RIVM Report 601506009/2008
A guidance document of the Dutch Platform
for the Assessment of Higher Tier Studies

Guidance for summarizing and evaluating aquatic micro- and mesocosm studies

F.M.W. de Jong[1], T.C.M. Brock[2], E.M. Foekema[3], P. Leeuwangh[4]

[1] RIVM

[2] Alterra, Wageningen UR

[3] Wageningen IMARES

[4] EC&C

National Institute for Public Health and the Environment,
PO Box 1, 3720 BA Bilthoven, The Netherlands.
Tel. 31-30-2749111, fax. 31-30-2742971

Photograpy cover: E.M. Foekema

Published by the
National Institute for Public Health and the Environment
PO Box 1
3720 BA Bilthoven
The Netherlands.

This publication can be found at www.rivm.nl/bibliotheek/601506009/pdf
RIVM report number 601506009/2008
ISBN-10: 90-6960-191-5
ISBN-13: 978-90-6960-191-5

ABSTRACT

Guidance for summarizing and evaluating aquatic micro- and mesocosm studies
A guidance document has been developed for ensuring that the test results required for the registration of pesticides be supplied in a uniform and transparent manner. This document is specifically directed at experiments carried out in artificial ecosystems in surface water (micro- and mesocosm studies). It has been developed by the Dutch Platform for the Assessment of Higher Tier Studies (PHTS), of which the Netherlands National Institute of Public Health and the Environment (RIVM) is the secretariat.

Within the framework of the rules and regulations governing pesticide registration in the Netherlands, applicants (e.g. manufacturers of crop protection agents) are required to supply all necessary information to the Dutch Board for the Authorisation of Crop Protection Products and Biocides (Ctgb). Based on this information the Ctgb assesses whether the use specified for a specific product is acceptable. Complex and often extensive data on micro- and mesocosm studies can be a necessary part of the information provided. The Ctgb, an independent administrative body, requests various external institutes to summarize and evaluate these studies. Potential differences in the evaluator's methodology may lead to a lack of uniformity in the form and content of the summaries and evaluations and – occasionally – in the conclusions.

These differences were the primary motivating factor for Ctgb to harmonize the evaluation reports of studies on ecosystems of surface water bodies. A secondary aim was to increase the transparency of the registration process.

Key words: pesticides, plant protection products, registration

RAPPORT IN HET KORT

Richtsnoer voor het samenvatten en evalueren van aquatische micro- en mesocosm studies

Er is een richtsnoer ontwikkeld om testresultaten voor de toelatingsprocedure voor gewasbeschermingsmiddelen eenvormig en transparant aan te reiken. Het richtsnoer geldt specifiek voor experimenten in nagebootste ecosystemen in oppervlaktewater (zogenoemde micro- en mesocosm studies).Het richtsnoer is ontwikkeld door het Nederlandse Platform voor de Beoordeling van Higher Tier Studies, waarvan het RIVM het secretariaat voert.

Bij de toelatingsprocedure voor gewasbeschermingsmiddelen leveren aanvragers (bijvoorbeeld de bestrijdingsmiddelenfabrikanten) informatie aan het College voor de toelating van gewasbeschermingsmiddelen en biociden (Ctgb). Aan de hand hiervan beoordeelt het Ctgb of een bepaald gebruik van een middel toelaatbaar is in Nederland. De geleverde informatie betreft onder andere complexe en vaak omvangrijke informatie over micro- en mesocosm studies. Het Ctgb laat deze studies vervolgens door verschillende externe partijen samenvatten en evalueren. Door verschillen in werkwijze kunnen de vorm van deze samenvattingen en evaluaties, en soms zelfs de conclusies, verschillen.

Vandaar de wens van het Ctgb om de evaluaties en samenvattingen van ecosystemen in oppervlaktewater te standaardiseren. Een aanverwant doel is hiermee het beoordelingsproces transparanter maken.

Trefwoorden: bestrijdingsmiddelen, gewasbeschermingsmiddelen, toelating

PREFACE

The present guidance document is an initiative of the Dutch Platform for the Assessment of Higher Tier Studies. The aim of the Platform is to improve and harmonize the assessment of higher tier studies. The guidance document, which was drafted by a working group of the Platform, has been discussed and approved in plenary platform meetings and then sent out for public consultation to European experts and stakeholders. We would like to acknowledge A. Aagaard (Danish EPA, DK), A. Alix (AFSSA, FR), A. Aldrich (ACW, CH), U. Hommen (Fraunhofer, DE), M. Bergtold, P. Dohmen, J. Kubitza, L. Weltje (BASF, DE) and J. Wogram (UBA, DE) for their comments on the draft report. The guidance document has been approved for publication by the plenary platform during the meeting of 4 September 2007.

The secretary of the Platform and the working group have been commissioned and funded by the Netherlands Ministry of Housing, Spatial Planning and the Environment in response to a request from the Board for the Authorisation of Plant Protection Products and Biocides (Ctgb). The working group was further funded by the Netherlands Ministry of Agriculture, Nature and Food Quality, Wageningen IMARES, Ctgb and Ecotox Consultancy & Constructions.

The Dutch Platform for the Assessment of Higher Tier Studies publishes practical and easy to use guidance documents for the evaluation of (semi-)field effect studies and other higher tier studies. A guidance document for summarizing earthworm field studies has been published earlier, and a guidance document for summarizing higher tier studies on terrestrial non-target arthropods is currently in preparation.

Bilthoven, November 2007

Dr. Mark H.M.M. Montforts
Chair

CONTENTS

1. INTRODUCTION

1.1 Background and motivation

The first step in the aquatic hazard assessment of pesticides in the EU ('tier 1') is based on a procedure in which the minimum data requirements are acute and chronic single-species toxicity studies for minimally an algal species, a daphnia and a fish as well as a bioconcentration factor (BCF) for in cases where compounds are potentially bioaccumulative (log $P_{ow} > 3$) (OECD, 1995; EC 2002, Directive 91/414/EEC). These toxicity data for aquatic organisms are compared with short-term and long-term exposure concentrations to generate toxicity-to-exposure ratios (TERs). If the TER for acute toxicity: exposure is ≥ 100, or chronic toxicity:exposure ≥ 10, then the risks to aquatic organisms are considered to be acceptable. Based on comparisons of the results of the first tier triggers with threshold concentrations of micro- and mesocosm experiments (see Brock et al., 2000a, b), it appears that for the vast majority of herbicides and insecticides evaluated, the preliminary risk characterization procedure is generally protective (Campbell et al., 1999). When the first tier trigger values are not met, no authorization shall be granted, *'unless it is clearly established through an appropriate risk assessment that under field conditions no unacceptable impact on the viability of exposed species occurs – directly or indirectly – after use of the plant protection product according to the proposed conditions of use.'* Such an appropriate risk assessment is commonly referred to as higher tier risk assessment, and (semi-)field studies are considered to be applicable here. The absence of an unacceptable impact has to be demonstrated in a risk-based assessment. It should be noted that in the phrase cited above, the protection goal is not clearly defined since the text gives no precise definition of the degree of impact that is acceptable, for example, it does not exclude that short-term impacts followed by recovery may be acceptable.

There are more options that qualify as an appropriate risk assessment apart from the approach that was taken in the first tier. The report from the Higher-tier Aquatic Risk Assessment for Pesticides (HARAP) workshop (Campbell et al., 1999) examined different types of higher-tier studies and developed guidance on how to apply these methods. Methods intermediate between laboratory and field tests may contribute to a more (cost) effective higher tier risk assessment. This set of tools can be used to specifically address the uncertainty associated with a certain pesticide, depending on the areas of concern that had been identified earlier. Additional single-species tests, population recovery studies, indoor microcosm experiments, outdoor micro- and mesocosm tests or a combination of these may reduce the uncertainty. In choosing an appropriate test, the higher tier risk assessment is tailored to the problem identified without the necessity of having to resort to a complex, expensive field test or the full set of tools. Guidance for such methods is currently available (e.g. Crossland and La Point, 1992; Hill et al., 1994; Campbell et al., 1999; Boxall et al., 2001). Although the step-by-step approach suggested by HARAP offers valuable tools for the aquatic risk assessment, in practice there seems to be a tendency to jump directly to micro- or mesocosm stud-

ies. However, given the costs of a micro- or mesocosm study, it will only be possible to generate a limited number of such studies.

Little guidance on the use of higher tier studies for risk assessment is given in the EU guidance documents (EU, 2002). A number of workshops have taken place with the aim of generating technical guidance on the design and conduct of outdoor micro- and mesocosm studies (Arnold et al., 1991; SETAC-RESOLVE, 1992; Hill et al., 1994; OECD, 1995) and providing guidance for the interpretation of the results of these experiments (CLASSIC, see Giddings et al., 2002). The guidance on the performance, evaluation and use of higher tier studies thus appears to be scattered over numerous documents dealing with a specific test type or organism group, while there is an increasing need for regulatory evaluation tools, particularly for (semi-)field studies (Campbell et al., 1999; Hill et al., 1994; Van Dijk et al., 2000). Based on discussions between authorities, scientists and industries in Europe (Giddings et al., 2002) and The Netherlands (De Jong et al., 2005), it is apparent that there is an urgent need to compile present knowledge and practices on the assessment and reporting of higher tier studies in general and (semi-)field studies in particular. Not only will systematic guidance for the evaluation of studies increase the consistency and transparency of the evaluations and their subsequent acceptance by all parties involved, the availability of a set of evaluation criteria will also facilitate discussions between authorities and applicants in the defining phase of the test protocol and will be helpful for improving the set-up of (semi-)field experiments.

The aim of this guidance document is therefore to provide specific technical instructions on the reporting and evaluation of ecotoxicological (semi-)field tests that are based on the study reports submitted with the dossier during the process of pesticide registration in the EU. The main focus is on field micro- and mesocosm experiments, although the guidance is also considered to be applicable for smaller scale studies carried out in the laboratory such as aquatic microcosm experiments.

1.2 Method

The present guidance was drafted using the procedure outlined in the following. Each of the members of a working group consisting of the authors of this report summarized a mesocosm study without any guidance. Although the conclusions drawn by the evaluators did not deviate from each other on the significant points of the study, the nature and the extent of the summaries varied considerably. The same was true for the detailed argumentation supporting the conclusions. It was therefore concluded that guidance for summarizing and evaluating was needed to harmonize the evaluation reports. The working group then drafted a guidance, which was subsequently tested with a second higher tier study. The result of this action is presented in Annex 3 of this report. The guidance was then adapted, using the experiences of the evaluators, and discussed in the Dutch Platform for the Assessment of Higher Tier Studies (PHTS). The

draft guidance was sent out for external consultation, and the comments were used to further improve the guidance document. The final version was approved by the PHTS.

1.3 Explanation of terminology used

The conclusions of the HARAP (Campbell et al., 1999) and CLASSIC (Giddings et al., 2002) workshops in terms of the evaluation and interpretation of micro- and mesocosm tests are (partly) incorporated in the EU Guidance Document on Aquatic Ecotoxicology (EU, 2002). In accordance with the recommendations of the HARAP and CLASSIC workshops, the EU Guidance Document on Aquatic Ecology states that transient effects may be acceptable. Important terms used in the EU Guidance Document on Aquatic Ecotoxicology are $NOEC_{population}$, $NOEC_{community}$, NOEAEC (no observed ecologically adverse effect concentration) and EAC (ecologically acceptable concentration). The concept EAC is avoided in more recent documents because of associated semantic issues (Crane and Giddings, 2004). From a philosophical point of view, it may be argued that ecology as a science has no moral principles and that, consequently, something like 'ecologically acceptable' does not exist and should not be confused with 'regulatory acceptable' (Brock et al., 2006). For this reason, the European Food Safety Authority (EFSA) Panel on Plant Protection Products and their Residues (PPR) panel have replaced the concept EAC with RAC (regulatory acceptable concentration) (EFSA-PPR, 2006b). In this guidance document we also use the concept of RAC.

In practice, the following procedure is proposed to handle the different terms:

- The $NOEC_{population}$ and $NOEC_{community}$ from a semi-field study are the highest concentrations tested at which no dose-related effects are observed for the population and community concerned, respectively. For the relevant taxonomic/ecological groups (e.g. zooplankton, phytoplankton, macroinvertebrates) in the study, a $NOEC_{community}$ is usually derived using appropriate multivariate techniques (e.g. Principal Response Curves). Where there are effects at the population or community level, the time taken for recovery to occur should also be reported. Treatment-related responses, including recovery, can be classified using the effect classes described in Table 1 of this report. In relation to the effect classes, the highest concentration belonging to class 1 is considered to be the $NOEC_{population}$ or the $NOEC_{community}$.
- The NOEAEC is the value at which effects observed in the specific study under evaluation are considered to be acceptable from a regulatory point of view. This can, for example, be interpreted as the concentration at which statistically significant short-term effects may occur, provided that recovery is seen within a certain acceptable period (see later in this report). In other words, effects on individuals resulting in no or only transient effects at the population or community level may be considered by regulators to be of minor relevance. Related to the classification of the effects (see Table 1) is the effect class at which a NOEAEC is set, which depends on such information as, the experimental design of the study, the exposure regime simulated, the species composition of the experimental community and the life-cycle characteristics of the affected populations in the specific study.

- While the NOEAEC is study-specific, the RAC is derived from an overall evaluation of a compound (for examples, see Campbell et al., 1999; Brock et al., 2006; EFSA-PPR, 2006b). The RAC can be seen as the regulatory acceptable concentration (for either short-term or long-term exposure) decided upon by the risk manager, taking all available data into account (e.g. the similarity of the test system with the ecosystems actually at risk; all lower and higher tier exposure and toxicity data). An appropriate assessment factor may be necessary for deriving a RAC for use in the spatio-temporal extrapolation of NOEAEC values.

According to the EU guidance document on Aquatic Ecotoxicology (EU, 2002), effects in micro- or mesocosm studies may be classified in five effect classes. This classification is based on two reports that review micro- and mesocosm studies performed with herbicides and insecticides, respectively (Brock et al., 2000a, b). In this report we propose using the adapted effect class system as described in Table 1.

The aim of the guidance for the regulatory evaluation of micro- and mesocosm tests is to avoid ambiguous interpretations. The criteria of Annex 1 should facilitate the interpretation of the test results. Experimental studies with insecticides and herbicides reveal that under similar exposure conditions in the laboratory and the (semi-)field there is no reason to believe that the sensitivity of the freshwater species being tested is consistently over- or underestimated when field and laboratory results are compared – if these results are caused by direct toxicity (Brock et al., 2000a, b). For insecticides, Maltby et al. (2005) show that this is also the case when comparing laboratory and species sensitivity distributions (SSDs). An important difference between laboratory and semi-field tests is that in micro- and mesocosm tests indirect effects and recovery may affect the long-term response of the populations and community of concern, despite the fact that the initial exposure concentrations and direct toxic effects were similar between these test systems. Another relevant difference is that in standard laboratory tests, the realistic fate of the test substance usually is not simulated. When the ecological threshold values for direct toxic effects observed in micro- and mesocosm experiments significantly deviate from tier 1 trigger values or SSD analysis, this should be explained on basis of the properties of the chemical and the test system. When the threshold values are significantly higher in the semi-field study than in the first tier, it is particularly necessary to have a satisfactory explanation in order to be able to accept that the higher tier study showed the absence of unacceptable effects.

The primary aim of this document is to provide guidance on summarizing and evaluating test reports on aquatic semi-field studies (micro- and mesocosm tests) as an integral part of the dossier evaluation process. In this document we distinguish three regulatory aspects: (1) the evaluation of the study, (2) the actual risk assessment and (3) risk management. Although in practice more than one aspect can be done by the same person, in this document we make a distinction between (1) the evaluator, who is the person summarizing and evaluating the particular study; (2) the regulator, who uses the results of the evaluation of the particular study in the risk assessment, taking into account all other information in the dossier; (3) the risk manager, which denotes the institution that defines the boundary conditions for the risk assessment, such as

the extent of effects that is deemed acceptable. The guidance is presented in Chapter 2. Comments on the usefulness of (semi-)field studies for risk assessment within the registration procedure of pesticides are given in Chapter 3.

2. GUIDANCE ON SUMMARIZING AND EVALUATING MICRO- AND MESOCOSM TEST REPORTS

When a micro- or mesocosm test is provided, the evaluator must verify the information presented and present the data used to reach a decision in a transparent, concise and consistent way. The evaluation report has the following structure:

1. Header table or abstract, which contains the decision-making information on the test results and the conclusions;
2. Extended summary of the study, including the test design, results and the conclusions of the authors of the report to be evaluated;
3. Evaluation (critical comments on the test, made by the evaluator) consisting of:
   3a. Evaluation of the scientific reliability of the field study;
   3b. Evaluation of the results of the study;
4. Suggestions for use in risk assessment (intended for the regulator).

The different items are elaborated below.

Item 1. The header table, abstract and remarks should provide the key endpoints, endpoint values and conclusions of the study and the evaluation, such as the NOEC, NOEAEC and the reliability of the study. The header table has two parts: a general part that is in accordance with the presentation criteria of EU monographs, and a second, in which specific information related to the particular study is summarized. An example of a header table and an abstract are given in Box 1 (page 16).
Essential information on such aspects as the exposure regime, type and duration of the study, type of ecosystem, among others should be found in this section so that the regulator can obtain an overall first impression of the study with one quick glance. Remarks essential to the specific study should be found in this section. The items mentioned in the header table (such as pH) can vary depending on the parameters important to the specific study. The reliability index (Ri) is assigned to assess the scientific quality of the study (see Annex 1). The reliability, among others, determines whether a study is acceptable for use in risk assessment (see below). If, based on the Ri, the study is rejected, the motivation for this rejection should appear in the remarks.
Item 2.The extended summary includes a description of the test design, measurement endpoints and results (as presented by the author) and should comprise all of the essential information that was used to reach to the conclusion of the author(s) and evaluator. Annex 1 can be used to check whether all relevant items are included in the summary. A factual representation of the study as well as the evaluation and the conclusion of the authors of the study report are given in the extended summary. The conclusions of the evaluator are given in item 3.

Item 2. The extended summary is needed because during the later stages of the risk assessment, the study report is no longer at hand. Decisions are consequently made on the basis of the evaluation reports, and the information in these reports should facilitate this process. As such, the summary of the study provided by the authors of the

Box 1 Example of header table and abstract (Ri, Reliability index; a.s. active substance)

Header table

Reference	: Smith et al. (2002)	GLP statement	: no
Type of study	: aquatic Outdoor microcosm	Guideline	: in accordance with HARAP
Year of execution	: 2000	Acceptability	: acceptable
Test substance	: formulation		

Substance	Method	Exposure regime	Endpoints based on	Duration (day)	Criterion	Value (µg a.s./L)	Ri
Formulation 250 mg a.s./L	Outdoor Microcosm	Repeated exposure 3× with a 10-day interval, test concentrations 0.5, 1, 5, 10, 50 µg a.s./L nominal	Periphyton Zooplankton, Macroin-vertebrates litter decomposition	20 treatment 50 day post-treatment	NOEC NOEAEC	5 (nominal) 10 (nominal)	1

Abstract

From the reliable outdoor microcosm study with formulation xxxx, 250 mg a.s. yyyy/L, it is concluded that for the species groups of periphyton, zooplankton and macroinvertebrates an overall $NOEC_{community}$ of 5 µg/L can be derived, and an effect class 3A No Observed Ecological Adverse Effects Concentration (NOEAEC) of 10 µg a.s./L can be derived.

Remarks

Lower tier risk assessment triggered a potential risk due to short-term exposure. Due to the short DT_{50} in water, long-term exposure did not occur in the test system. Actual concentrations were within 80% of nominal. The study was conducted in a mesotrophic, macrophyte-dominated system located in the UK. Given the duration of the post-treatment period (50 days), for some treatment levels recovery within 8 weeks after the last application could not be established (class 4).

study report may not be sufficient, since this summary has a different aim: stating the main results and conclusions. The guidance suggests that the design and the results of the study be presented as concisely as possible, preferably in the form of tables and figures. To this end, tables and figures may be copied from the study report. In the case of the evaluator constructing the tables and figures, this should be indicated. An example of an extended summary is given in Annex 2.

Item 3. For the evaluation, lower tier information should be taken into account as well: the mode of action of the substance, sensitive groups and other relevant information can focus the evaluation on relevant aspects (see below).

Item 3a. The following questions should be answered in the evaluation of the scientific reliability:

1. *Is the test system adequate and does the test system represent a relevant freshwater community?* [Trophic levels; taxa richness and abundance of (key and sensitive) species, representativeness of the biological traits of the tested species]
2. *Is the description of the experimental set-up adequate and unambiguous?* (ANOVA or regression design; overall characterization of the experimental ecosystem/community simulated; measurement endpoints; sampling frequency; sampling techniques)

3. *Is the exposure regime adequately described?* [Method of application of the test substance; concentration in the spray solution; dynamics in exposure concentrations in relevant compartments (e.g. water, sediment); detection limits]
4. *Are the investigated endpoints sensitive and in accordance with the working mechanisms of the compound, and with the results of the first tier studies?* (Compare selected measurement endpoints with the species potentially at risk as indicated by the lower tiers)
5. *Is it possible to evaluate the observed effects statistically and ecologically?* (Univariate and multivariate techniques applied; unambiguous concentration-response relationships; statistical power of the test; ecological relevance of the statistical output).

The above-mentioned questions can be answered with yes, unclear or no, and the answers should be substantiated with arguments. A further detailed checklist to assess the scientific reliability of the study is given in Annex 1, Table A1.2, which systematically lists the items important in evaluating a higher tier study. These items can be used for answering the above questions. The reliability of the study is lowered when items listed in Table A1.2 are inadequately reported or not reported at all. A reliability index is used to assess the reliability (see Annex 1, Table A1.1). Some items are deemed essential for the reliability of the study, and when these items are not (satisfactory) described, the test is principally unreliable. However, since higher tier studies are evolving and can be tailor made for specific problems identified in the lower tiers, it is possible that – in specific cases – the results of a study in which such items are lacking can still be used for risk assessment. In this case, the evaluation report should contain the argumentation. With the exception of these specific cases, a study with a Ri3 cannot be used for risk assessment. Other items can lower the reliability, but these are not deemed to render the test unreliable on their own, and the specific combination of different items will ultimately lead to a decision on the reliability. A study with a Ri2 can be used for risk assessment, but the regulator should be aware that some aspects of the study render the test less reliable. Given the lower tier information and other field data, the regulator then has to decide whether the aspects that render the test less reliable are essential to the specific case. Studies with a Ri1 can be used for risk assessment on scientific reliability. For further details, the reader is referred to Annex 1. Since the items listed are used to assess the reliability, these items must be described in the extended summary as well. Table A1.2 can also be used as a checklist for the extended summary.

There is a difference between reliability and usefulness. A study that is scientifically reliable is not automatically useful for risk assessment; for example, in a field study where the GAP (good agricultural practice) is not in accordance with the application. Although it is the scientific reliability of a particular study that is the primary aim of the evaluation (see, for example, Annex 1), it is not always possible to separate reliability and usefulness. The assessment of the usefulness is not seen as a part of the evaluation of the particular study but as a task of the regulators. Some aspects of usefulness are discussed in Chapter 3.

Item 3b. After these questions have been answered, the effects are described per concentration tested. In the Evaluation section the concentration–response relationships observed should be classified according to the effect-classes presented in Table

1. The occurrence of an effect at consecutive time points is more likely to be related to substantial damage to the ecosystem than an effect that is observed only once or even repeatedly but with time intervals in between. Indirect effects are treated the same way as direct effects. The duration of the ecological relevant period depends on the environmental compartment and the ecosystem/population involved. For aquatic mesocosm studies, a duration of 8 weeks may be chosen as an assessment endpoint for recovery (EU, 2002). (Brock et al. (2000a, b) chose a recovery period of 8 weeks in their review papers that introduced the Effect classes because in most of the micro- and mesocosm studies reviewed, invertebrates were sampled at intervals of 2–4 weeks. A study period of 8 weeks will therefore allow a few sampling dates to show consistent recovery. However, in practice, whether or not such an interval is sufficient to describe the effects in a proper manner, especially in the period directly following the applications, will depend on, among others, the mode of action of the compound, the DT_{50} in the water, the life-cycle strategy of the affected organisms, the size of the test system and the effects found. Choosing the acceptable time frame for recovery, which may differ for different taxonomic groups, should be a risk management decision (based on the consensus reached amongst risk managers, the Effect classes mentioned in Table 1 can be adopted accordingly). Table 1 is an adapted classification compared to the classification of the EU-guidance document. By taking into account not only the duration after the last application but also the total duration of the effects, a repeated application, each with a short-term effect, but with an overall total effect duration of > 8 weeks, is classified in another Effect class (3B) than a total effect duration of < 8 weeks. In Effect class 5, a distinction has been made between recovery within the study period (5A) and no recovery within the study period (5B).

All available information should be taken into account when deciding on recovery. Since the principal response curves (PRC) present effects relative to the control, it is theoretically possible that changes in the control suggest recovery. Therefore, the abundance of the individual populations should be considered as well. In the case that the numbers of a certain population in the control fall to the level of the treatment, no decision can be made on whether recovery occurred or not unless the test lasts long enough to observe an increase in the control. When this phenomenon occurs for relatively abundant species or for species that play an important role in the PRC, a decision on recovery cannot be made without additional data; it is therefore proposed that these effects be classified as Class 4 or 5.
The evaluator has to refer to the original data in the study report when describing treatment-related responses and assigning these responses to Effect classes.

In general, responses of the measurement endpoints are considered to be treatment-related when

- clear concentration–response relationships are observed
- statistically significant effects can be demonstrated on at least two consecutive sampling dates [except for endpoints characterized by a relatively low sampling frequency (e.g. chlorophyll *a*)]
- effects which became apparent during or directly after the treatment period

Table 1 Proposed Effect classes to evaluate the treatment-related responses observed in aquatic micro- and mesocosm tests

Effect class	Description	Criteria
1	Effects could not be demonstrated ($NOEC_{micro/mesocosm}$)	• No (statistically significant) effects observed as a result of the treatment • Observed differences between treatment and controls show no clear causal relationship
2	Slight and transient effects	• Effects reported as 'slight' or 'transient', or other similar descriptions • Short-term and/or quantitatively restricted response of one or a few sensitive endpoints, and only observed at individual samplings
3A	Pronounced effects; recovery within 8 weeks after first application or total period of effects < 8 weeks	• Clear response of sensitive endpoints, but full recovery within 8 weeks after the first application, or total period of effects < 8 weeks • Effects reported as 'temporary effects on several sensitive species', 'temporary effects on less sensitive species/endpoints' or other similar descriptions • Effects observed at some subsequent sampling instances
3B	Pronounced effects; recovery within 8 weeks after last application	• Clear effects of sensitive endpoints, but full recovery within 8 weeks following the last application. In the case of repeated treatments, a total duration of the effects of > 8 weeks is possible, • Effects reported as 'temporary effects on several sensitive species', 'temporary effects on less sensitive species/endpoints' or other similar descriptions • Effects observed at some subsequent sampling instances
4	Pronounced effects; study too short to demonstrate recovery within 8 weeks after the last application	• Clear effects observed as in Effect class 3, but the study is too short to demonstrate complete recovery within 8 weeks after the (last) application
5A	Pronounced effects; total period of effects > 8 weeks and no recovery within 8 weeks after the last application; full recovery within the test period	• Clear response of sensitive endpoints and recovery time is longer than 8 weeks after the last application, • Full recovery is reported before the end of the study • Effects reported as 'long-term effects followed by recovery on several sensitive and less sensitive species/endpoints' or other similar descriptions • On consecutive time-points
5B	Pronounced effects; no recovery within 8 weeks after the last application, and no full recovery demonstrated within the test period	• Clear response of sensitive endpoints, and recovery time is longer than 8 weeks after the last application, • Full recovery is not reported before the end of the study • Effects reported as 'long-term effects followed by recovery on several sensitive and less sensitive species/endpoints'a or other similar descriptions • On consecutive time-points

The latter condition does not mean, for example, that delayed effects are per definition not treatment related. Delayed effects and indirect effects do, however, suggest that the results be evaluated with extra care. A dose-related delayed effect is handled in exactly the same manner as an undelayed effect. The assessment period starts at the moment of the onset of effects.
In the case of temporal decreases in abundance followed by recovery and overshooting, (higher abundance than in the control), the duration of both the increases and decreases should be reported in the evaluation. If the sum of the periods for increases and decreases is more than 8 weeks, Effect class 5 would be assigned.
To assess whether recovery has occurred, the trend should be taken into account, and the effect parameters in the treatments should have returned to the level of the control, preferably at two consecutive sampling dates.
It is recommended that the results of the classification of treatment-related effects be presented in a table, of which an example is given in Box 2. The classification of the population effects is based on the most sensitive population of a certain group, or the most sensitive functional endpoint. When a table with the classified endpoints is already present in the study report, the evaluator should check the validity of the classification and modify the table if needed. The overall NOEC is the lowest concentration that has no significant effects on the population or community level on one or more consecutive sampling dates.
The evaluation of a mesocosm study is not a bookkeeping process; it is an evaluation carried out by scientists. In this sense, the criteria in Table 1 should be handled as guidelines; with good argumentation and a solid scientific basis, it should be possible – in special cases – to assign a certain effect to another class. One such example would be when the 8-week period for recovery was too long or too short, depending on the kind of effect and the life-cycle strategy of the organism). In such a case the argumentation should be given in detail.

Not only the duration of the effects but also the magnitude of the effects is important. It is proposed that the graph or table, in addition to including the information in Box 2, depicts the response of the most sensitive endpoint(s). This information allows the magnitude of the effects to be evaluated, which again enables the ecological/regulatory relevance of the observed effects to be interpreted.
The variation between replicates can greatly influence the detection of significant effects. In order to obtain an insight into this effect, the power of the statistical test or the variation between replicates should be clearly visible for the most important measurement endpoints. To this end, figures in which the variation is clearly presented can also be used.
The effects observed (as well as the derived NOECs and NOEAEC) should be expressed in terms of the ecotoxicologically relevant concentration (ERC). The nominal concentration should also always be given. The ERC is the concentration that correlates best with the treatment-related responses observed (for example, peak or TWA concentration in water of a depth-integrated water sample; peak, or TWA concentration in pore water in the top 10 cm of sediment). If the aim of the micro- or mesocosm experiment is to evaluate ecological risks of short-term exposure, nominal concentrations or

Box 2 Example of the summary of the Effect classes observed for several endpoints in the outdoor microcosm study with xxxx; ↓ indicates a downward trend; ↑ indicates an upward trend. *TWA*, Time weighted average, *PRC*, principal response curve

	Water concentration (µg a.s./L)				
Nominal concentration:	3	15	30	150	300
Measured peak concentration:	2.8	14.5	28	146	292
7-day TWA concentration:	2.5	12.9	24.9	130	260
21-day TWA concentration:	2.0	11.5	22.2	116	231
Species/group					
Chlorophyll a – periphyton	1	1	1	1	1↑
Chlorophyll a – phytoplankton	1	1	1	1	1↑
Periphyton (PRC)	1	1	1	1	1↑
Periphyton (populations)	1	1	1	1	2↑
Phytoplankton (PRC)	1	1	3A↓	3A↓	3A↓
Phytoplankton (populations)	1	1	3A↓	3A↑↓	3A↑↓
Zooplankton (PRC)	1	1	3A↓	5A↓	5B↓
Zooplankton (populations)	1	2↓	3A↓	5A↓	5B↓
Macroinvertebrate, sweep net (PRC)	1	1	1	1	4↓
Macroinvertebrate populations	1	1	1	1	5B↓

measured peak concentrations are commonly used to express the treatment-related effects observed. If the aim of the study is to evaluate long-term risks, peak to TWA concentrations may be used to express the treatment-related effects observed in the mesocosm test. The length of the time-window required to calculate this TWA concentration should be determined on the basis of the life-cycle and time-to-effect information of the species of concern. The choice of the length of the time-window of the TWA concentration is dependent on the toxic mode-of-action of the substance and the time-to-effect information available from the ecotoxicological tests (including latency). The subject of the choice of the TWA-length is described in more detail in (EFSA-PPR, 2005, EFSA-PPR, 2006b; Boesten et al., 2007).

A more in-depth discussion of the ERC is beyond the scope of this document. The reader is referred to the SETAC eLiNK workshop, which provides more guidance on the link between fate and effect.

If the lower tier information indicates that metabolites may potentially cause risk, it is recommended that the relevant metabolites be measured in the micro- or mesocosm study and that exposure concentrations be reported in the extended summary.

Item 4. The evaluation of a particular study is rounded off with the classification of the effects and, subsequently, the derivation of an assessment endpoint (NOEC, NOEAEC with the corresponding Effect class). Based on personal knowledge, however, the evaluator may add a separate Annex to the evaluation report in which he gives one or more suggestions for the use of these assessment endpoints in the risk management decision (for example, the use of an assessment factor, the meaning of the result of the higher tier study in relation to other test results etc., arguments for determination of

the RAC). A more detailed discussion on the derivation of the assessment endpoints is provided in Chapter 3 of this document. During the SETAC Ampere Workshop (Leipzig, April 2007), the subject of extrapolation from one mesocosm to another was discussed and a decision tree was proposed for the extrapolation from mesocosm to field.

An example of the process of deriving an NOEAEC from a study is presented in Box 3. This box illustrates how expert judgment can be formalized. As already stated, derivation of a NOEC does not mean that the parameters as formulated should always be used without any further consideration of the data. The overall power of the test, the possible occurrence of trends in the treatment-related effects and the sampling scheme should also be considered when deriving a (consistent) NOEC. One aspect worthy of attention when deriving the NOEAEC is whether the abundance in the treatment cosms recover to that of the control, or does the abundance of the control decrease to the level in the treatment cosms (see Item 3b). Furthermore, one should always question if a NOEC equal to the highest treatment level is not caused by the low numbers being too low to enable a statistical analysis.

Box 3. Example of assigning Effect classes, applying of expert judgment

A decision was made to express the treatment-related effects in terms of nominal concentrations, since the substance is a very fast dissipating pesticide, the measured peak concentrations resembled nominal concentrations and the aim of the study was to address risks due to short-term exposure.

NOECs measured in a study with macroinvertebrates (dosages applied: 0, 1,10 and 100 µg a.s/L); NOECs > 100 µg a.s./L are not listed). ↓ indicates the trend of effects observed at the next higher concentration; 10(↓) thus indicates that at 100 µg/L a decrease in the parameter was found compared to the control.

	Day after first treatment						**Day after second treatment (on day 29 after first dose)**							
Parameter	1	2	7	14	21	28	1	2	7	14	21	28	41	56
Species richness	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)		10(↓)	10(↓)		
Species diversity												10(↓)	10(↓)	
Multivariate analysis	1(↓)	1(↓)	1(↓)	10(↓)	10(↓)	10(↓)	1(↓)	1(↓)	1(↓)	1(↓)	10(↓)			<1(↓)
Taxon														
Taxon 1	1(↓)	1(↓)	1(↓)	10(↓)	10(↓)	10(↓)	1(↓)	1(↓)	1(↓)	10(↓)	10(↓)	10(↓)		
Taxon 2		10(↓)	1(↓)	10(↓)	10(↓)		1(↓)	<1(↓)	<1(↓)	<1(↓)	1(↓)		1(↓)	10(↓)
Taxon 3	<1(↓)	10(↓)	<1(↓)	1(↓)	1(↓)	<1(↓)	10(↓)	1(↓)	1(↓)		10(↓)	10(↓)		
Taxon 4	10(↓)	10(↓)			10(↓)	10(↓)	1(↓)		<1(↓)	10(↓)	10(↓)			
Taxon 5		1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)			<1(↓)	
Taxon 6									10(↓)					
Taxon 7	<1(↓)				1(↓)	1(↓)		1(↓)						

From the NOECs measured in the study (above) relevant/significant NOECs are derived below according to the following procedure: the NOEC is based on two or more statistical significant NOECs on consecutive sampling dates. When the consecutive NOECs differ, the highest value is taken, unless the lowest value is found on two or more consecutive dates. (e.g. taxon 2 and 4). One unique deviating value for a sample on the last sample date should be interpreted within the framework of the preceding results.

Derived consistent NOECs in the study with macroinvertebrates.

	Day after first treatment						**Day after second treatment (on day 29 after first dose)**							
Parameter	1	2	7	14	21	28	1	2	7	14	21	28	41	56
Species richness	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)	10(↓)		10(↓)	10(↓)		
Diversity												10(↓)	10(↓)	
Multivariate analysis	1(↓)	1(↓)	1(↓)	10(↓)	10(↓)	10(↓)	1(↓)	1(↓)	1(↓)	1(↓)	10(↓)			*
Taxon														
Taxon 1	1(↓)	1(↓)	1(↓)	10(↓)	10(↓)	10(↓)	1(↓)	1(↓)	1(↓)	10(↓)	10(↓)	10(↓)		
Taxon 2		10(↓)	10(↓)	10(↓)	10(↓)		1(↓)	<1(↓)	<1(↓)	<1(↓)	1(↓)		10(↓)	10(↓)
Taxon 3	10(↓)	10(↓)	1(↓)	1(↓)	1(↓)	1(↓)	10(↓)	1(↓)	1(↓)		10(↓)	10(↓)		
Taxon 4	10(↓)	10(↓)			10(↓)	10(↓)	10(↓)		10(↓)	10(↓)	10(↓)			
Taxon 5		1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)	1(↓)				
Taxon 6														
Taxon 7					1(↓)	1(↓)								

* the value of the multivariate analyses of <1 on day 56 is considered not relevant (unique value not in line with preceding values).

Summary of the Effect classes observed for several endpoints at test termination in a study with macroinvertebrates.

	Nominal concentration (µg a.s./L)		
Parameter/taxon	1	10	100
Species richness	1	1	3B(↓)
Species diversity	1	1	3B(↓)
Multivariate analysis community	1	3A(↓)	3A(↓)
Taxon abundance	3A(↓)	3B(↓)	5B(↓)

Summary of endpoints in a study with macroinvertebrates.

Group	**NOEC (µg a.s./L)**	**NOEAEC (µg a.s./L)**
Macroinvertebrates taxa	<1	1
Macroinvertebrates community level	1	10

The nominal NOEAEC of 1 (µg a.s./L) is used for risk assessment (two treatments at a 29-day interval). NOEAEC is based on the acceptance of full ecosystem recovery within 8 weeks after first application (Effect class 3A)

3. COMMENTS ON THE USE OF TEST RESULTS IN RISK ASSESSMENT

In this chapter no specific instructions are given on how the results of the (semi-)field study in risk assessment should be used, rather a number of subjects are discussed and suggestions for handling these items are given.

As a general starting point it is stated that the more detailed the formulation of the problem arising from the lower tier, the clearer it is whether the higher tier study answers the concern of the lower tier one. In order to increase the relevance of the results, lower tier information should be given the proper attention.

The higher tier assessment may focus on a typical area of concern that has been identified – for example, long-term exposure due to different emission routes (drift, drainage, surface run-off) or the typical physico-chemical properties of the compound. One option is to continue the original risk-based approach and to reconsider a number of the assumptions underlying the initial risk assessment. Another option is to deliver more data that will allow reducing the uncertainty in the assessment. When the focus is on a single source of uncertainty, the assumption is that the exposure assessment and the effect assessment performed earlier were either fully validated or relatively worst-case. The Scientific Committee of the European Food Safety Authority, the Panel on Plant Protection Products and their Residues (PPR Panel), has formulated the principle that 'there is a need to evaluate and describe how the various assumptions and refinements used in an assessment affect the overall level of protection, taking into account all elements of the assessment, including fate, exposure and ecotoxicology' (EFSA-PPR, 2006a). The appropriate risk assessment should pay sufficient attention to the link between exposure and effects in the different tiers (see, for example, EFSA-PPR, 2005; EFSA-PPR, 2006b; Boesten et al., 2007).

In the aquatic compartment, (semi-)field studies are directed towards population and community level assessments, although these often exclude fish. There are a number of guidelines currently available for the performance and reporting of (semi-)field studies in which the study design is described in detail (for example, for the number of replicates, dosages and the statistical elaboration of the results). The design of the field study may have a major impact on the sensitivity of the test, and, consequently, on the reliability and usefulness of the results (Liess et al., 2005). However, while the regulatory benchmark (consisting of the definition of endpoints and the fixed TER values) is still at hand for the refinement of the first tier assessment, in the (semi-)field study approach, the regulatory criteria to decide upon are less clearly described. When a different risk assessment paradigm is followed, some basic questions are raised:

1. what is an unacceptable impact on the viability of the exposed species – directly or indirectly; what are unacceptable effects?
2. how is the absence of these unacceptable effects going to be established under field conditions?

3. when has it been sufficiently demonstrated that these effects are really absent?

In light of these three questions, an important dilemma of the current risk assessment procedure should be mentioned. The validity of the assumption that the lower-tier risk assessment procedures (standard test species approach) result in concentrations that do not cause unacceptable effects on populations and communities of freshwater organisms can only be calibrated/validated by performing (semi-)field experiments. According to Van den Brink (2006) it is a blessing in disguise that the quest for the 'acceptability' criteria become clear in higher tier studies. In other words, aquatic micro- and mesocosm tests that focus on population and community responses (including indirect effects and recovery) correspond better with the Uniform Principles protection goal to avoid 'long-term repercussions for the abundance and diversity of non-target species'.

3.1 What are unacceptable effects?

When it is ultimately concluded that the higher tier study did address the first tier concern and the particular mesocosm study can be used as representative for the field situation of concern, the regulator has to decide whether the higher tier study did show that no unacceptable effects occurred in practice. In this document we will limit the discussion here to the remark that the effects found in the mesocosm study should be linked to the protection goals. A discussion on the principles that may be used for temporally and spatially differentiated ecological protection goals is provided by Brock et al. (2006) and Van der Linden et al. (2006).

It should be noted that the registration of pesticides is regulated in EU directive 91/414/EEC, while the Water Framework Directive (WFD) 2000/60/EC may set environmental quality standards for pesticides. At the EU level differences do exist between the methods used during pesticide registration and those used for deriving an Environmental Quality Standard (EQS). With the aim of harmonizing the demands of pesticide registration and standard-setting within the context of the Water Framework Directive, a working group in the Netherlands is drafting a decision tree for surface water. This working group initiated a study of the juridical relationship between both directives because it is this relationship that determines the question of whether the registration policy is WFD-proof. The results of this juridical research are described in Van Rijswick and Vogelezang-Stoute (2007). In the Netherlands the government has laid down in the pesticide regulation that, as a result of an application, the concentration in surface water should not exceed the specifically defined environmental standards (RUMB, 2000, RUUBg, 2005).

When the effects are classified in aquatic semi-field studies, it is common practice to also incorporate the recovery after perturbation. A standing practice that has been developing within the EU member states is that scientists of regulatory authorities decide which effects or Effect classes are acceptable (EU, 2002). There has been no public debate on these standards or on critical effect values (Crane and Giddings, 2004;

De Jong et al., 2005; Montforts and De Jong, 2007). The role of regulators and other stakeholders is to evaluate whether a benchmark has been reached and to function as scientists in contributing to the acceptability debate. Risk managers of governmental agencies and regulatory authorities have the role of determining where the benchmark ought to be (a regulatory competence). There is no guarantee that regulators and risk managers share the same notion of acceptability as other stakeholders in society. Although such a public debate has not taken place it can be anticipated that it would rely heavily on scientific data on the impact of chemical stress on the structure and functioning of ecosystems, on insight in the factors that influence the sustainability of ecosystems and on the role these ecosystems play for society (ecosystem services) (see Brock et al., 2006).

Several research areas have been identified, and the results of these may be used to inform risk managers (and the public) on principles that can be used to set protection goals, such as representativeness, extrapolation, recovery, amplitude and scale of impact (Brock et al., 2006; Montforts and De Jong, 2007). It is clear that recovery is an essential concept in the current strategies for the interpretation of field studies. Effects and subsequent recovery can be observed when samples are taken over a prolonged period. At least three sampling moments post-treatment are needed to observe effects and recovery. In Table 1, a recovery period of eight weeks is chosen to make a distinction between several Effect classes. This procedure is in accordance with the EU guidance document on aquatic ecotoxicology (EU, 2002). In their review papers that introduced the Effect classes, Brock et al., (2000a, 2000b) chose a period of 8 weeks because in most micro- and mesocosms evaluated, macroinvertebrates were sampled at intervals of 2–4 weeks. Consequently, a period of 8 weeks after the last application allows a few sampling dates to show recovery. It should be noted, however, that choosing the acceptable time-frame for recovery, which may be different for different taxonomic groups, is actually a risk management decision. One could also say that the job of the risk manager is to define the acceptable risk level for the occurrence of adverse effects on the sustainability of the populations of non-target species (for example, 70% temporal reduction within one season). Scientists should then define practical criteria, such as taxon-specific threshold values for the maximum duration of effect, which guarantee the desired level of protection. In future, the Effect classes mentioned in Table 1 can be adapted accordingly, based on the consensus reached among risk managers, possibly with different recovery periods for different (groups of) species. In addition, the interpretation of recovery in micro- or mesocosm tests involving a single pesticide should be evaluated in relation to the use of the total package of pesticides in the field – against the background of the integral agricultural management of a given area (see, for example, Van Wijngaarden et al., 2004; Arts et al., 2006).

The abundance of aquatic field studies has been instrumental in validating the TER approach. There is research indicating that the TER approach is protective at the effect level of slight transient effects for certain types of plant protection products (Brock et al., 2006). Has legislation thus codified the level of acceptability at this specific effect class (1 and 2)? Another, less inferential, procedure would be that risk managers

and other stakeholders define which effects are deemed acceptable; the test design (number of replicates) and reporting would then be derived from this definition. This procedure, however, would affect all tiers in the risk assessment, including the standard test species approach and the species sensitivity distribution (SSD) approach.

3.2 Has it been demonstrated that the unacceptable effects are actually absent?

Provided the research is addressing the problem formulation, there remains the question of covering spatio-temporal variability in sensitivity under field conditions. Apart from conducting (semi-)field studies over a whole range of conditions, the assessor faces the difficult task of extrapolating:

i) can the effects of the one mesocosm be extrapolated to the other?
ii) can the effects of the mesocosm be extrapolated to the real field situation of concern?

One practical solution to handling the spatio-temporal variability in sensitivity between different micro- and mesocosm experiments and between these test systems and the field is the application of an assessment factor that is dependent on the amount of information available (Van Wijngaarden et al., 2005; Brock et al., 2006; Montforts and De Jong, 2007). This aspect will be expanded upon below.

The results of several model ecosystem experiments performed with the same insecticide have revealed that the threshold level for no (Effect class 1) or slight (Effect class 2) effects are remarkably consistent – at least for short-term (single or repeated pulses) exposure regimes (see data on chlorpyrifos and lambda-cyhalothrin (Brock et al., 2006)). Whether this is also the case for compounds with other modes of action and for long-term exposure regimes needs to be investigated. Data available for the herbicide atrazine (Brock et al., 2006) suggest a larger variability in Effect classes 1–2 between experiments under long-term exposure regimes. Brock et al. (2006) reported that threshold levels for effects (Effect classes 1–2) can be predicted with lower uncertainty than, for example, Effect classes 3–5. One explanation is that factors such as indirect effects and recovery of affected endpoints are influenced by spatio-temporal variation in species composition and by the ecological infrastructure (for example, connectivity between water bodies) of the surroundings. The studies presented in Brock et al. (2006) for chlorpyrifos and lambda-cyhalothrin indicate that for short-term exposure regimes (single or repeated short-term pulses) and in the case of only a single high quality micro- and mesocosm study being available, an assessment factor of 3 may be necessary for the spatio-temporal extrapolation of Effect class 3 NOEAEC to ensure that at this short-term concentration level no class 4–5 effects will occur in various field situations. Effect classes 1–2 concentrations may be used without the application of an additional assessment factor. In case of the data presented for atrazine in Brock et al. (2006), an assessment factor of 3 may be necessary for the spatio-temporal extrapola-

tion of Effect class 2 NOEAEC in order to assure that at this chronic concentration level no class 3–5 effects will occur.

It should be noted, however, that the derivation of the RAC and the choice of the height of the assessment factor to be applied to a study specific NOEAEC are risk management decisions.

It should also be noted, however, that the above-mentioned assessment factors are based on a limited number of compounds, all of which are insecticides and herbicides. Other assessment factors may be required for other compounds, such as fungicides, that may have a less specific mode of action.

A number of aspects relating to the second question 'can the effects in the mesocosm be extrapolated to the real field situation of concern?' are discussed in detail here. In practice, there will be differences and similarities between the situation in the mesocosm and that in the field situation of concern. In general, the more similar the test system is to the field situation of concern, the higher its usefulness for risk assessment. The differences may result in either an over- or underestimation of the response of the field ecosystem.

- Species composition. The more similar the composition in the mesocosm is to that in the field, the more probable it will be that the effects are predicted in the right way. This, however, does not mean that the species composition in the micro- and mesocosm experiment should be exactly the same as that in the field; it is more important that a sufficient number of representatives of the sensitive taxonomic groups are present. For many pesticides this largely depends on the specific toxic mode-of-action of the substance. Maltby et al. (2005) revealed that taxonomy plays a more important role than habitat and geographical region in predicting the sensitivity of water organisms to pesticides with a specific toxic mode-of-action, Furthermore, the representativeness of the biological traits (for example, recovery potential) of the tested species for other relevant species is important. In general, vertebrates are not incorporated in the mesocosm studies. In the case vertebrates belong to the most sensitive group, it is clear that a mesocosm study without vertebrates is not the appropriate test system.
- External recovery. In most micro- and mesocosm tests, the migration and/or recolonization of organisms that complete their entire life-cycle in water is, in general, hampered because of the isolated character of these test systems. Under field conditions, migration/recolonization of the organism may compensate for potential effects in freshwater ecosystems such as streams and drainage ditches. The definition of taxon-specific and habitat-specific acceptable/critical effect levels may help to reach consistent management decisions.
- Avoidance. Examples are known from the literature (for example, *Gammarus pulex*; see Schulz and Liess, 1999) of organisms that temporarily avoid toxic substances by moving to parts of the compartment with lower concentrations. Other organisms do not have the possibility to avoid contact with the substance. In general, avoidance of toxic stress is hampered in isolated test systems that are treated for 100% of the surface (as is usually done in micro- and mesocosm tests).

- **Environmental conditions and system properties.** Aspects such as nutrient availability, climatic condition, weather conditions, substrate, multi-stress and mixture toxicity could influence the effects.

In the previous chapter a new proposal was made on how to measure the result of the (semi-)field test in terms of response (effect classes). The other crucial issue is the measurement of exposure. The PPR Panel concluded that the ecotoxicological endpoint from a study with a time-varying exposure should be expressed in terms of the ecotoxicologically relevant concentration (ERC) (EFSA-PPR, 2006a), which is the concentration that is most relevant for the risk assessment from the ecotoxicological point of view (for example, a peak or a TWA concentration in surface water for water organisms, in food, in the interstitial water or in the top centimetre of the total sediment). It can be defined using time-to-event information obtained from the available ecotoxicity studies as well as knowledge of the mode of action. After the ERC has been defined, it acts as the interface between the exposure and effect assessments, allowing flexibility with respect to the level of sophistication of both assessments. Additionally, the PPR Panel and Boesten et al. (2007) developed a generic procedure for comparing the time course of the ERC in ecotoxicological studies to that in the field. The proposed approaches appeared to work well for the dimoxystrobin and cyprodinil risk assessments (EFSA-PPR, 2005; EFSA-PPR, 2006b).

REFERENCES

Arnold, D., Hill, I., Matthiesen, P., and Stephenson, R. 1991. Guidance document on testing procedures for pesticides in freshwater static mesocosms - from the workshop "A meeting of experts on guidelines for static field mesocosm tests", Monks Wood Experimental Station, Abbotts Ripton, Huntingdon, UK. SETAC-Europe, Brussels, Belgium.

Arts, G.H.P., Buijse-Bogdan, L.L., Belgers, J.D.M., Van Rhenen-Kersten, C.H., Van Wijngaarden, R.P.A., Roessink, I., Maund, S.J. and Brock, T.C.M. 2006. Ecological impact in ditch mesocosms of simulated spray drift from a crop protection program for potatoes. *Integrated Environmental Assessment and Management* **2**, 105-125.

Boesten, J.J.T.I., Köpp, H., Adriaanse, P.I., Brock, T.C.M. and Forbes, V.E. 2007. Conceptual model for improving the link between exposure and effects in the aquatic risk assessment of pesticides. *Ecotoxicology and Environmental Safety* **66**, 291-308.

Boxall, A., Brown, C. and Barrett, K. 2001. Higher-tier laboratory aquatic toxicity testing. Cranfield Centre for Eco Chemistry Research, Cranfield, UK.

Brock, T.C.M., Arts, G.H.P., Maltby, L. and Van den Brink, P.J. 2006. Aquatic risks of pesticides, ecological protection goals and common claims in EU legislation. *Integrated Environmental Assessment and Management* **2**, E20-E46

Brock, T.C.M., Lahr, J. and Van den Brink, P.J. 2000a. Ecological risks of pesticides in freshwater ecosystems. Part 1. Herbicides. Alterra, Wageningen, The Netherlands.

Brock, T.C.M., Van Wijngaarden, R.P.A. and Van Geest, P.J. 2000b. Ecological risks of pesticides in freshwater ecosystems. Part 2: Insecticides. Alterra, Wageningen, The Netherlands. Report 089.

Campbell, P. J., Arnold, D. J. S., Brock, T. C. M., Grandy, N. J., Heger, W., Heimbach, F., Maund, S. J., and Streloke, M. 1999. Guidance document on higher-tier aquatic risk assessment for pesticides (HARAP), from the SETAC-Europe/OECD/EC Workshop, Lacanau Océan, France, SETAC-Europe, Brussels, Belgium.

Crane, M. and Giddings, J.M. 2004. "Ecologically Acceptable Concentrations" when assessing the environmental risks of pesticides under European Directive 91/414/EEC. *Human and Ecological Risk Assessment* **10**, 733-747.

Crossland, N.O. and La Point, T.W. 1992. Symposium on aquatic mesocosms in ecotoxicology. *Environ. Toxicol. Chem.* **11**.

De Jong, F.M.W., Mensink, B.J.W.G., Smit, C.E. and Montforts, M.H.M.M. 2005. Evaluation of ecotoxicological field studies for authorization of plant protection products in Europe. *Human and Ecological Risk Assessment* **11**, 1157-1176.

EFSA-PPR 2005. Opinion of the PPR Panel on a request from EFSA related to the evaluation of dimoxystrobin. *The EFSA Journal* **178**, 1-45.

EFSA-PPR 2006a. Opinion of the PPR Panel on the scientific principles in the assessment and guidance provided in the area of environmental fate, exposure, ecotoxicology, and residues between 2003 and 2006. *The EFSA Journal* **360**, 1-21.

EFSA-PPR 2006b. Opinion of the Scientific Panel on Plant health, Plant protection products and their Residues on a request from the EFSA related to the aquatic risk assessment for cyprodinil and the use of a mesocosm study in particular. *The EFSA Journal* **329**, 1-77.

EU 2002. Guidance document on aquatic ecotoxicology in the frame of the Directive 91/414. EU (DG Health and Consumer Protection), Brussels, Belgium. SANCO/3268/2001.

EU 2004. European Union System for the Evaluation of Substances 2.0 (EUSES 2.0). Prepared for the European Chemicals Bureau by the National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands, Available via the European Chemicals Bureau, http://ecb.jrc.it.

EWOFFT 1992. Summary and recommendations of the European workshop on freshwater field tests, Potsdam, Germany,

Giddings, J., Heger, W., Brock, T.C.M., Heimbach, F., Maund, S.J., Norman, S., Ratte, H.T., Schäfers, C., and Streloke, M. 2002. Community-Level Aquatic System Studies - Interpretation Criteria (CLASSIC), Fraunhofer Institute, Schmallenberg, Germany, SETAC, Pensacola, FL, USA.

Hill, I.R., Heimbach, F., Leeuwangh, P. and Matthiessen, P. (Eds.) 1994. Freshwater field tests for hazard assessment of chemicals. Lewis Publishers, Ann Arbor, USA.

Kersting, K. and Van Wijngaarden, P.A. 1999. Effects of a pulsed treatment with the herbicide afalon (active ingredient linuron) on macrophyte-dominated mesocosms. I. Responses of ecosystem metabolism. *Environmental Toxicology and Chemistry* **18**, 2859-2865.

Liess, M., Brow, C., Dohmen, P., Duquesne, S., Hart, A., Heimbach, F., Keuger, J., Lagadic, L., Maund, S., Reinert, W., Streloke, M. and Tarazona, J. (Eds.) 2005. Effects of pesticides in the field. SETAC, Brussels, Belgium.

Maltby, L., Blake, N., Brock, T.C.M. and Van den Brink, P.J. 2005. Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems. *Environmental Toxicology and Chemistry* **24**, 379-388.

Montforts, M.H.M.M. and De Jong, F.M.W. 2007. Field studies in pesticide registration: questioning the answers. *Integrated Environmental Assessment and Management* **3**, 150-153.

OECD 1995. OECD workshop on environmental hazard/risk assessment, **105**, OECD, Paris, France.

RUMB, 2000. Regeling uitvoering milieutoelatingseisen bestrijdingsmiddelen 2000. *Staatscourant* **114**, 18. 16 juni 2000.

RUUBg, 2005. Regeling uitwerking uniforme beginselen gewasbeschermingsmiddelen. *Staatscourant* **248**, 35. 21 december 2005.

Schulz, R. and Liess, M. 1999. Validity and ecological relevance of an active in situ bioassay using *Gammarus pulex* and *Limnephilus lunatus. Environmental Toxicology and Chemistry* **18**, 2243-2250.

SETAC-RESOLVE 1992. Workshop on aquatic microcosms for ecological assessment of pesticides, Wintergreen, Virginia, SETAC,

Van den Brink, P.J. 1999. Ecological and statistical evaluation of effects of pesticides in freshwater model ecosystems. Thesis Landbouwuniversiteit Wageningen.

Van den Brink, P.J. 2006. Letter to the Editor: Response to recent criticism on aquatic semifield experiments: Opportunities for new developments in ecological risk assessment of pesticides. *Integrated Environmental Assessment and Management* **2**, 202-203.

Van der Brink, J. and Ter Braak, C.J.F. 1999. Principal response curves: analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry* **18**, 138-148.

Van der Linden, A.M.A., Boesten, J.J.T.I., Brock, T.C.M., Van Eekelen, G.M.A., De Jong, F.M.W., Leistra, M., Montforts, M.H.M.M. and Pol, J.W. 2006. Persistence of plant protection products in soil; a proposal for risk assessment. RIVM, Bilthoven. Report 601506008.

Van Dijk, H.F.G., Brussaard, L., Stein, A., Baerselman, F., De Heer, H., Brock, T.C.M., Van Donk, E., Vet, L.E.M., Van der Gaag, M.A., Van Gestel, C.A.M., Van der Hoeven, N., De Jong, F.M.W., Van der Linden, A.M.A., Van Noort, P.C.M., Oomen, P.A. and Van Vliet, P.J.M. 2000. Field research for the authorization of pesticides. *Ecotoxicology* **9**, 377-381.

Van Rijswick, H.F.M.W. and Vogelezang-Stoute, E.M. 2007. De Kaderrichtlijn water en gewasbeschermingsmiddelen, Centrum voor Omgevingsrecht en Beleid/NILOS, Universiteit Utrecht.

Van Wijngaarden, R.P.A., Brock, T.C.M. and Van den Brink, P.J. 2005. Threshold levels of effects in freshwater ecosystems, a review. *Ecotoxicology* **14**, 353-378.

Van Wijngaarden, R.P.A., Cuppen, J.G.M., Arts, G.H.P., Crum, S.J.H., Van den Hoorn, M.W., Van den Brink, P.J. and Brock, T.C.M. 2004. Aquatic risk assessment of a realistic exposure to pesticides used in bulb crops: A microcosm study. *Environmental Toxicology and Chemistry* **23**, 1479-1498.

ANNEX 1 CHECKLIST FOR EVALUATING AQUATIC MICRO- AND MESOCOSM TESTS

In this annex a checklist is provided that can be used for summarizing an aquatic micro- or mesocosm study. An essential element in the checklist is the reliability index that can be used for valuing the different aspects that are deemed important for the interpretation of the study. The criteria listed in Table A1.2 should not be considered to be absolute since exceptions are always possible. However, if Ri3 (Reliability Index) is indicated for a specific item, any deviation from this classification should be explicitly described and substantiated with arguments. It should further be noted that despite the introduction of the reliability index, expert judgement will always be needed to value the reliability of the study as a whole.

Reliability index

The reliability is assessed by assigning a Reliability Index (Ri) to a particular (aspect of the) test: Ri1 indicates a reliable test, Ri2 indicates a less reliable one and Ri3 indicates an unreliable test (see Table A1.1). The definition of reliability is: the intrinsic quality of a test with respect to the methodology and the description (EU, 2004). Ri3 tests are not used for risk assessment.

Table A1.1 Definition of the three values of the reliability index

RELIABILITY INDEX (Ri)	DEFINITION	DESCRIPTION
1	Reliable	All data are reported, the methodology and the description are in accordance with internationally accepted test guidelines and/or the instructions, all other requirements fulfilled
2	Less reliable	Not all data are reported, the methodology and/or the description are less in accordance with internationally accepted test guidelines or the instructions, without motivation, or not all other requirements fulfilled
3	Not reliable	Essential data are missing, the methodology and/or the description are not in accordance with internationally accepted test guidelines and/or the instructions without motivation, or not reported, or important other requirements are not fulfilled

Both Ri1 and Ri2 tests can be used for risk assessment, but it depends on the overall data availability whether only Ri1 tests should be used, or whether Ri2 tests can be used as well.

A checklist for aquatic field studies is given in Table A1.2, followed by an explanation and specification. An ‘E’ in Table A1.2 indicates that expert judgement should be applied to judge the impact of the shortcoming on the reliability; a ‘Y’ indicates that the shortcoming renders an aspect of the test unreliable (Ri3) or less reliable (Ri2). In some cases, the shortcoming is deemed essential and the test is indicated as unreliable (Ri3)

based on the particular item. For example, if a test design does not allow a proper statistical analysis, this is sufficient to label the whole test as unreliable. All other items in the checklist with 'Y' render the test less reliable (Ri2), if not in compliance with the quality standards. However, it depends on expert judgement whether, for example, 1–4, 5–10 and ≥11 of such 'yes' items designate the test as a whole as being reliable, less reliable or unreliable, respectively. This is primarily because the test items may have a different "weight" in the overall judgement.

Checklist

The criteria in Table A1.2 are partly based on or derived from Hill et al. (1994) and Brock et al. (2000a, 2000b). The criteria below can be seen as minimum requirements for microcosm and mesocosms tests:

- The experiment and the methodology are well documented without vagueness.
- The test-system is composed of (parts of) a realistic community: relevant trophic levels are present, and these contain the target organisms and, preferably, decomposers and consumers as well.
- The abiotic composition of the test-system should include the relevant components of the ecosystem to be protected.
- The relevant exposure concentrations during the experiment are reported or can be deduced from at least the nominal exposure concentrations applied. No interference of toxic co-solvents can be expected.
- The endpoints that were taken into account are sensitive to the type of toxicant and are related to the mode-of-action of the substance; for example, primary producers are included for herbicides, arthropods for organo-phosphorous insecticides, etc.
- Effects are derived from an unambiguous concentration–response relation.
- Preferably, the lowest concentration tested should not show a consistent treatment related effect.
- Statistically significant effects should be found at least at the highest test concentration. In a test where no effects are found – even at the highest concentration – it cannot be excluded that some unmeasured circumstances cause the lack of effect.
- The decision on the acceptability of the effects is part of the risk management domain.

The set of tools for a refined higher tier risk assessment is described in the HARAP report. For an elaborate description of the methods, the reader is referred to Campbell et al. (1999).

For the interpretation and valuation of micro- and mesocosm tests a checklist is provided in Table A1.2.

Table A1.2 Checklist for the aspects that generally are considered to be of importance when evaluating aquatic micro- and mesocosm tests

TEST ITEMS		NOTES	RELIABILITY LOWER?
METHODOLOGY & TEST DESCRIPTION			
1.	Substance	Properly characterized and reported?	
	1.1Concentration	[identity and amount of a.s. per litre test water not reported]	Y [→Ri 3]
	1.2 Formulation and purity	[ingredients in the formulation influencing the working action of the a.s. should be reported]	E
	1.3 Vehicle	[in case a vehicle – other than in the formulation – is used, identity and concentration?]	E
	1.4 Chemical analyses	[method, LOQ, LOD, recovery, not reported]	Y [→Ri 2]
	1.5 Properties	[not reported]	Y [→Ri 2]
2.	Test site, duration	Properly characterized and reported?	
	2.1 Location	[necessary to make a link between the study and agricultural practice (effects, environmental conditions and the application method): representativeness]	E
	2.2 Soil type /substrate	[necessary to compare to the local conditions of concern; not reported?]	Y [→Ri 2]
	2.3 Test date / duration	[duration long enough to study recovery?]	E
	2.4 General climatic conditions	[necessary to make a link between the effects and local climatic conditions; not reported?]	Y [→Ri 2]
3.	Application	Properly characterized and reported?	
	3.1 **Mode of** application	[spraying or homogenizing the a.s. into the test medium; not reported]	Y [→Ri 2]
	3.2 **Dosage**	[actual concentrations during the test are most important; not reported?]	E
		[no chemical analysis of dosing solution and no actual concentrations]	Y [→Ri 3]
	3.3 **Application** scheme	[necessary to make a link between the test and the intended use of the pesticide; not reported]	Y [→Ri 2]
	3.4 **Condition of** application	[additional technical data, route under consideration; not reported]	Y [→Ri 2]
	3.5 **Climatological** conditions	[weather conditions during application, wind speed and temperature?]	E
4.	Test design	Properly designed and reported?	
	4.1 **Type & size**	[e.g. outdoor microcosm, outdoor pond or mesocosm; not reported]	Y [→Ri 2]
	4.2 **Test system**	[not properly reported?]	Y [→Ri 2]
	4.3 **Pre-treatment**	[no period reported, no proper equilibration?]	Y [→Ri 2]
	4.4 Post-treatment	[period, interval between treatments, not reported]	Y [→Ri 2]
	4.5 **Untreated control**	[insufficient number, invalid or improperly reported?]	Y [→Ri 3]
	4.6 **Replications**	[insufficient replications for proper statistical analysis?]	Y [→Ri 2]
	4.7 **Statistics**	[ECx's derived by regression, NOECs derived by ANOVA and preferably by multivariate techniques as PRC]	E
	4.8 **Dose-response**	[≥ 2 test concentrations for finding a dose-response relation (controls excl.)]	E

TEST ITEMS		NOTES	RELIABILITY LOWER?
	4.9 **Study under good** laboratory practices (GLP)	study not conducted under GLP ?	E
5.	Biological system	Representative and properly reported?	
	5.1 **Test organisms**	[e.g. species/taxa not reported?]	Y [→Ri 3]
	5.2 **Community**	[the community/ecosystem representative and complete?]	E
6.	Sampling	Is sampling adequate for risk assessment?	
	6.1 **General features**	[properties during test not monitored? E.g. pH, hardness, oxygen]	Y [→Ri 2]
	6.2 **Actual** concentration	[actual concentrations measured in medium and other compartments or biota?]	E
	6.3 **Biological** sampling	[no proper method, species, number, endpoints, frequency?]	Y [→Ri 2]
RESULTS			
7.	Endpoint	Properly reported?	
	7.1 **Type**	[relevant endpoints not chosen or specified]	Y [→Ri 3]
	7.2 **Value**	[results not based on measured data?]	Y [→Ri 3]
	7.3 **Verification of** endpoint	[test results are not verifiable? source data not reported]]	Y [→Ri 3]
8.	Elaboration of results	Are conclusions based on measured data? methodology correct?	
	8.1 **Statistical** comparison	[data do not meet requirements for method used?]	Y [→Ri 3]
	8.2 **Dose–effect** relationship	[not present?]	Y [→Ri 3]
	8.3 **Community level** statistics	[not reported? improper method?]	Y
9.	Control		
	9.1 Untreated control	[unexpected effects or disappearance of species?]	Y [→Ri 3]
	9.2 Positive control	[no clear effects in highest treatment or positive control]	Y [→Ri 3]
10.	Classification of effects	Not properly derivable?	Y [→Ri 2]
11.	Biological meaning of **statistically** significant differences	Insufficiently explained?	E
REMARKS			
The above-mentioned items concern the scientific reliability of a field study. The usefulness of a field test depends on the scientific reliability and purpose of use. One of the aspects for purpose of use is the similarity between the test situation and the situation of concern for the registration; for instance, the following test items must be checked: product, dosage/ concentration, application frequency, interval and type of ecosystem. The more similarity there is between the aspects found in the field test and the product under registration and its proposed conditions of use, the more likely it is that the field test is useful for risk assessment. However, general guidance is difficult to give, and expert judgement is therefore decisive, as the appraisal of the usefulness may differ from pesticide to pesticide. Although these considerations do not deal with the quality of the study, a remark can be made in the Remarks section. Other aspects that were not covered by the items mentioned above can also be introduced here.			

Item 1. The evaluator should take note of substances in the formulation that could influence the mode of action of the active substance. Chemical analyses of the active substance should be described in detail. Data on vapour pressure, water solubility, photolysis, hydrolysis, biodegradation, sediment and plant sorption can be useful in explaining the substance dissipation. The formulation should preferably be the same as the one under registration. However, when the formulation is different, or the technical substance is used, it should be shown that the test item is representative of the formulation under registration. This aspect does not affect the scientific reliability of the study, but it is of importance for the usefulness of the study. When a problem exists here, the evaluator should make a separate note in the 'Remarks' section.

Item 2. The test site and the conditions should be reported for two reasons. First, this information is needed to assess whether basic requirements are fulfilled and that no circumstances were present that would render the test less reliable. Second, these data are needed when, in a later stage of risk assessment, the usefulness has to be assessed; that is to say 'are the circumstances representative for the proposed use of the compound?'. For this reason these aspects should be present in the extended summary. The lack of these aspects, however, does not influence the scientific reliability of the study itself.

Item 3. A field test without any actual analyses of the active substance in the dosing solution and in the water of the treated micro- or mesocosm is considered to be unreliable, as the statistically or biologically significant effects cannot be linked with proper exposure analyses. The concentration should be measured initially and frequently during the test. (Large) differences between nominal and measured concentrations should be explained.
Whether, for example, concentrations in sediment or biota should be measured depends on the properties of the compound and the risks identified in earlier tiers (expert judgement), among others.
The exposure regime simulated in the test should be expressed in terms of the ERC. An important point that falls outside of the scope of this document is the link between the effects in the mesocosm and the exposure regimes found in practice (e.g. according to the FOCUS scenarios or measured field exposure concentrations.) Guidance for this aspect is currently being developed on the SETAC-eLiNK workshop.
The (micro)climatological conditions at the start and during the test are important. The presence of wind during the application may explain low recoveries in the water. Temperature and light conditions may influence dissipation processes to a large extent – if dissipation rates are temperature- or light-dependent. Heavy rainfall – when diluting the active substance substantially – may bias the test results.

Item 4. In terms of the test design, two workshops were organized in 1991 and 1992 that produced a guidance document for the design of microcosm and mesocosm studies (SETAC-RESOLVE, 1992, Arnold et al., 1991; EWOFFT, 1992, Hill et al., 1994). These documents can be used to obtain an idea of, for example, the size of the systems.

The test system should mimic a (semi-)natural outdoor ecosystem, with interactions at the intra-species, population, community and ecosystem level; the system should be pre-tested and proven to be 'stable': it should not collapse due to reasons other than applying a pesticide (for example, by overloading the system with fish).
Before treatment the system should have time for establishing a diverse and more or less stable community; the parameters of the system should be measured or monitored adequately.
Without untreated controls it is often impossible to interpret the results of a higher tier study. The results in the control are preferably represented with confidence limits, or when this is not possible due to a low number of replicates, the range should be given.
A dose–response design is recommended, with sufficient concentrations in a relevant range to obtain different effect levels and to analyse the dose–response relationship. There are preferably no effects at the lowest exposure level and clear-cut effects at highest exposure level.
An increasing number of field studies are conducted under good laboratory practice (GLP). The application of GLP places particularly high demands on the procedural aspects and the manner of reporting. This does not mean, however, that studies without GLP can per definition not be used for risk assessment. The acceptability of such a study depends on the description and the data in the study report.

Item 5. For a particular registration it should be checked whether the community/ecosystem in the mesocosm is representative and/or protective for the situation of the registration. It cannot be expected that a mesocosm study is conducted for every use of the product, or even for every country. The evaluator, however, should check whether the information presented can be used for local species in the area of concern. Both physiological and ecological aspects should be taken into account. Depending on the mode of actions of the substance studied, an indication could be given whether organisms not present in the mesocosm would be more susceptible. For example, in terms of the ecological aspect, the presence (or absence) of species with longer life cycles could influence the recovery. For these aspects more guidance has to be developed.
The presence of fish will have a strong impact on the ecosystem in the mesocosm and may ultimately affect the sensitivity of the study. When fish is the most sensitive taxon, a mesocosm study without fish is not suited as a higher tier; in this case only higher tier studies including fish are appropriate.

Item 6. Biological and physicochemical properties of the test system (including the water and sediment) should be monitored frequently as a monitoring of the pH, hardness, oxygen content (OC) and temperature will facilitate any interpretation of the test results. The OC content and eventually the thickness of a sediment layer should be reported. The frequency of sampling biota is preferably higher during and immediately after the application period and may depend on knowledge of the time-to-effect of the pesticide and life-cycle characteristics of the organisms of concern. For example, the frequency of sampling of a number of studies described in Van den Brink et al., (1999) are presented in Table A1.3.3. From this overview it is clear that, in general, the sam-

pling frequency during the first week following the last application is high, which is important for preventing short-term, acute effects from being missed. The table below is not a proposal for a sampling scheme; for example, although this was not carried out in the studies described by Van den Brink et al. (1999), it may be wise to monitor phytoplankton, even in the case of insecticides, in order to detect indirect effects.

Table A1.3 Example of a sampling frequency for different pesticides and groups of organisms

Type of pesticide	Type of exposure	Species composition Sampling date (weeks) for biota			
		Phyto-plankton	Periphyton	Zooplankton	Macroinverte-brates
Insecticide	Single dose; acute toxicity			-4, -1, 0.1, 1, 2, 4, 8, 12, 15, 19, 24, 42, 51, 55	0.1, 1, 2, 4, 8, 12, 15, 19, 24, 42, 51, 55
Herbicide	Chronic dose; chronic toxicity	-1, 0, 1, 2, 3, 4, 6, 8, 10	-1, 2, 4, 6, 8, 10	-1, 1, 3, 5, 7, 9, 11	1, 1, 3, 5, 7, 9, 11
Fungicide	Chronic dose; chronic toxicity	-3, -1, 0, 1, 2, 3, 4, 5, 6, 7, 9	-2, -1, 1, 3, 5, 7, 9	-3, -1, 0, 1, 2, 3, 4, 5, 6, 7, 9	-3, -1, 1, 3, 5, 7, 9

Item 7. Particularly relevant endpoints are those that are closely related with population, community or ecosystem dynamics [for example, mortality endpoints, endpoints respecting growth or reproduction, intraspecific (such as population growth), and interspecific endpoints and endpoints respecting biodiversity, primary and secondary production, food web interactions and resilience]. $NOEC_{population,\ community}$ values can therefore be based on different structural and functional parameters. The endpoint on which the $NOEC_{population,\ community}$ is based should therefore be clearly reported; consequently, all of the investigated endpoints and those that are the most sensitive – on which the $NOEC_{population,\ community}$ is finally based – should be reported.
As there are generally fewer replicates in the field than in laboratory tests, the variability between replicates is more important; consequently, information on the reliability and/or variability of endpoints should be reported, and this can be used to derive the minimum detectable difference.
Verification of conclusions by data recalculation increases the reliability of the field test; however, this is not possible when raw data are not reported.
It should be clear whether the effects are correlated with nominal or actual concentrations.

Item 8. Under normal conditions clear effects should be found in the highest treatment dosages, so that these dosages act as a positive control. The 'positive control' can be used to show that exposure took place and that effects were found. The test is not valid if no effects are found in a positive control because unforeseen circumstances may have been the reason that no effects are found.
Given the complexity of the communities in the mesocosms, it is desirable to include an up-to-date multi-species analysis (for example, a PRC; (Van der Brink and Ter Braak, 1999; Van den Brink, 1999).

Item 9. According to the EU guidance document on Aquatic Ecotoxicology (EU, 2002), effects in micro- or mesocosm studies may be classified according to Brock et al. (2000a, 2000b). For the adapted classification in which the duration of the effects is given particular attention as an extra criterion, see Table 1 (Chapter 2).

Item 10. The interpretation of a statistical significant effect should be explained using statistical, ecological and ecotoxicological data. For this interpretation use can be made of:

A data from lower tier studies
B the presence of a dose–effect relation: if such a relation exists it is much more probable that a consistent significant difference is actually due to the substance
C univariate and multivariate statistics of the micro- or mesocosm datasets
D knowledge of ecological relationships between species inhabiting the test system. Based on the available laboratory data, the use category (herbicide, insecticide, etc.) and/or the mode of action of the substance, it may be expected that some species or endpoints show a (temporary) effect. Unexpected effects on species or functional endpoints that were initially not identified as sensitive should, however, be considered as well and should be explained by ecological processes or the compound properties.

ANNEX 2 EXAMPLE OF AN EVALUATION OF AN AQUATIC MICROCOSM STUDY

Sensitivity of macrophyte-dominated freshwater microcosms to chronic levels of the herbicide linuron.

Disclaimer: the summary of the field study with linuron as presented below is an example of a summary as a result of the guidance for summarizing of Aquatic Micro/Mesocosm studies. The summary is not subject to the normal quality procedures handled at the registration procedure, and no consequences can be drawn from the conclusions of this evaluation.

A 2.1 Header Table and Abstract

Reference	: **Van den Brink et al. (1997); Cuppen** et al. (1997)	GLP statement	: **no**
Type of study	: **aquatic microcosm**	Guideline	: **not specified**
Year of execution	: **1994**	Aceptability	: **acceptable**
Test substance	: **Afalon**		

Substance	**Method**	**Endpoints**	**pH**	**Duration (days)**	**Criterion**	**Value (µg a.s./L)**	**Ri**
Afalon (xx g/l linuron)	Indoor microcosm	Phytoplankton, periphyton, zooplankton, macroinvertebrates, macrophyte biomass, decomposition, community metabolism, chlorophyll a	7.6–10.6	28 (treatment) 49 (post-treatment)	NOEC NOEAEC	0.5 > 0.5, < 5	2

Summary of

Sensitivity of Macrophyte-Dominated Freshwater microcosms to Chronic Levels of the Herbicide Linuron.

1. Primary producers. Authors: P.J. van den Brink, E.M. Hartgers, U. Fettweis, S.J.H. Crum, E. van Donk & T.C.M. Brock. Ecotoxicology and Environmental safety 38, 13-24 (1997)
2. Community Metabolism and Invertebrates. Authors: J.G.M. Cuppen, P.J. van den Brink, H. van der Woude, N. Zwaardemaker & T.C.M. Brock. Ecotoxicology and Environmental safety 38, 25-35 (1997)

Abstract

Effects of chronic concentrations of Afalon (a.s. linuron) 0, 0.5, 5, 15, 50, and 150 µg linuron/L were studied in indoor, macrophyte dominated freshwater microcosms. The concentration of linuron was kept constant during 28 days, and endpoints were meas-

ured during this period and 7 weeks after this period. The endpoints studied were community metabolism (dissolved oxygen, pH, conductivity, alkalinity, nitrate), decomposition, chlorophyll *a* (phytoplankton, periphyton, neuston), biomass *(Elodea nuttallii)* and abundance (periphyton, phytoplankton, zooplankton, macroinvertebrates taxa). The NOEC of 0.5 µg linuron/L is based upon the most sensitive endpoints, being the effect on photosynthesis as reflected in dissolved oxygen and pH, the abundance of *Chroomonas* and the effect on biomass of *Elodea nuttalli* as observed in a biotest. Since at 5 µg/L linuron (the treatment level above 0.5 µg linuron/L) long-term effects without full recovery were observed on the abundance of periphyton (particularly *Cocconeis*), it is not possible to derive a NOEAEC for long-term constant exposure, but it can be indicated that the NOEAEC from this study is < 5 µg linuron/L and > 0.5 µg linuron/L (the NOEC).

Remarks
The presented data set of the study is incomplete and therefore the reliability is judged as Ri2.

A 2.2 Extended summary

Reference
Van den Brink et al., 1997, Cuppen et al., 1997.

Guidelines
Not specified

Test design
Twelve indoor microcosms were used in a duplicated design with five doses and a duplicated control. Microcosms consisted of glass aquaria (l _h _ w = 1.1 _ 0.7 _ 1.1 m; water volume 600 L). The aquaria were filled with 10 cm of lake sediment and a 50-cm water column. Sediment, well water and organisms were added 3 months prior to treatment (15 February 1994). Organisms were added with the sediment and with the water column above the sediment, and characteristic organisms for Dutch ditches were introduced, as was the macrophyte *Elodea nuttallii*. Origin of the species and the species added are not described. Cosms were interconnected during the 3-month acclimatization period. Before the start of the experiment, microcosms were disconnected. Linuron concentration was kept constant during 4 weeks, and effects were measured for 7 weeks after the treatment period.

Application, concentrations, replicates
Linuron was applied as Afalon (no specifications). Nominal dosages of linuron (0.5, 5, 15, 50 and 150 µg/L) and an untreated control were used in a duplicated design. The test substance was distributed evenly over the water surface and mixed by stirring. Water was circulated in the cosms, and linuron was added twice a week to compensate for losses, as determined from measured concentrations.

Biological observations

Phytoplankton samples were depth-integrated, and per cosm several samples were taken using perspex tubes until 1 L of sample was obtained. Species were identified and cells were counted. Chlorophyll *a* was sampled using a second litre of water per cosm. For the sample scheme see Table A1.1.

Periphyton was obtained from an artificial substrate (six glass slides per sample), which was incubated for 8 weeks. Species were identified and cells counted. For sample scheme, see Table A1.1. From the description it is not quite clear whether all samples had the same incubation period or whether the incubation period was related to the first sampling date. The chlorophyll-*a* content was also analysed in the periphyton using another six slides – see Table A1.1. Periphyton for chlorophyll-*a* analyses was also obtained from the ten top 10-cm-long shoots of *Elodea nuttallii*; the results were expressed as milligram chlorophyll *a* per gram dry weight. As an indication for short-term effects, the bio-volume of the most dominant periphyton species, *Cocconeis* sp., was determined by measuring the length of the cells; see Table A1.1 for sampling scheme. A neustonic bloom occurred in the cosm with the highest two dosages, and the species composition was assessed qualitatively weekly from week 6 until the end of the experiment. The chlorophyll-*a* content was also determined in this period.

At the end of the experiment, dry weight of all macrophytes was determined, divided over *Elodea nuttallii* and other species. Bioassays were conducted using *Elodea nuttallii* bioassays in which 4 g wet weight of shoots were used per cosm. Dry weight at the start of the bioassay was estimated by drying four extra portions of 4 g wet weight shoots. The shoots were placed in a plastic beaker with sediment in a cage. The bioassays took place at the start of the first application up to week 3 following which time the dry weight was measured.

An extra laboratory experiment was added to the study in order to investigate the possible adaptation of algae to linuron. In this laboratory test, two samples of *Chlamydomonas reinhardtii* were cultivated in a medium with and without 150 µg/L linuron, respectively, and after cultivation for 5 days, a single species laboratory test was performed with both strains using 0, 15, 50, 150, and 500 µg/L linuron. The total algal bio-volume was estimated at day 0 and day 3.

Decomposition of particulate organic material was measured using litterbags [glass petri dish, diameter 11.6 cm, covered with stainless steel wire (mesh 0.7 _ 0.7 mm, two 0.5-cm holes)]. *Populus x canadensis* leaves and *Elodea nuttallii* shoots (dried for 72 h at 60^{o}C) were used as organic material. Per cosm two litterbags per substrate were placed for 2 weeks (starting 3 weeks before first application), following which time the material was removed, and new litterbags were placed. Invertebrates were separated from the material, identified, counted and returned to the microcosms.

Zooplankton was sampled according to Table A1.1, using a 40-cm perspex corer (diameter 4 cm). In total, 5 L was sampled per cosm. Zooplankton species were identified and counted.

Macroinvertebrates were sampled every 2 weeks using three multiplates and two pebble baskets per cosm and the litterbags as described above. The macroinvertebrates were collected from the substrates at 2-week intervals, identified, counted and released

into the same cosms. At week 13 a subsample of the macroinvertebrates was taken and analysed quantitatively.

Table A1.1 Timetable of the sampling during the treatment and post-treatment phase of the study

	Treatment phase	**Post-treatment phase**										
Week	0	1	2	3	4	5	6	7	8	9	10	11
Dissolved oxygen	x	x	x	x	x	x	x	x	x	x	x	x
Conductivity	x	x	x	x	x	x	x	x	x	x	x	x
pH	x	x	x	x	x	x	x	x	x	x	x	x
Alkalinity	x	x	x	x	x	x	x	x	x	x	x	x
Nutrient concentration	x		x		x		x		x		x	
Phytoplankton												
- Species composition	x	x	x	x	x		x		x		x	
- Chlorophyll a	x	x	x	x	x	x	x	x	x	x	x	x
Neuston												
- Species composition							x	x	x	x	x	x
- Chlorophyll a							x	x	x	x	x	x
Periphyton												
- Species composition			x		x		x		x		x	
- Chlorophyll a on glass	x	x	x	x	x	x	x		x		x	
- Chlorophyll a on Elodea	x	x	x	x	x	x	x		x		x	
- Bio-volume Cocconeis			x									
Elodea biomass				x								x
Zooplankton species		x		x		x		x		x		x
Macrofauna		x		x		x		x		x		x
Decomposition		x		x		x		x		x		x

Environmental conditions

The air (and water) temperature were kept constant (19 ± 2^{o}C) in a climate room. A light period of 14 h (120 µE/m^2.s at the water surface) was used. In the pre-treatment period P (0.05 mg/L) and N (0.30 mg/L) were added.

Verification of concentrations

Water samples for analyses of linuron were taken at 'several' moments, in duplicate, from the mid-depth of the cosm. For the lower dosages (0, 0.5 and 5 µg/L) the water samples were extracted using octadecyl solid phase extraction columns conditioned with methanol and water. The water samples for the higher dosages (15, 50 and 150 µg/L) were analysed without extraction. The samples were analysed using high-performance liquid chromatography (HPLC). The limit of quantitation (LOQ) was not specified.

Physical and chemical analyses

The sample scheme for dissolved oxygen, pH, alkalinity, ammonium, nitrate, ortho-P and Na+, K+, and Ca2+ is given in Table A1.1.

Calculations and statistics

Redundancy Analyses (RDA) was performed for the effects on the phytoplankton community with treatment and sampling date as explanatory variables. Univariate analysis (one-sided ANOVA, Williams test) was carried out to calculate NOEC levels at the taxon level. The EC_{50} value for growth inhibition of *Elodea nuttallii* was calculated using a logistic model. The laboratory test with *Chlamydomonas reinhardtii* were analysed with ANOVA.

RESULTS

Chemical analysis

Mean measured concentrations in the application solutions were within the range of 10% from target concentration in the lowest dosage and within 5% at the other dosages. Relatively more linuron had to be added at the lower dosages during the test to maintain the target concentration than at the higher dosages, indicating a faster disappearance from the water phase in the lower dosages. This is also found in the post-treatment period, where the disappearance of linuron ranges from 11 days for the 0.5 µg/L treatment to 49 days for the 150 µg/L treatment. As a possible cause the authors indicate that this can be explained by a pH effect: higher linuron concentrations cause a lower photosynthesis, which in turn causes a lower pH. This lower pH results in a slower hydrolysis of linuron.

Physical and chemical analyses

The dissolved oxygen and pH levels were lower in the treatments of more than 0.5 µg/L during the treatment period (see Table A2.2). In the two highest treatment levels these differences remained until the end of the experiment. The opposite effect was found for alkalinity and conductivity (see Table A2.2). No significant effects were found for ammonium and phosphate. From week 6 onwards an increase of nitrate was found in the highest dose. In the highest treatment a significant increase of Ca and K was found (data not provided).

Phytoplankton

Based on the RDA-biplot, in which only the 13 most discriminant taxa are shown, it is concluded that the samples of the 150 µg/L treatment and – to a lesser extent – the 50 µg/L treatment diverged from the controls. Only one taxon (*Chlamydomonas* sp.) positively correlated with the highest treatment, while nearly all other taxa had a clear negative correlation. The NOECs for the most abundant taxa are given in Table A2.2.

Chlorophyll *a*

In the case of phytoplankton, significant differences between the treatment and control are found for the highest treatment only, during the treatment period.

Test for adaptation

The result of the laboratory experiment with *Chlamydomonas* shows that the strain exposed to linuron in the culture period had a larger relative growth when exposed to 150 and 500 µg/L linuron than the strain cultivated in a linuron-free medium.

Periphyton

No RDA biplot is presented for periphyton and neuston. Table A2.2 shows the NOEC values for *Chlamydomonas* and *Cocconeis*, with the latter showing a concentration-dependent decrease in the post-treatment period at all doses except the lowest. During the treatment period (week 2) the bio-volume of *Cocconeis* was significantly smaller in the highest treatment.

Chlorophyll *a*

The chlorophyll *a* content of periphyton on the glass slides was significantly higher in the two highest treatments in the post-treatment period, and in the highest treatment during the treatment period. The chlorophyll *a* content of the periphyton on *Elodea* increased in the post-treatment period in the highest dose only.

Neuston

Effects are only described qualitatively; these indicate a dominance of *Chlamydomonas* in the two highest doses, and a dominance of *Nostoc linckia* in the control and the lowest dose.

Chlorophyll *a*

The chlorophyll *a* content of the neutonic algae was enhanced in the post-treatment period in the highest treatment only.

Table A2.2 NOEC values during and after the treatment period; the arrows indicate an increase or a decrease[1]

	Pre Treatment	Treatment	Post treatment
Physicochemical			
Dissolved oxygen	50↑	0.5↓	15↓
pH	50↑	0.5↓	15↓
Conductivity	-	5↑	5↑
Alkalinity	-	5↑	5↑
Nitrate	-	>150	15↑
Phytoplankton			
Chlamydomonas	0.5↑	50↑	15↑
Cocconeis	-	50↓	50↓
Chroomonas	-	0.5↓	>150
Phormidium foveolarum	-	50↓	50↓
Periphyton			
Chlamydomonas	-	50↑	15↑
Cocconeis	-	15↓	0.5↓
Zooplankton			
Cladocera	-	>150	15↑
Copepoda	0.5↓	>150	50↑
Rotatoria	50↓	5↓	50↓
Ostracoda	-	>150	15↓
Macroinvertebrates			
Physella acuta	-	50↓	50↓
Asellus aquaticus	-	>150	15↑
Dugesia	-	>150	15↑
Bithynia	-	>150	>150

[1] Normally the NOECs should be listed for the single sampling dates; in this case these data were not available in the publications used.

Decomposition
No significant treatment-related effects were found.

Zooplankton
The zooplankton was dominated by Cladocera, Copepoda and Rotatoria, while Ostracoda were present in low numbers. The most dominant species of the Cladocera were *Daphnia longispina, Simocephalus vetulus, Graptoleberis testudinaria* and *Cladocera* spp.. The Copepoda were dominated by nauplii, and the genera present were *Macrocyclops, Eudiaptomus* and *Canthocamptus*. *Synchaeta pectinata, Polyarthra remata* and *Mytilina bicarinat*a were the most dominant Rotatoria; while *Cipridiopsis vidua* was the only Ostracoda. The RDA-biplot indicates an effect on the community at the highest concentrations. The copepod *Macrocyclops* and nauplii showed the strongest positive correlation to the treatment, while the rotifers *Synchaeta pectinata* and *Polyarthra remata* showed the strongest negative correlation. The effects on the different species groups are summarized as NOECs in Table A2.2. Based on these data it was not possible to obtain a more detailed picture of significant differences between control and treatments.

Macroinvertebrates
The most dominant species were snails, crustaceans, triclads and oligochaetes; leeches and nemerteans were less abundant. Insects, with the exception of *Chaoborus obscuripes*, were scarce. Herbivorous snails were dominated by *Physella acuta* and *Lymnaea stagnalis*, and bottom and vegetation dwellers were dominated by *Bithinia tentaculata* and *B. leachi*. The crustaceans were shredders and included *Gammarus pulex* and *Asellus aquaticus, Proasellus meridianus* and *Proasellus coxalis*. The most abundant triclad was *Dugesia tigrina*. The most dominant species of the vegetation-inhabiting Oligochaeta were *Stylaria lacustris* and *Chaetogaster* spec., and the most dominant benthic species were *Dero digitata* and Tubificidae. According to the authors, the biplot shows an effect for the two highest treatment levels in particular. Species that appear to react strongly to the treatment (*Dugesia lugubris*, positive correlation; *Physella acuta*, negative correlation) also differed during the pre-treatment period. The NOECs of some individual species are given in Table A2.2.
The invertebrate samples at the end of the experiment showed a high variation between treatments and a significant decrease only for *Asellus aquaticus* in the highest treatment.

Macrophytes
Growth of *E. nuttallii* in the bioassays was significantly decreased in all dosages except the lowest (see Table A2.3). The total biomass of *E. nuttallii* was significantly decreased in the two highest dosages only.

Table A2.3 Effects on Elodea nuttallii biomass in bioassays and in the cosms at the end of the experiment.

Dosage µg/L	Bioassay mg d.w.	Cosm g d.w./m^2
0 (control)	518	91
0.5	481	118
5	416*	112
15	344*	90
50	292*	43*
150	217*	5*

*Significantly different from the control ($P \leq 0.05$)

Conclusion

The authors conclude that the NOEC for linuron in this study is 0.5 µg/l. This conclusion is based on the effects on the macrophyte *Elodea nuttallii*, on the algae *Cocconeis* and *Chroomonas* and on oxygen and pH levels that were observed during the treatment phase and indicate that overall primary production was affected.

A 3.3 Evaluation of the treatment-related effects observed in the indoor microcosm study

A 3.3a Evaluation of the scientific reliability of the field study

Criteria for a suitable (semi-)field study

The following five questions are important in determining whether the study can be used for an appropriate risk assessment.

1. Is the test system adequate and does the test system represent a relevant freshwater community?
   Answer: unclear

   The study was performed in indoor aquatic microcosms under controlled temperature and light conditions. In the field situation more variation can be expected in environmental conditions, but this should not necessarily lead to other study results. The origin and composition of the sediment used was not described in detail, nor was the origin of the species used. The systems housed an ecosystem that was representative of a freshwater community with the following restrictions:
   - the species composition at the start of the experiment is not fully reported,
   - insects were not present (or not recorded),
   - only one macrophyte species was included,
   - no fish present.

2. Is the description of the experimental set-up adequate and unambiguous?
   Answer: unclear

   The experiment is well designed, state-of-the-art and in accordance with recommendations in guidance documents, although the number of replicates is the mini-

mum mentioned in the guidance documents (e.g. Crossland and La Point, 1992; Campbell et al., 1999; Giddings et al., 2002). However, the presentation of the data is not adequate nor is it unambiguous. For example, the NOEC data refer to periods of 3 (pre-treatment), 4 (treatment) or 7 weeks (post-treatment). Within a specific period there is no information on the individual data for each individual sampling date. Neither is it clear whether the NOEC is the lowest value observed within a specific period or, rather, based on the mean of all values. Furthermore, to overcome the problem of false positives/negatives, in aquatic risk assessment studies an effect is considered to be significant when a statistically significant change occurs on at least two successive sampling dates. These raw data are not provided; only one NOEC per period is reported. The reliability of NOEC values could only be checked superficially from the RDA-plots of the phytoplankton, zooplankton and macroinvertebrates. The study was not performed according to general principles of GLP.

3. Is the exposure regime adequately described?
   Answer: yes

   The formulation in which the test substance was applied is only described as 'Afalon (active ingredient Linuron)'. Information on the content of the active ingredient or other substances is not given. Furthermore, the application of the test substance to the microcosm is not described in detail and, based on the text, it may only be assumed that the test substance was applied in pure form to the microcosms. Actual exposure concentrations of the active ingredient and the development in time are described adequately. The measured actual concentrations were, on average, ±10% of the nominal concentrations.

4. Are the investigated endpoints sensitive and in accordance with the working mechanism of the compound?
   Answer: yes

   Laboratory single-species tests performed with aquatic organisms indicate that of the aquatic primary producers, aquatic macrophytes and algae in particular will suffer acute toxic effects due to the photosynthesis-inhibiting action of the herbicide. The exposure regime is within the range of the acute toxicity for aquatic macrophytes and algae. The presence of only one macrophyte species (Elodea) could form a limitation of this study. Acute and chronic toxicity for crustaceans and fish is much less than acute toxicity for the primary producers. However, the indirect effects of exposure to linuron may affect non-sensitive taxa at lower concentrations than those that result in direct toxicity. One of the objectives of the present study was to describe the (presumably indirect) effects on the secondary producers (zooplankton, macroinvertebrates). In terms of risk assessment, these taxa were represented in the microcosms in sufficient quality and quantity.

5. Is it possible to evaluate the observed effects statistically?
   Answer: unclear

The recommendations of the Classic workshop are in line with the experimental design of the present study. Five levels of linuron and an untreated control were studied in the microcosm experiment, with a minimum of two replicates per treatment. The experimental design adopted allows for univariate and multivariate analysis of the responses observed. The techniques applied for univariate (ANOVA, one-sided Williams test) and multivariate analysis (PRC) of the datasets is state-of-the-art. However, the data are not reported in detail and, for example, no information on the variation between replicates is provided. The figures do not show which differences are statistically significant. Since no basic data were provided, the results and statistics could not be checked. In addition, treatment-related correlations were found before the actual treatment started. No further explanation for this phenomenon or the meaning for the experiment is provided.

It is concluded that the indoor microcosm pond study can be used for the ecological risk assessment of the test compound to aquatic primary producers and aquatic invertebrates. The microcosm study does not provide insight into the possible risks of linuron to fish since aquatic vertebrates were not present in the microcosms. However, the data provided in the two publications hardly allow for an acceptable ecological risk assessment.

Given the data provided and the items of Annex 1, a Reliability Index of 1 cannot be assigned, and for regular registration purposes the underlying data should be provided. Under the given circumstances, a large number of items listed in Table A1.2 result in a lower reliability; these include the description of the substance (item 1), description of the biological system (item 5) and the lack of original data making it impossible to evaluate the elaboration of the results (items 8 and 9). Were these data to be available (for example, in a study report), the test might be reliable. For the time being, Ri 2 is indicated.

A 3.3b Evaluation of the results of the study

According to the EU guidance document on Aquatic Ecotoxicology, effects in micro- or mesocosm studies may be classified. For this study an adapted classification is used in which particular emphasis is placed on the duration of the effects as an additional criterion.

Effect class	Description	Criteria
1	Effects could not be demonstrated ($NOEC_{micro/mesocosm}$)	• No (statistically significant) effects observed as a result of the treatment • Observed differences between treatment and controls show no clear causal relationship
2	Slight and transient effects	• Effects reported as 'slight' or 'transient', or other similar descriptions • Short-term and/or quantitatively restricted response of one or a few sensitive endpoints, and only observed at individual samplings
3A	Pronounced effects; recovery within 8 weeks after first application	• Clear response of sensitive endpoints, but full recovery within 8 weeks after the first application, • Effects reported as 'temporary effects on several sensitive species', 'temporary effects on less sensitive species/ endpoints' or other similar descriptions • Effects observed at some subsequent sampling instances
3B	Pronounced effects; recovery within 8 weeks after last application	• Clear effects of sensitive endpoints, but full recovery within 8 weeks post last application. In the case of repeated treatments a total duration of the effects of > 8 weeks is possible, • Effects reported as 'temporary effects on several sensitive species', 'temporary effects on less sensitive species/ endpoints' or other similar descriptions • Effects observed at some subsequent sampling instances
4	Pronounced effects ; study too short to demonstrate recovery within 8 weeks after the last application	• Clear effects observed as class 3 effects, but the study is too short to demonstrate complete recovery within 8 weeks after the (last) application
5A	Pronounced effects ; no recovery within 8 weeks after the last application; full recovery within the test period	• Clear response of sensitive endpoints, and recovery time is longer than 8 weeks after the last application, • Full recovery is reported before the end of the study • Effects reported as 'long-term effects followed by recovery on several sensitive and less sensitive species/endpoints' or other similar descriptions • On consecutive time-points
5B	Pronounced effects ; no recovery within 8 weeks after the last application; no full recovery demonstrated within the test period	• Clear response of sensitive endpoints and recovery time is longer than 8 weeks after the last application, • Effects reported as 'long-term effects followed by recovery on several sensitive and less sensitive species/endpoints' or other similar descriptions • Full recovery is not reported before the end of the study • On consecutive time-points

As a consequence of the reporting style of the study the possibilities to classify the reported effects was in some cases restricted. One reason for this is that the statistical significance of effects was only reported for two defined time periods (the treatment period and the post-treatment period) and not for individual time points. Therefore class 2 effects could not be assigned. Furthermore, the post-treatment period lasted 7 weeks and was thus too short to allow for any determination of effect classes 3B (pronounced effects, recovery within 8 weeks following the last application) and 5 (pronounced effects, full recovery not observed within 8 weeks following the last application).

Treatment level of 0.5 mg a.s./L
No treatment-related effects were observed in any of the measurement endpoints. Therefore, all parameters are classified as class 1 effects.

Treatment level 5 mg a.s./L
At this concentration, treatment-related effects could be observed for a few endpoints. In the bioassay with *Elodea*, a significant reduction in biomass was found (Van den Brink et al., 1997, Table 3) in the bioassay itself but not in the microcosm at the end of the experiment. This effect is therefore classified as 3B. A treatment-related response was found for the phytoplankton species *Chroomonas* in the treatment period (Van den Brink et al., 1997,, Table 4). Since no treatment-related effects are found in the post-treatment period, this effect is classified as a class 3A effect, and this effect causes a class 3A effect on the phytoplankton at this concentration. In the case of *Cocconeis* in the periphyton, the figures show an effect in the post-treatment period, and no complete recovery was demonstrated within 7 weeks after the last application of Afalon; therefore, this effect is assigned as class 4. Taken together, the overall classification for periphyton abundance is class 4. No significant or consistent effects were seen for neuston, zooplankton and macro-invertebrates (class 1). In terms of the chemical–physical parameters pH and dissolved oxygen, a treatment-related decrease was found during the treatment period (class 3B), (Cuppen et al., 1997, Table 1).

Treatment level 15 mg a.s./L
For macrophytes, phytoplankton and periphyton in the microcosms treated with 15 mg a.s./L, the same effects were found as in the 5 µg/L treatment, resulting in the same classifications. Consistent short-term treatment-related effects could be demonstrated for some phytoplankton and zooplankton taxa. For the zooplankton species, a decline is seen in the treatment period, but not during the post-treatment period; therefore, this effect is classified as a class 3A effect (Cuppen et al., 1997, Table 1). No effects were seen for the macroinvertebrates. However, an increase for *Asellus aquaticus* was found in the artificial substrates up to the end of the experiment; as this effect is not indicated as being significant, only the effect is indicated. In terms of the chemical–physical parameters, decreased levels of dissolved oxygen and pH are found during the treatment period (class 3A), and an increase of conductivity and alkalinity is found in the treatment and post-treatment period (class 4).

Treatment level 50 mg a.s./L
For the phytoplankton in the microcosms treated with 50 µg/L, *Chlamydomonas* showed an increase in abundance (Van den Brink et al., 1997,, Table 4) in the post-treatment period (class 4) and *Chroomonas* showed a decrease during the treatment period (class 3A). This resulted in a class 4 rating for phytoplankton abundance. For *Chlamydomonas* (Van den Brink et al., 1997,, Table 4) in the periphyton, an increase is found in the post-treatment period (class 4); for *Cocconeis*, a decrease is found during and after treatment (class 4). This results in a class 4 rating for periphyton. For the neustonic algae, there is a significant shift to a dominance of *Chlamydomonas* at the two highest dosages (class 4). Chlorophyll-*a* content is significantly increased in the periphyton in the

post-treatment period (class 4) (Van den Brink et al., 1997,, Table 6). Both the bioassays and the biomass show a significant reduction of biomass (class 4). For the zooplankton, *Cladocera* shows an increase during the post-treatment period, but at the last sampling date the differences are no longer significant (class 3B). There appears to be a decrease in rotatoria during the treatment period (class 3A). The *Ostracoda* show a decrease in the two highest treatment levels. It is not clear whether recovery actually occurred; therefore a class 4 effect is indicated for the zooplankton. For the macroinvertebrates, the biplot indicates an effect on the species composition: *Asellus aquaticus* and *Dugesia trigina* show an increase in the post-treatment period (class 4). All chemical–physical parameters are significantly affected by the treatment in the post-treatment period (class 4).

Treatment level 150 mg a.s./L

In the 150 µg/L treatment the PRC for the phytoplankton community differs from that of the control up to the end of the experiment. In terms of individual phytoplankton species, an increase in abundance is found for *Chlamydomonas* in the treatment and post-treatment period, without recovery (class 4), a decrease in abundance is found for *Chroomonas* during the treatment period (class 3A) and a decrease is found for *Cocconeis* and *Phormidium foveolarum* during the treatment and post-treatment period (class 4). Overall this results in a class 4 rating for phytoplankton. In the periphyton, a significant increase is found for *Chlamydomonas* during and after treatment and a significant decrease is found for *Cocconeis* (class 4). Chlorophyll-*a* content is significantly increased in the treatment and post-treatment period in the periphyton on the glass slides (class 4) in the post-treatment period on Elodea and in the Neuston (class 4). The chlorophyll-*a* content in the phytoplankton is increased in the treatment period only (class 3A). For the neustonic algae, a significant shift to a dominance of *Chlamydomonas* is found at the two highest dosages. Both the bioassays and the biomass show a significant reduction of biomass (class 4). The biplot indicates an effect on the species composition of the zooplankton. The *Cladocera* show an increase during the post-treatment period (class 4). The *Copepoda* show an increase during the treatment and post-treatment periods (class 4). The rotatoria shown a decrease during the treatment and post-treatment periods (class 4). For the *Ostracoda* a decrease is indicated in the two highest treatment levels (class 4). The biplot indicates an effect on the species composition of the macroinvertebrates. There is a decrease in *Physella acuta* abundance during and after the treatment period (class 4). *Asellus aquaticus* shows an increase in the artificial substrates until the end of the experiment, indicated as a class 4 effect. An increase is found for *Dugesia trigina* (class 4). All chemical–physical parameters are significantly affected by the treatment in the post-treatment period (class 4).

Summary of effects

A summary of the effects according to this classification is given in Table A2.4.

Table A2.4 Summary of the effect classes observed for several endpoints in the indoor microcosm study with linuron

Species/group	Nominal concentration (µg a.s./L)				
	0.5	5	15	50	150
Chemical-Physical	1	3A↓	4↓↑	4↓↑	4↓↑
Phytoplankton abundance	1	3A↓	3A↓	4↓↑	4↓↑
Phytoplankton Chlorophyll a	1	1	1	1	3A↑
Periphyton abundance	1	4↓	4↓	4↓↑	4↓↑
Periphyton Chlorophyll a	1	1	1	4↑	4↑
Neuston	1	1	1	4v↑	4↓↑
Neuston Chlorophyll a	1	1	1	1	4↑
Zooplankton abundance	1	1	3A↓	4↓↑	4↓↑
Macroinvertebrates abundance	1	1	1	4↑	4↓↑
Macrophytes biomass	1	3B↓	3B↓	4↓	4↓
Decomposition	1	1	1	1	1

↑=increase, ↓=decrease

On the basis of the univariate and multivariate analysis of the datasets, an overall $NOEC_{microcosm}$ of 0.5 mg a.s./L can be derived from the evaluated microcosm experiment.

A statistically significant effect on the abundance of the diatom *Cocconeis* was seen at the exposure level of 5 mg a.s./L. No complete recovery was demonstrated within 7 weeks after the last application of Afalon (effect class 4). Since the next lowest concentration (0.5 µg/L) is the NOEC, a NOEAEC cannot be derived, and it can be concluded that the NOEAEC of the evaluated microcosm study is therefore larger than the $NOEC_{microcosm}$ (0.5 mg a.s./L) but lower than 5 mg a.s./L.

A summary of endpoints as derived from this study is presented in Table A2.5.

Table A2.5 Summary of the several endpoints in the outdoor microcosm study with Linuron, values based on nominal concentrations

Group	NOEC (µg as/L)	NOEAEC (µg as/L)
Phytoplankton	0.5	
Periphyton	0.5	
Neuston	15	
Zooplankton	5	
Macroinvertebrates	15	
Macrophytes	0.5	
Ecosystem	0.5	>0.5 and < 5

Suggestions for use of the study for risk assessment

In the derivation of the RAC the results of the microcosm study with linuron should be considered in relation to the results of lower tier studies and in relation to other field studies. A further detailed consideration of the lower tier data is beyond the scope of this document and example. Other semi-field studies with linuron have been reported. Stephenson and Kane (1984) carried out a 42-day microcosm experiment with linuron. They reported adverse effects of linuron on macrophytes and a reduction in oxygen

and pH levels, but not an increase in the chlorophyll-*a* content of the phytoplankton. Since linuron was dosed only once at 1 mg/L in their study, the resultsdo not provide any insight into the variability of the response.

An additional experiment with linuron was performed in the experimental ditches of Alterra. As structural endpoints, the derived NOEC of the indoor microcosm experiment simulating a constant and chronic exposure (0.5 mg a.s./L) was a factor of ten lower than the NOEC (5 mg a.s./L) observed in the experimental ditches treated three times with linuron in order to simulate a pulsed exposure (1-month intervals, kept static for 1 week after each treatment) (Crum et al., 1998; Kersting and Van Wijngaarden, 1999; Van Geest et al., 1999). This result indicates that the NOEC of 0.05 mg a.s./L from the chronic toxicity mesocosm study is protective for the study with a pulsed exposure (NOEC of 5 mg a.s./L).

References

Campbell, P.J., Arnold, D.J.S., Brock, T.C.M., Grandy, N.J., Heger, W., Heimbach, F., Maund, S.J. and Streloke, M. (Eds.) 1999. Guidance Document on Higher-Tier Aquatic Risk Assessment for Pesticides (HARAP). From the SETAC-Europe/OECD/EC Workshop, Lacanau Océan, France, SETAC-Europe, Brussels, Belgium.

Crossland, N.O. and La Point, T.W. 1992. The design of mesocosm experiments. *Environ. Toxicol. Chem.* **11**, 1-4.

Crum, S.J.H., Aalderink, G.H. and Brock, T.C.M. 1998. Fate of the herbicide linuron in outdoor experimental ditches. *Chemosphere* **36**, 2175-2190.

Cuppen, J.G.M., Van den Brink, P.J., Van der Woude, H., Zwaardemaker, N. and Brock, T.C.M. 1997. Sensitivity of macrophyte-dominated freshwater microcosms to chronic levels of the herbicide linuron; II. Community metabolism and invertebrates. *Ecotox. Environ. Safe.* **38**. 25-35.

Giddings, J., Heger, W., Brock, T.C.M., Heimbach, F., Maund, S.J., Norman, S., Ratte, H.T., Schäfers, C., and Streloke, M. 2002. Community-Level Aquatic System Studies - Interpretation Criteria (CLASSIC), Fraunhofer Institute, Schmallenberg, Germany, SETAC, Pensacola, FL, USA.

Kersting, K. and Van Wijngaarden, P.A. 1999. Effects of a pulsed treatment with the herbicide afalon (active ingredient linuron) on macrophyte-dominated mesocosms. I. Responses of ecosystem metabolism. *Environ. Toxicol. Chem* **18**, 2859-2865.

Stephenson, R.R. and Kane, D.F. 1984. Persistence and effects of chemicals in small enclosures in ponds. *Arch. Environ. Contam. Toxicol.* **13**, 313-326.

Van den Brink, P.J., Hartgers, E.M., Fettweis, U., Crum, S.J.H., Van Donk, E. and Brock, T.C.M. 1997. Sensitivity of macrophyte-dominated freshwater microcosms to chronic levels of the herbicide linuron; I. Primary producers. *Ecotox. Environ. Safe.* **38**, 13-24.

Van Geest, G.J., Zwaardemaker, N.G., Van Wijngaarden, R.P.A. and Cuppen, J.G.M. 1999. Effects of a pulsed treatment with the herbicide afalon (active ingredient linuron) on macrophyte-dominated mesocosms. II structural responses. *Environ. Toxicol. Chem.* **18**, 2866-2874.

ANNEX 3 GLOSSARY

ACR	acute to chronic ratio
ANOVA	analyses of variance
a.s.	active substance
BCF	bio concentration factor
EAC	ecological acceptable concentration
EFSA	European food safety authority
ERC	ecotoxicologically relevant concentration
GAP	good agricultural practice
GLP	good laboratory practice
NOEAEC	no observed ecological adverse effects concentration
NOEL	no observed effect level
PRC	principal response curve
RAC	regulatory acceptable concentration
SSD	species sensitivity distribution
TER	toxicity to exposure ratio
TWA	time weighted average

Summarizing table for aquatic micro- and mesocosm tests

The column 'Test items' and 'Notes' specify the items to report in the summary. The column 'Reliability lower?' guides the evaluation of the reliability of the test, which should be discussed under the section Remarks. Y = Yes, and E = Expert judgment needed. How many of these demerits lead to an unreliable test should be justified case-by-case. [→Ri 3] indicates that this omission in itself is enough justification to consider the study unreliable.

TEST ITEMS		NOTES	RELIABILITY LOWER?
METHODOLOGY & TEST DESCRIPTION			
1.	Substance	Properly characterized and reported?	
	1.1Concentration	[identity and amount of a.s. per litre test water not reported]	Y [→Ri 3]
	1.2 Formulation and purity	[ingredients in the formulation influencing the working action of the a.s. should be reported]	E
	1.3 Vehicle	[in case a vehicle – other than in the formulation – is used, identity and concentration?]	E
	1.4 Chemical analyses	[method, LOQ, LOD, recovery, not reported]	Y [→Ri 2]
	1.5 Properties	[not reported]	Y [→Ri 2]
2.	Test site, duration	Properly characterized and reported?	
	2.1 Location	[necessary to make a link between the study and agricultural practice (effects, environmental conditions and the application method): representativeness]	E
	2.2 Soil type /substrate	[necessary to compare to the local conditions of concern; not reported?]	Y [→Ri 2]
	2.3 Test date / duration	[duration long enough to study recovery?]	E
	2.4 General climatic conditions	[necessary to make a link between the effects and local climatic conditions; not reported?]	Y [→Ri 2]
3.	Application	Properly characterized and reported?	
	3.1 **Mode of** application	[spraying or homogenizing the a.s. into the test medium; not reported]	Y [→Ri 2]
	3.2 **Dosage**	[actual concentrations during the test are most important; not reported?]	E
		[no chemical analysis of dosing solution and no actual concentrations]	Y [→Ri 3]
	3.3 **Application** scheme	[necessary to make a link between the test and the intended use of the pesticide; not reported]	Y [→Ri 2]
	3.4 **Condition of** application	[additional technical data, route under consideration; not reported]	Y [→Ri 2]
	3.5 **Climatological** conditions	[weather conditions during application, wind speed and temperature?]	E
4.	Test design	Properly designed and reported?	
	4.1 **Type & size**	[e.g. outdoor microcosm, outdoor pond or mesocosm; not reported]	Y [→Ri 2]
	4.2 **Test system**	[not properly reported?]	Y [→Ri 2]
	4.3 **Pre-treatment**	[no period reported, no proper equilibration?]	Y [→Ri 2]
	4.4 Post-treatment	[period, interval between treatments, not reported]	Y [→Ri 2]
	4.5 **Untreated control**	[insufficient number, invalid or improperly reported?]	Y [→Ri 3]
	4.6 **Replications**	[insufficient replications for proper statistical analysis?]	Y [→Ri 2]

TEST ITEMS		NOTES	RELIABILITY LOWER?
	4.7 **Statistics**	[ECx's derived by regression, NOECs derived by ANOVA and preferably by multivariate techniques as PRC]	E
	4.8 **Dose-response**	[≥ 2 test concentrations for finding a dose-response relation (controls excl.)]	E
	4.9 **Study under good** laboratory practices (GLP)	study not conducted under GLP ?	E
5.	Biological system	Representative and properly reported?	
	5.1 **Test organisms**	[e.g. species/taxa not reported?]	Y [→Ri 3]
	5.2 **Community**	[the community/ecosystem representative and complete?]	E
6.	Sampling	Is sampling adequate for risk assessment?	
	6.1 **General features**	[properties during test not monitored? E.g. pH, hardness, oxygen]	Y [→Ri 2]
	6.2 **Actual** concentration	[actual concentrations measured in medium and other compartments or biota?]	E
	6.3 **Biological** sampling	[no proper method, species, number, endpoints, frequency?]	Y [→Ri 2]
RESULTS			
7.	Endpoint	Properly reported?	
	7.1 **Type**	[relevant endpoints not chosen or specified]	Y [→Ri 3]
	7.2 **Value**	[results not based on measured data?]	Y [→Ri 3]
	7.3 **Verification of** endpoint	[test results are not verifiable? source data not reported]]	Y [→Ri 3]
8.	Elaboration of results	Are conclusions based on measured data? methodology correct?	
	8.1 **Statistical** comparison	[data do not meet requirements for method used?]	Y [→Ri 3]
	8.2 **Dose–effect** relationship	[not present?]	Y [→Ri 3]
	8.3 **Community level** statistics	[not reported? improper method?]	Y
9.	Control		
	9.1 Untreated control	[unexpected effects or disappearance of species?]	Y [→Ri 3]
	9.2 Positive control	[no clear effects in highest treatment or positive control]	Y [→Ri 3]
10.	Classification of effects	Not properly derivable?	Y [→Ri 2]
11.	Biological meaning of **statistically** significant differences	Insufficiently explained?	E
REMARKS			
The above-mentioned items concern the scientific reliability of a field study. The usefulness of a field test depends on the scientific reliability and purpose of use. One of the aspects for purpose of use is the similarity between the test situation and the situation of concern for the registration; for instance, the following test items must be checked: product, dosage/ concentration, application frequency, interval and type of ecosystem. The more similarity there is between the aspects found in the field test and the product under registration and its proposed conditions of use, the more likely it is that the field test is useful for risk assessment. However, general guidance is difficult to give, and expert judgement is therefore decisive, as the appraisal of the usefulness may differ from pesticide to pesticide. Although these considerations do not deal with the quality of the study, a remark can be made in the Remarks section. Other aspects that were not covered by the items mentioned above can also be introduced here.			