toxicants on phyto- and zooplankton (paper III-IV). Extrapolation of responses within these small and short-lived organisms to populations which are more complex (e.g. macrophytes (see paper V), macroinvertebrates) is still a matter of debate (Persoone & Janssen, 1994). Guidance may be provided by comparing the results of simple microcosm tests with more complex and larger microcosm experiments that were treated with similar concentrations of the same chemical.

Extrapolation across age and developmental stage must also be taken into consideration. Most organisms show differences in response to substances at different ages or stages of development. Commonly, younger organisms are more sensitive than older organisms, although, this is not always the case. Amphibians, for example, are often more sensitive to chemical stressors later

in development. Cyclic activity such as reproduction may make organisms more sensitive at certain times and seasons of the year. Extrapolations from one substance to another substance with a similar mode of action, extrapolation from a single substance to mixtures and experimental confirmation of extrapolation assessments could also be subjected to integrated microcosm studies (van den Brink, 1999).

All these approaches require careful considerations of statistical power, acceptable effect sizes and detectability of the designs (Peterman, 1990). Microcosm studies offer the ability to observe a number of responses that are not possible to investigate in single species tests under laboratory conditions. In addition, replicated microcosms tests provide the opportunity to perform ecosystem-level research and allow statistical interpretation of responses and comparisons with control systems. Toxicant-initiated interactions, recovery from effects, and redundancy of function are all more easily and realistically observed in ecosystem-level experimental models. Therefore, in order to satisfy our curiosity and need to understand more about the environmental impacts of manmade chemicals, research in microcosms will have to remain in the future. However, the funding might change and decrease if the environmental agencies are not able to implement microcosm data in the risk assessment of chemicals due to all the uncertainties present.

The perspectives for interpretation and application of the Precautionary Principle are political issues and thus follow inherent political changes. However, the perspective for scientific application of precautionary approaches is an issue of scientific epistemology, technology and practices. Dr. Gregory R. Biddinger of Exxon Mobile and I will be chairing a session on methods and perspectives for a scientific application of precautionary approaches at the SETAC North America 23rd annual meeting in Salt Lake City, Utah, 16-20 November 2002. We will focus the session around short term and readily applicable precautionary approaches in the risk assessment process. Moreover, we will involve more long term epistemic precautionary approaches, which transcends the hurdle from philosophy to actual possible implementation, like for example further development of post-normal science (Funtowicz and Ravetz, 1994) and analysis of benefits of products (analysis of unnecessary products and compounds) (paper VIII).

Overall conclusions

The experimental papers and reviews have all, context dependently, demonstrated and quantified the relative trade-offs between ecological and statistical significance of three spatially different microcosm designs. Typically, the more ecologically realistic the design, the lower is the statistical power of the design, this is due to increased natural variation. There is no scientific foundation for a direct numerical comparison of the three designs power. This is because of toxicological differences among the compounds used in the studies, the compounds fate and thus exposure and the sampling techniques and general maintenance of the microcosms. However, as mentioned in the introduction the aim was not to determine a correct design, nor to make direct context-free comparisons, as there is always a relatively unknown and variable extrapolation uncertainty, but to illustrate and context dependently determine the statistical power of three different microcosm designs. Depending on the effect size induced by the test compound and the sensitivity of the endpoint the relative ranges of required sample size (*n*) of the three designs were:

- 25m^2 earthen pond: 14-45 ponds
- 12m^3 outdoor PVC ponds: 2-19 ponds
- 30L indoor tanks: PFOS; 2-5 and PFOA; 5-31 tanks.

This indicates that in these tests the 30L indoor tanks produce the statistically best data and the 25m^2 earthen pond yield the statistically worst data. The 25m^2 earthen ponds and pelagic plankton were not applicable for higher tier risk assessment of Roundup. The 12m^3 outdoor PVC ponds are in between and yield relatively sound statistically and ecologically results, however, the treatment concentration is high compared to expected environmental exposure. This, in turn, masks and hampers the detection of subtle effects, because the treatment concentrations are increased in order to increase the effect size significantly breaking through the noise of the background variation. This can compromise the ecological realism and the validity of the experiment because the required effect concentrations (Δ), and hence predicted no effect concentration (PNEC), are frequently much higher than the predicted environmental concentrations (PEC), resulting in no unacceptable environmental risk, according to the risk equation PEC/PNEC << 1. In EU terminology, the trigger value of Toxicity/Exposure Ratio (TER) would be very low, because of the relatively high required effect size (Toxicity) compared to low exposure ratios resulting in TER << 10 (Dir. 91/414EEC). However, the risk of type II errors are only seldom taken into account in these analyses (paper I). Therefore, in order to design a conservative and well informed higher tier risk assessment microcosm study statistical power analysis needs to be implemented, preferably *a priori*, in order to determine the required sample size (*n*).

Ecotoxicology and ecology should quantify the ecosystem's carrying capacity in terms of, for example, recovery time and redundancy, to give scientific options regarding protection of system integrity, structure and function. In other words, supply the decision-makers with priority information regarding the ecological consequences associated to different levels of acceptable ecological effect size (Δ). The final political determination of a determining the acceptable effect sizes (Δ) and the acceptable risk of type I and II errors should be reached in consensus between stakeholders based on transparent, sustainable societal cost-benefit analysis and ethical considerations. Moreover, this procedure could be a preparatory point for a scientific application of the precautionary principle as a transparent and reflexive modernisation of environmental science.

References

Brink van den, PJ. 1999. Ecological and statistical evaluation of effects of pesticides in freshwater model ecosystems. Doctoral thesis. Wageningen, Netherlands.

CLASSIC. 2001. Community Level Aquatic System Studies Interpretation Criteria. In press.

Dir. 91/414EEC. EC Plant Protection Product Directive 91/414/EEC (Council Directive, 1991).

ECOFRAM. 1999. Aquatic report. By P. Hendley & JM. Giddings. www. Aqex_ecofram_Peer01_may499.doc.

Fisher, LJ. 1992. Decisions on the ecological and environmental fate data requirements task force. Memorandum. US Environmental Protection Agency, Office of Pesticide Programs, Washington, DC. USA.

Forbes, V. & T. Forbes. 1993. A critique of the use of distribution-based extrapolation models in ecotoxicology. Func Ecol 7:249-254

Funtowicz S. & J. Ravetz. 1994. Uncertainty, complexity and post-normal science. Environ Toxicol Chem 13:1881-1885.

HARAP. 1999. Guidance document on higher-tier aquatic risk assessment for pesticides. Eds. PJ. Campbell, DJS. Arnold, TCM. Brock, NJ. Grandy, W. Heger, F. Heimbach, SJ. Maund, M. Streloke. SETAC-Europe. Brussels, Belgium.

Hurlbert, SH. 1984. Pseudoreplication and the design of ecological field experiments. Ecol Monograph 54:187-211.

Levin S. & Kimball KD. 1984. New perspectives in ecotoxicology. Environ Managem 8:375-442.

Møhlenberg, F., Petersen, S., Gustavsson, K., Lauridsen, T., Friberg, N., 2001. Mesocosm experiments in the approval procedure for pesticides. A study on effects of mesocosm characteristic and validity of extrapolation methods to protect sensitive species. Pesticide Research No. 56 2001. Bekæmpelsesmiddleforskning fra Miljøstyrelsen, Danish EPA. Copenhagen, Denmark.

Kimball KD & Levin S. 1985. Limitations of laboratory bioassays: The need for ecosystem-level testing. BioScience 35:1-172.

Kraufvelin, P. 2000. Detecting change in marine zoobenthic communities by large-scale and long-term mesocosm experiments and field studies. Doctoral thesis. Åbo Akademi University Press Finland.

Persoone G. & Janssen CR. 1994. Field validation of predictions based on laboratory toxicity tests. Chapter 25 In Freshwater field tests for hazard assessment of chemicals. Eds. IA. Hill, F. Heimbach, P. Leeuwangh & P. Matthiesen. Lewis Publishers, CRC Press, Boca Raton, USA.

Peterman RM. 1990. Statistical power analysis can improve fisheries research and management. Can J Fish Aqua Sci 47:2-15.

Pratt, JR. & J. Cairns Jr. 1996. Ecotoxicology and the redundancy problem: Understanding effects on community structure and function. Chapter 12 In Ecotoxicology: a hierarchical treatment. Eds. MC. Newman & CH. Jagoe. Lewis Publishers CRC Press Boca Raton, USA.

Streloke, M. 2000. Objectives and design of aquatic microcosm studies for pesticide registration. Notes from technical workshop: Regulatory evaluation of aquatic microcosm studies with pesticides. Brighton, UK.

Thattersfield, L. 2001. Poster # M/FF224, at the 11th annual meeting of SETAC Europe, 6-10 May 2001 in, Madrid.

Weinberger, A.M. 1985. Science and its limits: The regulators dilemma. Issues in Sci Techn fall:59-72.

Selected work

Hans Sanderson

Cand.techn.soc., PhD Student

Email: hanss@teksam.ruc.dk
Telephone: 4674 2496 Office: 12.2

Research Interests

I was born 12. September 1968 in Ipswich, England. After my master thesis I started my PhD 1. February 1999.

My main field of work is risk assessment of toxic effects of manmade chemicals on the aquatic environment. I have focused on higher tier risk assessment using microcosm experiments at different spatial scales. I have conducted a mesocosm study in a series of artificial ponds involving total phytoplankton (chlorophyll *a* L^{-1}) and total zooplankton (total numbers and total species numbers L^{-1}) direct and indirect responses to the herbicide Roundup. My set-up was to reduce the extrapolation to real natural pond environments in the agricultural landscape as much as possible. The aim of the experiment was to test the applicability of microcosms test and pelagic plankton for higher tier risk assessment of Roundup.

Moreover, I have conducted a literature review on replicability of micro/mesocosm experiments from 1985-2000. In Canada I have participated in a microcosm experiment testing PFOS & PFOA. Furthermore, I have reproduced and validated the Canadian study at smaller spatial scale in 30 indoor 30L microcosms at University of Roskilde. I'm also involved in analysing a scientific implementation of the precautionary principle for the Society of Environmental Toxicology And Chemistry.

In my thesis I address methodological trade-offs in the microcosm approach on three different spatial scales and address problems and solutions for analysis (statistical), interpretation (accuracy vs. precision) and implementation (via power analysis and precautionary approaches) of ecosystem analysis. The result of my dissertation will be a Popperian critical evaluation of scientific and practical obstacles and challenges to the development of more sustainable and ecological risk assessment of chemicals and their regulation.

Professional and scientific interests:
Complex, multidisciplinary and statistical analyses on aspects and integration of; ecology, ecotoxicology, risk assessment, uncertainty management, cost-effective analysis, risk policy and legislation of chemicals.

Participation In Networks

Part of TSRI 0200 under Environment Canada on risk assessment of PFOS&A. Member of the Ecocouncil (2000-2002), chair for the chemicals subgroup (2000-2002). I have maintained good working relations at: Danish Hydraulic Institute, Alterra Institute (Holland), University of Guelph, University of Toronto (Canada), University of Åbo (Finland) and various universities and institutions in the US.

Administrative Tasks

Co-chairing a session at the 23rd SETAC NA meeting Nov. 2002 on scientific application of precautionary approaches. Planning, participation and execution of a national conference on the environmental effects of pharmaceuticals increasingly released into the environment. Member of the PhD board at the department.

Selected Publications

- Power analysis as a reflexive scientific tool for Interpretation and implementation of the Precautionary Principle in the European Union, *Environmental Science and Pollution Research, 2002. Doi:http://dx.doi.org/10.1065.espr2001.10.095*, Hans Sanderson, Søren Petersen.
- Ecological impact and environmental fate of Perfluorooctane sulfonate on the zooplankton community in indoor microcosms, *Jour. of Env. Tox. & Chem. in press accepted 15/1-2002*, Hans Sanderson, T. Boudreau, S. Mabury, Woo-Jay Cheong & KR. Solomon.
- Ecological impact and fate of Perfluorooctanoic acid on the zooplankton community in indoor microcosms, *Submitted to Aquatic Toxicology 2002*, Hans Sanderson, T. Boudreau, S. Mabury & KR. Solomon.
- SETAC position paper on the Precautionary principle, *Setac Globe, in prep. 2002*, Hans Sanderson, Greg Biddinger, Jerry Ravetz, Joel Tickner.
- Power analysis of Myriophyllum sp. microcosm toxicity data, *Environmental Toxicology and Chemistry, in review, 2002*, Hans Sanderson, Mark L. Hanson, Keith R. Solomon.
- International Summit on Science and the Precautionary Principle, *Environmental Science and Pollution Research 9, 1 2002*, Hans Sanderson.
- Design, analysis and interpretation of higher tier risk assessment of chemicals in aquatic microcosms, *PhD dissertation*, Hans Sanderson.
- Replicability and interpretation of Micro/Mesocosm studies, *Environmental Science and Pollution Research, vol. 8 p.1-7 2001*, Hans Sanderson.
- US EPA brings organic fluorine compound to a halt - but what about Europe?, *Chemical Awarenees, Vol. 15, 2001,pp 8-9*, Hans Sanderson.
- Statistical power analysis of phytotoxicity data on Myriophyllum sp. from microcosm experiments, *Poster at 11th. Annual meeting and SETAC Europe, 6-10/5-01. Madrid*, Hans Sanderson, Mark Hanson.
- US ahead of Europe in addressing PFOS, *Interview in "Ingeniøren NET, 26/4-2001"*, Hans Sanderson.
- Statistical power analysis of data from quasi-natural mesocosm ponds perturbed with Roundup2000, *Poster at SETAC 3. World Congress in Brighton May 2000.* Hans Sanderson.
- Håndtering af forsigtighedsprincippet, *Vand & Jord, Sept. 2000, pp. 88-90*, Hans Sanderson.
- Comparative Variability analysis in outdoor experimental ponds perturbed with Roundup, *Poster at DMU's anniversary conference 19-20/8-99*, Hans Sanderson
- Applicability of quasi-natural mesocosm ponds and pelagic plankton for higher tier risk assessment of treated with Roundup2000, *Journal of Environmental Contamination and Toxicology, 2002 in review*, Hans Sanderson, Søren Petersen.

Selected Work

- SETAC position paper on the Precautionary principle, 2001-2002, *Initiating and author of the SETAC position paper on scientific application of precautionary approaches*, Hans Sanderson, Greg Biddinger, Jerry Ravetz, Joel Tickner.
- Member of international advisory committee, Summit on Science and the Precautionary Principle, 2001-2001, *Member of committee leading up to the Summit in Lowell Massachusetts 20-22/-2001*, Hans Sanderson.
- Visiting student at University of Guelph, Canada, 2000-2001, *participating in microcosm investigation at Guelph University (Canada) 6 weeks during the summer 2000, and after that statistical analysis and interpretation of data. Revisited 2 weeks fall 2001*, Hans Sanderson.
- The PhD committee at the Department, 2000-2002, *Member of the committee*, Hans Sanderson.
- SETAC, 1999-2002, *Member of the Society of Environmental Toxicology and Chemistry*. Hans Sanderson.
- Ecocouncil, 1999-2002, *Member of the Danish Ecocouncil and chair of the chemical group under the council.* Hans Sanderson.

Selected Verbally Communicated Research

- Opening of conference on pharmaceuticals in the environment, *National meet addressing the environmental fate, effects and regulation of pharmaceuticals*, 2002, Hans Sanderson.
- Scientific application of precautionary approaches, *Co-chairing the session at SETAC NA. 23. Annual meeting in Salt Lake City, Utah, US*. 2002, Hans Sanderson, Greg Biddinger.
- Power of precaution, *Oral presentation at science seminar 2/3-01 at dept. of Env. Tech. & Soc. University of Roskilde*, 2001, Hans Sanderson.
- Managing the Precautionary principle and uncertainty with Statistical power analysis, *Oral presentation at 11th. Annual meeting of SETAC Europe, 6-10 May 2001 Madrid*, 2001, Hans Sanderson, Søren Petersen.
- PFOS. TV news, Friday 20/4, *Interview to national news broadcast (Danmarks Radio) on risks concerns about PFOS*, 2001, Hans Sanderson.
- Replicability and interpretation of micro- and mesocosm studies, *a platform presentation at SETAC 3. World Conference in Brighton UK*, 2000, Hans Sanderson.
- Progress report on analysis of limitations for interpretation of micro- and mesocosm experiments. *Platform presentation at Michigan State University at the 11th Annual Toxicology Seminar between Michigan State University and University of Guelph (Canada)*, 2000, Hans Sanderson.
- Risk of Type II error and power analysis of mesocosm data, *2. Environmental Science Day, November, 2000*, 2000, Hans Sanderson.
- A comparative study of ecological effects before and after perturbation with Roundup in experimental ponds, *1. Environmental science day Nov*, 1999, Hans Sanderson.

Resume på dansk

Design, analyse og fortolkning af akvatiske mikrokosmosforsøg til risikovurdering af kemikalier på højeste niveau

Hans Sanderson

RESUME

Mikrokosmosstudier har gennem årtier været anvendt i økologisk og økotoksikologisk videnskab og forskning. I løbet af 1980erne og 1990erne blev metoden gradvis udviklet og evalueret med henblik på implementering som et højt niveau i risikovurderingen af pesticider såvel i USA som i EU. Men i 1992 valgte den amerikanske miljøstyrelse, at den ikke længere vil forlange data fra mikrokosmosforsøg i registreringsproceduren af pesticider og kemikalier. Dette med henvisning til alt for store usikkerheder i forbindelse med design, udførelse, analyse og fortolkning af data fra mikrokosmosforsøg. Et dominerende videnskabeligt problem i forbindelse med mikrokosmosforsøg er stadig, at afgørelse den relative økologiske relevans kontra den statistiske signifikans, der knytter sig til designet af mikrokosmos forsøg. Målsætningen med nærværende Ph.d.-afhandling er at belyse de iboende økologiske og statistiske kompromiser i tre størrelsesmæssigt forskellige mikrokosmoseksperimenter. Desuden, at kvantificere usikkerheden i de tre forskellige designs ved hjælp af statistiske styrke betragtninger. Og endelig, at analysere en videnskabelig anvendelse af forsigtighedstilgange til at kontrollere og regulere videnskabelig usikkerhed fra et miljøregulerings og -politisk perspektiv.

Det indledende papir i denne afhandling giver i en aktuel litteraturbaseret oversigtsartikel (1985-2000) en analyse af replikationen af mikrokosmosstudier til anvendelse i risikovurderingen af pesticider. Oversigtsartiklen fokuserer på beregning af variationskoefficienter (CVs) fra studierne, som et mål for replikationen og dermed også muligheden for at detektere subtile effekter. En serie mikrokosmosstudier siden blev påbegyndt tre uafhængige økologiske og statistiske niveauer, der tentativt blev sat til: 25 m^2 jordlinede udendørs damme > 12 m^3 PVC udendørs damme > 30L transparente PVC-tanke. Den statistiske styrke på hvert niveau af økologisk signifikans blev bestemt, ihukommende at en direkte sammenligning af statistisk styrke er begrænset, dersom denne i høj grad er kontekstafhængig. Dette, blandt andet på grund af en ukendt og variabel ekstrapolationsfaktor mellem de forskellige forsøg og mellem stofferne, der anvendes i eksperimenterne. Til slut diskuteres en videnskabelig applikation af en refleksiv videnskab med udgangspunkt i statistisk styrke betragtninger, analyse af nødvendighed og post-normal videnskab i lyset af forsigtighedstilgange for regulering og håndtering af miljømæssige risici, usikkerhed og mangelfuld viden.

Afhandlingen præsenterer følgende nye originale videnskabelige informationer. Oversigtsartiklen dokumenterer, at hovedparten af mikrokosmosstudier lider under mangelfuld replikation og høje CVs, hvilket begrænser studiets følsomhed og dermed detektionsgrænse. Ergo, hvis den naturlige baggrundsvariations amplitude mellem replikater (s^2) er stor, så skal den toksisk inducerede amplitude (Δ) mellem koncentrationerne og kontrollerne være endnu større for at bryde signifikant ($p \leq 0.05$) gennem baggrundsstøjen fra den naturlige variation. Dette kunne igen føre til lav statistisk styrke ($1-\beta < 0.8$) og dermed en uacceptabel eller ukendt høj risiko for type II fejl (falsk negativ) (papir I). Denne tendens dokumenteredes endvidere i de tre mikrokosmosforsøg jeg var involveret i hvor, det generelle mønster var, at den statistisk signifikans var modsat proportional med den økologiske ditto. Anvendeligheden af akvatiske mikrokosmosforsøg og pelagiske planktoniske effektparametre til risikovurdering af Roundup er tvivlsomme da stoffet forsvinder fra vandfasen meget hurtigt (papir II). Afhandlingen rapportere de to hidtil eneste publicerede artikler omkring en høj niveau risikovurdering i mikrokosmos af to dominerende perfluoro overfladeaktive stoffer (papir III & IV). Papir V indeholder den eneste publicerede oversigtsartikel, der rangordner ti forskellige fyto-toksikologiske effektparametre i henhold til ASTM vejledningsprocedure # 1913-97 E i mikrokosmos. Endelig diskuteres i afhandlingen en videnskabelig applikation af forsigtighedstilgange via statistisk styrke betragtninger til håndtering og administration af usikkerheder i lyset af EU's aktuelle fortolkning af forsigtighedsprincippet, samt i relation til en refleksiv modernisering af miljøvidenskab (papir VI-VIII). Samlet set afdækker og kvantificerer afhandlingen nye aspekter angående usikkerheden forbundet med risikovurderingen af kemikalier på højeste niveau i akvatiske mikrokosmos, samt giver råd og vejledning omkring design, analyse og fortolkning af mikrokosmosforsøg.

Det konkluderes, at det generelt er mere overkommeligt og rentabelt at øge den statistiske styrke i mindre skalerede mikrokosmos ved at øge antallet af replikater. Desuden påpeges et paradoks mellem en øget økologisk relevans i mikrokosmosdesignet og behovet for deraf følgende høje effektstørrelser (Δ) og dermed urealistiske behandlingskoncentrationer i mikrokosmos, sammenlignet med mere realistiske behandlingskoncentrationer under urealistiske økologiske forhold i standardiserede laboratorium tests. Det konkluderes endvidere, at statistiske styrke betragtninger til en start kunne implementeres som en videnskabeligt baseret og refleksiv indgang til forsigtighedstilgang i miljøreguleringen. Dette skulle funderes på en udvidet samfundsmæssig godkendelsesprocedure af data, hvor den acceptable effektstørrelse (Δ) samt den acceptable risiko for henholdsvis type I og II fejl (α & β) skal bestemmes inden et mikrokosmosforsøg designes, samt at disse baseres på; grundige økologiske analyser til beskyttelse af økosystemets integritet, samfundsmæssige cost-benefit analyser og etiske overvejelser. Først herefter kan et pilotforsøg designes og udføres og det endelige design kan bestemmes med korrekt økologisk og statistisk relevans og med høj styrke. Det er vigtigt at ihukomme, at mikrokosmosdesign altid er afhængig af sin kontekst og sit beskyttelsesmål.

Til sidst reflekteres over mikrokosmoseksperimenters fremtidige rolle i miljøbeskyttelsen. Det konkluderes, på baggrund af papirerne I-VIII, at såfremt miljømyndighederne lægger vægt på den statistiske styrke af studierne og på forhånd og på baggrund af den udvidede samfundsmæssige godkendelsesprocedure af data og definerer enhederne i styrkeligningen, da kan mikrokosmos bidrage positivt til miljøbeskyttelsen, som et højt niveau i risikovurderingen af kemikalier. På den anden side, hvis dette ikke gøres, da kan anvendelsen af mikrokosmos virke modsat på grund af en høj eller ukendt risiko for type II fejl. Med hensyn til forskning og undervisning vil de stadig være værdifulde for en bedre forståelse af de miljømæssige effekter af menneskeskabte kemiske forbindelser i naturen på samfund og økosystemer, da disse kun kan testes og valideres i mikrokosmos. Desuden kan en korrekt anvendelse af mikrokosmos bidrage til udviklingen af teorier og begreber indenfor økologi og økotoksikologi, samt til vores viden om ekstrapolation af økologiske risici over tid og rum samt mellem systemer, arter og stoffer.

Nøgleord: Mikrokosmos, høj niveau risikovurdering, design, statistisk styrke betragtninger, type II fejl, *Myriophyllum*, ASTM # 1913-97 E, PFOA, PFOS, $Roundup_{2000}$, forsigtighedsprincippet.