CENTRE FOR CHEMICAL REGULATION AND FOOD SAFETY (EUROPE)
The Lenz, Hornbeam Business Park,
Harrogate. HG2 8RE UK
T (+44) 1423 853200 F (+44) 1423 810431
info@uk.exponent.com

CENTRE FOR CHEMICAL REGULATION AND FOOD SAFETY (US)
1150 Connecticut Avenue, NW
Ste. 1100, Washington, DC 20036
T (+1) 202 772 4900 F (+1) 202 772 4979
info@exponent.com

Please reply to the UK office

Literature reviews on ecotoxicology of chemicals with special focus on plant protection products. Reference: CFT/EFSA/PPR/2008/01

Lot 4 : Critical comparison of available and potential higher tier testing approaches for the risk assessment of plant protection products, considering at least field and semi-field experimental designs, extrapolation from dose-response relationships, and increased dosages (aquatic and terrestrial)

Prepared for:
European Food Safety Authority
Largo N. Palli 5/A,
43100 Parma, Italy

Prepared by:
Kevin Brown, Josie Tomlinson,
Jennifer Duncan and
Amelia Hinchcliffe
Exponent International Ltd.
The Lenz, Hornbeam Business Park
Harrogate, Yorkshire, HG2 8RE, UK

Katherine Palmquist
Exponent International Ltd.
15375 SE 30th Place, Suite 250
Bellevue, Washington, 98007

Date prepared:
August 2009
Project No: 0805947.UK0

.

Contents

1.0 General Introduction

This final report constitutes the final deliverable for the assignment contracted by the European Food Safety Authority (EFSA) to Exponent International Limited to perform a literature review on ecotoxicology of chemicals with special focus on plant protection products in order to perform a "critical comparison of available and potential higher tier testing approaches for the risk assessment of plant protection products, considering at least field and semi-field experimental designs, extrapolation from dose-response relationships, and increased dosages (aquatic and terrestrial)" (Lot 4 of CFT/EFSA/PPR/2008/01).

2.0 Literature search

2.1 Database search protocols

The following six groups of descriptors were combined in appropriate manners in the database search.

non-standard species	**ecotox***	**higher tier**	**aquatic**	**plant protection product**	**research**
additional species	**cnviron***	**field stud***	**bee**	**pesticide**	**procedure**
multi-species and multi species	**risk-assessment**	**semi-field stud***	**earthworm**	**herbicide**	**method**
time to event analysis		**semi-field test**	**lumbricid***	**acaricide**	**predict**
sensitive life stage		**cage**	**arthropod**	**insecticide**	**assess**
artificial stream		**tunnel**	**bird**	**fungicide**	
experimental ditch		**mesocosm**	**mammal**	**plant growth regulator**	

model ecosystem		**modeling**	**fish**	**miticide**	
species sensitivity		**population**	**invertebrate**		
distribution					
mesocosm		**microcosm**	**algae**		
microcosm			**sediment**		
			dwell*		
recovery			**soil organism**		
semi-field stud*					
field stud*					
higher tier testing					
higher tier approaches					
higher tier					

Note: * indicates any combination after the *, e.g. stud* would catch the words "study" and "studies"
Hyphenated terms should be run with and without the hyphen, e.g. semi-field and semi field

These terms were represented in the DIALOG search protocol run by Literature searchers at Exponent International as detailed in Appendix 1. Similarly the STN search protocol is shown in appendix 2. The DIALOG and STN databases searched and the periods which each database covers are listed at Appendix 3. No restriction was given to the period searched, covering thus as default the complete periods covered by the respective databases.

The search was done on 27 and 28 January 2009. System algorithms were used in the search to limit results to materials published in English. DIALOG sets s11, s16, s24 and s36 as listed in appendix 1 are considered to represent the critical sets for the search terms used. Initial DIALOG hits for sets s11 (1377 hits), s16 (356 hits), s24 (1060 hits) and s36 (399 hits) sum to a total of 3192 DIALOG hits. Additionally 310 hits were recorded for the STN databases. The grand total number of hits from the DIALOG and STN databases was therefore 3502 hits. The initial search results of 3502 hits were then subject to a preliminary filter performed by experienced literature searchers so that duplicates and clearly non-relevant items were removed. Non relevant items are

considered to be those where the keywords match, but the document is clearly not in context with the search.

2.2 References reviewed

The remaining items (2,779) following this preliminary filter were then considered by either one or other of the two ecotoxicologist experts. Relevance was based on the paper describing ecotoxicology (rather than environmental fate) methodology for higher tier testing. Higher tier testing was considered to include non-standard laboratory tests, as well as, semi-field and field methods. Higher tier methods for the various organisms considered under 91/414/EEC were included in the various sections of the 'Endnote' file. The Klimisch code[1] was used to determine papers that were well documented and scientifically acceptable and thus papers included in this document can be considered to be at least Klimisch code 2.

A total of 279 papers were identified as possibly relevant by the ecotoxicologist experts and are listed in the 'Endnote' file. Of the 279 references identified as outlined above, those that could be obtained in the timeframe of the project were reviewed and papers subsequently found to be relevant are discussed in this document. Additional references known personally to the authors including relevant posters from SETAC Europe 2009 were also added to the 'Endnote' file and used in the critical review. Based on the search terms and scientifically recognised databases used in the literature search we consider that the vast majority of testing approaches in the public domain will have been identified and reviewed.

[1] Klimish, H.J., Andreae, M. and Tillmann, U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology **25**, 1-5.

3.0 Literature review/assessment

3.1 Introduction

Higher tier testing approaches are triggered across all areas of ecotoxicology when the results from initial laboratory tests (often referred to as Tier 1) fail to show a clearly acceptable level of risk when combined with an application rate of the product to generate toxicity exposure ratios (TER's) or Hazard Quotient (HQ) values. According to EU legislation " no authorisation 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 proposed conditions of use" (European Commission, 1991).

Higher tier testing attempts to demonstrate the acceptability of any impact that would occur under field conditions, either by simulating more realistic conditions in the laboratory, semi-field or by sampling naturally occurring populations in the field. Whereas the initial tier is tightly defined, with precise methodology and clear procedures for using the results in a regulatory context, there are a range of higher tier options available and no agreed procedures for interpreting or accepting the appropriateness of the data.

All higher tier testing is confounded by the fact that there is no agreed definition as to what is or is not an acceptable impact. There is also some confusion as to whether species, communities or functional groups are the target of supposed protection goals.

In an attempt to refine the risk a greater degree of realism is included in the next tier of testing, either in terms of the exposure route or the relevance of the organisms tested. With each step to include more realistic exposure or more relevant organisms comes an increase in the amount of variability present within the data. The immediate consequence of this is to reduce the precision

of the study in question. This review considers current and potential testing approaches in ecotoxicology across the areas of birds and mammals, aquatic organisms, non-target arthropods, honey bees, soil organisms and non-target terrestrial plants. Whilst each subject area is presented separately there are many common threads which apply to all subject areas.

3.2 Birds and mammals

3.2.1 Background

The current European Union (EU) first-tier assessment of the risk of pesticides to birds and mammals (under Council Directive 91/414/EEC) is based on deterministic toxicity/exposure ratios (TERs). Toxicity values are based on the responses of individual organisms observed in controlled laboratory experiments, conducted to standard testing guidelines. The current standard laboratory studies are intended to be conservative and reflect a reasonable worst-case scenario.

Where first-tier toxicity exposure ratios do not demonstrate an acceptable risk to bird and mammals, a refined risk assessment is usually conducted, in line with the Guidance Document on Risk Assessment for Birds and Mammals Under Council Directive 91/414/EEC **SANCO (2000)** and more recently the revision of this document by the Scientific Panel on Plant Protection Products and their Residues (PPR), EFSA-Q-2006-064 (adopted June 2008). The refinements of the risk assessment for birds and mammals often involved revising the input parameters (e.g. focal species, diet composition and time in treated area) using publically available data. Higher tier testing approaches can also be a useful tool in refining the risk assessment. It is beneficial to start with simpler laboratory studies, followed by semi-field approaches, before scheduling a field study as the highest tiered approach. However, standardised test guidelines have not been developed for any of these tiers. Therefore,

assessment of the publically available literature on available and potential higher tier testing approaches for birds and mammals has been conducted.

3.2.2 Avoidance/palatability tests

The Guidance Document on Risk Assessment for Birds and Mammals SANCO (2000) provides limited guidance on avoidance tests. There are currently no internationally accepted guidelines for testing avoidance (repellency/palatability). Two national guidelines exist and a number of other protocols are under development. The **French guideline (INRA 1990)** measures consumption (on a daily basis) and effects in choice and no-choice conditions. Treated material is provided in pots or hoppers. The **German guideline (BBA 1993)** is intended for use with granular formulations, treated seeds and baits. The emphasis of the test design is on presenting the treated and untreated material in a realistic way, mixed together and spread on the floor of the test aviary. Feeding rate is not controlled.

There are a number of uncertainties regarding the design and interpretation of avoidance/palatability tests. In general, as with many higher tier testing methods, protocols should be considered on a case-by-case basis. One of the main problems with determining avoidance of birds and mammals to plant protection products, is that feeding rates under controlled laboratory studies can significantly differ from those in the natural environment. Some of the main factors which affect feeding rate are available feeding time, hunger level (food deprivation/restriction) and group size. The motivation of birds to feed is critical in determining whether they will ingest a lethal dose of treated seeds: birds which eat rapidly are more likely to ingest a lethal dose before the onset of an avoidance response **Fryday et al. (1999)**

The Central Science Laboratory (CSL, UK), contracted a **report (PN0914)** on validation of methods for testing the avoidance of treated seeds by birds, as well as an assessment of robust

tests for the acceptance of bait and treated seeds by birds **Fryday et al. (1999)**. It was agreed that the most pressing need for development of the guideline in this project was to refine methods for testing birds at different feeding rates and to validate the approach using treated seed, comparing the data with field data. The usefulness of the TAR (time to avoidance reaction) approach was also to be considered. In this case it was envisaged that if time to avoidance was relatively constant it may be possible to predict the dose if concentration and feeding rate was known.

An experiment was conducted using pigeons to measure the effects of varying group size, adding short periods of food deprivation and reducing feeding period on the rate of consumption of untreated wheat. The results of this study were then used to select test conditions for a validation study using wheat treated with fonofos to produce residues similar to those found on seed on the day after drilling.

The validation study predicted a low level of risk under normal field conditions. No mortalities occurred and where consumption levels were highest, significant levels of regurgitation were detected. It was considered that even at higher feeding rates than those tested, mortalities would be unlikely. The TAR approach was considered to be of limited use due to the difficulty in determining the point at which avoidance had occurred, the potential variation in the response and the influence of other factors such as intake rate, concentration and absorption rate. It was also considered that it would not be possible to include this as part of a standard dietary test due to the likely difficulty of obtaining the information and the potential difference between species. Overall it was concluded that further development of testing guidelines was required.

CSL also contracted a second report (**PN0909**) on robust tests for the acceptance of bait and treated seed by birds **Fryday et al. (2001**). This project tested whether the methods developed under PN0914 were suitable for another larger test species (pheasant) and aimed to develop a method for small birds that addressed different welfare requirements. It also included a further

validation study for pigeons using bendiocarb applied to peas and maize to mimic two uses that in the past have led to very different levels of risk to wild birds.

In the pheasant study, it was found that the birds could be motivated to consume more seed in a short test period by using the same approach as developed in PN0914. However, there was a less marked increase in peck rate with decreased feeding time than for pigeons. For small birds (house sparrows), it was found that removing food for progressively longer periods during the middle of the day did lead to increased consumption and feeding rate. While birds did not lose weight, they could not obtain their normal daily amounts of food under these conditions. However, given that the birds adapted to the conditions and maintained weight, it was considered that this approach was suitable for limited periods with appropriate care.

In the validation study bendiocarb caused similar mortality with both peas and maize when both were easily available. It was concluded that the lower mortality with maize in the wild is most likely due to feeding rate being limited by the density of seeds on the surface, which is much lower for maize than peas due to different seed rates and drilling methods. The results again confirmed that feeding rate should be controlled in tests of avoidance, and showed how this can be done for three key species.

3.2.3 Pen/cage "semi-field" studies

Pen and cage tests are only rarely conducted with birds and mammals, and there is no recognised standard method. Principally these tests follow a semi-field concept where the product is applied according to practical use conditions either by applying the substance within an aviary or pen or by setting up an open-bottom cage in the field after treatment. Evaluation is facilitated as replications are possible. Such a test design allows observations and measurements of individuals. In some regard the situation is severe as the animals are confined to the treated area,

on the other hand, however, energy expenditure and feeding rate may be reduced. Therefore care is needed in interpreting the results appropriately **SANCO (2000)**.

Few examples of semi-field testing methods for birds and mammals are available in the public literature. **Johnston et al. (1996) c**onducted a semi-natural field study, to assess the likelihood of a potentiation of toxicity between two compounds (an ergosterol biosynthesis inhibiting (EBI) fungicide and an organophosphorus (op) insecticide) in the red-legged partridge. Groups of partridges were kept in four large grassland enclosures, these were exposed to either prochloraz-treated (EBI) or control wheat for 7 days after which two of the enclosures were sprayed with malathion (OP) whilst the remaining two were sham-sprayed. This semi-field trial attempted to bridge the gap between laboratory studies and those carried out in the field. It has shown that there was no potentiation of toxicity between the EBI fungicide, prochloraz, and the OP insecticide, malathion, in what is in theory a worst case scenario.

Semi-field studies allow for more environmentally realistic conditions in comparison to investigations carried out under laboratory conditions. A semi-field study allows greater control over variables (including the species exposed, level and route of exposure) as opposed to full field studies, thus effects are simper to measure and the results easier to interpret. The semi-field approach is considered to provide a practical and more easily measurable half-way stage in attempting to assess the likelihood of interactions occurring in the field situation.

3.2.4 Field studies

Field studies should be designed on a case-by-case basis to address the issues which have been identified. There are no standardised protocols for avian and mammalian field studies, and guidance on potential study methods is limited. The published literature contains a number of field studies conducted to assess the effects of pesticide use on birds and/or mammals, although

it has to be noted that the data is strongly skewed towards effect on birds as opposed to mammals.

Many field studies use radio-telemetry, as well as manual searching and observations, to assess exposure and effects on natural populations of birds and mammals, after application of a test substance. The test substance is typically applied according to the correct GAP (good agricultural practice). This covers the correct application rate, time of application and application interval (where appropriate), crop (including appropriate growth stage) and member state where the product is intended to be used (i.e. appropriate climatic conditions). Therefore, the collected data is specifically relevant to the proposed practical use conditions. In this way, field studies are restricted to the proposed rate and use of the test substance and allow no (or very limited) extrapolation from dose-response relationships, and increased dosages.

Brown et al. (2008) conducted a farm scale field study to determine the risk to birds in vineyards from applications of an insecticide (chlorpyrifos). This involved establishing mist nests post-application in vineyards and radio-tagging and ringing all birds which were caught. The test substance was applied according to the intended GAP. The vineyard was searched for carcasses up to 7 days after application.

Searching efficiency was evaluated by placing a known number of artificial mice in the vineyards before each search date and then recording the number of these subsequently found by the carcass searching team. The rate at which dead animals would be taken from the site by predators or scavengers was also evaluated by placing a known number of dead mice in vineyards and returning after 24 hours to determine whether these animals had been eaten or moved.

The location of tagged birds was recorded for several days before treatment and for up to 10 days after treatment. Over the post treatment period the locations recorded for each bird were used to determine the proportion of time spent in-crop and to determine that they were still alive.

Wolf et al. (2009) also conducted a multiple field study program to assess the risk to birds from application of chlorpyrifos, in three major crops in three different European Member States (pome fruit in Italy, citrus in Spain and leafy vegetables in Poland). Two full-cover spray applications were applied to each crop. Following these applications, the treated crops were intensively monitored for 7 days in order to detect any lethal or sub-lethal effects on birds.

Individual birds were trapped inside the study areas and fitted with radio-transmitters in order to monitor their health status and potential exposure to the treatment. The radio-tracking of individual birds was combined with intensive visual observation and search for carcasses in the treated areas. Surveys of birds present on the plots were conducted at regular intervals by trained observers. This involved the observer walking slowly through the treated fields and recording each bird present and noting its health status. Additionally, birds entering the treated fields were observed throughout the application process. Finally, the treated fields were searched for bird carcasses on the day of the treatment application and again one and three days following application.

Poche et al. (1993) conducted a terrestrial field study to evaluate the potential for acute avian mortality resulting from application of a granular insecticide to turf grass. Survivorship of ground feeding birds was monitored using radio-telemetry at 8 golf courses. A total of 560 songbirds, 46 blue jays and 17 brown thrashers were captured in mist nets 3 - 21 days prior to scheduled chemical applications and fitted with radio transmitters. Each golf course was divided into 2 plots, 1 of which was randomly selected to receive the insecticide treatment. Radio signals from radio-tagged birds were monitored for a minimum of 2 stations per plot. Records were maintained on direction of signals, intensity of signals, and movement direction. When the signal intensity and direction appeared constant, an effort was made to find the bird *via* radiotracking and determine its fate. Personnel conducting the radio tracking underwent a series of method validation tests to assess their efficiency at determining bird movement.

Based on laboratory data, ingesting 50 to 100 pesticide granules (fewer than the number of granules applied to each square foot of the treatment area) would be potentially lethal to birds. However, this field study (i.e. using a realistic scenario) demonstrated no adverse effects.

Mineau (2002) reported that, a single directed field study by itself may not be sufficient to dispel a presumption of high risk that is place on a pesticide. This is because of the stochastic variability encountered in most field situations, as well as the inability to detect impacts every time they occur (in part because of the difficulty of finding evidence of an impact such as carcasses).

There are many uncontrollable variables in a field study. Thus, it is often difficult to determine the cause of death of an individual. Plus, due to movement and territory size of animals, it is difficult to have a true control in a field study **Poche** e**t al. (1993)**. Radio-telemetry has been demonstrated to be useful in many types of wildlife investigations. This can give an idea of the level of exposure of individuals, by tracking their movement within a treated area. Using radio-telemetry in conjunction with validated manual searching methods and observations can be a useful tool in gaining information on the exposure and effects of use of plant protection products under realistic environmental conditions.

However, field tests require extensive human and financial resources, yet they often provided very limited information on avian risk assessment **Tiebout (1995)**. The results and information obtained are very specific to a particular use scenario, therefore, allowing very limited extrapolation to other scenario, including different application rates.

In addition, very few field studies exist on the reproductive effects of pesticides on birds and mammals. The majority of available avian and mammalian field studies focus on effects on mortality.

3.2.5 Population assessment/modelling

As mentioned above, in the standard tiered assessment framework potential risk for birds and mammals is identified on the basis of responses of individual organisms observed in controlled laboratory experiments. However, ecological risk assessors have long argued that except in the case of threatened or endangered species, the abundance and persistence of populations of organisms are more relevant as endpoints as assessment than are responses of individual organisms **SANCO (2000)**.

Although the standard long-term toxicity exposure ratio (TERlt) is intended to be conservative and reflect a reasonable worst-case scenario, it is not clear how conservative it really is. It is also uncertain which stages of reproduction are likely to be affected, or what the consequences of those impacts may be on the overall reproductive success of individuals. The key question for the risk assessment process is whether there are likely to be impacts on population numbers, but it is not possible to make a link between the TERlt and population processes using the current guidance on risk assessment **Sibly et al. (2005**).

Sibly et al. (2005) referenced reviews which show that populations cannot be reliably safeguarded by analysis of single individual endpoints, the reason being that some risk assessments have been found to give varied results depending on which individual endpoint was used. Current regulatory practice, however, assesses avian and mammalian risk at the level of the individual rather than the population because methods for assessing risk at population level are not sufficiently established. This causes difficulties in the regulatory decision-making process, particularly in the frequent cases where individual effects are of uncertain ecological significance.

In addition, to date no field studies have been able to incorporate the temporal- and spatial-induced differences in crop types, crop rotation, and crop management together with specific ecological relationships at a landscape scale **Topping et al. (2004).**

Due to the limitation of field studies, more emphasis is being placed on understanding ecosystem level dose-response patterns by increasing mathematical modelling efforts, applying existing laboratory data, and generating new data to define model parameters **Tiebout et al. (1995).** Population models of all types have an important role to play in pesticide risk assessment. This role is translating the impacts at the individual level to impacts at the population and community levels. Assessing impacts at these higher levels integrates effects of pesticides on different biological processes (survival, mortality, dispersal) and on different life history stages, and thus is ecologically more relevant **Topping et al. (2005)**.

3.2.6 Discussion and summary

Higher tier testing approaches can be a useful tool for the assessment of risk to birds and mammals from plant protection products. In contrast to laboratory studies under controlled conditions, standard guidelines and recommendations for higher tier methods for birds and mammals are extremely limited. Higher tier trials should rather be designed individually, on a case-by-case basis, addressing the problems that have been identified.

With regards avoidance and palatability studies, there are a number of uncertainties regarding the design and interpretation of these tests. One of the main problems in determining avoidance of birds and mammals to plant protection products, is that feeding rates under controlled laboratory studies can significantly differ from those in the natural environment. Under natural conditions, variables such as competition and food availability affect feed rate and motivation to feed. Therefore, the reliability of results on avoidance and palatability obtained under laboratory conditions can be questionable.

Semi-field studies allow for more environmentally realistic conditions in comparison to laboratory studies, but with more control over variable in comparison to field studies. In general,

the semi-field approach is considered to provide a practical and more easily measurable half-way stage in attempting to assess the likelihood of interactions occurring in the field situation.

In principle, field data obtained under practical use conditions add a further level of realism to a risk evaluation, particularly if data are more focussed on a particular application regime, crop stage or geographical area. Also, data from field studies may be suitable to describe effects over time under natural conditions, which are very difficult to obtain from laboratory studies. However, full field tests require extensive human and financial resources. The results and information obtained are very specific to a particular use scenario, therefore, allowing very limited extrapolation to other scenarios, including different application rates.

Blus et al. (1997) reviewed the advantages and disadvantages of experimental and field studies for determining effects of pesticides on birds. They concluded that, although there are limitations with field investigations, particularly uncontrollable variables that must be addressed, the value of a well-designed field study far outweighs its shortcomings.

Few field studies exist on the reproductive effects of pesticides on birds and mammals. The majority of available avian and mammalian field studies focus on effects on mortality. **Hart et al. (2005)** identified the main areas of difficulty in conducting assessments of the long-term risks to birds and mammals. These are stated to be grouped under the following headings: toxicity endpoints, extrapolation of chronic toxicity between species, exposure assessment, mismatches between exposure in the laboratory and the field, and how to evaluate effects at the population level.

In the current tiered approach for risk assessment, potential risk for birds and mammals is identified on the basis of responses of individual organisms observed in controlled laboratory experiments. However, ecological assessors have long argued that except in the case of threatened or endangered species, the abundance and persistence of populations of organisms are more relevant as endpoints for assessment than are responses of individual organisms **SANCO**

(2000). Population-level assessments generally require the use of models to integrate potentially complex data about the effects of toxicants on life-history traits, and to provide a relevant measure of ecological impact. Modelling is often considered a much more cost effect approach compared to conducting field studies.

Overall, ecotoxicological studies for pesticide risk assessments for birds and mammals strive to develop cause-and-effect relationships between pesticide application and adverse effects and to determine the mechanisms by which observed effects occur **Fairbrother (1993)**. Higher tier testing methods can be a useful tool for pesticide risk assessment. However, rigid protocols are not usually appropriate and test methods and approaches should generally be considered on a case-by-case basis. In planning higher tier tests, the importance of defining the specific objectives, optimal study design and appropriate analysis of the data, should be highlighted **Ganio (1994)**.

3.3 Aquatic Ecotoxicology

3.3.1 Background

Aquatic 'tier 1' risk assessments for the evaluation of plant protection products use laboratory single species data (fish, *Daphnia* and algae). When the Toxicity/Exposure ratio (TER) values using these data are shown to be lower than the trigger values of 100 (acute fish and invertebrates) and 10 (algae, aquatic plants/bacteria and chronic fish and invertebrates) then acceptable higher tier risk assessment will be required before use of the product may be authorised. The report from the Higher-tier Aquatic Risk Assessment for Pesticides (HARAP) workshop **Campbell et al. (1999)** considered different types of higher-tier studies and developed guidance on how to apply these methods. The workshop noted that higher tier studies included further single species studies, indoor multi-species tests (microcosm) and field tests (microcosm and mesocosm). The Guidance document on Aquatic Ecotoxicology in the context of Directive 91/414/EEC **SANCO (2002)** states that the term 'microcosm' can be used for small-scale studies, whereas the term 'mesocosm' generally refers to larger outdoor tests.

SANCO (2002) also specifies the following uncertainties that need to be addressed when extrapolating single-species laboratory data to a multi-species ecosystem as follows:

- Intra- and inter-laboratory variation of toxicity data
- Intra- and inter-species variation of toxicity data
- Short-term to long-term/chronic toxicity extrapolation (temporal)
- Extrapolations of mono-species laboratory data to field impact on ecosystems (spatial and temporal)

It is stated in the Guidance Document that there are data to show the uncertainty in the first three bullet points but relatively little data if any to support that detailed in bullet point 4.

It is considered that higher tier aquatic data for refined risk assessment should provide additional information on effects seen from realistic exposure (i.e. in presence of sediment, macrophytes etc) and/or to address some of the uncertainties listed above. The literature search was designed to capture published data using higher tier aquatic testing methods. The studies found in the literature which are detailed in the sections below relate to modified exposure tests, Species Sensitivity Distributions (SSDs), indoor and outdoor microcosms, indoor artificial streams, as well as, outdoor mesocosms and artificial streams. These general method types are the same as those not in the public domain which have been used in the support of plant protection product authorisations.

3.3.2 Laboratory studies

3.3.2.1Tests with additional species

Laboratory data on additional species may exist already for older chemicals or may be generated in order to perform a species sensitivity distribution (SSD). A number of different approaches have been proposed for selecting additional test species. The preferred approach will depend on a number of factors including whether the substance has a known mode of action **Boxall et al. (2002)**. For substances without a mode of action specific to a particular tax on, the Aquatic Dialogue Group (SETAC) proposed that the test species should include at least two species of fish, one invertebrate and one aquatic plant (macrophyte of algae) plus four other species. The US EPA recommended that additional data should include invertebrate acute and chronic tests, sediment toxicity tests, rooted plant testing and amphibian testing. Based on an understanding of

the mode of action of a compound, it may be possible to identify and group sensitive and less sensitive organisms. This allows the testing strategy to be focussed on the groups at high risk.

SSDs' are known to have been performed in support of plant protection products; however, these data are not in the public domain. It should be noted that an opinion of the Scientific Panel on Plant health, Plant protection products and their residues was reported regarding the possibility of lowering the uncertainty factor if additional species were tested (EFSA 2005). Of note in this opinion is the fact, due to the legislation requiring testing of two fish species instead of only one for aquatic invertebrates, it was considered that a different procedure for fish was required when additional species were tested. However, the authors of a more recent presentation note that some species (e.g. the rainbow trout) are observed to violate the assumption of exchangeability i.e. they are non-exchangeable **Hickey et al. (2009)**. The authors stipulate that if a species such as rainbow trout is non-exchangeable and sensitive, then it will result in increased conservatism. In light of this, it is recommended that the procedures for use of SSD data and their relation to uncertainty in the risk assessment should be based on the known information for the species tested (i.e. consider if the critical species is exchangeable or not).

As noted in (van den Brink, Blake et al. 2006)**,** the taxonomic composition of the species assemblage used to construct the SSD does have a significant influence on the assessment of the hazard (e.g. only sensitive primary producers should be included for the risk assessment of herbicides). No systematic difference in sensitivity between standard and non-standard test species was observed. Hazardous concentrations estimated using laboratory-derived acute and chronic toxicity data for sensitive freshwater primary producers were compared to the response of herbicide-stressed freshwater ecosystems using a similar exposure regime. The lower limit of the acute HC5 and the median value of the chronic HC5 were protective of adverse effects in aquatic micro/mesocosms even under a long-term exposure regime. The median HC5 estimate based on acute data was protective of adverse ecological effects in freshwater ecosystems when a pulsed or short-term exposure regime was used in the microcosm and mesocosm experiments.

There was also concordance between the predictions from the effect model PERPEST and the concentrations at which clear effects started to emerge in laboratory and field studies. However, compared to the SSD concept, the PERPEST model is able to provide more information on ecological risks when a common toxicological mode of action is evaluated as it considers both recovery and indirect effects.

Similarly, **Maltby et al. (2005)** noted that the species assemblage used to construct the SSD for insecticides does have a significant influence on the assessment of the hazard, however, it was noted that habitat and geographical distribution of species do not. Hazardous concentrations estimated using laboratory-derived acute toxicity data for freshwater arthropods (the most sensitive taxonomic group for insecticides) were compared to the response of freshwater ecosystems exposed to insecticides. The sensitivity distributions of freshwater arthropods were similar for both field and laboratory exposure, and the lower HC5 (95% protection with 95% confidence limits) estimate was protective of adverse ecological effects in freshwater ecosystems. The corresponding median HC5 (95% protection level with 50% confidence limits) was generally protective of single applications of insecticide but not of continuous or multiple applications.

A study by **Hose et al. (2004)** compared Australian and non-Australian laboratory species based SSD curves and compared them to local mesocosm experiments and field monitoring data. The SSD curves indicated that the sensitivities of Australian fish and arthropods were not significantly different from those of corresponding non-Australian taxa. Arthropod taxa in the mesocosm were less sensitive than taxa in laboratory tests, which suggests that laboratory-generated single-species data may be used to predict concentrations protective of mesocosm systems. SSDs based on laboratory data were also protective of field populations.

Boxall et al. (2002) noted that extrapolation procedures mostly lead to lower ‘safe’ values than NOECs from multi-species studies demonstrating that field effects can be predicted provided uncertainties related to mode of action are accounted for. There are a number of limitations to

additional species testing: requirement for large dataset; disagreement on the number and taxonomic distribution of taxa to be tested; test results may not be comparable; test guidelines not available for some species; organisms may not be from the same sensitivity distribution and lack of knowledge on the ecology and physiology of a test organism may mean that extrapolations are difficult.

3.3.2.1.1 Testing of additional laboratory species – conclusion

In conclusion, testing of additional laboratory species is a valuable higher tier method. As discussed above it appears that use of appropriate HC5 values from SSD using appropriate taxonomic data will generally be at least if not more protective than NOEC values from multi-species studies. The reduction in uncertainty will depend on the number and quality of the studies used in the SSD.

3.3.2.2 Modified exposure studies

A summary of potential additional realistic exposure scenarios that can be utilised in laboratory studies is given by **Boxall et al. (2002)**. These include time-to-event analyses, variable and pulsed exposure and inclusion of dissipation processes. The published papers summarised below include examples of a variable/pulsed exposure design for algae and two examples including dissipation processes (leachate from soil cores and sediment and water from vegetated and non-vegetated greenhouse microcosms). It should be noted that, although no time-to-event analyses are summarised from the literature, data to address this may be present in the data recorded during standard laboratory tests.

Various authors have described modified exposure methods for algae. A flow-through method with *Selenastrum capricornutum* is described by **Grade et al. (2000)** based on modification of the standard OECD guideline 201. It is stated that digitally controlled pumps mean that the system is suitable for testing substances with any required exposure regime and in combination with metabolites. Additionally, posters describing a chemostat system (flow-through) to allow continuous culture of algae with time variable exposure and OECD 201 culture medium is presented in **Weber et al. (2009)** using the algal species *D. subspicatus and P. subcapitata.* Both methods specify that more realistic exposure patterns can be simulated over longer exposure durations. Modification of the standard algal study according to OECD guideline 201 with addition of sediment was described by **Shillabeer et al. (2000)**. This approach allows more realistic exposure but it was found that only certain sediment types are appropriate for the test to avoid interference with algal growth.

Another type of modified exposure test design described by **Abrantes et al. (2008)** incorporated leachate from a terrestrial model ecosystem (soil core) in standard aquatic laboratory tests with algae/cyanobacteria and *Daphnia*. The advantage of this type of study is that organisms should be exposed to realistic levels of compound and its metabolites that could result from compounds leaching from soil. The disadvantage of the study design is that the results could be deemed soil (or site) specific. Additionally, the authors noted up to one week could elapse between leachate collection and use in the bioassays. Confirmation of lack of degradation or more rapid use of the leachate may be necessary on a case by case basis.

In another study, **Bouldin et al. (2005)** set up vegetated and unvegetated greenhouse microcosms using ditch sediment and dechlorinated tap water. Microcosms were either populated with monocultures of *Ludwigia peploides* (water primrose) or with monocultures of *Juncus effusus* (soft rush) or with no vegetation. Following treatment with either atrazine or lambda-cyahlothrin, water and sediment were extracted from the microcosms (0 h, 3 h, 24 h, 7 day, 14 day, 28 day and 56 days) and used in standard toxicity tests: 48 h acute tests with *Ceriodapnia dubia* (water

flea) and *Pimphales promelas* (fathead minnow) and *Chironomus tentans* (midge larvae) survival and growth in solid-phase 10 day sediment tests. The authors note that vegetation in drainage ditches are a possible mitigating factor for run-off exposure to other organisms from pesticides.

3.3.2.2.1 Modified exposure studies – conclusion

Matching the exposure regime with that expected in the environment could provide more realistic effects data. This approach would comply with the recommendations of the ELINK workshop **ELINK (2008)**. However, it is important to remember that realistic exposure will be site specific depending on whether exposure is from spray drift and/or run-off and the type of water body exposed (lentic, lotic, vegetated or not etc.). In terms of risk assessment, it would be best to select any appropriate additional testing based on the worst case surface water scenario determined by the FOCUS surface water modelling. Additional toxicity data using modified exposure techniques may lead to higher toxicity endpoints but it should be noted that the same trigger values could apply unless the species used in the modified test was known to be the most sensitive species from an SSD type approach.

3.3.2.3 Population level studies

As noted in **Boxall et al. (2002)**, both modelling and experimental approaches can be used to determine population level effects.

3.3.2.4 Tests with sensitive life-stages

No specific published methods of tests with sensitive life stages were found using the literature search. However, the brief summary in **Boxall et al. (2002)** notes that standard ecotoxicity studies generally focus on neonate or juvenile animals as these are likely to be the most sensitive life stage. However, in cases where it is known that a substance is likely to be more toxic to a life stage not studied in the standard tests, it could be appropriate to do additional tests. There are examples of older organisms being more sensitive than younger ones, e.g. older daphnids were shown to be more sensitive to chlorpyrifos than neonates. Also, two-week old tadpoles were shown to be more sensitive to endosulfan than newly hatched tadpoles because endosulfan affected the post-hatch development of the neuromuscular system. However, in general it should be noted that the testing of sensitive life stages could be problematic due to the lack of test and culture methods for some species.

3.3.3 Laboratory multi-species tests

3.3.3.1 Indoor defined microcosm tests comprising well-defined assemblages of organisms at different trophic levels to assess critical ecological threshold levels

Sugiura (1992) tested different compounds for effects in a multi-species microcosm containing green algae (*Chlorell*a and *Scenedesmus*), a filamentous blue-green alga (*Schizothrix*), a ciliate protozoa (*Cyclidium*), two rotifers (*Philodina* and *Lepadella*), aquatic oligochaetes (*Aeolosoma*) and bacteria (> 5 species) in the early stages of succession. Population densities and community metabolism were measured for 25 days.

In another paper, **Williams et al. (1992)**, constructed indoor microcosms with a vertical biological filter in the centre of each aquarium containing nitrifying bacteria. Plants,

invertebrates and fish from 6 phyletic groups were selected based on ease of accessibility and to represent different trophic levels. In both experiments, fish were separated from the shrimp and plants by the biological filter to prevent predation. Effects of genetically modified *Pseudomonas* were tested (3 replicates) along with three control systems. Each aquarium had 15 of each non-target test species and the study duration was 15-29 days. Survival, water quality and fate of the test compound (genetically modified *Pseudomonas putida*) were monitored. Advantages of the system were that it is easily set up and replicated and fish separation meant that they did not disrupt the system. However, not all snail species were considered suitable for the test design.

Leeuwangh et al. (1994) described three trophic levels kept in separate sub-systems, connected by recirculating flow. An autotroph sub-system containing algae (*Scenedesmus*), a herbivore subsystem (*Daphnia*) and a decomposer sub-system (bacteria on a sand filter). Steady state was achieved which improved statistical analysis. Strengths of these test system are high replication potential in both time and space and low cost.

An aquatic indoor microcosm was used by **Liebig et al. (2008)** to study effects of the pesticides parathion-methyl and prometryn on phototrophic flagellates (*Cryptomonas* sp.), predatory ciliates (*Urotricha furcata*) and bacteria. Three trophic levels were represented: producers (autotrophic flagellates), consumers (algivorous ciliates) and decomposers (unspecified bacterial community). The combination of these organisms in the same aquatic medium under defined conditions was defined as an indoor multispecies microcosm test system representing a canonical community (can-com). Canonical in this context means 'the simplest representative that still has all essential properties of the microcosm system' e.g. nutrient assimilation, growth, degradation, and nutrient cycling. Objectives were to generate effects data for species in 3 trophic levels and to obtain data for modelling effects of biological and toxicological stressors based on assumptions of the Dynamic Energy Budget (DEB) theory. DEB-Tox models allow integration of all the data produced during the test period resulting in an overall 'no effect concentration' (NEC) which is independent of the evaluated time point. In contrast the NOEC is derived

statistically for a certain parameter and for a certain time point of exposure. The advantage of such a system is that it is simple and low in cost, whilst assessing both direct and indirect effects across several trophic levels.

3.3.3.2 Defined laboratory microcosm tests – conclusion

Defined laboratory microcosm tests allow controlled conditions, dose response and replication; however, full ecosystem complexity (species sensitivity) and realistic exposure are not covered by the design. Although such testing systems do not provide much additional information for risk assessment, they may be useful to better understand indirect effects and to better define complex field studies (Campbell, Arnold et al. 1999).

3.3.4 Indoor semi-realistic microcosms comprising complex natural assemblages

3.3.4.1 Lentic systems

A systematic series of experiments to determine the optimal design and procedures for including turbulence and benthic materials in lentic microcosms were performed **Harte (1984)**. Microcosms (50 L) were set up with water and sediment from a reservoir in the San Francisco bay area housed in a controlled environment microcosm facility with run-times of 2-3 months. Additionally, the publications considered 4 litre microcosms with water and sediment cores from a sub-alpine pond placed semi-submerged on a wooden structure floating on the pond. The chemical and taxonomic variables (phytoplankton and zooplankton) of the microcosms were compared with those in parent water bodies. The data confirmed that more realistic conditions (i.e. inc. a benthic layer) lead to better simulation in the microcosm. The benthic core experiments showed that the greatest similarity of microcosm to natural water body was

achieved using ratio of benthic core area to the overlying water volume roughly equal to the ratio of the lake sediment to the lake volume. **Harte (1984)** concludes that microcosm containers should be 4 litres or larger, and that 50 litres, if practical, is best. Polyethylene is the preferred material and cylindrical geometry for tank shape is recommended. Water body temperature and lighting patterns should be mimicked in the microcosm. It was noted that less agitation is required in the microcosm compared to the actual water body. It was recommended that the entire water stock should be taken from water body of interest, with use of benthic sediment and triplicate replication. The authors recommend prevention of algal wall growth by replacing the container every week. This type of set up allows increased realism for assessing the fate and effects of chemicals to phytoplankton and zooplankton.

Landner et al. (1989) undertook community testing with natural associations of periphyton and phytoplankton. Small samples were derived from natural communities of periphyton or phytoplankton. The sensitivity of toxicants was estimated using short-term measurements of photosynthesis in laboratory experiments. The basic assumption was that changes in metabolic activity precede and are indicative of structural and functional changes that occur in the community during prolonged exposure. Advantages of the test are that it is easy and rapid so that many replicates can be easily handled. It may provide useful information for comparative ecotoxicology e.g. for determination of seasonal or regional differences in algal sensitivity. However, the use of metabolic process as a test parameter is dependent on mode of action of the toxicant. It is not clear if photosynthesis is an appropriate test parameter, however, use of integrating parameters (e.g. growth or community structure) would be independent of mode of action and may be worth investigating. Additionally, the method does not enable detection of long-term effects. Functional endpoints alone will not reduce the uncertainty in the tier 1 risk assessment; however, such data could be useful to help in the design of higher tier studies.

Leeuwangh et al. (1994) described indoor derived microcosms designed to simulate the community of Dutch drainage ditches. Twelve microcosms (1.1 long x 1.1 m wide x 0.7 m high)

were made from glass aquaria. Natural sediment from a lake was introduced (0.1 m) and overlying water from outdoor mesocosms (0.5 m). The sediment also provided many freshwater species, including micro-organisms, algae, zooplankton, snails and oligochaete worms. Several mobile macroinvertebrates (e.g. isopods, amphipods, insects) characteristic of Dutch drainage ditches were also deliberately introduced. Acclimation was for three months with interconnection of microcosms by tubes and recirculating of water until four weeks prior to application. It was noted that it was difficult to take samples of macro-invertebrates and macrophytes at regular intervals without disturbing the system significantly. This problem was overcome by using artificial substrates where macro-invertebrates were collected and counted before putting back in the system and macrophytes were only harvested at the end of the experiment. *In-situ* cages were therefore used with susceptible arthropods to gain insight on their recovery. The advantage of this type of system is that medium complexity is achieved at relatively low cost. A disadvantage is that the system is relatively small and thus less complex than outdoor ditches they are designed to mimic. Additionally, some insect species cannot be maintained for long periods as they are lost on emergence and there is no source for re-colonisation. The system could potentially reduce uncertainty related to multi-species and spatial parameters.

Six microcosms were set up with a volume of 300 L and a bottom surface of 4800 cm^2 **Traunspurger et al. (1996)**. They were maintained in a greenhouse with controlled temperature and lighting. Sediment and water were obtained from an extensively cultivated fish pond. The population density of the zooplankton was controlled by 4 encaged fish (*Puntius semifasciolatus* Schuberti). Five snails (*Appolaria* sp.) were placed in the fish cages to prevent proliferation of epiphytic filamentous green-algae. The microcosms were allowed to acclimatise for several weeks under slight aeration and were interconnected to allow uniform conditions between the microcosms. The connections were removed one day before application of isoproturon (IPU) in acetone. Two water controls and one control with acetone and one aquarium for each of the three IPU concentrations. Sampling of water, sediment, phytoplankton, zooplankton, nematodes and microbial degradation was undertaken over the 56 day period. The advantage of this type of

microcosm is that it can assess the fate and effects of chemicals on multi-species from several trophic levels. However, this study design included no replication of treatments and thus the statistical power is low.

Barry et al. (1998) detailed the establishment of aquatic indoor microcosms using Australian sediment from the basin of dried temporary ponds. When the sediment was flooded with 3L of distilled water to make microcosms, the resting stages or eggs of zooplankton, phytoplankton, macrophytes and filamentous algae activated into a community in the period of a few weeks. A total of twenty 3L microcosms were established. Of these, 16 were selected after 6 weeks for treatment with endosulfan at 3 application rates plus a control (four replicates per treatment). The pH, conductivity, dawn and dusk oxygen levels, ammonia, nitrite, orthophosphate, chlorophyll a, zooplankton and phytoplankton were measured at weekly intervals. After 10 weeks, the total composition of each microcosm was determined, as well as, test substance residues in sediment and macrophytes. The advantage of the design is that small reproducible experimental units can be easily and quickly produced. One disadvantage was that due to the small size, sampling lead to dilution of the system. In addition there may be fewer microcrustacea species in the system as compared to the environment, as well as the absence of insect fauna.

Relatively simple laboratory microcosm experiments were conducted in HDPE microcosms (7L capacity) filled with synthetic, moderately hard dilution water under controlled light and temperature **Pratt et al. (1998)**. Three different nutrient regimes were applied using phosphate and nitrate (low, medium and high). Naturally derived microbial populations were collected on polyurethane foam (PF) substrata that had been placed within 50 cm of the surface of a eutrophic local pond for 21 days. Two colonised substrata were randomly placed in the centre of each microcosm and served as sources of micro-organisms (“epicentres”). Four sterile, barren “island” PF substrata were placed around the epicentres in each microcosm. Over time, the “island” substrata became colonised by microbial species from the “epicentres”. Sediments were not

included in the microcosms in order to allow longer exposure to the test material diquat as well as a simpler design. Microcosms were developed under the different nutrient conditions for 25 days before dosing 3 microcosms at each nutrient level with a single application of diquat. Nutrient and diquat levels were monitored during the experiment. Island substrata were sampled before and after diquat dosing (up to 23 days after) by removing one substratum from each microcosm. Microbial communities were harvested by squeezing each substratum into a beaker. Total protein (microbial biomass), chlorophyll a (algal biomass), alkaline phosphatase (APA) and electron transport system activity (ETSA) were measured. Glutaraldehyde fixed aliquots were used for algal enumeration. Gross photosynthesis and respiration were estimated. Relative abundance of dominant algal taxa was reported for Cyanophyta, *Golenkinia* sp. and *Scenedesmus* sp. It was concluded that nutrient treatments had a small influence on toxicant effects; the magnitude of the herbicide effects was comparable across nutrient levels. However, the capacity for recovery was lower in low nutrient microcosms where the herbicide persisted longer. The advantages of such a system are that it is relatively cheap and easy. It is worth noting that appropriate nutrient levels could be selected to mimic either an average or worst case situation. Disadvantages of such a system are lower realism than larger more complex systems and variation in "epicentres". This type of system could potentially reduce uncertainty associated with species sensitivity for algal species if sufficient numbers of algal species are confirmed to be present in the system. Additionally, the system could be useful as a screening test prior to more complex microcosm/mesocosm studies in order to determine relevant dose rates.

Simple indoor (12 x 8 litre) microcosms with controlled light and temperature were formed using water taken from a pool North of the Alterra building in the Netherlands **Daam et al. (2003)**. Water was sieved to exclude *Chaeborus* larvae (zooplankton predator). Nutrients were added to stimulate growth of phytoplankton, and zooplankton from the pool were then added three days before the experiment start. Measurements were taken of water parameters, chemical concentrations, chlorophyll a, decomposition of particulate matter and phyto/zooplankton and snails. Differences in structure of zooplankton communities were visualised by Principal

Component Analysis (PCA). The advantages of this simple system is the relatively low cost and potential to determine the appropriate concentration range for testing in larger studies, as well as, identification of sensitive phyto- and zooplankton species. The disadvantages of such a study design are lack of ecological complexity and possible overestimation of exposure due to absence of sediment and macrophytes. Realism can be considered to be intermediate between laboratory single species tests and microcosm/mesocosm tests that include sediment etc. However, this type of study could be considered to represent worst case exposure of phyto- and zoo-plankton and thus results could be useful to support argumentation for reduced assessment factors being applied to other higher tier data.

Another study carried out in the Netherlands used twelve indoor freshwater microcosms of approximately 600 L in volume **Van Wijngaarden et al. (2004)**. Each microcosm had a sediment layer (sandy loam from a Netherlands lake) of 10 cm and a water column (unchlorinated tap water) of 50 cm. Light and temperature were controlled. *Elodea nuttallii* shoots, plankton and macroinvertebrates collected from uncontaminated drainage ditches were introduced to develop a macrophyte-dominated freshwater community. Acclimation and interconnection of the microcosms was undertaken for two months. Circulation between microcosms stopped and then after 15 days the first pesticide applications were made. Microcosms were aerated slightly to maintain some water movement and low levels of nutrients were added weekly to support plant growth. A rack with glass slides was used to study periphyton. Pebble baskets and multiplates served as artificial substrates for macroinvertebrates and a petri dish containing leaf material was used to study decomposition Treatments were performed in duplicate plus four controls. Monitoring of pesticide levels, physicochemical properties, phytoplankton, periphyton, zooplankton, macroinvertebrates, macrophytes and decomposition (poplar leaves) was undertaken for thirteen weeks (92 days). The advantage of this system is a high level of realism under controlled environmental conditions. However, disadvantages are that the laboratory design prevented certain recovery processes (immigration etc.) and may represent worst case exposure compared to field (lack of sunlight induced

photolysis and water movement and dilution expected in the natural environment). It is considered that reduction in uncertainty for species sensitivity and some spatial and temporal attributes may be possible using this type of study design.

Laboratory (indoor) microcosms (water volume approx 14 litres) with temperature, light regime and nutrient levels that simulated cool 'temperate' and warm 'Mediterranean' environmental conditions were compared **Van Wijngaarden et al. (2005)**. The fate of chlorpyrifos in the water column was monitored and the effects on zooplankton, phytoplankton and community metabolism were followed for four to five weeks. Sediment and water were collected from an uncontaminated eutrophic ditch in the Netherlands and used to provide a sediment layer of approximately 2 cm and a water layer of 30 cm. The systems were seeded with zooplankton and phytoplankton from uncontaminated waterbodies from the same location and from a pond in the Netherlands. Some of the experiments used *Daphnia* gr *galeata* from a laboratory culture. Conditions for phytoplankton growth were provided by addition of nutrients. To suppress periphyton growth, five snails (*Lymnaea stagnalis* L) were introduced into each system. The fate of the test item chlorpyrifos was followed for 28 days. The species composition of the phytoplankton and zooplankton was determined to the lowest practical taxonomic level. Chlorophyll a and community metabolism were also monitored. Both univariate analysis and multivariate analysis (PRC) were undertaken.

The authors discussed that the organisms tested were from the temperate climate zone and may not be the same as those from warmer climates. However, it was noted that representatives of the major zooplankton groups (cladocerans, copepods and rotifers) are expected in freshwater systems all over the world. More specifically the same representatives of the groups sensitive to chlorpyrifos (cladocerans, i.e. *D.*gr *galeata, S. vetulus* and copepod nauplii) can also be found in the Mediterranean. The data provided by Van Wijngaarden, et al. (2005) and others indicates that there is no or only minor differences in sensitivity distributions to be expected for chlorpyrifos between plankton communities in temperate and warmer freshwater systems.

Additionally, these indoor microcosm experiments showed that the $NOEC_{community}$ for chlorpyrifos was 0.1 µg/L. The same $NOEC_{community}$ was derived from an outdoor mesocosm study with a single application of chlorpyrifos. It can therefore be concluded that in this situation with a substance with a short DT_{50} and where a small system can maintain the most sensitive organisms (in this case cladocerans) then there is no significant difference in the safe threshold levels determined from a small scale experiment compared to a large scale one. It was noted that above the $NOEC_{community}$ threshold level, responses and effect chains differed between experiments. Advantages of this simple test design are that it is of relatively low cost and easily replicated. It is appropriate for organisms that can be maintained at high levels in a small scale experiments e.g. plankton. These data indicate that a relevant $NOEC_{community}$ can be derived if the most sensitive organisms are included and the toxicant has low persistence. Alternative environmental parameters could be assessed. A disadvantage may be that indirect effects may not be determined in a small scale system and it may not be appropriate for species that are not easily maintained in small scale systems and/or for persistent compounds. Overall this type of study has an intermediate level of realism that could be highly useful in risk assessment if the most sensitive species are plankton. Reduction in uncertainty associated with species sensitivity and temporal parameters could be possible for appropriately designed studies.

Finally, **Chang et al. (2005)** set up indoor microcosms in 20L cylindrical tanks in a temperature and light controlled environment. The microcosms used 1 kg of bottom mud from a eutrophic Japanese Lake. The green alga *Chlorella* was added and then a predator (*M. pehpeiensis*) was introduced at two different densities (at day 24). Zooplankton communities were then monitored until day 33 and then carbaryl was added with further zooplankton measurements for 13 days. Assessment of zooplankton community structure and food web analysis was undertaken. This type of study is of intermediate realism but is relatively easy to set up and includes replication. Such a design may be useful to understand specific effects on zooplankton prey seen in larger studies.

3.3.4.2 Lotic systems

Four artificial laboratory stream designs **Pontasch et al. (1989)** with and without flow-through and current were evaluated (static; static with current; flow through no current and flow-through with current). Flow-through and current when provided were 12 ml Min^{-1} and 30 cm^3 sec -1 respectively. Oval artificial streams (1.7 x 0.24 x 0.13 m channel) were constructed of moulded fibreglass. Each design was evaluated in triplicate. Daylight equivalent lighting and temperature were provided and each stream was covered with an emergence trap. Test organisms were derived from a relatively unimpacted riffle habitat in the US. Macroinvertebrate communities were sourced following 30 day colonisation of artificial substrates placed in the riffle. Periphyton communities were collected on polyurethane foam (PF) artificial substrates. At experimental start, two colonised PF substrates were placed in each artificial stream and squeezed to initiate growth of periphyton as a food source for the macroinvertebrates. Three colonised macroinvertebrate substrates were assigned randomly to each of the 12 streams (a further 9 substrates were sampled to provide an estimate of macroinvertebrate abundances). Samples taken using a 350 µ mesh were taken from the source riffle for comparison. Adult insects were collected from each artificial stream every 48-72 hours throughout the experiment. After 30 days all organisms in the artificial streams were sampled, identified to the lowest possible taxonomic unit and enumerated. Total density (adults and young combined) per taxon in each artificial stream were determined. Data were analysed by one-way ANOVA in conjunction with the Least Significant Difference Criterion for the separation of means.

It was found that certain species e.g. *Isonychia* (filter feeder) and the mayfly genus *Baetis* (Baetidae) require a current for long-term maintenance in laboratory streams. Artificial streams supplied with current were able to maintain all mayfly taxa at or above initial levels for the entire 30 day experiment. It was noted that a supplemental source of food may be necessary during long-term experiments with large numbers of hydropsychids. The artificial substrates did collect the same relative or absolute abundance of some taxa compared to the source riffle, however, the

number and kinds of species collected in the two sample types were nearly identical. The advantages of the system are some control of environmental conditions whilst enabling a relatively long multi-species bioassay that includes some moults and complete life cycles for some species. Disadvantages of the laboratory system are lack of recovery due to e.g. recolonisation from up-stream. It is considered that such a design provides high realism useful for refinement of the risk assessment for macro-invertebrates in the lotic environment. Appropriately designed experiments should lead to reduced uncertainty as a result of multi-species data for macro-invertebrates and some reduction in spatial and temporal uncertainty.

Fifteen laboratory stream microcosms of the same dimensions as detailed in the publication above were used to investigate effects from continuous exposure to fenvalerate to a riffle insect assemblage **Breneman et al. (1992)**. Periphyton were collected on polyurethane foam units from a riffle area in the Volga River headwaters left in place for seven days before removal and extraction. Periphyton slurry was added to the microcosm and allowed to develop for five weeks before test initiation. Macroinvertebrates were collected over six weeks in rock-filled plastic containers. Macroinvertebrates were allowed to acclimate for two days in the artificial streams before application of the test substance. Each stream was covered with an emergence net. Light/dark, temperature and water volume was maintained. Streams were pulse dosed to nominal concentrations by addition of 550 mL of stock solutions and then continuous dosing at the 4 treatment rates of fenvalerate plus control (three replicates per treatment). Macroinvertebrate data were used to quantify mortality during transportation and avoidance to initial fenvalerate exposures. After 30 days, the contents of each microcosm were sampled and macroinvertebrates enumerated. The advantage of this design is a high level of realism for risk to stream macroinvertebrates (multi-species) when continuous exposure is anticipated. A disadvantage of this design is that continuous exposure may be overly conservative and not relevant to pulsed exposure that is usually considered in the current risk assessment scheme for plant protection products. As noted above, appropriate design should enable reduced uncertainty in relation to

species sensitivity for stream macro-invertebrates, as well as, possibly some reduction in spatial and temporal uncertainty.

Lowell et al. (1995) describe a series of short-term (48 hour) toxicity experiments with the mayfly *Baetis tricaudatus* Dodds to determine the effect of current velocity on mayfly response to the reference toxicant sodium chloride. The tests were run at three substratum-level velocities: low (0 cm/sec), medium (6 cm/sec) and high (12 cm/sec). The endpoints measured for the mayflies were immobilisation and number of moults. Mayflies were collected from a Canadian creek and kept in the laboratory under controlled conditions prior to the start of experiments. Artificial streams were circular Plexiglas artificial streams: diameter 8.8 cm, stream bottom area 50 cm^2. The bottom of each stream had been roughened with sandpaper to enable the mayflies to remain attached. Current was produced by small water jets driven by pumps drawing from a supply reservoir, water was returned to the reservoirs *via* a central standpipe drain in each stream. Each experiment utilised thirty artificial streams (10 x *B. tricaudatus* per stream): five replicate streams at each current velocity per control and experimental concentration. No refugia were provided in the streams so that the mayflies could not move into microenvironments. To determine the LOEC, each of the experiments was analysed using ANOVA. Initial results show that short-term toxicity tests using lotic organisms should ensure that the animals are exposed to at least some flow (possibly 6cm/sec and above). The advantage of this type of study laboratory design is the ability to generate replicated effects data in a lotic environment for a sensitive ephemeropteran species. Data from this type of study could be useful in determining dose rates for micro or mesocosm studies if mayflies are known to be sensitive. The disadvantage of the study design is a low level of realism with only one species in short-term test without the presence of environmental refugia and other species. Reduction of uncertainty in the risk assessment is unlikely from this type of data in isolation.

3.3.4.3 Indoor semi-realistic microcosms – conclusion

In comparison to single species tests, indoor semi-realistic microcosms with relevant assemblages of organisms known to be sensitive should enable reduction in uncertainty due to species sensitivity. Additionally, uncertainty due to spatial and temporal factors could also be reduced depending on the set up (i.e. similar environmental conditions to the field and/or sufficient duration to show recovery). Laboratory semi-realistic microcosm data may be useful in their own right in the risk assessment or they may provide useful information for the appropriate set up of additional higher tier data (e.g. outdoor micro/mesocosm). Additionally such studies may provide supporting argumentation for reduction of assessment factor in conjunction with other higher tier (micro/mesocosm) data.

The appropriate design of semi-realistic microcosms will be specific to the plant protection product both in terms of its expected environmental fate and its expected aquatic toxicity effects.

3.3.5 Field studies

Field and semi-field studies of plant protection products in multi-species lentic systems were first conducted in the late 1960s and early 1970s but only used to support registrations in the early 1980s. The studies allow assessment of organism interactions, as well as, more realistic exposure approaching that occurring in the environment.

The book 'Aquatic Mesocosm Studies in Ecological Risk Assessment' (Graney 1994) presented the collected papers from a symposium on 'Utilization of Simulated Field Studies in Aquatic Ecological Risk Assessment' held at the 11th Annual meeting of the Society of Environmental Toxicology and Chemistry on November 11-15, 1990, in Arlington, Virginia. This book noted by way of introduction that field testing can either be natural field studies or simulated field

studies. Natural field studies are site specific and designed to evaluate the impact of chemicals on specific ecosystems and are not designed to be predictive. Simulated field studies are composed either of an isolated subsection of the natural environment or a man-made physical model of a lotic or lentic ecosystem.

Various test systems were discussed in the book, including:

- Large pond systems (volume 100,000 to 1000,000 L) – artificially constructed earthen ponds which are allowed to colonise for a predetermined period and fish are stocked prior to treatment. These studies were historically required by the EPA.

- Outdoor microcosms (volume 2000 to 20,000 L) – fabricated tanks large enough to be representative of lentic ecosystems and not greatly influenced ambient environmental conditions.

- Limnocorrals (< 100L to > 100,000 L) – artificial enclosures placed in the pelagic region of ponds, lakes or marine environments. These may or may not be in contact with the profundal region. Fish are generally excluded from these systems.

- Littoral Enclosures (approx. 50,000 L and maximum depth 2 m) – plastic dividers are used to isolate the littoral region (shoreline) of ponds.

- Lotic systems – artificial streams of various sizes, no standard design.

Chapter 18 of the book (Graney 1994) notes various factors to consider when establishing an 'artificial' aquatic ecosystem: (1) system construction (2) source of inoculum (3) time required for colonisation (4) control of system components such as macrophytes (5) exclusion of undesirable components such as tadpoles (6) fish stocking and management. Uncertainty associated with the transport of chemicals to the aquatic ecosystem is great and this component

of exposure must be addressed independent of the mesocosm study. For a photolabile compound, researchers must consider that time of day of application may affect exposure, especially as it may take all day to dose all the replicates.

Following two expert meetings,[SETAC-Europe meeting at Monks Wood in the UK in July 1991 **SETAC (1992)** and SETAC Foundation for the Environmental Education at Wintergreen, Virginia, USA in October 1991 **SETAC (1992)**], flexible guidance for testing plant protection products in outdoor lentic freshwater systems was outlined **Matthiessen (1994)**. The motivation for this guidance stemmed from the US EPA mesocosm guideline **Touart (1988)** which was considered by many to be too prescriptive. The thrust of the revised guidance was to allow the individual design of tests to meet the particular needs of each situation. The EPA guideline **Touart (1988)** requiring the presence of fish lead to the requirement for very expensive (large (>300 m^3) systems and long (6-18 month) test durations. Although assessment of effects on fish reproduction may sometimes be necessary, it was considered that studies at lower trophic levels are usually just as sensitive and much more cost-effective. Additionally, the requirement to protect all species (biodiversity) for conservation reasons is increasing in importance and thus sole attention on fish as a resource was considered to be in decline. Both workgroups proposed similar guidance allowing for the individual design of a test to address specific problems of the test item and need not always include fish. This flexibility in design meant that there could also be flexibility in the number of exposure concentrations required (although multiple exposures considered essential) and in the number of replicates per concentration. Both guidelines emphasise that the type of statistical treatment will be influenced by the experimental design. Neither workgroup specified system size precisely, but it was clear that few, if any, applications would require a volume greater than 50,000 L or a depth greater than 1.0m. Monitoring for up to 6 months was considered sufficient.

3.3.5.1 Published Lentic field studies: pond, mesocosm, microcosm and enclosure studies

Publications relating to the design of outdoor pond, mesocosm and microcosm studies are detailed below in date order. Enclosure studies are included in this section.

A limnocorral design **Landner et al. (1989)** combined both benthic and pelagic communities using enclosures in a lake. Continuous contact with the mother system can be obtained using a flow-through device; however, this requires care with flow rate otherwise plankton may be washed out of the system. The limnocorral experiments were carried out in Lake Sömorgen in 1985 and 1986 with a duration of about 5 months. Four limnocorrals, arranged in two pairs and kept together by a central working platform, were placed in the lake about 400m from the shore at a water depth of 4m. The corrals had glass fibre reinforced polyethylene walls and were 10m in diameter and 4m deep, giving a volume of approximately 300,000 L. The enclosures were open to the sediment, into which the walls were embedded. Peristaltic pumps were used to pump water into and out of the corrals, yielding a theoretical turnover time of the water of 100 days. The advantage of this type of system is a high level of realism for exposure and effects on multi-species that can be maintained over a relatively long time period. Disadvantages are the high expense coupled with low replication potential. Additionally, it was noted that certain parameters in the enclosures differed to those in the lake: lower level of phosphorus in the enclosures compared to lake due to phosphorus binding to the enclosure walls and lowered density of fish lead to increased zooplankton levels in the enclosures. Also, each lake has its own specific parameters and the choice of lake should be made carefully. Overall it is considered that data from this type of study should lead to a reduction in uncertainty relating to species sensitivity, spatial and temporal factors.

Twelve 0.1 ha (700,000 L) outdoor mesocosms were used to monitor effects on fish (bluegill), benthos, zooplankton, phytoplankton, macrophyte biomass, diurnal oxygen and water quality over a 5 month period **Fairchild et al. (1992)**. Additionally, single species lab tests were done

for fish and *Daphnia* with and without sediment. Single species lab studies with sediment and mesocosm studies are both useful as they provide more realistic exposure. Comparisons of field and lab data showed that nominal concentrations causing adverse field effects in fish were closely approximated by the results from standard lab tests (for esfenvalerate).This type of mesocosm study reduces uncertainty relating to intra/interspecies sensitivity, short-term to long-term toxicity extrapolation (temporal) and lab to field extrapolation (spatial).

In another publication, twelve x 0.1 ha rectangular ponds (volume approx. 1,100,000 L) were used as mesocosms **Webber et al. (1992).** A shallow-water area (littoral zone), ranging from about 0.1 to 0.5 m deep, extended about 6 m from one end of each mesocosm. Selected macrophytes were planted in the littoral zone. The bottom and sloping sides of each pond consisted of packed clay soil (approx 15 cm of topsoil mixed with lime and fertiliser). In April 1987 acclimation of the mesocosms began following pumping of filtered water from a nearby reservoir. Coarse mesh filters (approximately 1 mm pore size) screened out fish while allowing the natural bacteria, fungi, algae, zooplankton, insects and other aquatic invertebrates to pass through. Water was recirculated among all mesocosms every other week before pesticide application in an effort to assure similar water quality in each pond. Inorganic nutrients were added to stimulate plankton growth at a rate recommended for US sports fish ponds although these additions were suspended in the latter phase of the acclimation period due to excessive macrophytes found in five mesocosms.

Experimental design included four treatments (3 rates of esfenvalerate and control) each with 3 replicates. To reduce variability among treatments due to habitat differences, 4 of the 5 mesocosms with excessive macrophyte growth were randomly assigned to each treatment. On 4 April 1988, a mixture of 225 sexually mature Bluegill sunfishes was stocked in each mesocosm at a rate of approximately 2,250 per hectare, typical of managed farm ponds in the SE USA. Eight grass carp (*Centopharyngodon idella*) ranging in length from 15-30 cm were stocked in each mesocosm to control excess macrophyte growth. Nutrient additions were started again (late

June 1988) once macrophyte abundance was under control. Esfenvalerate was applied in 10 x drift (18 m boom) applications and 5 x run-off (soil slurry) simulations starting in July 1988. Monitoring of various parameters was undertaken until February 1989.

Measurements include residue analysis, physicochemical analysis, phytoplankton (density, species composition and biomass (chlorophyll a and photosynthesis), community metabolism (simplified diel oxygen method), zooplankton community, macrovertebrate community and fish populations. Macroinvertebrates were sampled with artificial substrates (stationary plastic artificial substrates or SPAS) and a standard Ekman dredge (15 x 15 cm). Emerging adult insects were captured using modified floating traps. Means of most variables were estimated and tested for significance using ANOVA.

Ecosystem changes in structure and function were observed in the mesocosms as a result of both bluegill predation and esfenvalerate toxicity. The evaluation of esfenvalerate effects was enhanced by sampling the natural substrate, because data from SPAS samplers were inconclusive. It was noted that identification down to generic level showed effects not seen at the family and higher level identifications. The conclusion notes that bluegill overcrowding could be addressed by addition of proper levels of a piscivorous species (e.g. largemouth bass). An advantage of this type of design is the high level of species, inclusion of variable habitat (sloping area with littoral species) and long duration (approx 7 months). Disadvantage of this type of system is the variation between mesocosms due to varying levels of macrophytes and high fish abundance levels. The very large size of these systems makes such studies very expensive. Uncertainty due to species-sensitivity, spatial and temporal variation could all potentially be reduced although care is needed to distinguish the indirect effects caused by high level of fish predation.

A comparison of the effects of cyfluthrin in sixteen microcosms (1900 L) and fourteen mesocosms (634,700 L) was undertaken **Gregg Morris et al. (1994)**. All the cosms were established using water, sediments, biological inoculums and bluegill sunfish from the same

sources. Two microcosm controls were established one with and one without young bluegill sunfish. It is assumed that the smaller microcosms will be easier and cheaper, however, it was noted that certain effects were seen in the microcosm systems that were not seen in the mesocosm systems (i.e. slight growth effect on bluegill seen in microcosm but not in mesocosm). Bluegill affected zooplankton populations in the microcosms. Reduction in uncertainty from species sensitivity, spatial and temporal factors are all possible with both these microcosm and mesocosm designs but again care is required as effects from the fish populations on zooplankton could obscure toxicant effects.

Another publication compared small outdoor microcosms (5000 L) and large mesocosm ponds (75,000 L), in which benthic organisms, all trophic levels of planktonic organisms and caged rainbow trout were examined **Heimbach et al. (1994)**.The larger mesocosms also included aquatic plants. Microcosm data were shown to be similar to mesocosm data except for non-inclusion of macrophytes. The disadvantage from inclusion of macrophytes was increased differences in phyto- and zooplankton communities between the mesocosms. Fish in the microcosm required the addition of food during the experiment. Reduction in uncertainty is considered feasible; however, the inclusion of fish and macrophytes may have lead to variations in plankton communities.

A comparison of earthen ponds (470,000L) and fibreglass tanks (11,000 L) exposed to sulprofos by both spray drift and run-off simulation was also undertaken **Howick (1994)**. Both had benthic and limnetic communities from the same source. The ponds were stocked with adult Bluegill sunfish and the tanks with juvenile Bluegill sunfish. Dissolved oxygen, hardness and conductivity were found to be higher in tanks than ponds; total suspended solids were higher in the ponds; sulprofos concentration in the water was slightly lower in tanks; sulprofos concentrations in the sediment were lower in the ponds. The major difference in the zooplankton communities between the tanks and ponds was the persistently greater densities of crustacean zooplankton in the tanks due to predation by the many thousands of juvenile bluegill in the ponds

but not in the tanks. However, responses of zooplankton were similar despite differences in density. Tanks and ponds showed the same response for ephemeropterans and gastropods, however, chironomids were impacted at lower nominal concentrations in the tanks presumably due to the higher concentrations of sulprofos in the tank sediments. Greater effects on fish were seen in the ponds possibly due to juvenile bluegill recruitment, predation by bluegill and the toxicity of sulprofos to zooplankters and macroinvertebrates. Despite their physical differences, the tanks produced many dose-related effects that were essentially the same as those found with the ponds particularly for the zooplankton and benthic macroinvertebrate communities. The study concluded that the microcosms (tanks) were much cheaper than larger mesocosms (20% of mesocosm cost). Replication is easier with the smaller microcosms whilst they are still larger than laboratory microcosm and thus could include fish and some habitat heterogeneity. However it was noted that further design refinement in the area of fish stocking would be required. Uncertainty reduction may be possible for the multi-species components of the microcosm (but not fish as only one species) as the uncertainty due to species sensitivity will be reduced. Additionally some reduction in uncertainty due to spatial/temporal variation could be possible from this type of design.

Another publication compared the fate and effects of cyfluthrin in pond mesocosms (634,000 L) and concrete microcosms (1900 L) **Johnson (1994)**. Sampling was for three weeks pre treatment and nineteen weeks post treatment. Aqueous cyfluthrin concentrations were similar between systems, but the aqueous half-life and sediment concentrations were influenced by system scale. Biological effects (zooplankton, macroinvertebrate colonisation and aquatic insect emergence) showed parallel response patterns in both systems. Sexually mature fish used in mesocosms lead to large fish populations and subsequent reduction in zooplankton densities. It was noted that microcosms should include emergence routes linking sediments to the surface for emerging odonates. As noted previously microcosms are cheaper and easier to construct to include a reasonable level of replication. The main disadvantage of microcosms was the difference in fate. It was noted that bluegill predation and habitat (macrophytes) may affect the absolute numbers

and dominant taxa, however, sensitive and insensitive taxa were similar among the two systems. Uncertainty reduction may be possible for the multi-species components of the microcosm (but not fish as only one species) as the uncertainty due to species sensitivity will be reduced. Additionally, some reduction in uncertainty due to spatial/temporal variation could be possible from this type of design.

Leeuwangh et al. (1994) described outdoor mesocosms designed to simulate drainage ditches. Twenty uniform mesocosms (40 m long, volume 60 m^3) were created with a water depth of 0.5m and with 0.25 m of sandy loam sediment with moderate nutrient content. The sediment served as a source of benthic and pelagic organisms. Water was sourced from a local well and stored in a supply reservoir for several months before introduction into the mesocosms. One year after construction the mesocosms contained biological communities typical of macrophyte dominated shallow ponds and ditches. After two years acclimation, duplicate mesocosms were used to test 4 concentrations of chlorpyrifos with 4 control mesocosms. *In-situ* cage experiments were also included. Advantages of the outdoor design include the provision of biological complexity and the ability to study species that cannot be maintained or do not complete their lifecycle in the laboratory. Additionally, long-term primary and secondary effects and recovery of community can be studied. However, the systems afford less replication and high cost compared to laboratory data. The high level of realism means that reduction of uncertainty (multi-species, spatial and temporal) is likely to be possible.

Juettner et al. (1995) and **Peither, et al. (1996)** used compartments of approximately 1000 L within a Bavarian pond to assess the effects of pesticides on plankton for periods of up to forty-seven days. The advantage of such an outdoor enclosure design is greater realism for community effects on plankton where both direct and indirect effects can be determined. A disadvantage is that seasonal variation in planktonic populations needs also to be considered in interpretation. This type of study could reduce uncertainty related to species sensitivity (plankton species) and due to spatial and temporal variation.

In another publication, twelve outdoor mesocosms (volume approx 12,000 L) were formed by digging holes and then lining with black PVC film **Caquet et al. (1996)**. The mesocosms were filled with tap water and 10 cm of silty sediment and then left to stabilise over 8 months. Macrophytes, free and caged snails, wood lice (*Asellus aquaticus* L.) and goldfish in net tunnels were introduced into the mesocosms. The mesocosm was stated to be conceptually based on SETAC-Europe (1992) recommendations. In addition to voluntary and accidentally introduced organisms, many animal species spontaneously colonised the artificial ponds where they developed and reproduced. It was considered that the resulting communities resembled those living in natural lentic systems of the surrounding area. The realism of this system is considered to be relatively high and thus lowered assessment factors should be possible due to reduction in uncertainty due to species sensitivity (multi-species component of the mesocosm) with possible reductions due to spatial and temporal elements.

Shaw et al. (1996) conducted four month studies in eighteen (17,000 L) outdoor microcosms in 1993. Copper sulfate was applied on three occasions at five treatment rates. There were three replicates of each treatment rate and untreated controls. Microcosm enclosures consisted of open-ended, fibreglass cylinders inserted into the sediment and clay lining of flat-bottomed ponds ($1m^3$ depth). Macrophytes were allowed to develop naturally and artificial refuges were provided (four per microcosm). Each microcosm was stocked with 30 juvenile bluegill sunfish.

Bioassays were performed *in situ* in three replicates of the control and the highest treatment. The water column bioassay used two genera of water boatman (*Notonecta (*Linnaeus) and *Buenoa* (Kirkaldy), Notonectidae, Hemiptera). In addition, two epibenthic bioassays were conducted using the mayfly *Caenis* (Ephemeroptera) and laboratory-reared *Hyalella azteca* (Saussure) (amphipoda). Bioassay cages were placed *in situ* 3 days prior to introduction of the organisms to allow accumulation of natural organisms, macroinvertebrates (including potential predators) were removed from the inside of the cage prior to introduction. After 72 hours, the cages were removed and the numbers of dead and living organisms were recorded. Assessments of the

natural populations of macroinvertebrates, emerging insects and water column organisms were also made. Macroinvertebrate community structure effects were analysed using canonical discriminant analysis. The advantage of the relatively small size of the microcosm allows replication and good statistical analysis. The bioassays allow detail on potential recovery. In this study Notonectidae and *Caenis* were easy to collect for bioassays due to the close proximity of untreated ponds. That *Caenis* survival in bioassays was high and effective sampling of populations in the microcosms was also possible indicates that the *Caenis* bioassay was the most useful. Bioassays of *Hyalella* required more effort because a laboratory culture had to be maintained and since the animals had to be slowly acclimated to pond conditions. Reduction of the assessment factor for the macroinvertebrate risk assessment may be possible due to lowering of uncertainty due to species sensitivity (community data), spatial (more realistic environment for both community and bioassay) and temporal (community - actual recovery) and bioassay (potential recovery).

Van Wijngaarden et al. (1996) described outdoor experimental ditches (mesocosms): length 40 m, width at water surface 3.4 m; water volume 60,000 L. The ditches had a 0.25 m sediment layer of sandy loam and a water column that was 0.5 m deep. The ditches were lined with a water tight PVC layer. Sediment was used as the source of benthic and pelagic organisms. Prior to experimental start, the mesocosms were allowed to develop for more than two years and became dominated by macrophytes. Eight months prior to the experimental start twenty to thirty individuals of *Asellus aquaticus* L and *Gammarus pulex* (L.) were introduced into each mesocosm as these species usually appear in drainage ditches in the Netherlands. *In-situ* 48 hour bioassays in cages were undertaken to assess effects on fixed numbers from relatively constant exposure. Chlorpyrifos was sprayed at four nominal concentrations (duplicate for each treatment) plus four control mesocosms. Chlorpyrifos levels were measured and invertebrates sampled for one week after application. It was noted that short-term effects (ECs) from the laboratory and mesocosms differed by less than a factor of three for the seven species studied. The cage studies confirmed this similarity. The advantage of this design is a high level of realism as well as a

regression approach allowing quantification of an ECx outside the tested range. The disadvantage is that high biological variation could restrict the usefulness of the approach. Uncertainty due to species sensitivity and spatial factors for zooplankton can be reduced. Experiments would need to be extended to consider long-term and recovery issues (i.e. temporal uncertainty).

Forsyth et al. (1997) used twenty seven enclosures in a twelve hectare permanent pond in Canada to investigate herbicide effects on two submerged macrophyte species. Each enclosure of 1 m square was placed in water 50-70 cm deep. Enclosures were separated by about 4 m and arranged in five rows of five or six each with wire mesh around all enclosures. Macrophytes were collected from the study pond as young rooted plants. Twelve plants of each species of macrophyte (*Potomogeton pectinatus* and *Myriophyllum sibiricum*) were selected for similarity in size and root development, weighed, then planted individually in numbered 10-cm diameter plastic pots and placed on the surface of the substrate in each enclosure. The sediment for potting the plants was collected from one part of the study pond and mixed to ensure homogeneity prior to use for potting. Each of the nine treatments (pesticide x 4 each at 2 rates and control) was replicated three times. Herbicides were applied by pouring 1 litre of the appropriate solution into each of the enclosures and mixing into the water. AT thirty days post application, the plants (including all roots, rhizomes and new shoots) were carefully removed from their containers, washed and blotted dry and weighed. Plants were examined for signs of injury and numbers of floral spikes (inflorescences), then replanted in the fresh sediment and returned to their enclosures. This process was repeated at 60 days post herbicide application, at which point the number of tubers produced by each *P. pectinatus* plant was also recorded.

Preliminary studies had indicated that repeated measurements of fresh weight of the plants would not injure the plants. However, there was considerable variability of growth in plants within enclosures and within treatments during the first 30 days in treated and control groups. Severely stunted plants were deemed to have failed to adapt to transplantation to pots and resulting data

were deleted. No further deterioration or mortality due to handling was apparent in control plants at 60 days post-treatment. The advantage of such a system is that macrophytes can be tested in more realistic conditions. However, the disadvantage of this study design is that plants did not always adapt to transplantation and thus could not be used to assess the effects of the toxicant. If the study was improved to enable higher success rates for transplantation there may be scope for reduction of the uncertainty due to spatial and temporal parameters. Also, only two species of macrophytes were tested and unless it is known that these do represent the most sensitive species (i.e. information from an SSD) significant reduction of uncertainty due to species sensitivity will not be possible.

Burdett et al. (2001) published results from a replicated field pond experiment to test effects of three herbicides on aquatic invertebrates. Shallow experimental ponds were constructed at Yanco Agricultural Institute (Australia) on land not used for cropping for eight years. Two parallel rows of ten ponds with earthen banks were divided by an irrigation ditch. Channel water was siphoned from this central ditch into the twenty ponds through a PVC pipe. Each pond was approximately 38 m^2 in area and was filled to depth of 11 cm. The ponds at the ends of each row were flooded but excluded from treatment and sampling in case the immigration of insect species favoured the end ponds. After flooding the ponds were left untreated for one week to allow natural recruitment of invertebrates.

Of the sixteen remaining ponds, three replicate ponds were treated with each of three herbicides and three untreated controls. Application by spraying at the maximum spray rates to assess effects in rice growing. After one week, ponds were sampled for invertebrates using a PVC cylinder 24 cm in diameter and 60 cm high pushed into the soil. Water was bailed out into a fine gauze sieve (500 µm mesh) and due to bailing invertebrates in surface sediment layer were also collected. Further sampling occurred 5 weeks after spraying. Plant sampling was also carried out at the same time: all macroscopic plant matter above the soil surface was removed from inside the sampling cylinder and dried and weighed.

Only the taxonomic groups with more than 100 individuals collected from all samples at one of the sampling periods were analysed for differences between the four treatments. A nested ANOVA was used to test of the effect of herbicide treatment and pond at the sample level. The advantage of the test system is a high level of realism for effects to aquatic invertebrates in a rice growing situation following flooding. However, as noted by the authors more detailed and frequent sampling needs to be undertaken to check for species level effects. Uncertainty due to species sensitivity, temporal and spatial parameters could be possible if more detailed sampling was undertaken.

A 2003 publication **Hanson et al. (2003)** details outdoor microcosms (water volume of approx. 12,000 L) with rooted and floating macrophytes (*Myriophyllum spicatum*, *M. sibiricum* and *Lemna gibba*). The experimental design was for five different treatments each with three replicates. Plants were exposed to dichloroacetic acid (DCA) and assessed for a variety of endpoints including plant growth, root growth, number of nodes, wet and dry mass, chlorophyll-a, chlorophyll-b, carotenoids, and citrate levels. EC10, EC25 and EC50 values were calculated for each endpoint that showed a concentration response.

Each microcosm bottom was covered with 46 plastic trays (approx. 52 x 25 x 7 cm^3 deep) filled with sediment (1:1:1 sand, loam and organic matter with carbon content of 12.8%). Water came from an irrigation pond fed from a well. Water was circulated between ponds until two weeks prior to treatment. The microcosms also contained breeding fish kept in cages which were part of a separate evaluation. The microcosms were open to aerial colonisation by insects and the polyvinylchloride sides provided a substrate for periphyton growth.

Myriophyllum spicatum and *M. sibiricum* were obtained from axenic lab cultures. Every microcosm was supplied with 8 plants of each species evenly spaced in each tray. Trays placed at random in the centre of the microcosm to ensure maximum light and reduce edge effects. After a one day acclimation period. the microcosms were treated. Sampling occurred at day -1, 4,

7 and 21 days post-treatment. At each sampling point, two plants of each species were removed and evaluated, except for day -1 when ten plants were evaluated.

L. gibba was obtained from a laboratory culture and was introduced immediately after exposure by the toxicant for a 14-day exposure duration. They were contained in floating wooden cages (38 x 14 cm^2). The top and bottom of the cages were covered with a black plastic mesh (4 x 3 mm^2) to ensure containment.

The most sensitive endpoints were wet biomass and plant length especially for *M. spicatum*, followed by root endpoints. The authors conclude that if the 14 day distributions of endpoints for the 3 species are compared then *M. sibiricum* was the most sensitive. It was noted that *M. spicatum* produces algicidal allelopathic compounds and thus may not be suitable for use in microcosms where algal populations are also being evaluated. The authors conclude that *M. sibiricum* may be the species of choice due to this and since tissue samples can be taken for biochemical analysis without affecting overall growth. The advantage of this test design is the ability to expose laboratory derived organisms in realistic outdoor conditions. Appropriate design may allow the reduction in uncertainty related to the risk assessment for macrophytes although questions of relevant species for testing and the appropriate duration to assess recovery should be considered.

In situ single species exposure (*C. riparius* – fourth instar larvae) and biomarker analysis (anti-cholinesterase (AChE)) were evaluated in a microcosm study **Maycock et al. (2003)**. The microcosms were situated in rubber lined ponds (5x5 m) in UK with a natural sediment layer and river water. The plants and invertebrates present in the ponds were from natural colonisation, although *Elodea Canadensis* was mostly removed one month before treatment. Each individual test chamber consisted of a polyvinylchloride pipe (68mm diameter and 2mm wall thickness). Three pipes per microcosm were driven into the sediment to a depth of 5-10cm and secured to a supporting metal pole placed across the pond. *C. riparius* were placed for 48 hours on nine different occasions ranging before and after treatment. Surviving larvae were analysed for AChE

activity. It was noted that reduction in toxicity within the sediment was detected earlier that seen with standard macroinvertebrate monitoring. The advantage of this study type is partial standardisation (effects on specific growth stage) coupled with realistic exposure whilst burrowing in sediment. This type of study can be used to assess the potential for recovery. Disadvantages are that the AChE biomarker is only relevant for specific compounds (i.e. AChE inhibiting insecticides). Additionally, problems with indigenous Chironomids caused some confusing results. Appropriate *in-situ* tests may enable uncertainty associated with recovery to be reduced.

Outdoor microcosms (volume 30,900L) were used to assess the fate and effects of chlorfenapyr on zooplankton, macroinvertebrates, phytoplankton and fish in a freshwater system with exposure to simulate surface runoff and/or spray drift **Rand (2004)**. The bottom of each tank was covered with 10-15 cm layer of sand followed by 28-35cm layer of pond sediment (mixed) from a local farm pond. Tanks were filled with pond water containing natural assemblages of biota (zoo- and phyto-plankton). Each tank contained a volume of approximately 17,000 L of water at an operating depth of 1.5m. Tanks were filled mid June and left for four weeks to settle. In mid-July juvenile bluegill sunfish were added. Microcosms were aged for ten weeks prior to treatments. A regression design was used in which five treatments and one control was randomised among six microcosm tanks. Biological, chemical and physical monitoring occurred randomly in each microcosm during each phase using a quadrat system. Zooplankton, phytoplankton, macroinvertebrates and emergent insects were monitored. Fish were monitored daily for mortality and measured and weighed. Abundance data of biota before and after application were analysed by simple linear regression. It was noted that product was more hazardous to fish and zooplankton if it enters an aquatic system by spray drift rather than surface run-off. The advantage of this design is increased realism compared to the laboratory with respect to exposure and interaction of species with relatively low cost compared to a larger mesocosm. However, a more detailed analysis of taxa would require replication. If the design included the relevant species of concern, then some reduction in uncertainty could be possible

although lack of replication may mean that the study is not considered valid without other supporting data.

Roessink et al. (2005) undertook experiments using five concentrations of Lambda-cyhalothrin in mesotrophic (macrophyte dominated) and eutrophic (phytoplankton dominated) ditch microcosms (approx. 500 L). The test system was comprised of Macrophyte dominated and phytoplankton dominated ditches contained by polycarbonate, light permeable cylinders (enclosures) of diameter 1.05 m and height 0.9 m. In each ditch system fourteen enclosures were pressed into the sediment (depth 15 cm) and had the same water level as the ditch (0.5 m). Lambda-cyhalothrin was applied three times at weekly intervals by very gentle stirring. Average macrophyte biomass in the macrophyte-dominated enclosures was 117 ± 47 g/m^2. No macrophytes occurred in the enclosures in the phytoplankton-dominated ditch. Macroinvertebrates were sampled from each enclosure pre and post application using litter bags and two types of artificial substrates (multiplates and pebble baskets) with identification to the lowest practicable taxonomic level. Zooplankton and phytoplankton were sampled using a Perspex tube. Additionally, *in-situ* bioassays of two crustaceans *Asellus aquaticus* and *Daphnia pulex* and the insect *C. obscuripes* were performed. Measurements of physical chemical properties, community metabolism and decomposition were performed. Data were generated for up to 45 days post application. NOEC calculations were done at the parameter of the taxon level using the Williams test (analysis of variance Effects at the community level were analysed using the principal response curve (PRC) method.

The two systems differed in macrophyte biomass, phytoplankton densities and invertebrate composition. Dissipation of the test chemical was shown to be similar (rapid). Community effects in the mesotrophic macrophyte dominated enclosures were longer-lasting than those in the eutrophic ditch. The *in-situ* bioassays allowed a distinction between potential and actual recovery. Potential recovery is defined as the decline of the chemical to a concentration at which

it no longer has adverse effects on sensitive arthropods. The bioassays showed that potential recovery of even the most sensitive invertebrate in the present study (*Chaeoborus* sp.) occurred earlier than actual recovery (abundance). Direct effects on sensitive invertebrates were consistent with short-term laboratory toxicity data for the same species. No major differences were found in threshold levels for direct effects between the two systems. The differences seen were rate of recovery and indirect effects at higher concentration levels. The advantage of this system is that it is relatively small and of low cost with a dose response design replication (2 x each concentration) enabling different ditch environments to be compared. Additional information on recovery potential is gained from the *in-situ* bioassays. It is considered that the application used approximates worst case exposure. A disadvantage of such a design is the inherent variability in ditch organisms. Reduction of uncertainty due to species sensitivity (relevant species assemblages), spatial and temporal variation (45 day post application effects) should be possible.

Coors et al. (2006) carried out macrophyte *in-situ* bioassays in three outdoor mesocosms (2.5 m diameter, 1m water depth and about 0.1 m sediment layer). Five species of submersed macrophytes were planted in plastic pots (80 mm diameter) with sediment from the mesocosms. The length of planted sprouts (two per pot but three per pot in the case of *C. globularis*) was recorded and the sum of sprout length per pot calculated. Macrophytes were inserted by means of plastic pot holders which fixed the macrophytes at a depth of 0.2-0.3 m. Plants were allowed to acclimatise for 14 days prior to 1st application of test material. During the study macrophytes were visually inspected (e.g. for chlorosis) and at day 56, pots were harvested and macrophytes dried and dry biomass per pot measured. In the case of *M. spicatum* and *P. lucens* plant length was also recorded at day 56. Number of replicates (pots per pond) was either three or four depending on species. Statistical tests used ANOVA.

Three of five planted (plastic pot) submersed macrophytes (*M. spicatum, P. lucens, E. canadensis*) showed satisfactory growth in the control pond and therefore demonstrated their suitability for *in-situ* bioassays.

The *in-situ* bioassay with *Lemna minor* exposed in hand-made floating devices (circular aluminium plastic tubes kept afloat in vertical position by air-filled plastic tubes). The device enclosed the plants on a surface circle area with a diameter of 100 mm and was open both to the water and the atmosphere. Bioassays lasted either 28 days (after 1st application) or 21 days (after 2nd application). Number of replicates (No. of floating devices per pond) was three in all bioassays and each replicate consisted initially of 10 *L. minor* at 3 frond stage. *Lemna* development was photographically documented every 7 days. Using the known diameter of the circle, it was then possible to derive the absolute frond area of *Lemna*, summed for each replicate, and relate this measure to the initial total frond area at the first day of exposure. The authors recommended the use of *L. minor* from high-nutrient laboratory cultures to assess direct effects in realistic environmental conditions. Phytoplankton and zooplankton were also sampled and enumerated in the study following exposure to terbuthylazine (TBA).

It is considered that a model ecosystem is more representative of typical shallow water body if it includes macrophytes. The design described allows the generation of toxicity data for several species of macrophyte under semi-field conditions whilst including the interaction of various parts of the aquatic community. The design reduced variance by the controlled introduction of macrophytes (bioassay in plastic pots or floating devices). However, one disadvantage of this type of design is that the presence of macrophytes can influence aquatic communities and their metabolism and can therefore obscure the detection of effects on phytoplankton. In conclusion, this study design enables realistic exposure of macrophytes and scope for reduction of uncertainty due to species sensitivity, spatial and temporal attributes although number of species tested may need to be increased and the length of time necessary to assess recovery would need to be determined.

3.3.5.2 Published Lotic field studies

Comparison of macroinvertebrate communities in stream microcosms with those in the field was undertaken by **Schulz et al. (2002)**. Macroinvertebrate communities were obtained from an uncontaminated control site in the Lourens River (S. Africa) and established in outdoor stream microcosms. The effects on invertebrate taxa of azinphosmethyl seen in the microcosms were compared to the distribution of the same taxa at the control and orchard exposed sites of the Lourens River. Levels of exposure and duration were monitored in the river.

The outdoor artificial stream system consisted of fifteen static stainless steel circulating streams (1.5 x 0.2 x 0.2 m). Water taken from the control site of the river was used in the microcosms. Each stream contained a volume of 30 L. Stream microcosms were established two days before introduction of test organisms. Rocks (8-10 cm diameter) and associated drifting invertebrates (caught by hand net) were collected from the control site and placed in microcosms with the same orientation. Exposure to pesticide for 1h was done one day after introduction of the rocks by addition in 100 mL of water to the circulating stream. Each of the five treatments was replicated three times. Emergent insects were caught in gauze placed over the stream system. After six days, the rocks were removed from the microcosms and all animals were counted and identified. Water quality and chemical analysis were done. Additionally, field sampling was done at the control and contaminated sections of the river. The advantage of this design is the high realism with validation of the microcosm data with field data. A disadvantage was that the study duration was relatively short and thus recovery was not assessed. This type of study could reduce uncertainty due to spatial and species sensitivity parameters.

Heckmann et al. (2005) described a lotic method involving the placement of channels in the shallow part of a stream riffle from which macroinvertebrate drift and benthic samples could be taken following pulsed exposure. The advantage of this type of test is the ability to examine effects on invertebrates at the community level in a lotic system and to assess recolonisation

potential from short pulsed exposure. A potential disadvantage is the variation in regional and physical-chemical factors in-stream that may cause considerable variation in the impact among different stream ecosystems, although this could be overcome by reasonable selection criteria as would also apply for any field lentic systems. It is considered that uncertainty due to species sensitivity, spatial and temporal parameters for macroinvertebrates could be reduced using this type of study.

Beketov et al. (2008) tested pesticide effects (thiacloprid) in sixteen outdoor artificial streams (approx. 1000 L). Each stream was designed as a closed circulation system. Water flowed from the upstream to the downstream sections of the stream as a result of gravity, and then fell into the 200 L reservoir installed below the downstream margin of the stream from where it was pumped back to the upstream section. At the end of each stream, a dam with a polyester net filter (1 mm mesh) was installed to prevent loss of the animals to the reservoirs. The bottom of the streams was covered with a mixture of fine gravel and sand. The streams were located as parallel lines with the 0.8 m distance in between channels having riparian vegetation to provide refuge for emerged insects, reduce amount of direct sunlight and increase ecological realism. The experimental design included four treatments with two replicates for each concentration level and ten for the control (regression experiment design). The high number of control replicates was used to allow usage of the Monte Carlo permutation test following multivariate ordination techniques. Experiments were run for seven months. The streams were constructed in the summer of 2003 and planted with watercress *Nasturtium officinale* in late 2003 and early 2004. Sediment and associated macroinvertebrates were extracted from an uncontaminated small stream in E. Germany using a surber sampler and added to the streams. Further additions of sediment were added to the streams during the winter of 2004/2005 and also in October 2005 to mimic the natural influx of species by drift. Physicochemical parameters were measured in the streams every four months from June 2005 and no significant differences between the streams were found.

Thiacloprid stock solutions were added to the reservoirs below the stream in order to dilute the toxicant and to make the input gradual. Aquatic invertebrates were sampled using a metal frame designed to cover a 15 x 15 cm area of the stream bottom. During sampling macrophytes were removed by hand and washed and checked for macroinvertebrates. The water column was sieved and the sediment examined for macroinvertebrates. Except for the first sampling, thirty-four weeks before contamination, the animals were identified *in situ* and returned to the stream. To assess effects on emergence of merolimnic insects, six emergence traps were installed on each stream mesocosm. An overview of effects was given using the univariate parameters abundance and taxa richness. To test for significance of the toxicant's effect on particular species, only two species including the stonefly *N. cinerea* were monitored. Community response was analysed using Principal Response Curve (PRC) method and a set of Redundancy Analyses (RDA) performed for the different sampling time-periods. The advantage of this type of study is the high level of realism for macroinvertebrate exposure in a lotic environment over a long duration of seven months allowing recovery to be investigated. A possible disadvantage is that real streams are not recirculating and there would be expected to be immigration from upstream and therefore the system could be considered to be conservative with respect to both exposure and recovery potential. Uncertainty due to species sensitivity (macroinvertebrate), spatial and temporal attributes could all be reduced by this type of study.

3.3.5.3 Field studies – conclusion

Historically it can be seen that in the early day's very large pond studies as required by the US EPA predominated whilst later studies tend to use smaller micro or mesocosms. Comparisons of effects in concomitant microcosm and mesocosm studies showed that effects on communities were very similar. Both small and large outdoor systems provide greater realism compared to laboratory studies and have the potential for more species (i.e. flying insects etc.) with more

possibilities for recolonisation than indoor systems. The inclusion of fish and macrophytes in microcosms and mesocosms was shown to be problematic.

Design of micro and mesocosms in later years can be seen to be more targeted based on known sensitivity (e.g. to zooplankton, macrophytes etc.) and has included relevant *in-situ* bioassay techniques. This is in line with the main conclusion from HARAP **Campbell et al (1999)** that the design of field studies should be on a case by case basis. Many microcosm and mesocosm studies focus on plankton effects, whilst specialised designs have been developed for macrophytes (e.g. *in-situ* bioassays) plus stand alone higher tier laboratory studies for fish (e.g. SSD and modified exposure). Additional designs for lotic systems in order to assess effects on macroinvertebrates have also been developed in more recent years. It was noted in Ian R. Hill, et al in Chapter 24 of the book **Graney (1994)** that both "enclosed" (e.g. pond enclosures) and "flowing" (e.g. artificial streams) systems are feasible, the latter are described as being least well developed and understood. It seems possible that extrapolation from an "enclosed" system to both static and flowing bodies of water will be easier than from a "flowing" design. Also enclosure design will offer most severe test of effects as organisms will be exposed for a longer period of time

It is considered that uncertainty due to species sensitivity, spatial and temporal (recovery) can be reduced using appropriate higher tier field testing. Although it is worth noting that there is much debate on the extrapolation of meso- and microcosm results to natural systems. It was noted **Crane (1997)** that further research is required on the repeatability, reproducibility and predictive ability of such systems.

The proceedings of the Community-Level Aquatic System Studies – Interpretation Criteria (CLASSIC) SETAC workshop held at the Fraunhofer Institute –Schmallenberg, Germany 30 May- 2 June 1999 **Giddings et al. (2002)** provided a number of recommendations relating to design of community level aquatic system studies: dosing regime; dosing methods; timing of application; level of taxonomic resolution; inclusion of fish and macrophytes; univariate versus

multivariate statistical methods; derivation of acceptable concentrations; structural and functional endpoints; population recovery; addressing uncertainty about recovery; data representativeness of microcosm and mesocosm results; need for database development and landscape ecology.

The CLASSIC workshop **Giddings et al. (2002)** recommended an exposure-response experimental design with replication as being preferable. This type of design has been used for many of the published field studies summarised in the section above. It is noted in (Giddings, Solomon et al. 2001) that consistent exposure-response relationships were discerned from four cypermethrin mesocosm studies evaluated as a group. Pooled results from three esfenvalerate studies also showed consistent exposure-response relationships for the major taxa.

The *OECD series on testing and assessment N*o. 53: Guidance document on simulated freshwater lentic field tests (outdoor microcosms and mesocosms) OECD (2006) takes into account the general recommendations given in HARAP **Campbell et al. (1999)** and CLASSIC **Giddings et al. (2002)** which in turn followed recommendations set out in the earlier SETAC workgroups including SETAC-Europe 1991 and SETAC/RESOLVE 1991. The guidance provides a rough outline of the issues and reporting detail to be considered in outdoor lentic microcosm and mesocosm design. The guidance notes that generally meso- and microcosm size between 1000 and 20,000 L are appropriate. Smaller 100 to 1000 L microcosms are stated to be appropriate for studies where planktonic species are the major concern. Larger mesocosms may be used but they may be much more resource intensive. The size selected depends on the objectives of the study but it is noted that in general, studies in smaller systems (approx. 1000 to 5000 L) are more suitable for shorter studies (3-6 months) and for studies with smaller species (e.g. planktonic species). Larger systems are more appropriate for longer studies (6 months or more).

SANCO (2002) states that data from microcosm and mesocosm studies should be used to determine the following endpoints: ecologically acceptable concentration (EAC). For the

relevant taxonomic groups in the study, a no observed effect concentration at the community ($NOEC_{community}$) should be derived using appropriate statistical techniques (e.g. Principal Response Curves). In addition, NOECs for populations of relevant organisms should be reported ($NOEC_{population}$). Where there are effects at the community or population level, the time taken for recovery to occur should also be reported. The $NOEC_{community}$, the $NOEC_{population}$ and the time taken for recovery should then be used to determine a no observed ecologically adverse effect concentration (NOEAEC). The NOEAEC is defined as being the concentration at or below which no long-lasting adverse effects were observed in a particular higher-tier study. An NOEAEC is study specific but an EAC is derived from the overall evaluation of the compound. EAC values may be defined directly using studies or by the addition of an appropriate uncertainty factor. More recently, the ELINK workshop **ELINK (2008)** referred to effects assessment endpoints as regulatory acceptable concentrations (RACs) and the interface between the exposure and effects assessment as the ecotoxicologically relevant concentration (ERC).

The overall conclusion is that higher tier field studies must be designed to address the problems identified at lower tiers, with consideration of the recommendations outlined in **OECD (2006), SANCO (2002, Giddings et al (2002) and Campbell et al (1999).** In higher tier tests with refined exposure, it is important to be able to easily compare field exposure (PEC) with the test concentration profile in the tests ELINK (2008).

3.3.6 Aquatic monitoring

Monitoring studies were not specifically targeted in the literature review, but it is worth noting that where available such data should be examined to check the scale and duration of effects in the environment. However it should be noted that effects on populations and communities in the environment may be due to a number of different stressors alone and in combination (e.g. different chemicals, nutrients, flooding or drought etc.). Various systems for assessing the quality

of water by the presence or not of various species have been documented in the literature over the years. **Landner et al. (1989)** discussed the Pollution-Induced Community Tolerance (PICT) approach. In a PICT study the communities are established under ecologically realistic conditions, in a contaminated ecosystem or a set of micro- or mesocosms. At locations where the toxicant is present at high enough concentrations to exert a selection pressure, the sensitive organisms will be excluded due to lethality, avoidance or decreased competitive ability, and replaced by tolerant ones. It is essential for the PICT concept that this pollution-induced community tolerance can be quantified. In relation to pesticides, a recent paper by **Liess et al. (2005)** identified stream invertebrates according to their vulnerability to pesticides. Species were classified as species at risk (SPEAR) and species not at risk (SPEnotAR). Ecological traits used to define these groups were sensitivity to toxicants, generation time, migration ability and presence of aquatic stages during the time of maximum pesticide application. It was shown that the number and abundance of SPEAR in disturbed stream sections are greatly increased when undisturbed stream sections are present in upstream reaches. The results indicate that ecological traits and recolonisation processes are important at the landscape level for ecotoxicological risk assessment. Although this type of assessment is rarely currently used in the regulation of pesticide authorisations, this could change in the future as a result of the requirements under the Water Framework Directive (2000/60/EC).

3.3.7 Aquatic modelling

Traas et al. (2004) focuses on a freshwater food web model for the combined effects of nutrients and insecticide stress and subsequent recovery. The microcosm data was based on simulations of Dutch drainage systems, dominated by the macrophyte *Elodea nuttallii* and were stocked with phytoplankton, planktonic grazers and macro-invertebrate species. Endpoints investigated were responses of phytoplankton, planktonic grazers, macrophytes, macroinvertebrates, periphyton,

community metabolism and litter decomposition. Sampling of biota occurred bi-weekly. It was stated that modelling can extend the use of microcosms as a link between laboratory and field as this allows the prediction of effects and recovery of ecosystems for concentrations that have not been tested.

van den Brink et al. (2006) noted concordance between the predictions from the effect model PERPEST (a model that predicts the ecological risks of pesticides in freshwater ecosystems) and the concentrations at which clear effects started to emerge in laboratory and field studies. However, compared to the SSD concept, the PERPEST model is able to provide more information on ecological risks when a common toxicological mode of action is evaluated as it considers both recovery and indirect effects.

Hanratty et al. (1994) and Hanratty et al. (1996) used littoral enclosures to assess the accuracy of the Littoral Ecosystem Risk Assessment Model (LERAM) to predict effects from diflubenzuron. It was noted that the LERAM model requires fine tuning and that the availability of a bigger toxicity dataset would lead to better modelling.

3.4 Non-target Arthropods

3.4.1 Background

Because its origins lie in the assessment of suitability of pesticides for use in Integrated Pest Management, non-target arthropod testing differs from other areas of ecotoxicology. In the early 1990's testing was conducted to determine the compatibility of products with the use of

predatory and parasitic arthropods as biological control agents. The International Organisation for Biological Control, West Palearctic Regional Section (IOBC/WPRS) working group on "pesticides and beneficial organisms" **Hassan (1992)** developed a series of laboratory tests for a wide range of organisms potentially useful as biological control agents.

Because these tests were designed to identify harmlessness they involved initial testing on a glass or quartz sand substrate, intended to represent worst case exposure. The IOBC ran a joint testing programme and its' members generated tables of data describing laboratory effects of a range of products on beneficial arthropod species. The IOBC programme included a sequential testing scheme, intended to move from laboratory to semi-field and field methodology. For most species only laboratory methods were developed and usually on the basis of testing one species per researcher, with little quality control and no peer review process. These data were not intended as a regulatory tool and were generated to assist in the selection of suitable products for use in IPM.

When a requirement arose to test terrestrial arthropods in regulatory risk assessment working groups were established to ring test and validate methods, all of which were derived from the existing IOBC procedures. At that time the concern was for beneficial arthropods and not non-target arthropods, so the choice of predators and parasites as test species was appropriate.

The shift to non-target arthropods occurred at the first Escort workshop **Barrett et al. (1994)**, when reference to risk in the off-crop habitats was also made. Despite these changes testing approaches remained focussed on beneficial species, largely because a considerable effort had been invested in developing and validating laboratory test methods. The IOBC booklet **Candolfi et al. (2000)** presents tier 1 (worst case substrate of glass or quartz sand) laboratory methods for eight species, *Aphidius rhopalosiphi* (Hymenoptera; Braconidae), *Typhlodromus pyri* (Acari; Phytoseiidae), *Aleochara bilineata* (Coleoptera; Staphylinidae), *Chrysoperla carnea* (Neuroptera; Chrysopidae), *Coccinella septempunctata*, (Coleoptera; Coccinellidae), *Orius laevigatus* (Hemiptera; Anthocoridae), *Poecilus cupreus* (Coleoptera; Carabidae), *Trichogramma*

cacoeciae (Hymenoptera; Trichogrammatidae) and *Pardosa spp*. (Araneae; Lycosidae). Whilst all of these methods appear to have equal status in the guideline document the level of validation differs considerably between them, with the greatest amount of validation, in the form of ring testing, being conducted for *T. pyri* and *A. rhopalosiphi*.

Following the second Escort Workshop **Candolfi et al. (2001)**, the initial tier of testing with non-target arthropods was revised to consist of glass plate rate/response studies to determine LR_{50} values with two sensitive indicator species, *A. rhopalosiphi* **Mead-Briggs et al. (2000)** and *T. pyri* **Bluemel et al. (2000)**. The term rate/response is used to refer to multiple rate contact tests since it is not possible to determine what dose an individual organism was exposed to. The 48 hour LR_{50} value for *A. rhopalosiphi* and 7 day LR_{50} value for *T. pyri* are divided by the field application rate to generate a Hazard Quotient (HQ) value. Comparison with an industry-owned data set of field studies showing the presence and absence of effects **Campbell et al. (2000)** led to acceptance of a trigger value of 2.0 at the Escort 2 workshop **Candolfi et al. (2001)**. If an HQ of greater than 2.0 is derived for either species then there is the potential for harmful effects on non-target arthropods and further testing is required to quantify this risk.

In the initial tier 1 studies, the test organisms are confined over a treated surface and there is no opportunity for them to avoid contact exposure. The glass plate is considered to represent worst case exposure since it offers little opportunity for a molecule to bind to the surface and thus contact bioavailability is extremely high. Most of these tests involve contact exposure and do not take into account potential exposure that may occur due to direct overspray of the organism or oral routes. For some products, such as those with a mode of action as a stomach poison, oral contamination may be particularly relevant. For products with a specific mode of action, such as insect growth regulators, specialist tier 1 tests on immature life stages are required.

Higher tier non-target arthropod testing it is conventionally sub-divided (within the context of a sequential testing scheme) into extended laboratory, aged residue, semi-field and field studies. As the scheme progresses a greater degree of realism is introduced in terms of exposure of the

organisms to dried residues of the pesticides. All except field studies concentrate on the fate of individuals of single species, usually by exposing laboratory reared organisms from the recognised indicator species. Mortality is usually the primary endpoint although a sublethal endpoint, such as reproduction of survivors, is usually also reported. Because of the small number of surviving individuals involved and the high natural variability in reproductive parameters, these studies are only able to detect extreme sub-lethal effects. The studies are not true "reproduction" studies and have limited precision with respect to the sub-lethal endpoint.

Where there is a strong likelihood that such results represent false positives, it is logical to demonstrate acceptable risk through the most cost effective route, usually a combination of extended laboratory and aged residue studies. Insecticides and acaricides are unlikely to show acceptable risk in extended laboratory studies (similar to the tier 1 tests but using natural substrate) and semi-field or field studies are usually the most appropriate to demonstrate the nature of the risk to non-target arthropods.

Only field studies have the potential to investigate effects on populations and can investigate effects on a community of naturally occurring arthropod species.

3.4.2 Extended laboratory studies

Although they represent a higher tier than the glass plate studies, extended laboratory studies typically introduce realism only in terms of the substrate used. Extended laboratory test designs generally mimic those of the tier 1 tests in terms of duration, temperature and life stages exposed but involve the exposure occurring on plant substrate for foliar dwellers and on a standard soil for soil dwellers. This approach can be problematical for herbicides, which often kill the plant making it difficult to assess the risk to foliar dwelling arthropods. In some cases (as for *T. pyri*), the test system remains 2 dimensional with animals exposed to residues on flat leaf surfaces

whereas in others (e.g. *A. rhopalosiphi*) the test system is three dimensional. When a three dimensional system is used the organism is able to avoid the test item to some extent, for example in the case of the *Aphidius rhopalosiphi* method **Mead-Briggs et al. (2009)**, by resting on the inner surfaces of the untreated glass enclosures around the barley seedlings. This was not of particular significance when the results were being used to assess the risk to parasitic wasps as potential biological control agents but may be relevant when the results from tests with *A. rhopalosiphi* and *T. pyri* are being used to represent the risks to arthropods as a whole. Avoidance of a foliar residue in a cage arena may be appropriate for species that can fly but such a response may not be found in apterous (wingless) arthropods.

The list of species commonly tested in the laboratory, either on glass or a realistic substrate, is heavily biased in favour of beetles (Coleoptera) and parasitic Hymenoptera. Major orders of Insecta are unrepresented. There are existing methods, for example **Tornier et al. (1992)**, that describe an extended laboratory method with adult hoverflies *Episyrphus balteatus* (Diptera; Syrphidae). Adult hoverflies (8 male and 8 female) were released into cages over treated buckwheat plants containing aphids. Mortality was assessed for 4 days together with oviposition and behaviour. A residual toxicity test was started after 8 days and lasted for a further 6 days. Whilst this approach provides some assessment of the risk to flies it does not include exposure of larvae which, since they are soft bodied with a large contact area, would be expected to be more sensitive and which would often be exposed by contact with treated foliage.

Extended laboratory studies generally involve a high level of replication and test individual species. There is no data-base of field studies against which to compare extended laboratory results so no hazard quotient should be derived from these studies. Although there remains some uncertainty as to the suitability of the tested species to cover potential effects of plant protection products on non-target arthropods as a whole the extreme nature of the confined exposure is considered to make products appear more harmful than would occur under field conditions.

3.4.3 Semi-field methodology

A workshop on field and semi-field methodology held in Versailles, France in 1999 **Candolfi et al. (2000)** defined semi-field testing as being a single species study where the test system is maintained outdoors. This publication provided generic guidance for all semi-field studies, including the selection of standard test species derived from the ESCORT 1 list **Barrett et al. (1994)** as opposed to crop or nationally-relevant ones. Semi-field tests were recommended as being conducted with standard crops and the authors advised that off-crop risk should be assessed at reduced rates to simulate drift. A toxic standard is required in a semi-field study so as to demonstrate the potential to have effects and the test design is required to have sufficient replication to have an 80% power to reject the null hypothesis. The use of an indiscriminate toxic standard, rather than a reference compound with expected effects occurring within a prescribed range (and less than 100%), means that such tests may not be directly comparable between testing facilities. Although **Candolfi et al. (2000)** proposed production of a database across testing laboratories to define acceptability criteria this was never generated. Ideally, multi-application products require multi-application studies although the guidance allows the use of an accumulated residue study if this is not possible.

Semi-field methods have been devised and published for a number of individual non-target arthropod species. In the 1990's the most widely used semi-field method was that of **Heimbach et al. (1992)** for the carabid beetle *Poecilus cupreus*, although the method was also considered suitable for other ground beetle species such as the smaller *Bembidion lampros* which was found by **White et al. (1990)** to be more sensitive to lambda-cyhalothrin than *P. cupreus*. This method involves the confinement of adult beetles within 0.5 m x 0.5 m metal enclosures dug into crop situations, (typically cereals, grass or vegetables). Exposure is realistic in that the enclosures are over-sprayed. Mortality is assessed by the recapture of released beetles and sub-lethal effects are assessed by measurement of consumption of *Drosophila* spp. pupae as prey items. A criticism of these enclosure studies is that *Poecilus cupreus* is a medium sized Pterostichine carabid beetle,

adapted for burrowing with a wedge shaped body and a relatively thick cuticle. Beetles that burrow beneath the soil surface for all or part of the exposure period of a study will be less likely to be exposed than those that remain on the soil surface. Larger beetles, with thicker cuticles might be expected to be less likely to be adversely affected than smaller beetles with thinner cuticles which might experience more uptake into their body through contact exposure. Although results from such studies demonstrate the potential effects on adult Pterostichine carabids they may not necessarily represent potential effects on smaller surface dwelling (epigeal) beetles. Additionally, whereas many of the early IOBC derived laboratory tests focussed on testing of the most sensitive life stage the method with *P. cupreus* does not. In conclusion, adult carabid beetles must be considered to be considerably more robust than their larvae and probably more robust than most epigeal arthropod species.

Jepson et al. (1992) described a series of four semi-field methods for use in cereals, comprising barriered enclosures (similar to those of Heimbach *et al* (1992)), 2m cube cages suitable for *Coccinella septempunctata,* sleeves and barriered large plots. In the first three methods laboratory reared organisms were released into cages or enclosures shortly after treatment and their survivorship recorded after periods of time. The greatest drawback of these three methods, and common to all semi-field methods involving release and catch back, was the non-recovery of a large number of the released organisms. The use of sleeves, typically mesh tubes over individual stems or parts of a plant, means that test organisms can avoid exposure by transferring themselves to the untreated sleeve itself.

The fourth semi-field method described by Jepson and Mead-Briggs (1992), the use of large barriered plots, was different from the others in that it involved sampling of naturally occurring organisms as found within the crop rather than re-capture of released individuals reared in the laboratory. Considering the definition of **Candolfi et al. (2000)**, in that the test organisms are not standard species and are not laboratory reared, these barriered plot studies will be discussed further in the 'Field Studies' section of this report. **Brown et al. (1990)** compared different

exposure routes of three insecticides in the laboratory (topical application and contact within small enclosures) with results from 1m x 1m field enclosures, similar to but larger than those of **Heimbach et al. (1992)** using adults of a lycosid hunting spider (*Pardosa*) and the large carabid beetle *Pterostichus melanarius*. This methodology generated very similar mortality results for both the spider and the beetle in the laboratory contact test and the field enclosures but demonstrated that for those organisms topical treatment with a micro-applicator under-represented mortality that would occur in the field. This suggests that for surface active predators contact appears to be the most relevant route of exposure.

Tornier et al. (1992) described a semi-field method with adult hoverflies (Diptera; Syrphidae), which is essentially the same as their extended laboratory method with hoverflies **Tornier et al. (1990)** but conducted outdoors. The method as described only includes two enclosures per treatment rate but this could be modified to include adequate replication. The method is suitable for multi-rate testing with a view to generating a NOEC for this test organism. This method was originally intended to assess the risks to pollinators but, in the absence of a semi-field method for *Aphidius* spp. it could be a potentially useful semi-field method for evaluating effects on flying insects with foliar activity.

A field method for determining the effects of pesticides on the green lacewing **Vogt et al. (1992)** is considered here as a semi-field approach since it involves the release of laboratory bred larvae of a single species, *Chrysoperla carnea*, (400 per tree), onto fruit trees before application of treatments. At intervals after treatment, bait cards are placed on the trees and surviving larvae are recaptured. This approach attempts to assess mortality of lacewings under realistic exposure and the authors cite success with conventional insecticides and insect growth regulators. Although lacewing populations in orchards will not comprise solely of larvae the mortality assessed using this approach is indicative of the likely effects of products on this species of foliar predator under commercial use. No information is provided to suggest how the results from this study relate to potential effects on non-target arthropods as a whole.

3.4.4 Aged Residue Studies

After the Escort 2 workshop **Candolfi et al. (2001)**, recovery was considered to be an acceptable end-point in regulatory risk assessment for non-target arthropods. Recovery within one year was considered to be acceptable for in-crop risk and within an ecologically relevant time frame for off-crop risk assessment. There is no guidance on how to determine what constitutes an ecologically relevant time-frame. However, recovery can only occur when the residues of a particular product have declined to levels which are no longer harmful *and* when the non-target arthropod populations are able to immigrate or increase in number due to reproduction of survivors. The aged-residue study became an increasingly useful approach to demonstrate the potential for recovery after a given period of time. **Candolfi et al. (2000)** describe an aged residue study as being a hybrid, where pesticide deposits are aged under field conditions but exposure of test organisms occurs in the laboratory or under semi-field conditions. The protocols developed for extended laboratory and semi-field studies also apply to aged residue studies. A water control and a toxic standard must be used and the product should be applied at the maximum application rate and with appropriate drift rates.

3.4.5 Field Studies

General guidance on the design, conduct and interpretation of non-target arthropod field studies was given by **Candolfi et al. (2000)** following the workshop held in Versailles in 1999. Advice on studies with predatory mites was excluded since this was provided separately by **Bluemel et al. (2000)**. The authors' work is cited in the Escort 2 workshop report **Candolfi et al. (2001)** as the source of guidance for regulatory field studies and therefore represents current regulatory advice. Field studies should be conducted on the basis of lower tier results **Candolfi et al. (2000)** and may therefore be focussed on all or part of the non-target arthropod fauna. Recovery is

usually considered to be a key end-point and the effects of plot size on this are discussed. Although current field test methodologies are only designed to allow within season recovery to be assessed (except for organisms of low mobility such as Collembola) these studies are often being extended to one year or more in duration so as to try to show acceptable recovery.

Candolfi et al. (2000) suggest that off-crop risk is better evaluated in semi-field studies but consider that it is acceptable to study off crop risk using in-crop situations by using reduced rates to simulate spray drift on in-field taxa. Field studies should ideally take place in one of two model crop systems, arable or orchard, with realistic worst case exposure and an assessment of effects on phytophagous, detritivorous and predatory arthropods. Orchard studies should have a minimum plot size of 0.2 ha and outside row of trees should not be used for sampling. Since orchards have limited size it can be acceptable to reduce the number of replicates of the reference item to two. **Candolfi et al. (2000)** describe a range of possible sampling methods and advise on sufficient sub-sampling. Field studies should have a water treated control and a toxic standard (now generally referred to as a reference item).Taxonomy should be conducted to species level where appropriate, except for difficult groups such as Collembola, Hymenoptera and Aleocharinae where sub-order or family may be appropriate.

Where a product is to be used in the North and South of Europe, **Candolfi et al. (2000)** consider that studies should be conducted in both regions. Within each climatic zone a worst case crop use should be selected. In interpretation of the results from field studies there is no simple threshold for acceptability of effects and each study has to be addressed on a case by case basis. The ecological role of any population that is affected should be assessed.

Publications describing field testing approaches can be divided into those for predatory mites, (typically in tree crops or vines), those for arable crop systems and those for the non-acarine fauna of fruit orchards or tree crops.

3.4.5.1 Predatory mite field studies – vineyard and orchards

The predatory mite field method of **Bluemel et al. (2000)** is the only published guideline that has been ring tested and shown to generate reproducible results. This method evaluates the short and long term effects of plant protection products on phytoseiid mites in vineyards and orchards by sampling population density with respect to a water-treated control at different time intervals after application. The study design includes five replicates of each treatment except the reference item (which requires two or more replicates). Crop plants are sprayed with treatments so as to simulate commercial practice and naturally occurring populations of test organisms are exposed both directly and indirectly. Mites are sampled by collecting leaves (and either counting directly or using an extraction method, such as washing or heat extraction) to give a mean number in the control of 30 mites per sample. The number of motile stages per sample is recorded before and at intervals after treatment. Ring tests in apple orchards and vineyards showed that the test design identifies effects of plant protection products of greater than 50% as statistically significant in 90% of cases.

Other publications that evaluate effects on predatory mites (e.g. **Duso (1994)**,**Miles et al. (2003**)) are similar to the method of **Bluemel, et al. (2000)**. In addition to leaf sampling, **Gyorffyne Molnar et al. (1994)** sampled overwintering populations of mites by dissecting buds collected in February. **Sterk et al. (1994)** described field studies with the mites *Euseius finlandicus* and *Typhlodromus pyri* in orchards in Belgium using methodology similar to that of **Bluemel et al. (2000)** and considered that populations can be very heterogeneous within trees and within an orchard. Densities can sometimes be low making it difficult to achieve high enough numbers of mites in the samples. **Sterk et al. (1994)** considered that sensitivity of mite populations can vary between fields due to previous exposure of the test organisms to pesticides.

Without knowing the response of the population of mites at any given test site to a range of pesticides from different classes, it is not possible to be certain that resistance is not present in the test organism population. This would mean that the results from a single study would not extrapolate to other situations where more sensitive mites would be exposed.

Although there is a well recognised and ring tested method for conducting predatory mite field studies, it is not clear to what extent such studies reflect the response of non-target arthropods as a whole. Organisms such as predatory mites, which may be sensitive as individuals, have relatively short generation times and their populations may be able to recover more rapidly than species with only one or two generations per year.

3.4.5.2 Arable field studies

The earliest published guideline for arable studies **Carter (1993)** is essentially the same as the UK guideline **PSD et al. (1995)** written by a working group drawn from industry and academia in the early 1990's. This guideline recognised the importance of replication (requiring at least 4 replicates) and gave the researcher the option to use large un-barriered plots (>1 ha in size) or smaller barriered plots (>10m x 10m). This guideline was intended to determine the within-season effects of insecticides on non-target terrestrial arthropods in cereals in summer and specified that data from two or more site-years was necessary so as to cover a range of conditions. The large plot study gives a high degree of realism, has no requirement to erect barriers, had no risk of over sampling and provides data from a wide range of taxa, especially polyphagous predators. The large plot study requires > 20 ha of cereals which need to be homogeneous with respect to arthropod populations. The **PSD et al. (1995)** guideline requires that the test item is applied at the maximum field rate, there should be an untreated control and a reference item, applied at the normal field rate.

The barriered plot approach is practical in that it only requires a site 1 ha in size. Because the total area is small it is possible to examine several possible sites and choose the one with a high and comparatively uniform population of arthropods. At least seven days before the application of treatments, each plot is surrounded by a barrier of polythene (typically 60 cm high with the lower edge buried to a depth of at least 15 cm). The guidelines specifies sampling in both the open and barriered plot studies to investigate effects on spring and autumn breeding carabid beetles, staphylinid beetles, spiders, and aphid-specifics (parasitoids, coccinellids, neuropteran larvae, syrphidae and game-bird chick food insects).

For both the open study and the barriered plot study, the guideline recommends pitfall trap sampling (at least 9 traps per plot) and visual counting of tillers for aphids and parasitised mummies. Night time sweep net sampling is recommended for the large plots whereas to avoid damage to plants, suction sampling (for example with a D-vac or more recently a Vortis sampler) is recommended in the smaller plots.

This guideline reflects the focus interest in the early 1990's on predators and parasites as potential pest control agents in crop as well as concern for bird food insects linked to the decline of the grey partridge. Nocturnal sweeping can be highly effective at catching a wide range of non-target arthropod taxa. There is no mention of sampling for Collembola (probably the most numerous organism in a cereal agro-ecosystem) or mites and there is no consideration for the investigation of off crop effects.

It is interesting to note that the 1 ha plot study was only considered by its authors to be suitable for within season effects, being too susceptible to immigration for use over a longer period, and that data from two site years were required.

Brown et al. (2004) presents experience from conducting large plot (>1ha) replicated field studies and makes the point that taxonomy at the family level can mask effects at the species level. For example, the absence of effects on total Carabidae (with all individuals of the family

being combined) does not mean that all species of the family Carabidae were unaffected. After the Escort 2 workshop **Candolfi et al. (2001)**, recovery became the recognised endpoint for acceptability. Without an adequate and widely-accepted standard definition, there are many forms of potential recovery that could occur experimentally that may not occur following field or farm scale treatment. Whilst larger experimental plots generate more realistic data with respect to recovery this often comes at the expense of adequate replication.

Several field studies **Inglesfield (1989), White et al. (1990)** used large experimental plots, up to 4 ha, but with only one plot per treatment. Although these studies ran for more than one year they have limited ability to assign differences in arthropod abundance to treatment without adequate replication. The White *et al* (1990) study also used hedgerow searching and attempted to assess immigration and emigration by placing pitfall traps either side of a polythene barrier parallel with the field boundary. **Wick et al. (2000)** sampled 10 ha plots within a 100 ha field over two seasons and refer to the data from individual samples from the same field as being replicates. Whist this approach is moving closer to commercial monitoring, this study suffered from pseudo-replication and with one plot per treatment the results could easily reflect the field differences or heterogeneity of arthropod distribution and not treatment effects.

Several authors do not describe an entire field approach, but instead generate supporting evidence that could be valuable in refining existing approaches. **Naranjo et al. (2005)** compared methods to evaluate the effects of Bt maize and insecticides on spider assemblages, concluding that suction sampling was the most effective method to sample spiders in maize crops. **Meissle et al. (2005)** also focused on the risk to spiders in a genetically modified crop and provide base line data on spider assemblages in maize. Although looking at a GM crop the study used plots 30 m x 50 cm in size within a single maize field per replicate farm. **Meissle et al. (2005)** recommended suction sampling or beating to sample spiders in maize and not stem eclectors or plant removal. **Thacker et al. (1993)** conducted a risk analysis based on the potential for ecological recovery, linking recovery with experimental design in large scale arable field studies. They found that

recovery in linyphiid spider populations was positively correlated with proximity to field margins.

Since field studies generate thousands of specimens from a wide range of taxonomic groups it is extremely time consuming and costly to identify all of the collected arthropods. **Cilgi (1994)** suggested that a non-target arthropod field study in cereals should be confined to the study of a few relevant indicator species. The author proposes species with poor dispersal, sensitive life history and names potential indicators together with proposed sampling methods. This approach is consistent with the view of **Brown et al. (2006)** in that certain species, those with consistently high species scores in the first order PRC analysis, provide more information about the effects of a pesticide on non-target arthropods.

Jansen (2000) presented an alternative "in-field" approach looking at foliar dwelling predators with a design comprised of three replicates of small plots (3m x 10 m) per treatment, including an untreated control. Treatments were applied in late June or July over three years when aphid numbers were high. Plant dwelling predators sampled three days after treatment using beating method (1994) and sweep nets (1995 and 1997). This approach doesn't generate information about non-target arthropods as a whole and leads to the question, "when should the taxa under investigation broaden from the predators and parasites listed in Escort 1 **Barrett et al. (1994)** to those occurring in the community?" Guidance could reflect this by requiring that at "field level" relevant not laboratory species should be studied. Due to the limitation to foliar dwelling predatory taxa, this approach has very little relevance for non-predatory taxa that may occur off-crop.

Since the publication of the **PSD et al. (1995)** guideline, many non-target arthropod field studies have been conducted by agrochemical companies as part of their regulatory submissions. There has been a gradual trend to include the testing of a reduced rate of the test item so as to simulate off crop risk and to look at effects on a wider range of organisms, including Collembola and soil mites. Most of these studies are not in the public domain. A large scale field study in winter

wheat **Brown et al. (2006)** was published without naming the products being evaluated and looked at the results of using univariate techniques at family level and at species level for carabid beetles, staphylinid beetles, linyphiid spiders and Collembola. The conclusions of this analysis were compared with those made using Principal Response Curves (PRC). When data was summarised at the family level genuine effects on arthropod species could be overlooked. PRC analysis indicates that some non-target arthropods, particularly small carabid and staphylinid species, may be more important indicators of treatment effects than others when a 1 ha plot experimental system was used.

Most regulatory field studies attempt to evaluate the off crop risk in an in-crop situation using a reduced application rate. In crop sampling, particularly in cereal fields where there is very limited botanical diversity, tends to be focused heavily on beetles, spiders, springtails and mites. Are these representative of arthropods as a whole? Some authors have attempted to study the effects of genuine spray drift in off-crop situations, either by sampling naturally occurring populations or by conducting bioassays. **Langhof et al. (2003)** studied the effects of spray drift on weed strips 1, 2 and 3 m from a sprayed field margin, mostly using bioassays with *Aphidius colemani* as a species with a sensitive adult life stage. Although only a limited number of drift rates could be achieved in the field, it would be possible to combine data from different distances to generate a median lethal drift distance for organisms such as *A. colemani* as part of an off-crop risk assessment.

Freier et al. (2001) conducted an off-crop study in the margin of a 50 ha wheat field in two consecutive years. The 650 m long grass strip margin was divided into eight plots, four with spray drift and four without. Sampling was by biocoenometer surveys (recording all arthropods in a $1m^2$ area), pitfall traps (one per plot) and recording of number of grasshoppers in quadrats, eight per plot. Malaise traps were also used but these were quickly abandoned. Fluorescent marker was used to assess contamination together with residue analysis of carabid beetles collected from margins together with adult *P. cupreus* held in cages. The study reported a low

density of arthropods in grass dominated crop margins. However, with a single pitfall trap per plot this may reflect the nature and number of subsamples per plot actually taken. Movement between plots may have masked effects and the authors cited the demand to identify a wide range of arthropods as being a limitation of this approach. Surprisingly, the authors detected no adverse effects of lambda-cyhalothrin drift on any non-target arthropod taxa, which is surprising since given the results from other field studies spiders would have been expected to have shown some effects.

3.4.5.3 Field studies in fruit orchards

Fruit orchards, whether pome fruit, stone fruit or citrus, can contain populations of non-target arthropods associated with the crop plant, the under-storey plant cover or the soil beneath the trees.

There are relatively few publications describing testing approaches for non-target arthropods in orchards. This may reflect the fact that orchards are relatively high value crops that are typically sprayed with many plant protection products and thus have low non-target arthropod populations. Whereas most cereal fields contain substantial populations of non-target arthropods and are therefore suitable as potential field study sites, this cannot be said for most fruit orchards. It should be noted that the orchard studies that have been conducted were either done in abandoned orchards or in orchards with uncharacteristically low pesticide inputs.

Reboulet (1994) describes methodology for studying the impact of plant protection products on beneficial arthropods in orchards. The principal method uses large plots of 6 rows of trees and 50 m in length each with beating (striking 50 branches per sample with a stick whilst holding a collecting funnel beneath the branch) to sample foliar insects and visual observation for mites and psyllids. The study design involves use of a soft reference, a toxic reference and the products

under test but no untreated or water treated control. There was no replication in the plot layout so it is not possible to assign differences to treatment with an associated probability.

For orchard studies, replication is much more difficult to achieve than for arable studies. Whereas it is often possible to find 20 ha of cereals in a single field or a set of adjacent fields, it is much more difficult to find a large enough orchard (at least 1000 trees) and containing high enough number of arthropods, in which to conduct a field study. **Brown (1998)** considered the design of experiments to assess the effects of pesticides on beneficial arthropods in orchards, in particular the dilemma between replication and adequate plot size. In a comparison of a replicated small plot study with 30 trees per plot and an un-replicated large plot study with 150 trees per plot at the same orchard site, he found that effects in the small plots were extremely short lived for all except the most sessile of taxa. The same taxa appeared to be reduced in number for much longer in the larger plots. Sampling of the small plot study was by suction sampler (modified D-vac), whereas the large plot study was sampled using inventory sprays of selected trees with a volatile insecticide.

Candolfi et al. (2000) give some guidance on conduct of studies in orchards and specify a minimum plot size of 0.2 ha when inventory sampling is used. Even this plot size will be unlikely to be appropriate for the most mobile insect taxa, such as adult flies and Hymenoptera. Although 4 replicates is desirable three replicates is considered acceptable if plot size is a constraint. Sampling methods for orchards recommended by Candolfi et al (2000) include inventory sprays, beating, the provision of refuges, malaise traps, water traps and visual observations. Although non-target arthropods had become the focus of attention by 2000, as opposed to beneficial arthropods which were the main interest when **Reboulet (1994)** published his methodology, there is no inclusion of sampling methods for arthropods present on the soil surface of orchards or the importance of taxa that may live predominantly off-crop in the Candolfi et al. (2000) guidance. Recovery is considered to be an endpoint but there is no recognition that recovery in 0.2 ha plots may be an artefact of the study design and may not

represent recovery that may occur in commercial scale orchards or in their adjacent off-crop habitats.

3.4.6 Modelling

Modelling to predict the likely impact of pesticides of non-target arthropods occurs at different levels, from those based on single species testing, multi-species testing and meta-population modelling. Whilst there is apparent usefulness in all these approaches they all require considerable validation to serve as useful tools in predicting risk to non-target arthropods.

Jagers et al. (1999) present an integrated toxico-kinetics based hazard model. The approach involves a bioassay with a herbivorous chrysomelinid beetle (*Gastrophysa polygoni*) sprayed on its host plant, black bindweed (*Fallopia convolvulus*) together with a model to allow prediction of effects at different temperatures. Since the larvae of *G. polygoni* are herbivorous and live on the undersides of leaves, it was assumed that the main uptake of toxicant was oral and not *via* contact. Residues were measured at different temperatures (12, 17 and 25 °C). The model successfully predicted larval morality with dimethoate but not with cypermethrin, possibly due to repellency.

Whilst there could be advantages in using a plant based bioassay together with modelling in attempting to quantify the risk to non-target arthropods, the comparative sensitivity of this test organism is unknown. The inability of the model to predict the effects of cypermethrin means that the approach requires refinement before it could be considered as being useful in quantify risk to non-target arthropods.

Sherratt et al. (1993) describe two meta-population simulation models to predict the likely long term impact of pesticide use on non-target arthropods. Both models assume that invertebrates disperse at particular rates over a matrix of fields and that each field experiences a specific

pesticide regime. One model investigates the dynamics of a polyphagous predator, which experiences direct mortality from the pesticide but which is not affected by the availability of the target pest. The second model investigates a similar system but also considers the dynamics of the pest, which is influenced by predation and by pesticide exposure. The first model predicts that the chances of a polyphagous predator population persisting in a field are greater if few other fields are sprayed. The second model predicts that regular applications of pesticides could eventually cause prey resurgence, with meta-populations fluctuating at higher densities than would occur in the absence of the pesticide.

This approach attempts to take into account the pattern and frequency of the intended use to predict the likely impact of a pesticide on non-target arthropods on a landscape scale. Whilst the models presented apply to predatory beetles it would be possible to construct similar models for a wide range of different taxa with differing susceptibilities and dispersal characteristics. The authors consider that such models could be used along-side field trials to determine acceptability of a pesticide under a proposed use pattern.

Stark et al. (1995) attempted to predict potential field effects from laboratory derived selectivity ratios with a range of species (pea aphid *Acyrthosiphum pisum*, convergent ladybird *Hippodamia convergens*, parasitic wasp *Aphidius ervi* and three bee species) to potential field effects. The authors calculated hazard ratios to determine the theoretical number of toxic doses per unit area. The authors draw a parallel with honey bee testing by concluding that hazard ratios of <50 are considered harmless. Probit substitution was used to demonstrate percentage mortality of non-target species at doses equivalent to LD_{90} in the target (pea aphid). Although aimed at IPM and the use of specific named products, this methodology comprises a parallel approach to the current regulatory approach where laboratory data for two species is used to predict harmlessness.

3.4.7 Non-target arthropod approaches - Conclusions

Extended laboratory studies with non-target arthropods are usually set up in a similar ways as for tier 1 tests in terms of the duration, temperature and life stage exposed but involve exposure on plant substrate for foliar dwellers and on a standard soil for soil dwellers. The aim of the studies is to provide more realistic exposure as part of the test system. A possible disadvantage of some test systems using 3 dimensional surfaces and untreated enclosure materials may be that the organisms can avoid exposure to some extent. Although the relevance of selected test species to assess the sensitivity of non-target arthropods as a whole is not known, it is considered that the confined exposure in the extended laboratory tests will still provide more conservative results than would occur in the field.

This review considers 'semi-field' methods to be as defined by Candolfi et al (2000) with use of standard laboratory reared species exposed in either barriered enclosures, cube cages or sleeves. The advantage of this methodology is the realistic exposure in a crop situation (usually cereals, grass or vegetables). Possible disadvantages are the appropriateness of the species/life stage tested and the non-recovery of a large number of the released organisms.

Aged residue studies involve the aging of pesticide residues in the field but with exposure of the non-target arthropods either as part of a extended laboratory or semi-field enclosure study. These studies utilise the appropriate protocols for either extended laboratory or semi-field studies and enable the potential for recovery of non-target arthropod populations to be assessed.

The reviewed papers describing higher tier approaches address different questions, depending on the time of their publication. Early papers were largely concerned with biological control and the effects of products on beneficial arthropods within the crop. Later approaches considered the non-target arthropod community as a whole and subsequently included an assessment of the off crop risk by using reduced rates but still in a crop situation. Virtually all approaches involve application of the pesticide under investigation at the maximum proposed field rate and a

reduced rate to simulate drift. No publications considered using a higher application rate than the proposed commercial rate to attempt to offset uncertainty over extrapolation to other sites or to untested organisms.

Few papers describe what constitutes an acceptable higher tier study and, whilst there is reference to variability in field data and low numbers of test organisms there is a reluctance to specify how homogeneous or at what level of abundance a population needs to be for a study to be considered to be valid. Higher tier studies with non-target arthropods offer surprisingly low levels of precision. Instead of the studies being designed to meet a required regulatory precision it appears that the regulatory interpretation is being driven by the typical precision that current study designs can realistically achieve.

Bluemel, et al (2000) consider that the predatory mite field method will detect effects of 50% in magnitude 80% of the time. No level of precision is quoted for studies covering a wide range of species. **De Jong et al. (2009)** presented draft guidance on how to interpret field studies with non-target arthropods. Although intended for interpretation of studies, this presentation provides lists of taxa which should be present in field studies conducted in arable and orchard crops, as well as, off-crop. The desirable taxa list is no longer focussed on beneficial groups and includes herbivores. The authors suggest eight classes of effect, from no effects observed to pronounced effects and no recovery within the study period.

The most logical higher tier test system for non-target arthropods must surely be one with the greatest taxonomic diversity (so as to reduce uncertainty over unrepresented taxa) and the opportunity to provide the maximum amount of information as to the extent of initial impact at a range of exposure rates. This is unlikely to occur in a monoculture but most applicable to a model off crop system with standard but high plant diversity. Historically, higher tier approaches have had a strong focus on predatory and parasitic taxa whereas an approach to protect all non-target arthropods must include representatives of different trophic levels and different functional groups.

Recovery, or the potential for recovery, is currently considered as the target end-point in higher tier studies with non-target arthropods. However, since recovery has been shown to be dependent on the scale of application **Pullen et al. (1992)**, it may not be appropriate to consider 1 ha plots as being representative of recovery from field or farm scale applications. The consequence of the selected scale of a higher tier approach is rarely considered but will determine the extent to which a study is relevant for different types of arthropods. Too small a scale for the organisms in question will generate an artificial impression of recovery which in reality is transient immigration.

3.5 Bees

3.5.1 Background

Although there is a regulatory requirement to evaluate the effects of pesticides on honeybees (*Apis mellifera*) in the Terrestrial Guidance document **SANCO (2002)**, there is also a general concern over the assessment of possible risk to bumble bees (*Bombus* spp.) due to their commercial use as pollinators in glasshouses. There is uncertainty over the validity of extrapolation of the results obtained for honey bees to risks to bumble bees. In this review testing approaches pertaining to both honey bees and bumble bees are considered. Approaches suitable for insect growth regulators are discussed.

The initial tier of testing for honey bees involves laboratory contact and oral tests (in accordance with OECD 213 and OECD 214 methodology) to generate topical and oral LD_{50} values. For insect growth regulators testing is recommended to follow the bee brood feeding methodology of **Oomen et al. (1992)**. A parallel initial tier for bumble bees has been developed with an acute oral test including 24, 48 and 72 hour observations of mortality and a contact LD_{50} test with 72 hour mortality **Gretenkord et al. (1997): Steen et al. (1996)**. Bumblebees do not share food and there is great variation in the amount of food taken up by individuals. This, together with the fact that bumble bees frequently fall into torpor during tests, means that group feeding is not suitable for assessment of effects of pesticides on bumblebees.

Whereas for honey bees there is a large body of data from field tests to validate the use of a Hazard Quotient of 50 **OEPP/EPPO (1992),** little such data exists for bumble bees and no HQ value is validated.

Higher tier approaches for honey bees fall into the following four categories.

1) Aged residue tests, which may involve assessment of repellency;

2) Tent, tunnel or cage tests which involve confinement of individual colonies over a crop which may be wholly or partly treated with test item. Cage or tent tests generally use smaller mesh covered enclosures whereas tunnel tests use commercially available mesh covered tunnels;

3) Field studies with single plots, with or without a control plot;

4) Monitoring field studies involving many field sites including untreated controls.

Aged residue, tent, tunnel and cage tests typically evaluate field rates of a test item in comparison with an untreated or water treated control and a reference item and use mortality and colony condition as endpoints.

3.5.2 Aged Residue tests

Lewis et al. (1990) describe an aged residue test, similar in principal to those commonly conducted for non-target arthropods. Field planted strips of relevant crop (oilseed rape and lucerne) were treated when plants were mature but not yet in full flower. After aging for 3, 8, 24, 48 and 96 hours foliage samples from the top half of the plants were returned to the laboratory for bioassays. Chopped foliage from each crop was placed in ventilated enclosures 50 mm high and 140 mm in diameter with the top and bottom formed by a Petri dish. Thirty bees were added to each enclosure and there were three replicates for each treatment, including a water treated control. Sucrose was added on a cotton wool wick at the bottom of the cage so bees had to crawl through foliage to reach it. Mortality was assessed after 24 hours exposure and results plotted over time to determine a time after which mortality was 50% (LT_{50}) and 25% (LT_{25}). The results demonstrate extremely large confidence intervals for these endpoints which make it difficult to use these in a meaningful manner in assessment of the risk to bees.

3.5.3 Cage, tent or tunnel tests

Schmidt et al. (2003) introduced a sequential testing scheme for bees with indices for the evaluation of tent tests and field test with honeybees. Products where the laboratory test generated a hazard quotient (HQ) of <50 are considered to be harmless whereas those with an HQ of > 50 require further testing in tent tests. If they are not found to be harmless in tent tests (where exposure is forced and deemed to be worst case) then they need to be the subject of full scale field trials to determine whether they are safe for bees or hazardous for bees.

Tent tests are considered to have a number of advantages. The substance is applied to flowering plants and the bees live in a real but small colony containing a queen. Sub-lethal effects on behaviour or on pollination can be observed and the health of the whole colony can be assessed. The tent tests enforce higher exposure than would occur in the field so if they show a product to be harmless then that would be likely to occur under the reduced exposure of field conditions.

Schmidt et al. (2003) propose calculating indices, taking for example the results of a variable such as number of dead bees after treatment divided by the number of dead bees before treatment. Where the control would be expected to remain close to 1.0 and index of greater than 2.0 (i.e. twice as much mortality as in the control) would be considered by the authors to indicate a cause for concern. Whilst the indices factor out absolute numbers to provide values useful for comparison they do not include a statistical evaluation of the power of the data. Two dead bees in a treatment versus one in the control represents a very different response from 100 dead bees in a treatment and 50 in the control yet both would have the same indices. Similar indices can be calculated for foraging activity and for brood development and serve to condense the data into a form which is relatively easy to interpret.

Tunnel tests with honey bees normally follow a standard protocol, similar to that described by **Delabie (1984)**. **Lewis et al. (1990)** established four tunnels on a 50 m x 40 m plot of oilseed rape, with one half planted 6 weeks later than the other to provide flowers for a prolonged period. A single queen-right colony was placed inside each tunnel and left to acclimatise. Mortality and foraging activity were assessed within each tunnel before treatment and for one week after treatment. Tunnel tests were particularly useful for assessment of pyrethroid insecticides which were found to be highly toxic in the laboratory when exposure was forced, but which showed some level of repellency when wet (residue) in the field and lower toxicity when the residue had dried.

Gretenkord et al. (1997) describe a tent test with bumblebees. A healthy queen-right colony of at least 100 workers is placed in a cool box in the ground outside the tent, to protect it from overheating. The box is connected to a mesh tent 3m x 4m x 2 m high over *Phacelia tanacetifolia*. When a constant foraging activity of approximately 10 workers per day is reached, the connection tube is closed during the day and the crop is sprayed. The test colony remains in the cage for 2-3 weeks and is kept in the laboratory for a further two weeks to check for abnormalities. **Van Der Steen et al.(2001)** reviewed bumble bee testing methods and considered that the main problem with tent tests was that the crop area and size of colony are not proportionate. There is not enough pollen or nectar available in the tent for a colony of normal size. **Gretenkord et al. (1997)** reduce colony size artificially, but this affects the structure of the colony.

According to the EPPO Guideline **OEPP/EPPO (1992)**, a brood test is required if it is presumed that a product affects bee brood development. Several workers report different methods for evaluating effects of pesticides on brood in honey bee colonies. **Brasse et al. (2003)** reported a ring tested method for assessing the side effects of plant protection products on the honey bee brood under semi-field conditions based on the work of **Oomen et al. (1992)** and **Muhlen (1996)**. The method involved a water treated control and Insegar 25 WG (active substance

fenoxycarb) as a reference item. Small polystyrene hives (Mini-Plus-Beuten) with synchronised nuclei and sister queens and approximately 6000 worker bees were exposed in tents of at least 40 m^2 established over a *Phacelia tanacetifolia* crop four days before the proposed application of treatments. Application was made at full flowering and during flight so as to ensure exposure of bees and their brood to treated nectar and pollen. Mortality was assessed daily using dead bee traps attached to each hive. Flight intensity was evaluated before application, 4 times during the first hour after application, 2, 4 and 6 hours after application, 3 times during the day after application and daily during the remaining exposure period in the tents. Brood development and brood termination rate were determined. Results from 5 trials were comparable and demonstrate that this method is appropriate for evaluating the risk of insect growth regulators to honey bees.

In an attempt to provide an adequate supply of food within the enclosure to support the colony, **Leymann et al. (2000)** proposed a semi-field brood test with large flight cages (4m x 12m x 2m) containing flowering *Phacelia* and *Sinapis*. One cage was used for each treatment (control, test item and toxic reference item). Each tent contained a small bee colony in an observation hive with about 100 eggs and young larvae marked on a clear sheet taped to the window of the hive. Windows made it possible to study the development of the individual brood without disturbing the hive. The reference item Alsystin WP25 (active substance, triflumuron) at 800 g a.s./ha resulted in 94.9% dead larvae and the control experienced 14.5% mortality. Whilst the results of this test are extreme and appear clear without the need for further analysis, with only one tunnel per treatment rate it is not possible to analyse these results statistically. This approach with three or four tunnels per treatment rate would be more statistically powerful.

Except when testing insect growth regulators, the majority of approaches record bee mortality at the hive (using dead bee traps) and colony condition by examining the combs (recording numbers of cells present at different life stages). **Colin et al. (2004)** considered the impact of sub-lethal doses of two insecticides within insect proof tunnels. Two semi-cylindrical tunnels, each 4 m high and with an 8 x 20 m surface area were each divided into two compartments to

provide four testing units. For each run of the test over time one control and three treated units were used. Concentrations of imidacloprid and fipronil 70 times more dilute than the oral LD_{50} concentrations were provided in feeders within each unit containing 600 g of syrup, enough to feed 150 bees. A video camera mounted above each feeder was used to record the number of bees present at the feeder every 3 minutes. Numbers of bees observed feeding (active) were recorded together with the number observed to be inactive. The ratio of inactive to active bees at each time point was determined. Both insecticides resulted in significant differences in inactivity to activity ratio by the fourth day of observation. Fipronil in particular induced a marked decrease in numbers of foragers coupled with an increase in inactivity at the feeder. Whilst this study indicates that low level exposure to insecticides could lead to reductions in foraging activity, it does not mean that these effects would be observed in the field, where bees foraging freely would be taking a mixture of contaminated and uncontaminated pollen.

3.5.4 Field studies

Although field studies provide the greatest level of realism in terms of exposure, they are complicated by the fact that bees will not necessarily forage on the flowering crop closest to their hive. In order to generate data with free flying bees, it is often necessary to use large plots of flowering crops. This makes replication and appropriate statistical analysis of the results difficult to achieve. Bees experience higher numbers of deaths in the colony during periods of cold or wet weather so it is preferable to conduct the critical phases of such studies during dry settled weather.

Whilst small open plots are not suitable for determining effects of a treatment on bees, they have been used successfully to investigate repellency. **Reet et al. (2007)** established twelve plots of oilseed, each 1 m x 10 m in size, within a field of wheat. Four were untreated, four were sprayed with the insecticide alphacypermethrin once and four were sprayed twice. Observations of bee

foraging were made and reported as numbers of bees per 1000 flowers. The experiment was repeated on three consecutive years. The relative number of honey bees foraging was found to be connected to floral density and the foraging results indicated no evidence of repellency after either a single or repeated application.

Inglesfield (1990) reported a field study using a single 206 ha block of flowering oilseed rape. No other rape was grown within flying distance of the test plot. Five bee hives were positioned at two locations adjacent to the crop, fitted with either dead bee traps or pollen traps. A further 30 hives were placed adjacent to the crop and monitored for colony condition. Hives were monitored and dead bee traps and pollen traps sampled daily before and for seven days after spraying the entire block of rape with the pyrethroid insecticide alphacypermethrin. Bees actively foraged the crop and pollen analysis was >90% rape. Low numbers in dead bee traps together with observations of healthy colonies over were used to propose that this product was of low risk to bees. Whilst these conclusions are plausible, because mortalities were low in this study, it would have been impossible to interpret such results if there had been intermediate levels of mortality, higher than caused by bad weather alone. The single large isolated block approach would be greatly enhanced if there were more than one block of the test item treatment and if there was an untreated block of oilseed rape to serve as a control.

Although brood development was observed in studies with conventional insecticides it has become the primary focus of studies with insect growth regulators. One of the main advantages of the open field approach is that the individual bees and the colonies receive realistic exposure to the test item. The method of **Oomen et al. (1992)** for honey bee brood feeding tests with insect growth regulating insecticides involves feeding colonies with one litre of sugar solution contaminated with the test item at the concentration used for spraying. Control colonies are fed uncontaminated solution and others are fed with a reference item. Whilst the bees can also forage freely, the exposure of the larvae in this system to the test material would be considerably higher than would occur following commercial practice. As such this method represents a screening to

identify harmlessness rather than an approach to quantify harmfulness that might arise from realistic exposure. **Steen et al. (1990)** present a similar feeding test in which brood is studied in colonies fed contaminated sucrose by marking brood cells on a translucent sheet and checking those cells weekly. They also describe the next stage in a sequential scheme, a field test in commercial orchards. One orchard was used as an untreated control and a second was treated with the insecticide phenoxycarb when the trees were in full bloom. Bee colonies within the orchard were fitted with traps to collect dead pupae and brood cells were marked as described for the feeding test. Oviposition dates were determined and brood development of successive oviposition dates was checked until brood cells were empty.

The observation of bee brood, marking the state of a given cell on a translucent overlay, and following its fate through to emergence is the standard technique for determining effects of IGR's on bees. Field studies can be conducted to determine the effects of treatments at different timing of application with respect to the flowering crop **Steen at al. (1990).** Whilst these field tests include an untreated control, particularly helpful in determining background mortality due to the weather, there is no replication and results for a treatment are derived from a single orchard or plot.

In response to the possible implication of a widely used product in reported bee mortality a large monitoring study was conducted **Steen et al. (2007)** involving 39 orchards with at least 1 km between them. Products known to be harmful to bees were not applied in any of the orchards. Nine orchards served as controls and 20 were assigned to treatment with the test item. Two bee colonies (queen right and healthy with brood covering at least 10 simplex frames) were placed in each orchard at the start of flowering. The test item, indoxacarb, an insecticide used to control Lepidoptera, was applied by fruit growers using their own equipment and according to local practice but without mixing with other plant protection products. All hives were fitted with Munster dead bee traps that were emptied before and after treatment and before application of any other products to the orchards. Dead bee counts were presented as the mean number of dead

bees per colony per day. Hive inspections were used to investigate colony development and growth. With such a large number of orchards the data were analysed statistically (ANOVA) and the conclusions concerning mortality are clear. There was no evidence of any mortality induced by the test item.

3.5.5 Bees; Conclusions

The sequential testing scheme from laboratory to semi-field (tent/tunnel or cage) through to field testing is intended to identify harmless products and limit testing to the lowest appropriate level. The cost and technical demands of conducting these studies increases considerably as the sequence progresses. In reality insecticides applied to flowering crops are likely to pose a risk to honey bees and there may be merit in stepping from the laboratory to a sequence of field studies.

Whilst there are advantages of the semi-field approach, (replication and standardisation), the fact that cages rarely provide sufficient food for the numbers of bees introduced adds uncertainty to the validity of their findings. Tunnel tests with a larger area of flowering crop than the smaller tents or cages will have fewer problems with colony behaviour and food shortages.

Field studies without any form of replication are dependent on finding harmlessness whilst confirming exposure for their results to be credible. If there is no control plot then the results are even more difficult to interpret.

Plant protection products which are believed to be harmful, but which might be expected to pose some risk to bees could be subject to commercial monitoring. The commercial monitoring study of **Van Der Steen et al. (2007)** had a large number of sites and generated a large body of data with little uncertainty and the ability to conduct meaningful statistical analysis. The concern from any one field study that the results could be specific to that location was addressed by the use of 29 locations.

All of the published approaches for honey bees and bumble bees focus on the short term and use immediate observations of mortality and brood development to evaluate the impact of products shortly after their application. The fact that Colin et al. (2004) found sub-lethal doses affected foraging suggests that mortality alone may not be a particularly robust endpoint for regulatory studies. Whilst it is clearly difficult to measure sub-lethal effects as they occur, since there are a multitude of possible types of effect, there has been no attempt to monitor hive condition over the longer term to detect the consequences of such effects. It would not be difficult to manage hives in such a way as to include meaningful evaluations of hive condition several months or after overwintering following their exposure to a test item in an experimental study.

3.6 Soil Organisms

3.6.1 Background

Approaches to evaluate effects of pesticides in soil can either have functional or structural endpoints. Functional studies, such as those which measure microbial activity through respiration or the extent of organic matter breakdown were the subject of the EPFES workshop in Lisbon in 2002 **Rombke et al. (2002)**. Structural endpoints typically involve investigating the effects of a pesticide on a single class of organisms, such as earthworms, Collembola or Enchytraeidae with no assessment of other ecological groups. Since organisms do not exist in isolation there has been growing interest in the use of multi-species systems in evaluating effects of pesticides, either as relatively simple small microcosms (e.g. **Burrows et al. (2002)** or larger Terrestrial Model Ecosystems (e.g. **Edwards et al. (1998))**.

Jansch et al. (2006) conducted a systematic review, comparing the laboratory toxicity data with the effects of pesticides on soil invertebrates reported from model ecosystem and field studies.

Adequate first tier methods are available for assessing the effects of pesticides on four soil invertebrate species, *Eisenia fetida* (earthworm, Lumbricidae), *Folsomia candida*, (springtail, Collembola), *Enchytraeus albidus* (potworm, Enchytraeidae) and *Hypoaspis aculeifer* (mite: Acari). Only for earthworms is there a field test guideline, which has been recognised since 1994.

3.6.2 Collembola

Collembola feature in higher tier testing both as soil organisms and as non-target arthropods. Epigeal Collembola species are readily sampled in pitfall traps and suction samples in large scale field studies conducted for non-target arthropod risk assessment **Frampton (1994)**.

Although predicted environmental concentrations (PEC values) in soil are calculated for products based on their application rates, **Houx et al. (1996)** consider that the interstitial concentrations in the pore water may be the primary route of exposure for Collembola. These authors present an acute toxicity study for *F. candida* and pesticides in an aqueous medium using 100 ml sample vials as test chambers. Four adult *F. candida* were exposed to a range of six or seven test concentrations of each test item within the vials for a period of 4, 7 and 14 days. *F. candida* walks on the water surface and cannot drown because it is not wetted and has no tracheal system. Death was found to very difficult to determine, since moulting individuals appeared dead. **Houx et al. (1996)** used the inability to respond to stimulation with a hair as a suitable endpoint for intoxication. Measurement of test item concentrations during the test was considered to be an essential element.

Wiles et al. (1996) describe a field bioassay approach to assess the toxicity of insecticide residues to Collembola. Three insecticides were sprayed onto a sandy clay loam soil (present at the field site) and a commercially available sandy soil (Lufa 2.2) with and without a winter

wheat canopy. Individuals of four Collembola species were confined for 24 h in the laboratory on field sprayed soils collected 1, 2, 3, 8 and 15 days after treatment. On each date, 10 individuals of each of the field collected species *Sminthurus viridis, Isotomurus palustris* and *Isotoma viridis* and 20 individuals of the cultured *Folsomia candida were added to* each of three replicate enclosures per treatment. No food was provided during a 24 hour exposure period after which time mortality was assessed in each chamber using a flotation technique.

Both soils were de-faunated using heat (70°C for 2 h) and stored at 5 °C in the dark. Field-collected Collembola were stored in conditions of 16 h light 8 h dark at 20 (±2°C) on plaster of Paris. The field collected species were fed with 10 g of baker's yeast, grassy vegetation and 10 g of de-faunated soil. *F. candida* was given only baker's yeast.

Collembola placed in order of sensitivity to chlorpyrifos were *Sminthurus viridis, Folsomia candida, Isotomurus palustris* and *Isotoma viridis.* As would be expected, probably due to increased bio-availability, Collembola were more sensitive on the sandy soil than on the clay loam. The presence of a cereal crop canopy reduced spray penetration to ground level and resulted in markedly lower toxicity than on bare soil. This approach has good potential as a higher tier test method and the inclusion of three species reduces uncertainty over the responses of different taxa. The approach would be suitable for multi-rate dose response testing.

Potential drawbacks of the field bioassay approach are that there is no potential for repellency to occur as the test organisms are given enforced exposure and the use of de-faunated soil reduces the potential degradation of the pesticide residue by micro-organisms.

The most widely used Collembola species in laboratory bioassays are edaphic species, especially *Folsomia* spp.. In comparison with epigeal Collembola, these species possess morphological adaptions to a subterranean environment, having smaller appendages and lacking setae or scales. These morphological differences may affect their susceptibility to pesticides as will their overall size and behaviour. There was a clear relationship between body size and susceptibility of

Collembola with the smallest, *S. viridis* being the most susceptible and the largest *I. viridis*, being the least susceptible.

3.6.3 Earthworms

Since the Annex VI (uniform principles) refer to risk to earthworms under field conditions it is not surprising that earthworms have been the subject of a range of higher tier methodology.

The initial tier for earthworms is the laboratory acute test with *Eisenia fetida* using artificial soil (OECD Guideline 207 (OECD 1984) and ISO11268-1). The sub-lethal so called "reproduction" study (ISO 1998) also with *E. fetida*, also represents a lower tier test, being a laboratory test with a sub-lethal (reproduction) endpoint.

3.6.3.1 Field studies

Higher tier testing with earthworms has conventionally consisted of field studies, either conducted in a relevant crop or in a model crop system. **(Heimbach 1992; Heimbach 1993)** reported the results of many field studies in permanent pasture using 10m x 10 m plots treated with the test item at field rate and at 4 times field rate. Worms were sampled using the formalin extraction method and abundance was estimated 4-6 weeks after treatment, in the autumn and in the spring of the following year. The Heimbach approach only included two experimental plots per treatment which did not give sufficient replication for adequate statistical analysis. The Federal Biological Research Centre for Agriculture and Forestry, Germany (BBA) guideline **BBA (1994)** provided a more rigorous description of requirements for an earthworm field study. A field site in grassland or orchards has to have at least 100 animals /m^2 whereas a site in arable

crops must contain at least 20 individuals /m^2. There must be representatives from each ecological group present and two important species must have a dominance of at least 10%.

The studies must have a randomised bock design with 10m x 10m plots with at least 4 replicates of each treatment, comprising of control, toxic standard, and test item. Four samples are taken per replicate at the pre sampling date. Formaldehyde or electrical extraction is proposed with a 30 minute period from the start of sampling to the end for each sample. Whatever extraction method is used it is necessary to conduct an efficiency assessment (by digging and hand sorting) to show at least 60% efficiency. Benomyl, (carbendazim) is proposed as a reference item at 2-4 kg/ha in 400-800 L/ha water. Sampling should occur 1 month, 4-6 months and 12 months after application. All sampling periods are to lie within the activity period of the worms. Biomass as well as species abundance should be determined. This guideline became the industry standard for field studies with earthworms.

Field testing approaches for earthworms were standardised by the International Institute for Standardization (ISO) with the publication of the guideline **(**ISO-11268-3) in 1999 (ISO 1999). However this guideline was originally designed for the evaluation of contaminated soil and modified for use in evaluation of the effects of plant protection products. The methodological requirements of ISO-11268-3 are essentially the same as the BBA (1994) guideline **(BBA 1994)**. The major advantages of the field study method are the inclusion of a range of different species with different routes of exposure to a toxicant in a single study and the realism of that exposure. Drawbacks of the field study approach are that the exposure may not represent worst realistic case, due to the soil type or due to other factors, such as interception by foliage, and the fact that weather conditions can adversely affect the results, with dry conditions yielding lower numbers of worms resulting in less precision in a study.

So as to develop a common regulatory language, **De Jong, Montforts et al. (2009)** published guidance for summarising earthworm field studies. This guidance considers the level of detail and precision in a study and assigns reliability criteria, from (1) reliable, (2) less reliable to (3)

not reliable. Studies in class 1 or 2 may be used in risk assessment but not those in class 3. Features which could render a study "not reliable" include improper description of the test substance, improper test site description, the absence of a negative control, very low worm numbers in the negative control and fewer than 50% effects on at least one sampling date in the positive control.

3.6.3.2 Earthworms in microcosms

Microcosms appear to be a particularly useful approach for studying the effects of a substance on earthworms in a specific situation. The risk to earthworms in forestry, where there is a thick litter layer and particularly high organic matter content, **Addison et al. (1995)** is not predicted by studies conducted on agricultural soils with a sandy loam or clay loam particle size. A forest organic microcosm study was developed **Addison et al. (1995)** using 10 cm long acrylic tubing approx 7 cm in diameter filed with forest litter to investigate effects of an insecticide on *Dendrobaena octaedra*. Substrates were frozen to -15°C for two weeks to kill any cocoons in the collected soil and soil was maintained at a moisture content of 70% wet weight, relevant for spring in Canada where the relevant compound would be applied. Although also included in the study *Eisenia fetida* did not thrive in the forest litter microcosms. Time to burrowing (following exposure), weight change and cocoon production were used as endpoints in the method. Cocoon production was found to be the most sensitive variable.

Choo et al. (1998) carried out a similar field cage experiment using sections of PVC pipe (30 cm diameter and 15 cm in length) driven 15 cm into soil to make their enclosures. In autumn, when worms are aestivating, the sections were removed and mesh was strapped to the underside to contain worms. In June (in Australia), 30 adult *Aporrectodea trapezoides* (Lumbricidae) were added to each enclosure and mesh was attached to the top of the pipe to prevent escape. After 4 days, treatments were applied to the surfaces of each cage. 38 days after treatment the soil in

each cage was hand sorted and individual worms were held for 24 h then weighed. The results indicated that cocoon production is a more sensitive indicator of pesticide effects in earthworms than growth.

Reinecke et al. (2007) conducted an earthworm microcosm investigation into cholinesterase inhibition in worms exposed to different levels of an organophosphate insecticide using stainless steel enclosures 20 cm deep and 12 cm in diameter with 5 replicates per treatment. There was a correlation between levels of biomass change and levels of cholinesterase inhibition in worms of the species *A. caliginosa.*

Whilst microcosms provide a level of control over the test system, using a known number of introduced individuals to generate data with much less variability that would occur in the field there is less realism and they only produce information for a single species of worm. Therefore whilst single-species microcosms represent a useful and valid approach there remains uncertainty over the effects on earthworms with a different behaviour than the one tested. Many workers have reported certain species of earthworm appearing to be more sensitive than others in certain systems. In a laboratory study, (Addison and Holmes 1995) found *D. octaedra* to be 8 times more sensitive to fenitrothion than *E. fetida*. In a mathematical model, **Baveco et al. (1996)** found *Lumbricus terrestris* to be more sensitive to pesticide effects than *L. rubellus* because it is constrained by the long duration of its juvenile stage.

Though the effects of earthworms can be studied within microcosm,s it is not possible to include a range of species or a range of life stages in such tests. Relatively few individuals are present for each treatment so effects on earthworm abundance are relatively coarse. The duration of most microcosm studies is relatively short and such tests would not necessarily detect effects on cocoon production which may be particularly sensitive to pesticide treatment.

3.6.4 Terrestrial microcosm, multispecies assemblages

National and international regulations on plant protection require data to evaluate the effects of plant protection products on terrestrial ecosystems. Usually this evaluation is based on the results derived from laboratory systems, although these tests may not be representative of ecosystems. Alternatively, field tests are undertaken which do represent ecosystems but are difficult to reproduce, are expensive to conduct and are time consuming. An alternative approach is the use of microcosms or model ecosystems.

The American Society for Testing and Materials **ASTM (1991)** defined a microcosm as being an intact soil core containing the natural assemblages and biota surrounded by the boundary material . Their guideline ASTM (1991) involved soil cores 60 cm deep with a 17 cm diameter mounted in a High Density Polyethylene (HDPE) tube with a collecting funnel beneath to collect leachate. The test takes place in a controlled environment room or a greenhouse. The cores are mounted on a cart with insulating beads between them to maintain a cool temperature in the soil. Whilst plants are included, either planted specially if an agricultural soil or grasses if pasture is used, this method does not refer to the soil fauna and does not consider the role of Collembola, Nematodes, Enchytraeidae, Mites or soil bacteria.

Mothes-Wagner et al. (1992) described a terrestrial microcosm as an intermediary between controlled laboratory studies on single species and the open field on a wide range of naturally occurring species. The bean plant, *Phaseolus vulgaris* was selected as the producer in the system because it is preferred food for phytophagous organisms, simple to rear and has a short growth phase. Other organisms were selected based on their ecological importance, being representatives of different trophic groups, being groups which naturally combine in their ecosystem, being organisms about which there is adequate knowledge of their ecological and ecophysiological variability, being organisms occurring at high density with short generation times and a small size, suitable for mass rearing and measurement of chosen parameters.

Bacteria, fungi, nematodes and enchytraeids were selected as reducers in the microcosm. Free living nematodes, enchytraeids and spider mites were selected as consumers. This approach uses histo-enzymology, histopathology and morphology to evaluate a wide range of endpoints. Responses of the test system to a chemical were determined by measuring soil parameters, cellular indicators, indicators at the organism, population and community level. Microcosms were 30 x 46 x 20 cm in size and succession was observed in the laboratory, greenhouse and in the field. Whilst this represents an interesting approach with potential usefulness in regulatory risk assessment, the procedure needs to be validated.

Edwards et al. (1998) described two methods, which were considered to represent different tiers in a potential testing sequence. Integrated Soil Microcosms involved small units (15 cm high x 7.5 cm internal diameter) filled with sieved field collected soil containing endogenous micro-organisms, introduced and indigenous invertebrates (micro-arthropods and nematodes) and a single plant species. Each microcosm was filled with 1 kg weight of air-dried soil per unit and there were six replicates of each of a range of doses of test item. Ten wheat seedlings were sown in the top 0.5 cm of soil and these were thinned to 1-2 per microcosm after 1-2 weeks. Three earthworms *Aporrectodea tuberculata* (Eisen) (1.5 g total weight) were added to each microcosm.

Soil moisture was maintained at 40-60% of field capacity, excess water was added weekly and leachate collected for nutrient and pesticide residue analysis. Samples of soil were taken 0, 7, 14, 28 and 56 days after treatment. Microbial biomass, litter decomposition, enzyme activity, bait lamina tests, nutrient leaching and pesticide degradation measurements were taken at each time point. Number of micro-arthropods, nematodes and earthworms were assessed at the end, after 56 days.

Edwards et al. (1998) also described larger Terrestrial Model Ecosystems (TME's) containing intact soil cores collected from the field maintained under lab conditions. These contained a greater diversity of indigenous invertebrates and mixed plant flora. Each TME consisted of a 40

cm deep x 17 cm diameter soil core encased in an HDPE tube resting on a funnel. In the lab the TME's are on carts to ensure that the soil and leachate temperature is lower (12-15 °C) than that above ground (20-22°C in day and 16-18°C during night). Artificial rainwater is added to TME's and those which produce small or large volumes of leachate are discarded. Treatments are applied so as to mimic agricultural practice. TME's are watered once per week to match weekly rainfall from the original site.

In assessing usefulness of the method, a model compound was applied at field rate (TI), TI x 6, TI x 36, TI x 216. The same endpoints as described for the Integrated small microcosm were measured 7 days before treatment and 1, 4, 8 and 16 weeks after treatment in the TME. The endpoints were designed so as to measure ecosystem structure, (e.g. microbial biomass and numbers of nematodes) ecosystem processes (e.g. microbial respiration and soil chemistry) and the fate of the pesticide (e.g. amount in soil, earthworms and plants).

Different workers have devised microcosms of differing sizes and with differing number of earthworms added from different species. Not all workers used intact cores and generally the smaller test systems used sieved soil added to HDE pipes. **Bogomolov et al. (1996)** used 15 cm deep units with a 5 cm diameter and included 1 earthworm (*Aporrectodea trapezoides*) as well as a mesh bag containing litter to measure organic matter decomposition. Studying the effects of copper, **Bogomolov et al. (1996)** used substrate induced respiration (SIR), soil urease activity and total nematode numbers as endpoints together with earthworm mortality, growth and body accumulation of the toxicant. For Copper, SIR was found to be the most sensitive endpoint. **Burrows et al. (2002)** used the Integrated Soil Microcosm approach of Edwards et al (1998) but included three *Lumbricus rubellus* in each of the test units. They assessed nematode populations before treatment and at the end of the test (after 56 days). Earthworms were sampled and individually weighed 7, 14, 28 and 56 days after treatment. One of the advantages of using soil cores removed from the field is the ability to collect leachate from beneath the column for

residue and nutrient analysis. **Hantschel et al. (1994)** describe an automated system for sampling leachate, conducting irrigation and analysing carbon dioxide from microcosms.

The term TME has come to represent a specific microcosm system developed and ring tested by **Knackerl et al. (2004)** as part of an EU project to evaluate their use to assess environmental risks (Project ENV4-CT97-0470). This project involved conducting pre-experiments, TME studies with carbendazim as a model test compound and field validation studies in four different countries, U.K., Germany, Portugal and Netherlands. The Portuguese study was conducted in an arable system but the others were all performed using grass. In each case the TME's comprised of 40 cm deep intact cores 17.5 cm in diameter mounted on carts. Different aspects of the project using carbendazim were published separately. **Rombke et al. (2004)** describe the use of TME's to evaluate the effects of carbendazim at rates of 0.36, 2.16, 13.0 and 77 kg a.s.ha^{-1} on earthworms. Grass cover was cut before spraying the pesticide and each column was artificially irrigated with 100 ml of artificial rain after treatment. In the ring test worms were sampled 1, 8 and 16 weeks after treatment. The inhomogeneity of earthworm distribution in the field appears to be have been realistically reflected by the TME's. Effects on single species could not be statistically evaluated since absolute numbers were too low. The authors conclude that the abundance and biomass of earthworms are suitable endpoints for assessment of chemicals within TME's but at sites where abundance is low data interpretation may be difficult. Predictability of biomass results derived from TME's is restricted if the number of large earthworms, such as *L. terrestris* or *L. rubellus*, is high.

Koolhaas et al. (2004) describe the soil micro-arthropod element of the TME ring test project. Sampling area within the TME's was relatively small consisting of a single sample 5-6 cm in diameter. Collembola communities showed large variations in numbers and no effects of carbendazim on species diversity were seen. Mites (identified to Astigmata, Cryptostigmata, Mesostigmata and Prostigmata) also showed large variations and the authors did not report consistent effects of the carbendazim treatment. Differences in vegetation between the TME's in

the four countries were thought to have led to variations in soil moisture which in turn would affect arthropod abundance.

The nematode element of the TME ring test was reported by **Moser et al. (2004)** Effects caused by the chemical treatment were observed on the number of nematode families, on the trophic structure of the nematode community and on the maturity index. The same effects were observed in the TME's as in the validation field sites performed at the locations from which the original soil cores were taken. Due to the higher sensitivity of omnivorous nematodes it was recommended to use their abundance as the main endpoint for nematodes in TME's.

Organic matter breakdown as a functional endpoint within TME's was investigated by **Forster et al. (2004)** using cellulose filter paper placed on the surface or inserted into the topsoil as standardised organic matter. Faunal feeding activity was assessed using a bait lamina method. The carbendazim induced effects on organic matter decomposition were the same in the TME and the field study and followed a dose response relationship. Effects on decomposition were correlated with effects on earthworms and enchytraeids but not with effects on bait lamina consumption.

Sousa et al. (2004) describe the effects of carbendazim on soil microbial parameters in the TME's and in the validation field studies. Control values for SIR, DHA and thymidine incorporation were similar in TME's and field studies. Phosphatase activity revealed more differences but results from TME's and field studies were of the same order of magnitude. Effects of carbendazim on SIR and DHA were observed early in the post-treatment phase of the study whereas those on phosphatase and thymidine incorporation were found 8 and 16 weeks after treatment. The responses to the model chemical were similar in most cases across the four countries and the authors conclude that TME's are promising as an integrative higher tier testing tool.

Nutrient cycling was investigated in the EU TME project by **Van Gestel et al. (2004)**. In the first series of tests, carbendazim at rates up to 77.8 kg as ha^{-1} did not affect sulphate or phosphate concentrations in the top 15 cm soil layers so these nutrients were not included in the second TME or the field validation study. Ammonium concentrations in the top layers of soil in the field study and in the TME's as well as in the TME leachate did not show any effect of carbendazim treatment. Nitrate concentrations in both soil and leachate did show some reduction at the highest carbendazim treatment level but this may be related to moisture content or earthworm activity. It is concluded that nutrient levels in TME's and field tests showed similar patterns, confirming the predictive value of the TME test system.

Considering the results of the EU validation of TME's as a whole, it appears that the results are more conclusive for variables which are unlikely to be affected by aggregation of organisms or to have high inherent heterogeneity. Soil micro-arthropods are unlikely to be homogeneously distributed in a field site, resulting in very different levels of abundance in subsequent cores taken from the same field. Whilst TME's provide many advantages, control over the system and the ability to sample leachate, they also appear to suffer from many of the difficulties associated with field studies, high variability resulting in low precision.

Weyers et al. (2004) considered the use of data from TME's in environmental risk assessment of biocides and industrial chemicals, and considered that the high degree of realism resulted in reducing the assessment factor applied to the endpoint down to 5. Given that the assessment factor applied when there is one long term NOEC is 100 this is a reduction in the assessment factor of 20-fold. Areas where TME data could be applicable for refinement of the risk assessment for plant protection products are the sections on other arthropods, earthworms and effects on soil non-target micro-organisms. Other arthropods could include data for gamasid mites or staphylinid beetles which are soil dwellers and could be tested within a TME.

TME methodology was assessed at the SETAC workshop PERAS (Semi-field Methods for the Environmental Risk Assessment of Pesticides in Soil) held at Coimbra, Portugal, 8-10 October 2007 (http://www.gaiac.rwth-aachen.de/peras). A review of this workshop was published by Schäffer et al (2008). TMEs were considered to be a suitable tool at the semi-field level to assess structural effects on the soil community. The TME should contain undisturbed soil cores, e.g., from an established grassland, containing natural communities, e.g. microarthropods, enchytraeids, nematodes, microorganisms. Efforts should be made to link and quantify exposure and effect in the TME systems, e.g. by chemical analyses and modelling. General requirements for TMEs were stated to be sufficient abundance of sensitive organisms in the soils used, measurement of soil moisture, optimised TME size, appropriate sampling, and appropriate statistics.

3.6.5 Functional Endpoint Studies

The EPFES workshop **Rombke et al. (2002)** considered five possible methods that could have relevance for the functional process of organic matter breakdown. These are the litter bag test **Kula et al. (2001)**, the mini-container **Eisenbeiss et al. (1999)**, the cotton-strip assay **Harrison et al. (1988)**, stable C and N isotopes **Nagel et al. (1995)** and the bait lamina assay **Torne (1990)**. Only the litter bag test was considered to be sufficiently well developed and relevant to be suitable as a technique for assessing pesticide effects on organic matter breakdown in the field.

The litter bag approach involves burial of mesh bags containing dried organic material (normally straw) in the soil of a field site which is treated with the test substance so as to represent realistic worst case agricultural use. The litter bags are removed from the soil at intervals over time (usually up to 12 months) and the mass loss of the control and treatment groups is determined for each sampling date. A draft test guideline for this approach is provided in **Rombke et al. (2002)**.

Early studies to detect the effects of pesticides on organic matter breakdown used the soil sterilant methyl bromide as a toxic reference. When this product was withdrawn from use in Europe it was not possible to find an alternative product that would reliably impact on the breakdown of organic material buried in the soil. In the absence of a suitable reference item the EPFES workshop, **Rombke et al. (2002)** introduced the need to demonstrate exposure by conducting residue analysis on the soil from within treated plots. The litter bag test very rarely resulted in significant differences in organic matter breakdown between a control and a test item treatment. Since the mesh bags recommended for use in this approach were of large mesh size, breakdown of buried material could occur due to the combined actions of earthworms, micro-arthropods or microbial activity. An effect on any one of these functions could be masked by the actions of the other. The extent to which these three functions were responsible for organic matter breakdown would be variable both within and between studies.

In the absence of a reliable reference item, the organic matter degradation approach may be considered to be of limited value.

3.6.6 Modelling

Modelling represents a potential higher tier approach in conjunction with the results of laboratory, TME or field studies. **Baveco et al. (1996)** propose an individual based modelling approach using the effects of pesticides on growth, maturation and reproduction to predict changes in the population size and structure of earthworms. The model suggests that *Lumbricus terrestris* is more susceptible to the effects of pesticides than *L. rubellus* which the authors suggest may be due to the duration of its juvenile stage. Whist it is possible to model the effects of a range of doses of a toxicant, the model is un-validated and the authors recognise that the validity of the Kooijman-Metz model used to represent the behaviour of single individuals is critical to the usefulness of this approach.

3.6.7 Soil organisms – Conclusions

Collembola higher tier data may be used to refine both the soil organisms and the non-target arthropod sections of the plant protection product risk assessment. Higher tier methods described in the literature include a laboratory method for *Folsomia candida* in aqueous medium **Houx et al. (1996)** and field assays using a number of different species **Wiles et al. (1996).** It was noted that edaphic and epigeal species may have different sensitivities due to their morphological differences and that this should be taken into account in the risk assessment.

Standardised methods exist for both laboratory and field earthworm testing. However, microcosms with earthworms are considered to be a useful approach for assessing the effects of a chemical in specific situations (e.g. forestry where there is a very thick litter layer). However, studies in microcosms are usually restricted to one species with only a few individuals with only a short duration time. Thus the level of precision and information on long-term effects from such studies is likely to be low.

TMEs are considered to be a useful system to investigate the impact of chemicals on the structure of the soil community. Considering the results of the EU validation of TME's as a whole, it appears that the results are more conclusive for variables which are unlikely to be affected by aggregation of organisms or to have high inherent heterogeneity. Soil micro-arthropods are unlikely to be homogeneously distributed in a field site, resulting in very different levels of abundance in subsequent cores taken from the same field. Whilst TME's provide many advantages, control over the system and the ability to sample leachate, they also appear to suffer

from many of the difficulties associated with field studies, high variability resulting in low precision.

The EPFES workshop **Rombke et al. (2002)** considered five possible methods to measure the functional process of organic matter breakdown, however, only the litter bag test Kula et al. (2001) was taken forward. This test design is considered to have a number of disadvantages. Firstly lack of an appropriate reference compound following the removal of methyl bromide sterilant from the market in Europe. Additionally any degradation seen could not be assigned to a specific group of organisms

Modelling represents a potential higher tier approach for soil organisms. An earthworm model described by **Baveco et al. (1996)** considered the possible difference in sensitivity for different species of earthworm. However as for all modes the validity of the model would need to be ascertained before it could be used for risk assessment purposes.

3.7 Terrestrial Non-target Plants

3.7.1 Background

Non target terrestrial plants may be exposed to pesticides *via* overspray, runoff, drifting, or leaching outside the intended sprayed areas. Many of the off-crop areas that are repeatedly exposed to plant protection products are ecologically important. Field boundaries are an important refuge for wildlife and play a key role in the maintenance of many wild species as well

as acting as corridors in which species can move from one natural area to another **Boutin et al. (2000)**. Therefore, the assessment of the risk to non-target terrestrial plants from the use of plant protection products (herbicides in particular) is important not only to protect the non-target (i.e. off-crop) plant species themselves, but also the plants as an important natural environment for other wild species.

The risk assessment of the terrestrial phytotoxicity of plant protection products, under the EU Directive 91/414/EEC, as described in the Guidance Document on Terrestrial Ecotoxicology **SANCO (2002)**, requires appropriate data on the toxicity of the substance of concern to a number of different plant species. At present, the assessment of effects on terrestrial non-target plants is based on studies of short-term effects on annual plant species. These standard tests include effects on seedling emergence and growth, and vegetative vigour (OECD 208 and 227 (2006)). Additional standard tests include ASTM E1963 (1994), ISO 11269-1 (1993), ISO 11269-2 (1995) and US EPA OPPTS 850.4000 (1996). There are number of concerns of the current testing approach (further details of which are given below). However, available and potential higher tier approaches for terrestrial non-target plants are extremely limited and not well documented in publically available literature.

3.7.2 Species selection

Routinely, tests are performed with a maximum of 10 species at the seedling stage or on seed germination on a case-by-case basis. Although the OECD Guidelines 208 **OECD (2006)** and 207 OECD (2006) include an extensive list of potential non-crop species which could be tested, as well as a list of crop species, the choice of test species is not specified in any regulatory guidance. Therefore, the test species are usually crop species. The choice of annual plant species as test species for risk assessment is primarily based on the ease of testing and the economical importance of crop plants. However, in contrast, most natural and semi-natural

habitats are dominated by perennial plant species. Perennial species differ from annual species because they can directly carry effects from one year to the next, whereas effects on annual plant species are only manifested the year after exposure if seedling recruitment is seed limited in the habitat of interest **Kjaer, et al. (2006)**.

Although it is important to protect agronomically important species from accidental herbicide drift, it is equally important to protect the habitats bordering areas where herbicides are used because of the ecological importance of the many wild plant species found there **White et al. (2007)**. It is possible that the current suite of species prescribed in current guidelines will not be adequate for the protection of habitats, e.g., field margin species, in agricultural areas. The non-randomness in the current selection of species favoured in these tests could cause an unacceptable bias which could mean that risk is underestimated **Boutin et al. (2004)**.

Boutin et al. (2000) examined the sensitivity of different plant species to a range of chemicals, in order to determine: 1. the optimal number of plant species that should be tested, and 2. the type of species that should preferably be tested to assess the phytotoxicity of a pesticide. On the whole it was found that crop species were not consistently more, or less, sensitive to the herbicide tested than non-crop species. Conclusions drawn from the study include the fact that grasses tend to respond in a similar way to various chemicals, thus the number of grasses tested could be minimized relative to broad-leaved species. It was recommended that more broad-leaved species should be tested than the number currently requested in the U.S. EPA guidelines and that more than ten species should be considered, although the ideal number could not be determined.

Boutin et al. (2004) reported the results from a unique greenhouse experiment in which 15 non-crop plant species were sprayed with 6 herbicides. The plants favoured were species commonly found in field margins of Europe and/or North America. One of the objectives of the study was to explore the feasibility of using non-crop plants commonly found in field boundaries as test species for herbicide risk assessment. The results from this study showed that: in general, the

selected plant species in the Danish/Canadian database were easy to grow and maintain in the greenhouse. The Danish/Canadian plants were overall more sensitive than the species tested in the US EPA data, yielding to a 5% protection threshold (HC5(50)) that was always more conservative. There was a large variability in plant responses among herbicides. Recommendations were provided on species that should and should not be used for risk assessment of non-target plants.

Cole et al. (1993) conducted analysis of the relationship between the number of endangered species and family size; this indicated that the most species abundant families are also representative of threatened species. This enabled construction of a prioritised list of plant families from which test species can be selected. Further selection of species from this list was based on their performance under glasshouse conditions and resulted in a pool of 14 species from 10 families suitable for regulatory testing.

3.7.3 Short term versus long-term effects

Currently, no tests are required at the adult stage to assess effects on reproductive growth and yield. Standardised tests do not use plant reproduction as a measure of effects, although it is recognised that the fitness of annual plant species largely depends on the reproductive output **Boutin et al. (2000)**. In addition, the plants used in the standard dose response experiments are all of similar age and size, although a much wider range of age and growth would be encountered naturally **Breeze et al. (1992)**.

The available data demonstrate that some herbicides may be equally or even more harmful depending on the growth stage of the plant. For example, **Boutin et al. (2000),** conducted a study to investigate the most sensitive phenological stage of a range of plant species to metsulfuron methyl. The seedling stage was the most sensitive period for all species tested,

although surviving plants sprayed at later stages showed considerable effects on the reproductive parts. Other examples (also referenced in **Boutin et al. (2000)** include glyphosate which is more toxic to hard to- kill perennial species in the fall than in the spring when plants are fully grown and thus have a large contact surface for penetration of the herbicide that can be translocated into the storage organs. Glyphosate has been shown to affect seed germination when parent plants were sprayed during the seed development. Pendimethalin may be more toxic to germinating seeds when applied pre-emergence or pre-planted incorporated. For most herbicides, effects at the vegetative and reproductive stages are largely unknown **Boutin et al. (2004)**.

A standard test guideline for chronic toxicity testing in higher plants is available: ISO (International Organisation for Standardisation) 22030:2005. This guideline describes a method for determining the inhibition of the growth and reproductive capacity of higher plants by soil under controlled conditions. The test can be performed either with *Brassica rapa* (turnip) or *Avena sativa* (oat). Its duration is 35 to 64 days with OECD artificial soil and a German standard field soil acting as controls. Besides measuring biomass and shoot length, the number of pods, seeds and flowers are applied as chronic measurement endpoints Kalsch et al. (2006).

The chronic plant test is considered to be a useful addition to the battery of existing plant tests and of ecotoxicological tests in general, for the evaluation of single chemicals. However, this test guideline only covers exposure *via* soil and testing of two species.

3.7.4 Exposure regime and scenario

The exposure methods used in the standard test guidelines (i.e. overspray and soil incorporation) may not be worst-case scenarios. Drifting droplets of pesticide in the field may land on various surfaces of the plant. Application to plant parts, such as old leaves, may cause less damage than

the same dose on a different part of the plant. Furthermore, greater toxicity may result from a dose being distributed over the plant in many droplets, compared with the same amount of herbicide in one droplet **Breeze et al. (1992)**.

It has been reported **Boutin et al. (2000)** that when spraying the herbicide aminotriazole directly on barley plants, the effect was 10 times less pronounced than damage caused by the same amount of the herbicide reaching plants through drift. The explanation for this was that leaves retained the smaller droplet sizes produced by drift better than those that were directly applied. It has also been found that fine droplets gave comparatively more biological activity than did large droplets in an experiment performed with chlorsulfuronon and cherry trees (*P. avium* L.) **Boutin et al. (2000)**.

The environmental conditions in which the plants are exposed are also a key factor which should be considered. **Dixon et al. (2005)** reviewed available date on the effects of humidity on responses of plants to pesticide exposure. In a series of experiments with *Populus* spp. some leaf distortion of sprayed leaves followed all clopyralid applications but the severity varied. Treatment in early spring to shoots from newly planted cuttings caused relatively little damage whereas one application in July to well-grown plants caused severe shoot distortion. A likely reason for this variation in response is the effect of humidity at the time of spraying on clopyralid phytotoxicity. In experiments on plants grown in controlled environments, greater toxicity of clopyralidon to *Tripleurospermum inodorum* (scentless mayweed) was found when grown in 80 percent compared with 63 percent mean relative humidity. It has also been found that absorption and translocation of [14 C] clopyralid in *C. arvense* was doubled in >95 per cent relative humidity compared with 40 per cent.

Differential root uptake is another possible factor causing variable symptoms. Examples, again referenced by **Dixon et al. (2005),** include clopyralid which is not strongly adsorbed on soil and may therefore have the potential to be leached by rain or irrigation down the soil profile where it can be absorbed by roots and have phytotoxic effects. In pot and field experiments on recently

germinated forest tree seedlings, irrigation after foliar spraying of clopyralid resulted in severe leaf distortion.

3.7.5 Controlled versus realistic environmental conditions

Testing is typically performed in the greenhouse with single species grown per pot, under controlled environmental conditions. Whether results from these tests are representative of field situations where plants undergo more adverse conditions (such as wind, occasional drought, insect damage, competition) is debatable and poses questions on the legitimacy of extrapolating from greenhouse tests to natural ecosystems. For example, it has been reported that the combined effect of several stressors in the field increased the sensitivity of *Polygonum convolvulus* to copper compared to laboratory tests **Boutin et al. (2004)**.

In addition, tests are not generally performed with non-target plants to investigate possible effects at the population, community, or ecosystem levels. Recurrent sub-lethal effects occurring on a few plant species may have repercussions at the community or ecosystem levels **Boutin et al. (2000)**. There are currently no standardised test guidelines for conducting non-target plant field or semi-field studies. However, such studies could be very useful higher tier testing approaches for the risk assessment of plant protection products. Higher tier studies could be conducted in line with a specific pattern of use of a substance, and involve exposure to a natural population of non-target plants. Such studies would therefore, provide a much more realistic indication of the lethal and sub-lethal effects on non-target plants. However, due to the natural variability of non-target plant communities and the number of variable environmental parameters, field studies are difficult to conduct and the results difficult to analyse and interpret.

de Snoo et al. (1999) referenced a microcosm approach for detecting the effects of herbicides on ditch-bank boundary vegetation. Bioassay studies show that a 2 m wide buffer zone is generally

sufficient to prevent the death of plants adjacent to fields sprayed with herbicides, although it is sometimes necessary to maintain a 6 m zone. The results from this study concluded that despite the variation occurring in the natural vegetation, it appeared feasible to draw conclusions from the study, probably in part because of the consistently pairwise comparison of ditch banks of the same field.

3.7.6 Summary

Herbicides will inevitably impact non-target species due to their extent of use and limitations in selectivity. Currently the risk assessment for terrestrial non-target plants (under Council Directive 91/414/EEC) is based on the results from standard first tier greenhouse studies. These studies assess short-term effects on standard test species under relatively homogeneous conditions.

However, the available literature highlights that current pesticide registration guidelines may not be adequate at predicting the effects of herbicides on wild plants and habitats, and several components of current phytotoxicity testing have been identified as areas of potential weakness that require further investigation **White et al. (2007)**.

Due to their availability and ease of culturing and maintaining in the greenhouse, crop species are typically tested. It would be considered more realistic to conduct laboratory studies on plants, which are representative of the natural environment (inc. crop plants, species in field margins and habitats interspersed within the agricultural landscapes). Many questions remain unresolved as to the adequate type and number of species to be tested **Boutin et al. (2004).**

Field and semi-field studies could provide very useful higher tier testing approaches for the risk assessment of plant protection products. However, it is often difficult to measure and predict the effects of herbicide use on natural communities in the field, because of the high variability

inherent in natural populations. This is especially true under conditions of spray drift, when the doses received by the organisms downwind of the sprayer may be sublethal **de Snoo et al. (1999)**. In addition, it is difficult to understand how persistence and reapplication timing interact with native plant demography **Crone et al. (2009)**.

Further work in developing higher tier testing approaches for terrestrial non-target plants is crucial. Also important are experiments investigating differences that may exist between data produced in greenhouses compared to the field, single-species tests compared to multiple-species tests (micro- or mesocosms), and whether longer test periods would yield different results regarding phytotoxicity of any given herbicide **White et al. (2007)**.

3.8 General Conclusions

Higher tier testing approaches across all areas of ecotoxicology have evolved and continue to change in response to changing regulation, developing scientific knowledge and the shifting of emphasis in the scientific community. For all areas of ecotoxicology there is a continuum from laboratory testing with high control, replication but limited realism (in terms of test organisms and their exposure), through semi-field/cage and outdoor microcosm/mesocosm with increased realism but still adequate replications and control, through to the field with populations and higher realism but with high variability. The literature shows that most higher testing has been done to assess risks to aquatic organisms, followed by non-target arthropods, bees and soil organisms, and finally with the least testing having been done for non-target birds, mammals and terrestrial plants. An overview table of the published methodology available for higher tier methods is included in Chapter 11.

It can be concluded based on the large amount of data available for aquatic organisms that laboratory tests on additional species and the appropriate use of modified exposure studies on species known to be sensitive can be very useful higher tier methods. Fish and aquatic arthropod data showed that the sensitivities of Australian fish and arthropods were not significantly different from those of corresponding non-Australian taxa **Hose et al. (2004)**. Additionally, arthropod taxa from a mesocosm were less sensitive than taxa in the laboratory tests, which suggests that laboratory-generated single-species data may be used to predict concentrations protective of mesocosm systems. SSDs based on laboratory data were also protective of field populations. This type of approach with appropriate validation could be very useful for the other areas of ecotoxicology.

The current approach to higher tier testing remains highly focussed on mortality and survivorship. Work with honey bees suggests that behaviour can be altered at exposure levels many times lower than the LD_{50} value. Longer term population studies would go some way to addressing this.

The pros and cons of the various types of laboratory microcosm and semi-field and cage studies are discussed in the various sections. It should be noted that design of such studies which are intermediate in complexity to those of standard laboratory tests or field trials should be undertaken on a case by case basis. The results of these studies can provide useful information and in some cases they can be used to enable the best design of follow-on field studies.

The move from semi-field to field approach usually incorporates the move from laboratory bred individuals to naturally occurring populations of multiple species with multiple life stages present. This represents a major increase in complexity and the results from such studies may often disappoint, since it is rarely possible to provide answers for all of the groups of concern from a single field study. The increased complexity in field studies leads to increased uncertainty so that results may have less precision than desired to answer all the questions raised.

No one field study design is appropriate for all the species found in a given test system. Are the taxa for which the study is "not suitable" relevant and of concern? If so then an additional approach is also necessary. For example, Terrestrial Model Ecosystems (TME's) appear from the ring testing papers to be ideal for soil microbial and enchytraeid study, but not ideal for the study of naturally occurring earthworms. Field studies with arthropods conducted in-field generally sample abundant and active taxa. These taxa are often specialists of disturbed habitats and have high resilience, being able to recover relatively rapidly from treatment effect. Effects on the less abundant species, with specific feeding or climatic requirements may not be detected in the studies currently being undertaken. It is clear in the aquatic area that later designs of micro and mesocosms have been targeted to assess the most sensitive organisms (e.g. zooplankton, macrophytes) and to include relevant *in-situ* bioassays.

Results from most field studies, particularly those when just field rate and/or drift rate are tested, present the decision maker with uncertainty over extrapolation to other locations with differing taxonomic composition and climate. The extent of the uncertainty over extrapolation to other taxa is different across the different areas of ecotoxicology. Low uncertainty occurs for bees, where the concern is largely limited to one species but large uncertainties remain for non-target arthropods, where the representativeness of the response for unrepresented organisms has never been demonstrated.

The basic principles developed for aquatic higher tier testing by CLASSIC **Giddings et al. (2002)** could be considered in the development of testing in other areas, namely dosing regime; dosing methods; timing of application; level of taxonomic resolution; species to be included; univariate versus multivariate statistical methods; derivation of acceptable concentrations; structural and functional endpoints; population recovery; database development and landscape ecology. It is noted that an exposure-response experimental design with replication was stated as being preferable for aquatic micro- and mesocosm studies. Field studies without sufficient

replication to conduct meaningful statistics are not useful to inform regulatory decision making. Pseudo-replication should be recognised as being unacceptable.

There will be a distribution of field responses and the results from a single field study could fit anywhere on the relevant curve. There is a distribution of possible responses that can occur in the field and a single field study from a single location will produce a response that could lie anywhere on that distribution. Workers can attempt to show that the particular study is worst case (and therefore protective of other situations) by using techniques such as discriminatory use of a reference item, recording of climatic conditions at the site itself and the use of field sites with greater bioavailability (e.g. sandy soil). Results from more than one higher tier study will serve to reduce this area of uncertainty.

Post registration commercial scale monitoring across a large number of real field sites, as conducted for honey bees by **Steen et al. (2007),** is a potentially useful tool across all areas of ecotoxicology since it removes uncertainty over location and has the potential to confirm the relevance of particular organisms. Monitoring data for active substances in conjunction with the presence or absence of relevant aquatic species in the environment is an evolving area with regard to the Water Framework Directive (WFD); however, the complex interactions in the environment may make clear conclusions on the causal agents of any species' absence difficult.

It is clear that higher tier testing methods can be a useful tool for pesticide risk assessment. In many cases especially for field studies it is clear that study design should be undertaken on a case by case basis. In planning higher tier tests, the importance of defining the specific objectives, optimal study design and appropriate analysis of the data, should be carefully considered **Ganio (1994).** It is also of note that clearer protection goals would greatly aid both the notifier and the risk managers to assess whether a data package demonstrates acceptable risk. At present, there remains large scope for different decision makers to apply different concerns to the same data.

3.9 Overview Tables

Birds

Organisms/ Taxa	Testing Approach	Literature Reference	"Pros": added value of design	"Cons": drawbacks of design
Avoidance/ palatability tests				
Bird: Columbidae	The paper (PN0914) describes a validation method for testing avoidance of treated seeds and assessment of robust tests for the acceptance of bait and treated seeds by birds.	Fryday, et al. 1999	The results stressed the importance of testing birds at the correct feeding rate. Time to avoidance reaction was also measured in bird species, and feeding rate determined to be highly influenced by ecological factors.	It is unclear whether test birds were able to select from contaminated and non-contaminated food.
Birds: Galliformes, P. domesticus	This reference (PN0909) tested whether methods developed under PN0914 were suitable for a larger test species (pheasant) and aimed to develop a method for small birds that addressed different welfare requirements. Plus validation work for exposure of pigeons on peas and maize seeds.	Fryday, et al. 2001	Characterisation of the feeding behaviours and rate of three important tests species in relation to contaminated food sources. It was confirmed that feeding rate should be carefully controlled while measuring avoidance.	Feeding rates may be species specific and knowledge of the appropriate species may be required.

Pen/cage "semi-field" studies				
Bird: *Alectoris rufa* cross	Semi-field study with red-legged partridge.	Johnston, et al. 1996	Increased environmental realism. Semi-field approach allowed greater control over variables, easier to interpret statistical significance, and provides a practical half-way stage between field and laboratory studies.	This design may be more expensive to implement than laboratory-based studies.
Field studies				
Birds*: Phoenicurus ochruros, Turdus sp., Sylvia atricapilla, Cyanistes caeruleus, Cettia cetti, Emberiza cirlus, Parus major, Picus viridis, Passer domesticus,* Corvidae, *Hippolais polyglotta, Erithacus sp., Certhia brachydactyla, Saxicola rubicola, Streptopelia turtur, Motacilla alba*	Birds were captured by mist net, tagged and released as part of this farm-scale study. Carcass searching following insecticide sprays was then monitored *via* tagged birds.	Brown, et al. 2008	High level of realism of acute effects in the environment.	There is potential for uncertainty with regards the efficacy of carcass searching and loss of carcass (e.g. by predators).
Foraging birds in the United Kingdom	Individual birds were fitted with radio-transmitters and tracked to determine potential long-term and sublethal effects following full cover spray applications. This was combined with visual searches for carcasses in the treated areas.	Wolf, et al. 2009	High realism, and therefore useful for risk assessment purposes	The field study design is costly and labour-intensive

Birds: *Turdus migratorius, Cyanocitta cristata, Toxostoma rufum*	Survivorship of ground feeding birds was monitored using radiotelemetry at 8 golf courses.	Poche, et al. 1993	While laboratory data suggested ingestion would be lethal to birds, this field study (i.e. using a realistic scenario) demonstrated no adverse effects.	There are many uncontrollable variables in a field study. Due to movement and territory size of animals, it is difficult to have a true control in a field study. Conducting study is likely to be time consuming and labour intensive, and using radiotelemetry also involves the use of expensive equipment.
Population assessment/ modelling				
Birds: *P. domesticus, Quelea quelea, Anas platyrhynchos, Agelaiius phoenicens*	Results of avian field studies were examined to model the likelihood of mortality in regards to the type of pesticide application and bird guilds. Variables tested for their explanatory power were: acute oral toxicity and application rate; oral to dermal toxicity of the pesticides; Henry's law constant; and the possible avoidance of contaminated food items.	Mineau 2002	Modelling is more cost-effective compared to numerous field studies, and can be used to estimate the direct losses of birds as a result of pesticide usage	This model only addresses acute lethal effects resulting from pesticide exposure (provides no information concerning reproduction effects, indirect effects, or even delayed mortality)

Bird: *Alauda arvenis*	Agent-based simulation model of skylarks in agricultural landscapes and its use to assess the impact of pesticides relative to changes in landscape structure and mortality assumptions.	Topping and Odderskaer 2004	Modelling approach considers spatio-temporal factors in population dynamics and the impact on risk assessment techniques. Landscape structure, crop diversity, or migration mortality was shown to significantly affect skylark populations (more so than pesticide exposure), so factors other than pesticides are likely to be limiting.	Input data requirements and model validation.

Aquatic Ecotoxicology

Organisms/ Taxa	Testing Approach	Literature Reference	"Pros": added value of design	"Cons": drawbacks of design
Modified exposure studies				
Algae: *Selenastrum capricornutum*	Refinement of the static algal growth inhibition study (OECD guideline 201) to a flow-through test system.	Grade, et al. 2000	Method provides a realistic exposure regime and comprises an Intermediate in complexity compared to standard lab studies and field testing.	Comparison to endpoints from static tests would be difficult and it is noted that the SSD approach should use studies undertaken with the same exposure regime.
Algae: *Desmodesmus subspicatus, Pinnularia subcapitata*	A chemostat was used to produce continuous culture, nutrient influx at constant rate, and constant environmental conditions (light, temp, aeration, CO_2) in testing toxicity of compounds to algae species	Weber, et al. 2009	Standard conditions can be easily and quickly established, increasing the appropriateness of comparisons. Because of this, the length of study can be longer and recovery potential can be assessed.	The design does not incorporate environmental interactions from other species and variations arising from habitat heterogeneity and climate conditions.
Algae: *S. capricornutum, Navicula pelliculosa*	Refinement of the static algal growth inhibition study (OECD guideline 201) to include sediment.	Shillabeer, et al. 2000	Environmental realism increased.	The inclusion of some sediments interfered with algal growth and its measurement. It was noted that only certain sediments are suitable for use in this design.
Algae: *Pseudokirchneriella subcapitata, Aphanizomenon flosaquae*; Cladocera: *Daphnia magna, Daphnia longispina*	Algae and cladocera exposed to leachate from agricultural field soil cores	Abrantes, et al. 2008	The approach incorporates a realistic exposure to relevant compounds in leachate.	The soil type chosen could render the design very site-specific. Storage of leachate before use may be problematic if breakdown of relevant compounds is rapid.

Cladocera: *Ceriodapnia dubia* Invertebrate: *Chironomus tentans* Fish: *Pimphales promelas* Macrophyte: *Ludwigia peploides, Juncus effuses*	Ditch sediment was transferred to indoor microcosms and populated with and without monocultures of macrophytes (i.e. vegetated and unvegetated). These microcosms were then exposed to pesticides. Standard test organisms were then exposed to water and sediment taken from these microcosms.	Bouldin, et al. 2005	This approach incorporated realistic concentration exposures to standard test species. Allows some standardisation of the effects assessment.	The exposures were dependant on vegetative density, and this design does not model effect in lotic systems. Population- and community-level impacts were not assessed (no information on interactions/indirect effects).
Laboratory multi-species tests Indoor defined microcosm tests				
Algae: *Chlorell*a, *Scenedesmus*, *Schizothrix* Protozoa: *Cyclidium* Rotifera: *Philodina*, *Lepadella* Oligochaeta: *Aeolosoma*	Enclosures contained green algae, a filamentous blue-green alga, a ciliate protozoa, two rotifers, aquatic oligochaetes and bacteria. Population densities and community metabolism were assessed	Sugiura 1992	The design incorporated controlled conditions and the use of multiple replicates was possible.	Such microcosms cannot describe or assess full ecosystem complexity.
Macrophytes: *Salicornia biglovii, E. canadensis* Invertebrates: *Turbo fluctosus, Bunodosoma californica, Ephydra sp., Crassostrea gigas, Palaemonetes pugio, Tubifex tubifex, Gyraulus sp, Margaritifera margaritifera, Palaemonetes kadiakensis* Fish: *Cyprinodon variegatus, Poecilia latipinna*	Indoor aquaria microcosms containing vertical biological filters cultured with bacteria and either marine or fresh water. Invertebrates and fish were added to each. Length of study 15-29 days.	Williams, et al. 1992	This design is easily replicated and set up, so that new studies can be set up within 2-3 days.	Indirect effects could not be monitored due to division of species within set-up. Not all species in the test were considered suitable and others may need to be defined. Study length may need to be extended to assess recovery.

Algae*: Scenedesmus* Cladocera: *D. magna* Macrophytes: *Elodea nuttallii*	Three trophic levels kept in separate sub-systems, connected by recirculating flow. An autotroph sub-system containing algae (Scenedesmus), a herbivore subsystem (Daphnia) and a decomposer sub-system (bacteria on a sand filter).	Leeuwangh, et al. 1994	High replication potential in both time and space. Low cost	Low complexity and no ability to see multi-species interactions.
Phytoplankton: *Cryptomonas* sp., Protozoa: *Urotricha furcata* Unidentified bacterial community	Indoor aquatic microcosms were stocked with phototrophic flagellates, predatory ciliates, and bacteria, representing three trophic levels.	Liebig, et al. 2008	Simple multi-species test that can assess both direct and indirect effects across many trophic levels with easy implementation, handling and low cost. Despite low realism such studies may be useful to understand indirect effects and to better define complex field studies	Because it is a laboratory system and run for only 10-13 days, this design provides a low realism. Also, recovery of populations and individual species sensitivity are not addressed.
Laboratory multi-species tests **Indoor semi-realistic microcosms comprising complex natural assemblages (lentic systems)**				
Phytoplankton: *Stephanodiscus astraea, Fragilaria crotonensis, Ceratium hirundinella, Chroomonas sp.* Zooplankton: *K. cochlearis, K. quadrata, Polyartha sp., Calanoida, Cyclopoida, Daphnia schodleri, Alona spp., Bosmina spp.*	Multiple experiments were conducted to determine optimal sediment composition and turbulence in lentic mesocosms. Additionally a technique for floating microcosms in subalpine ponds was investigated.	Harte 1984	Details operating features and appropriate scaling properties for experimental design of microcosms. Increased realism for plankton and chemical behaviour modelled either in microcosms in lab or floating in natural water body.	Would need to determine the appropriate water body to use for sourcing water and benthic material so that study can be reliably used in the risk assessment.

Periphyton and phytoplankton communities	Small samples are derived from natural communities of periphyton or phytoplankton. The sensitivity of toxicants is estimated using short-term measurements of photosynthesis in laboratory experiments.	Landner, et al. 1989	A flexible design that can be used in both the field and laboratory, and provides community level information. Simple design allows for many replicates, and so is good for comparative ecotoxicology.	Use of metabolic process as test parameter is dependent on toxicant's mode of action. Not suitable for assessment of long-term effects.
Algae: *Chlamydomonas rheinhardi*, *Secenedesmus subspicatus* Protozoa: *Tetrahymena pyriformis* Cladocera: *D. magna* Nematoda: *Caenorhabditis elegans* Fish: *Puntius semifasciolatus* Gastropoda: *Appolaria sp.*	Microcosms set up as aquaria contained in greenhouse, stocked with sediment and water from cultivated fish ponds. Caged fish and snails added to control zooplankton and algae	Traunspurger, et al. 1996	Microcosm design allows assessment of effects on multi-species from several trophic levels. Reduction in uncertainty for plankton and microbial species sensitivity, and spatial and temporal parameters may be possible.	This experiment was conducted without replication of treatments and thus statistical power is low.
Zooplankton, Phytoplankton, Macrophytes and Filamentous algae	Replicate aquatic indoor microcosms were seeded with sediment from dried temporary ponds (containing resting stages/eggs of organisms), and exposed to plant protection product.	Barry and Logan 1998	This approach describes the use of a small, easily reproducible experimental unit that adequately modelled conditions observed in field	Small size lead to dilution of nutrients following sampling. Not all microcrustacea species found in the field were found to be represented. Insect fauna which may dominate temporary ponds in late successional stages were completely absent.

Microbial populations including: Algae: *Golenkinia sp., Scenedesmus sp.* Cyanobacteria: *Oscillatoria sp., Microcytis sp.*	Laboratory microcosm systems filled with synthetic, moderately hard dilution water. Three different P and N nutrient regimes were applied and naturally derived microbial populations on substrata were added ("epicentres"). These had been colonized in natural environments.	Pratt and Barreiro 1998	The design in a relatively cheap and simple system to set up and replicate. It is possible to select appropriate nutrient level to mimic average or worst case situation.	Incorporates less realism than larger more complex systems, and so produces only general results (in terms of protein, chlorophyll and biomass, as well as, photosynthesis and respiration parameters). Variability of the microbial populations on the "epicentres" could hamper interpretation of the results.
Zooplankton: Rotatoria: *Lecane lunaris, Lecane bulla, Keratella cochlearis, Kereatella quadrata, Lepadella patella* Cladocera: *Chydorus spp, D. magna, Daphnia galeata, Alona rectangular* Copepoda Insecta (Cloeon spp.) Ostracoda Phytoplankton	Indoor microcosm design was stocked with field-collected zooplankton and phytoplankton prior to exposure	Daam and van den Brink 2003	The simple, presumable low-cost laboratory mesocosm design may prove useful for initial concentration range-finding and for identification of sensitive phyto- and zooplankton species.	Design lacks ecological complexity since no macrophytes, macroinvertebrates or sediment were included and exposure may therefore be overestimated.
Macrophyte: *E. nuttallii* Plankton Macroinvertebrates: *G. pulex, A. aquaticus, Proasellus meridianus*	Indoor microcosms were constructed with aquaria containing natural lake sediment and tap water. These were stocked with *E. nuttallii* shoots, plankton and macroinvertebrates, and allowed to acclimated for four weeks	Van Wijngaarden, et al. 2004	This indoor microcosm design permits a high level of realism in controlled environmental conditions.	The restrictive laboratory design prevented certain recovery processes and so may represent worst-case exposure impacts (compared to impacts observed in the field).

Zooplankton Phytoplankton (plus lab sourced *D. galeata*,	Indoor laboratory microcosms were seeded with natural sediments, water, and phyto- and zooplankton. Effects on abundance and community metabolism were assessed.	Van Wijngaarden, et al. 2005	Relatively simple and low cost which can be easily replicated. Alternative environmental parameters can be assessed. Good for organisms that can be maintained at high levels in a small scale experiment eg. Plankton. Relevant $NOEC_{community}$ can be derived if the most sensitive organisms are included.	Indirect effects may not be determined in small scale system. May not be relevant for certain species that are not easily maintained in small scale systems.
Zooplankton Cladocera: *Mesocyclops pehpeiensis* Algae: *Chlorella*	Two densities of primary producer chlorella (green algae) and consumer *M. pehpeiensis* were added to simple indoor mesocosms containing field-collected sediments. Assessment of zooplankton community structure was made and food web analysis.	Chang, et al. 2005	This relatively small indoor microcosm is simple to set up and replicate for statistical analysis	The simplistic design does not encompass all the complexity of the natural environment.However, such data may support argumentation for reduction of assessment factor for other higher tier (micro/mesocosm) data.

Laboratory multi-species tests Indoor semi-realistic microcosms comprising complex natural assemblages (lotic systems)				
Macro-invertebrates including *Isonychia spp. Stenonema spp., Baetis sp., Caenis sp.*, Corydalidae, Hydropsychidae, Chironomidae, Elmidae, Periphyton	Indoor mesocosms with either flow-through or no-flow conditions and or current were constructed and seeded with field-collected (riffle) organisms prior to exposure.	Pontasch and Cairns 1989	The design enabled a relatively long bioassay that included complete life cycles of some species. Artificial streams supplied with current were able to maintain all mayfly taxa at or above initial levels for the entire 30 day experiment. High realism may be useful for refinement of risk assessment for macro-invertebrates in the lotic environment.	Laboratory system lacks the ability to provide a full assessment of recovery due to recolonisation or immigration. Additionally, there may be a limit on the duration of such microcosm experiments.
Invertebrates: Baetidae, Oligoneuriidae, Hydropyschidae, Philopotamidae, Helicopsychidae, Limnephilidae, Perlidae, Elmidae, Chironomidae, Simuliidae, Oligochaeta	Laboratory stream microcosms were seeded with field periphyton slurry and macroinvertebrates and exposed to continuous concentrations. Toxicity to riffle insect assemblage was assessed.	Breneman and Pontasch 1992	This approach comprised a realistic model for risk of continuous exposure to stream macroinvertebrates and provides realistic assessment of bioavailability.	A continuous exposure assessment is conservative and not relevant to pulsed exposures.
Invertebrates: *Baetis tricaudatus*	Indoor artificial stream systems were stocked with field-collected mayflies. Acute toxicity was assessed at different levels of current: low (0 cm/sec), medium (6 cm/sec) and high (12 cm/sec). The endpoints measured for the mayflies were immobilisation and	Lowell, et al. 1995	Provides insight into effects for a sensitive Ephemeropteran species in a lotic environment. Low cost of replication.	Low level of realism since only one species, short-term test without presence of environmental refugia and other species.

	number of moults.			
Lentic field studies: pond, mesocosm, microcosm and enclosure studies				
Fish and benthic and pelagic communities	Limnocorral design (approx. 300 m^3) combines both benthic and pelagic communities using enclosures in a lake with a duration of about 5 months.	Landner, et al. 1989	High level of realism enables assessment of ecosystem structure and function. Realistic fate can be modelled over relatively long time period.	Expensive, lower replication possible. Lower levels of phosphorus in the enclosures compared to lake due to phosphorus binding to the enclosure walls. Lowered density of fish lead to increased zooplankton levels. No established guidelines and validation. Each lake has its own specific parameters and choice of lake should be made carefully.
Macrophytes: *Chara sp, Naja sp, Typha sp, Sagittaria sp.* Cladocera: *D. magna* Invertebrates: Diptera, Ephemeroptera, Gastropoda, Coleoptera Fish: *Lepomis macrochirus*	Outdoor mesocosms (700 $m^{3)}$containing bluegill, macrobenthos, zooplankton, phytoplankton, and macrophytes were held for a 5 month time period. Additional laboratory microcosms were set-up containing either bluegill or *D. magna*	Fairchild, et al. 1992	Mescocosm with sediment provided more realistic exposure and assessment of indirect effects.	Increased cost compared to lab data.
Macrophytes: *Typha latifolia, Sparganium americanum, Eichhornia crassipes, Potamogeton diversifolious* Zooplankton: Cladocera, Copepoda, Rotifera, Invertebrates: Diptera, Odonata, Trichoptera, Gastropoda,	Large outdoor rectangular ponds (1100 m^3) with varying depths (including a shallow littoral zone) were constructed with reservoir water and natural bacteria, fungi, algae, zooplankton, insect communities. Cultured fish	Webber, et al. 1992	Variable habitat (sloping area with littoral species) and high diversity of species increases realism, as does the seven month duration of study	Significant variation between mesocosms resulting from different macrophyte and fish densities. Further, the large size makes these systems relatively expensive.

Coleoptera, Ephemeroptera Fish: *Ctenopharyngodon idella*	populations also introduced into systems			
Fish: *L. macrochirus* Macroinvertebrates: Ephemeroptera, Diptera, Odonata Zooplankton: *Macrothrix rosea, Diaphanosoma brachyurum, Chydorus sphaericus, Alona rustica, Brachionus spp, Monostyla bulla, Filinia longiseta*	Microcosms were stocked with bluegill, zooplankton, phytoplankton, macroinvertebrates, and then exposed to plant protection product. Microcosms (1.9 $m^{3)}$ were compared with larger mesocosms (634.7 m^3).	Morris, et al. 1994	Microcosm design described in this reference is less expensive to set-up and monitor than comparative mesocosms. Production of dose response data would be more feasible using microcosms.	Specific effects were observed in microcosm but not observed in mesocosms (i.e. slight growth effect on bluegill). Presence of bluegill negatively affected zooplankton populations in microcosms.
Phytoplankton, Zooplankton, Macrobenthos, fish	This method compared outdoor microcosm (5 m^3) and mesocosm (75 m^3). set-ups. Microcosms contained phyto-and zooplankton, benthic species, and caged fish. Mescosms contained the same organisms, plus macrophytes	Heimbach, et al. 1994	Microcosm data were shown to be similar to mesocosm data except for the non-inclusion of macrophytes. Both microcosm and mesocosm have increased realism – better simulation of environmental exposure. Microcosm design cheaper and easier to replicate.	Macrophyte inclusion in mesocosms lead to differences in the phyto- and zooplankton communities between the mesocosms. Fish required the addition of food.
Fish: *L. macrochirus* Zooplankton: Cladocera, Rotifera, Copepoda Invertebrates: Ephemeroptera, Gastropoda	Both earthen ponds (470 m^3) or fiberglass tanks (11 m^3) were stocked with benthic invertebrates and adult bluegill (ponds) or juvenile bluegill (tanks), and exposed to plant protection product.	Howick, et al. 1994	The tank design is larger than laboratory microcosm and smaller than most outdoor mesocosms, and so provides good reliable data at less of a cost (e.g. may be good design for obtaining preliminary data). This design is able to incorporate fish and have some	The design may benefit from further refinement in selection of fish species and the possible inclusion of piscivorous fish.

			habitat heterogeneity.	
Phytoplankton: Bacillariophyceae, *Fragilaria sp., Cryptomonas marsonii, Cryptomonas erosalovata* Zooplankton: *D. longispina, Simocephalus serratulus, Synchaeta* sp., *Polyarthra* sp.	The test methodology describes the enclosure of 1 m^3 compartments in a natural pond to assess the effects of pesticides on phyto- and zooplankton for up to 47 days.	Juettner, et al. 1995	Inclusion of multiple species in realistic environment provides a high level of realism.	Significant seasonal variation in planktonic populations must be considered in interpretation. Relatively large mesocosm size means relatively large cost, restricting number of replicates.
Zooplankton: *D. longispina, Eucyclops serrulatus, C. sphaericus, K. cochlearis, Ascomorpha sp, Asplanchna sp., Synchaeta sp., Polyarthra, sp*	Compartmentalized mesocosms (1 m^3) were inserted in a natural pond, treated and observed for 47 days.	Peither, et al. 1996	This design provides a realistic exposure scenario and allows for assessment of interspecies interactions in a dose response study. Duration of study also allows for definition of recovery period.	The design lacked replication, and there was a potentially significant loss of lindane in study. Presence of non-planktonic organisms.
Macrophytes: *Typha angustifolia, Elodea canadensis* Invetebrates: *Lymnaea palustris,* Diptera, Ephemeroptera Amphibians: *Bufo bufo, Rana temporaria, R. esculenta, Triturus helveticus* Fish: *Carassius auratus*	Outdoor mesocosm systems (12 m^3) were seeded with natural silt, macrophytes, woodlice, goldfish and snails (both caged and free), followed by 8 months of natural insect and amphibian colonization prior to exposure.	Caquet, et al. 1996	This test system simulates realistic ecological conditions.	The complexity of test system may render it difficult to set up.
Fish: *L. macrochirus* Insecta: *Notonecta sp., Buenoa sp., Caenis sp., Callibaetis sp.* Amphipoda: *Hyalella azteca* Macrophyte: *Ludwigia uruguayensis*	Microcosms (17 m^3) enclosed within outdoor ponds, with natural sediments and provided refugia in each. Water column bioassay with *Notonecta (*Linnaeus) and *Buenoa* (Kirkaldy), Notonectidae, Hemiptera. Two	Shaw and Manning 1996	The relatively small microcosm size allows for replication, better statistical analysis, and assessment of potential recovery. Bioassays provide more detail on potential recovery. *Hyalella* is a well characterised species	Not all possible ecosystem interactions can be assessed using this design. Bioassays of *Hyalella* required more effort because a laboratory culture had to be maintained and since the animals had to be slowly acclimated to

	epibenthic bioassays using *Caenis* and laboratory-reared *Hyalella azteca* .		used routinely in the US to assess toxicity in sediments.	pond conditions.
Benthic and pelagic organisms from sediment. Invertebrates: *A. aquaticus* L , *Gammarus pulex, D. magna* Macrophytes: *E. nuttallii, Chara sp., Ranunculus circinatus*	Outdoor experimental ditches 40m in length were used to assess the effects on aquatic communities. These were allowed to develop for two year prior to pesticide application, and invertebrates introduced eight months prior to application.	Van Wijngaarden et al. 1996	This design incorporated a high level of realism, and caged studies allowed for the conduction of regression approach and ECx values could be calculated outside the tested range.	High biological variation could lower the precison of results.
Macrophytes: *Potamogeton pectinatus, Myriophyllum sibiricum*	Enclosures were constructed within a natural pond (part of wetlands system interconnected by canals) and populated with transplanted macrophytes. Effects of toxicant on plants were assessed after 30 and 60 days respectively.	Forsyth, et al. 1997	Macrophytes were exposed under semi-natural environment and conditions.	Environmental variability. Problems with plants adapting to transplantation, although such effects seemed to stabilise over time.
Invertebrates: Ephemeroptera, Chironomidae, Cladocera, Odonata, Dytiscidae, Cyclopoida, Notonectidae, Leptoceridae, Ceratopogonidae, Corixidae, Ostracoda	Shallow (11 cm) experimental ponds (38 m^2) were constructed and stocked by natural macroinvertebrate colonization . These were exposed through spray simulation	Burdett, et al. 2001	A relevant measure of the realistic impact of rice field flooding on aquatic invertebrates community can be determined through this methodology.	The assessment of species-level effects will require more detailed and frequent sampling than utilised in this study.
Macrophytes: *M. spicatum, M. sibiricum, Lemna gibba*	Macrophytes from laboratory cultures and transferred to outdoor mesocosms containing fish and insects.	Hanson, et al. 2003	Laboratory-cultured organisms exposed in outdoor conditions. *M. Sibiricum* was the most sensitive species and may therefore be useful as an indicator test	*M. Spicatum* produces algicidal allelopathic compounds and thus may not be suitable for use in microcosms where algal populations are also being evaluated.

			species.	
Invertebrates: *Chironomus riparius*	Sediment enclosures were placed in artificial pond systems and seeded with *Chironomus riparius* larvae. AChE activity measured as an endpoint.	Maycock, et al. 2003	Standardised dimensions and inexpensive construction allow for efficient recovery and adequate replication. Design enables test organisms to burrow directly into the sediment.	Relevance of the specific AChE biomarker. Indigenous Chironomids may cause confusing results.
Fish: *L. macrochirus* Zooplankton: *Rotifera, Copepoda, Cladocera, Ostracoda* Phytoplankton: *Chlorophyta, Bacillariophytceace, Cryptophyta, Euglenophyta, Pyrrhophyta* Invertebrates: Insecta, Annelida, Nematoda	Rectangular fibreglass outdoor mesocosms (30.9 m^3) were stocked with natural pond water, stocked with bluegill, and aged 10 weeks.	Rand 2004	High realism for exposure and interaction of species. Relatively low cost compared to larger mesocosms.	A more detailed analysis of taxa would require replication.
Macroinvertebrates: *G. pulex, Chaoborus obscuripes, Asellus aquaticus*, Ephemeroptera, Tubellaria Zooplankton: *Daphnia pulex, K. cochlearis, K quadrata, Anureopsis fissa*, Copepod*a, D. galeata* Macrophyte: *E. nuttallii*	Mesotrophic (macrophyte dominated) and eutrophic (phytoplankton dominated) outdoor ditch microcosms (approx. 0.5m^3) formed by pressing plastic tubes into the sediment. Native macroinvertebrates, zooplankton, and phytoplankton were enclosed within the tubes and dosed.	Roessink, et al. 2005	The design is a relatively small and low cost dose-response testing methodology, with some replication. Additional information was made available from *in-situ* bioassays. The application method utilized simulated worst case exposure	Natural variability in ditch organisms.
Macrophytes: *Lemna minor, Myriophyllum spicatum, Potamogeton lucens, Chara globularis*	Macrophytes were planted in plastic pots using mesocosm sediments and inserted in pot holders, or in the case of *L. minor*, placed in floating containers,	Coors, et al. 2006	Use of standardised laboratory species exposed under more realistic conditions (shallow water body). Three of the five	Macrophytes may obscure the detection of effects on phytoplankton

	and then exposed to plant protect product.		planted (plastic pot) submersed macrophytes (*M. Spicatum, P. Lucens, E. Canadensis*) showed satisfactory growth in the control pond and thus demonstrated their suitability for *in-situ* bioassays. Similarly L. minor can be used to assess direct effects in realistic environmental conditions.	
Published Lotic field studies				
Invertebrates: Ephemeroptera, Diptera, Odonata, Trichoptera, Plecoptera, Coleoptera, Turbellaria	Outdoor stream microcosms were constructed of stainless steel and stocked with river water and field-collected invertebrates and rocks. Concurrent field sampling in uncontaminated and contaminated sections of river for comparison purposes.	Schulz, et al. 2002	High realism was demonstrated for effects on and distribution of taxa, and data from microcosm studies was validated with field data. Design incorporated a reasonable level of replication and dose response with multiple concentrations applied as pulses.	Test methodology did not include an assessment of recovery
Macroinvertebrates: *Gammarus pulex, Baetis sp., Ephemerella sp., Dytiscidae, Chironomidae, Leuctra sp, Hydropsyche, Oligochaeta*	Lotic mesocosm systems were established in the shallow part of a stream riffle. Macroinvertebrate drift and benthic samples were taken following exposure by	Heckmann and Friberg 2005	The method enables examination of effects at the community level, and assessment of recolonisation potential following pulsed exposure.	Variation in regional and physical-chemical factors in-stream may cause considerable variation in the impact among different stream ecosystems.

	the toxicant.			
Invertebrates: Ephemeroptera, Odonata, Plecoptera, Trichoptera, Heteroptera, Coleoptera, Isopoda, Amphipoda, Oligochaeta	Closed circulation outdoor artificial streams with downstream reservoirs were established using sediment and macroinvertebrates collected from an uncontaminated stream.	Beketov, et al 2008	The determination of macroinvertebrate species sensitivity in a lotic environment was accomplished with a relatively high degree of realism. The study was of reasonably long duration (7 months) allowing recovery to be investigated.	A re-circulating stream system is unrealistic and does not allow for expected immigration from upstream thus the system could be considered to be conservative with respect to exposure and recovery potential.

Non-Target Arthropods

Organisms/ Taxa	Testing Approach	Literature Reference	"Pros": added value of design	"Cons": drawbacks of design
Extended laboratory studies				
Parasitic wasp: *Aphidius rhopalosiphi*	Standardised extended lab test to evaluate the effects of plant protection products on *A. rhopalosophi* both in terms of acute (mortality over 48 hours) and sub-lethal effects (reproductive capacity). Treatment of a 3-D test substrate (barley seedlings) over which the insects are confined for 48 hours.	Mead-Briggs, et al. 2009	Standardised and ring-tested method.	Possible avoidance of exposure by resting on the untreated surface of the glass enclosures.
Hoverfly: *E. balteatus*	Adult hoverflies are released into cages over treated buckwheat plants containing aphids. Mortality assessed for 4 days together with oviposition and behavior. A residual toxicity test is started after 8 days and lasts for a further 6 days.	Tornier and Drescher 1992	Enables assessment of risk to the lifestages tested.	This method does not assess the effects of larval exposure.
Semi-field methodology				
Carabid beetle: *Poecilus cupreus*	Adult beetles confined in enclosures dug into crop situations. Enclosures are over-sprayed. Mortality assessed by recapture of beetles and sub-lethal effects assessed by consumption of Drosophila pupae as prey.	Heimbach, et al. 1992	Exposure scenario is realistic.	*P. cupreus* is a burrowing beetle with a relatively thick cuticle. Those individuals that burrow beneath the soil surface for all or part of the exposure period of a study will be less exposed than those that remain on the soil surface.

Invertebrates: Carabidae, Staphylinidae, Coccinellidae, Lycosidae, Chrysopidae, Anthocoridae	Four methods for use in cereals: barriered enclosures, 2m cube cages (*C. septempunctata*), sleeves and barriered large plots. In the first 3 methods laboratory reared organisms were released into cages or enclosures shortly after treatment and their survivorship recordedafter periods of time.	Jepson and Mead-Briggs 1992	Confinement and exposure of laboratory-derived insects in semi-field conditions increases realism of exposure while minimizing the interspecies response variability.	A large rate of non-recovery of released organisms. Where sleeves were used organisms may avoid exposure by clinging to the untreated barriers.
Carabid beetle: *Pterostichus melanarius* Lycosid hunting spider:*Pardosa sp.*	1m x 1m enclosures.	Brown, et al. 1990	Realistic exposure especially for surface active predators where contact most relevant route of exposure.	
Green lacewing: *C. carnea*	This semi-field approach incorporated laboratory-cultivated insect larvae released into orchard trees immediately prior to spraying. Larvae were re-captured with bait cards.	Vogt, et al. 1992	Realistic exposure . Authors cite success of method with conventional insecticides and insect growth regulators.	No information on how to extrapolate results to predict effects on other non-target invertebrates.
Field studies				
Non target arthropods	General guidance on design, conduct and interpretation of non-target arthropod field studies. Advocates use of either arable or orchard as model crop system. Use realistic worst case exposure and assessment of effects on phytophagous, detritivorous and predatory arthropods. Taxonomy to species level where possible.	Candolfi, Bigler et al. 2000		

Predatory mite field studies – vineyard and orchards				
Predatory mite: *T. pyri*	This method assesses the short and long term effects of products on phytoseiid mites in vineyards and orchards by sampling population density compared to that in a water treated control at different time intervals after application.	Bluemel, et al 2000	Realistic exposure of naturally occurring populations both directly and indirectly. Ring tests in apple and vineyard showed that effects of greater than 50% were statistically significant in90% of cases.	Relatively expensive and variability of natural populations may reduce precision. Not clear how effects on mites reflect the response of non-target arthropods as a whole.
Predatory mite: *T. pyri*	This methodology describes the addition of overwintering sampling of predatory mites to existing methods (by dissection of leaf buds collected in February).	Gyorffyne and Polgar 1994	Enhancing sampling to include surveys in an additional season generates useful data concerning effects at different life stages and the potential for recovery following exposure.	Increased costs may be associated with sampling through leaf bud collection and dissection.
Mites: *Euseius finlandicus* and *T. pyri*	Assessment of effects on mites in orchards using similar methodology to Bluemel et al 2000.	Sterk, et al. 1994	Realistic exposure.	Sensitivity of mite populations can vary due to previous exposure to pesticides. This could make extrapolation of results to other situations difficult.

Arable field studies				
Spring and autumn breeding carabid beetles, staphylinid beetles, spiders, aphid-specifics (parasitoids, coccinellids, neuropteran larvae, syrphidae and game-bird chick food insects).	Early reference that outlines some necessary components of field and semi-field studies in cereals, including at least four replicates of either small barriered or large unbarriered plots, and that data should be collected in two or more site-years.	Carter 1993	Large plot: a high degree of realism, with no need to erect barriers and no risk of over sampling; also provides data from a wide range of taxa, (especially polyphagous predators). Small barriered plot is more practical as only 1 ha required. Due to smaller size selection of site with high and relatively uniform population is more achievable.	Large plot: requires > 20 ha of cereals with homogeneous arthropod populations. This early guideline focused on predators, parasites and bird food insects. There is no mention of Collembola or mites and there was no investigation of off-crop effects.
Main taxa:Braconidae, Empidoidea, Carabidae, Staphylinidae, Linyphiidae	Experiment was carried out in two 4 hectare fields, with one sprayed with a synthetic pyrethroid and one with a positive control substance, yearly for five years. Foliage and soil invertebrates were collected with D-vac and pitfall traps.	Inglesfield 1989	The multiple year duration of the study increases realism and allows for observation of long-term effects on several invertebrate populations	Methodology incorporated only one plot per treatment which severely restricts statistical power of conclusions.
Predatory taxa plus specific assays of *Nebria brevicollis*, *Bembidion obtusum*, *Trechus quadristratus*	Movement of fauna between sprayed and unsprayed areas was estimated using a different method in each of the 2 years (traps on either side of barrier and then surface searches of fields and hedgerows). Enclosures were used to assess the mortality of key beneficial species.	White, et al. 1990	Assessment of immigration and emigration. Significant effects were detectable in about 50% of species tested when this method was utilized. The semi-field enclosures provided additional information for the key beneficial species.	Lack of replication.

Invertebrates: Aphidae, Araneae, Carabidae, Chrysopidae, Coccinellidae, Entomophthorales, Staphylinidae, Syrphidae, Cicadina, Diptera, Heteroptera, Hymenoptera, Nematocera, Symphyta, Thysanoptera	Two 10 hectare plots were established in existing crop fields (one control, one treatment), and multiple within plot invertebrate samples were collected over two years.	Wick and Freier 2000	Sampling over multiple years allowed researchers to track long-term effects and recovery of invertebrate populations.	Replicates utilized during the course of this study were actually pseudo-replicates, and this design is therefore lacking in statistical rigor.
Theridiidae, Linyphiidae, Tetragnathidae, Araneidae	Methodology was developed to determine the effects of BT corn and pesticide spraying on spider populations using 30 x 50m sub-plots	Meissle and Andreas 2005	Suction sampling was determined to be the most efficient and cost-effective methodology.	
Carabidae Linyphiidae	3 year field trial. The data from the first 2 years were used to develop models of the recovery process whild data from the third year was used to validate the models. Pitfall traps were used to sample non-target epigeal invertebrates.	Thacker and Jepson 1993	Model was able to pinpoint distance from field as crucial variable in determining population-level recovery.	Need information on level of variation of recovery rate in families to assess effects on individual species.
Invertebrates: *Coccinella septempunctata, Propylea quatuordecimpunctata, Episyrphus balteatus, Chrysoperla carnea*	This paper presents an in-field method of examining susceptibility of foliage-dwelling invertebrate predators. Three replicates were utilized per treatment, and invertebrates were collected *via* beating and sweep nets.	Jansen 2000	Realistic exposure.	Methodology is limited in that it specifically assesses foliar dwelling predators, and may not be useful in assessing off-crop communities.

Invertebrates: Carabidae, Staphylinidae, Linyphiidae, Collembola	This describes a large-scale field study in winter wheat using univariate analysis to identify changes at the family and species level for carabid, staphylinid beetles, linyphiid spiders and Collembola.	Brown and Miles 2006	Certain indicator species (identified through first order PRC analysis) may provide the most information on non-target arthropod effects in 1 ha plots.	High cost.
Syrphidae, Chrysopidae, Coccinellidae Bioassay: *Aphidius colemani*,	Weed strips one to three meters from the margin of a sprayed field were surveyed for resident insects following field spraying.	Langhof, et al. 2003	Design of experimental methodology allowed for the calculation of median lethal drift rates for invertebrate taxa.	The numbers of beneficial non-target arthropods sampled from plots was low and may decrease power of statistical assessment.
Invertebrates: non-target agroecosystem arthropods	Off-crop study in the margin of a wheat field. The 650 m long grass strip was divided into plots with spray drift and controls (no spray drift). Resident arthropods were sampled *via* biocoenometer surveys, pitfall traps, and grasshopper counting in quadrats following spray drift	Freier, et al. 2001	Realistic off-crop exposure.	Low density of arthropods reported in the grass margin, however, this may be due to the nature and number of subsamples taken per plot. Identification difficult due to wide range of arthropods.
Field studies in fruit orchards				
Invertebrates: Mites, psyillds	Large plots containing six rows of orchard trees (50 m length) were exposed *via* spray. Mite and psyilld populations were then sampled *via* beating methods	Reboulet 1994	Realistic exposure.	Untreated or water-treated controls are not included. No replication in the study. Replication of orchard studies is difficult to achieve. No sampling methods for arthropods on the soil surface or taxa that may predominantly live off-crop.

Invertebrate: Hemiptera, predatory Heteroptera, Coleoptera, Neuroptera, Hymenoptera, Diptera, Araneae, Dermaptera, Lepidoptera, Orthoptera, Thysanoptera,	This reference detailed the comparison between a small plot study (30 trees per plot) and an un-replicated large plot study (150 trees per plot) at the same orchard site, with respect to measuring invertebrate following exposure to plant protection products.	Brown 1998	Both field set-ups were able to detect effects in arthropods communities from realistic exposure.	The effects observed in small plots were short-lived and transitory in nature. Effects seen in the large plot may be difficult to interpret with any statistical certainty.
Modelling				
Bioassay: *Gastrophysa polygoni* L. on host plant *Fallopia convolvulus*	This reference describes the development of a laboratory bioassay and a model to predict toxic effects at different temperatures.	Jagers op Akkerhuis, et al 1999	Laboratory bioassay and model could allow standardization.	Sensitivity of G. polygoni is not known. While the model successfully predicted larval morality with dimethoate, it did not predict responses to cypermethrin, possibly due to repellency therefore refinement of the approach would be required before this approach could be used for non-target arthropods as a whole.
Invertebrate: Carabidae, Staphylinidae	Modeling techniques were developed to predict the likely long term impact of pesticide use on a polyphagous predator. Model 1 considered effects on the predator independent of pest populations, while Model 2 incorporates pest population dynamics resulting from both predation and the pesticide.	Sherrat and Jepson 1993	Model parameters are designed to take into account the pattern and frequency of the pesticide use in order to predict the likely impact. Model can also be adapted for use with a number of non-target predator taxa.	Models need to be validated with field studies.

Invertebrates: *Acyrthosiphum pisum, Hippodamia convergens, Aphidius ervi, Apis sp*	Laboratory-derived selectivity ratios were used to predict effects in the field.	Stark, et al. 1995	Although constructed for use in IPM programs, this method can be adapted for any use in which laboratory data for two species can be used to predict harmlessness in the field.	The approach does not take into account bioavailability and exposure in the field. Field validation would not be as easy as for bees where population effects can be assessed by consideration of mortality at the hive.

Bees

Organisms/ Taxa	Testing Approach	Literature Reference	"Pros": added value of design	"Cons": drawbacks of design
Aged Residue studies (bees)				
Invertebrates: *A. mellifera*	This methodology describes an aged residue test utilizing honeybees. Field foliage was sprayed, allowed to age for up to 96hr, cut and placed in laboratory honey bee enclosures.	Lewis, et al. 1990	Realistic exposure of foraging bees following natural weathering of the residues.	The confidence levels generated for the results were extremely large therefore relatively difficult to use the results for risk assessment.
Cage, tent or tunnel tests				
Invertebrates: *A. mellifera*	Tent test: Toxicant applied to flowering plants and bees live in a real but small colony containing a queen. Sub-lethal effects on behavior and pollination and health of colony can be assessed	Schmidt, et al. 2003	Realistic exposure and effects on whole colony can be observed.	The indices proposed in the paper could be misleading. Full statistical analysis of the data would be preferable.
Invertebrates: *A. mellifera*	Tunnel test: Single queen right colony placed in tunnels on plots of flowering oilseed rape. Mortality and foraging activity assessed.	Lewis, et al. 1990	Realistic exposure allowing for repellency effects and lower toxicity of dried residues.	Variability in the results.
Invertebrates: *Bombus sp.*	Tent test over flowering Phacelia. Test colony remains in cage for 2-3 weeks and then in laboratory for another 2 weeks.	Gretenkord 1997	Deployment of colony in semi-field conditions, followed by laboratory observation, led to increased realism of effects, while also enhancing the quality of observations of individual bee behaviours.	Artificial reduction in colony number is may have altered the structure of the hive.

Invertebrate: *Apis mellifera*	Tents: Honeybees were kept in polystyrene hives in an agricultural field sprayed at full flowering during flight (to ensure exposure of both workers and brood). Worker behavior and mortality, as well as brood development were quantified as endpoints.	Brasse, et al. 2003	Methodology has been validated through ring testing, and deemed an appropriate tool in determining effects of insect growth regulators on honeybee populations and hive health.	Repeated, careful field observations may require significant time and cost.
Invertebrate: *A. mellifera*	Tent brood test: small bee colonies (with approx 100 eggs and young larvae marked on the clear window of the hive) deployed in large flight cages with flowering *Phacelia* and *Sinapis*.	Leymann, et al. 2000	The clear window system on the hive that allowed for observation of developmental progress of honeybee larvae without disturbing the hive.	Only one replicate was utilized at each treatment level, reducing the statistical strength of the data.
Invertebrates: *A. mellifera*	Tunnel: honeybee enclosure that allows for the observation of sublethal behavioural effects following dietary exposure to plant protection products.	Colin, et al. 2004	Bee behaviour could be tracked and quantified every three minutes using video system, and so the progression of effects could be accurately tracked and measured.	This design did not allow bees to select from contaminated and non-contaminated food sources, a more probable scenario in field situations.
Field Studies (Bees)				
Invertebrate: *A. mellifera*	Plots were established within a flowering oilseed rape field (1m x 10m), and either used as control, sprayed once or twice. Bee foraging was observed following application.	Reet, et al. 2007	Small plots appear to be suitable for investigating repellency.	

Invertebrates: *A. mellifera*	Thirty five hives were placed in a field of flowering oilseed rape, and subjected to pyrethroid spray. Hives were monitored for dead bees, pollen collection, or general hive condition.	Inglesfield 1990	A realistic exposure scenario and low pyrethroid-related mortality increased validity of conclusions concerning field toxicity of compound to honeybees.	No use negative or positive control or replicate treated plots severely restricts usefulness of the data generated.
Invertebrates: *A. mellifera*	Colonies fed with sugar solution containing the insect growth regulator (IGR) at the test concentration used for spraying. Control and reference item treatments. Bees can also forage freely. Effects on adults and brood monitored.	Oomen, et al. 1992	Allows a screen for IGRs that are harmless.	Exposure is considerably higher than that expected in the field following spraying.
Invertebrates: *A. mellifera*	field test in commercial orchards. Mortality and brood condition monitored	Steen and Ruijter 1990	Effects at different timings can be assessed. Untreated control enabled background mortality due to weather to be assessed.	No replication of orchard/plot.
Invertebrates: *A. mellifera*	Monitoring at 39 orchards with no history of pesticide spray toxic to honeybees and at least 1km apart were selected and populated with 2 bee hives each. 9 controls and 20 treated orchards. Following spraying bee mortality and brood condition were monitored.	Steen and Dinter 2007	The large number of test sites allowed for a high degree of statistical certainty.	Field plots covered a large area, and likely require extensive monitoring effort.

Soil Organisms

Organisms/ Taxa	Testing Approach	Literature Reference	"Pros": added value of design	"Cons": drawbacks of design
Collembola				
Folsomia candida	The methodology describes an acute toxicity test for Collembola in an aqueous medium in 100ml sample vials	Houx, et al. 1996	Effects expected from exposure in pore water can be assessed using this test method. Concentrations were measured several times which allows for the generation of reliable and accurate dose-response curves.	Mortality was difficult to determine.
Sminthurus viridis, Folsomia candida, Isotomurus palustris, Isotoma viridis.	Bioassay: Collembola were contained in the laboratory for 24 hr on pre-sprayed soils aged for varying times in the field.	Wiles and Frampton 1996	Assessment of the sensitivity of different species. The approach could be adapted to for multi-rate dose response testing.	This methodology did not provide an option with which to assess repellency.
Earthworms in microcosms				
Invertebrates: *Dendrobaena octaedra, E. fetida*	A forest soil microcosm (forest litter)was used to characterises pesticide toxicity to earthworms Following exposure, burrowing time, weight change and cocoon production were utilized as endpoints.	Addison and Holmes 1995	Enables risk in specific circumstances to be evaluated. In this case effects in forest areas with thick litter and high organic matter was evaluated.	*E. fetida* did not thrive in this system, indicating that this methods may not be applicable for all earthworm species.
A. trapezoides	Enclosures were made from PVC pipes. Earthworms were added and then treatments made. 38 days after treatment numbers of earthworms and cocoons assessed.	Choo and Baker 1998	Smaller system that may enable dose rates to be assessed.	*A. trapezoides* is a shallow-burrowing earthworm, and so may not appropriately model the susceptibility or exposure of deeper-burrowing worms.

Aporrectodea caliginosa	Earthworm microcosm system, constructed of using stainless steel enclosures 20 cm deep and 12 cm in diameter (5 replicates per treatment)	Reinecke and Reinecke 2007	Acetylcholinesterase inhibition in earthworms was shown to strongly correlate with biomass changes.	Low natural earthworm densities may confound results.
Terrestrial microcosm, multispecies assemblages				
Plants: *P. vulgaris* Soil microorganism communities Invertebrates: *Pelodera strongyloides*, Enchytraeidae	Microcosm comprised of bean plants, phytophagous organisms, soil bacteria, fungi, and micro-invertebrates. Assessed endpoints include soil parameters, cellular indicators, indicators at the organism, population and community level	Mothes-Wagner, et al. 1992	The assessment of multiple taxa at different trophic levels utilizing cellular to community endpoints provides a thorough accounting of potential effects.	Methodology requires validation. Ability to use in risk assessment needs to be confirmed.
Aporrectodea tuberculata	Micrososms were filled with field-collected soil, planted with wheat seedlings, and three earthworms were added to each system. Microbial biomass, litter decomposition, enzyme activity, bait lamina tests, nutrient leaching and pesticide degradation measurements were taken periodically, and abundances were quantified at the end.	Edwards, et al. 1998	Treatment regime was designed to mimic natural spray events, adding realism to microcosm studies, and endpoints were selected to give insight concerning ecosystem processes.	Collection and containment of natural soil communities may cause problems with non-homogenous replicate communities.
Aporrectodea trapezoides	This description of earthworm mesocosm methodology includes the use of mesh bags with organic matter buried in 15 cm deep units with one earthworm.	Bogomolov, et al 1996	Toxicity thresholds produced using this methodology was very similar to those produced using other methods, indicating that this method has been	Soil chemistry, including pH, may alter sensitivity of soil invertebrate to plant protection products. Use of a single earthworm could limit statistical analysis.

	Earthworm survival, growth, and body accumulation, as well as organic matter decomposition, substrate induced respiration, soil urease activity and total nematode numbers were measured as endpoints.		verified.	
L. rubellus, Nematoda	Integrated Soil Microcosm approach (using soil cores taken from field), in which three *L. rubellus* were added to each enclosure	Burrows and Edwards 2002	The use of soil cores allows for the collection and analysis of leachate (for plant protection product concentrations and nutrients.	Methods for interpretation and extrapolation of microcosm results for use in risk assessments have not yet been developed.
Soil community	Reference describes methodology for the automated collection of soil core leachate, irrigation, and analysis of CO_2 production.	Hantschel, et al. 1994	Automation streamlines the process of conducting soil microcosm studies.	Unexpected differences detected in CO_2 production of soil core microcosms were not explained.
E. fetida, Enchytraeus albidus	Both laboratory and field studies were conducted to determine the efficacy of using lab experiments to predict impacts in the field. Soil cores were exposed in greenhouse conditions, while field communities were subjected to overspray.	Rombke, et al. 2004	Inhomogeneity of earthworm distribution in field was realistically reflected by the TMEs.	The authors conclude that the abundance and biomass of earthworms are suitable endpoints for assessment of chemicals within TME's but at sites where abundance is low data interpretation may be difficult. Predictability of biomass results derived from TME's is restricted if the number of large earthworms, such as *L. terrestris* or *L. rubellus*, is high.

Collembola, Astigmata, Cryptostigmata, Mesostigmata and Prostigmata	Microarthropod TME: Soil cores were collected from multiple fields, irrigated, acclimated for two to four weeks, and treated with compound. Sampling was conducted at weeks 1, 4, 8, and 16 following exposure.	Koolhaas, et al. 2004	Conclusions in TME mirrored those in the field study and thus predictive value of TME is illustrated.	Large variations in both Collembola and mite communities. Differences in the vegetation in the TMEs in the 4 countries possibly caused variation in soil moisture, which may have affected soil micro-arthropod communities independently of exposure.
Nematodes	Methodology was developed concerning the impact of pesticides on soil-dwelling nematodes as a part of the ring-testing of Terrestrial Model Ecosystems in the EU.	Moser, et al. 2004	Effects on trophic structure and individual populations can be assessed. Validation of TME since field studies showed the same responses to exposure	High variability of data may conceal effects and increase the likelihood of misinterpretation.
Enchytraeidae, earthworms	A Terrestrial Model Ecosystem was used to measure the breakdown of organic matter. The breakdown of cellulose inserted into a soil column or on the soil surface. Faunal feeding was measured with a bait lamina method.	Forster, et al. 2004	Effects on organic matter decomposition were the same in the TME and the field study and showed a dose response relationship.	The feeding activity of the soil fauna showed a large variability.

Soil microbial community	A coupled set of experiments (one laboratory and one field) were conducted to describe the impact of a pesticide on soil microorganisms. Various microbial parameters measured.	Sousa, et al. 2004	Comparisons on data variability also revealed the absence of significant differences between experiments in all parameters in most cases, indicating that TMEs were able to represent the spatial variability found in the field. Measured responses to the model chemical in TMEs were similar to the field study.	Soil moisture lead to some of the variability in microbial parameters.
Nutrient cycling	Terrestrial Model Ecosystem soil columns were used to quantify the impact of plant protection products on soil nutrient cycling. Soil and leachate ammonium and nitrite concentrations were measured following application.	Van Gestel, et al. 2004	Field data showed similar patterns in nutrient levels and thus the TME's predictive value was confirmed.	Variability in moisture or invertebrate activity may confound results. Because soil invertebrates are not homogenously distributed in field soils, columns may contain significantly different community structures or abundances.
Modelling				
Earthworm: *Lumbricus terrestris, Lumbricus rubellus*	This reference describes a mathematical modeling technique to determine the relative sensitivity of different earthworms to pesticides. It incorporates some life history data.	Baveco and de Roos 1996	Model was able to detect differences in species susceptibility based on life history strategy *(L. terrestris* was determined to be to be more sensitive to pesticide effects than *L. rubellus,* most likely as a result of the long duration of its juvenile stage.)	Model has not been validated

Terrestrial Non-target Plants

Organisms/ Taxa	Testing Approach	Literature Reference	"Pros": added value of design	"Cons": drawbacks of design
Species selection				
Plants: *Mimulus ringens* L., *Bidens cernua, S. arvensis, Phaseolus vulgaris, Echinochloa crusgalli*	Two wetland plant species, two terrestrial species, and one species found in both habitats were exposed to 1% and 10% of recommended label rate of metsulfuron methyl to investigate the effects of exposure and determine the most sensitive phenological stages.	Boutin and Rogers 2000	Additional species tested.	Overspray simulation scenario may not be a worse-case exposure scenario.
Plants: *Bellis perennis, Centaura cyanus, Inula helenium, Rudbeckia hirta, Solidago canadensis, Leonorus cardiac, Mentha spicata, Nepeta cataria, Prunella vulgaris, Polygonum convolvulus, Rumex crispus, Anagallis arvensis, Digitalis pupurea, Sinapis arvensis, Papaver rhoeas*	This paper presents methodology in which testing was performed with 15 non-crop plant species sprayed with 6 herbicides (representing various modes-of-action) in a green house setting. Plant species were selected based on prevalence in field margins of Europe and/or North America.	Boutin, et al. 2004	Generally the selected plants were easy to grow and maintain in the greenhouse and they provided a more conservative HC5 value than standard US EPA data.	Testing performed in greenhouses with single species potted individually may not be representative of field situations where plants undergo more adverse conditions (wind, occasional drought, insect damage, competition). It is possible that environmental stressors could increase sensitivity to herbicidal compounds.
Exposure regime and scenario				
Plant: *Cardamine pratensis* L., *Centaurea nigra* L., *Cynosurus cristatus* L., *Galium mollugo* L., *Geum urbanum* L., *Hypericum perforatum* L., *Leontodon hispidus* L., *Lolium perenne* L., *Lotus corniculatus* L., *Lychnis flos-cuculi* L.,	The dose responses of 14 plant species to four herbicides were measured in glasshouse experiments. Exposure mimicked spray drift at field application rates and to predict the distances travelled by	Breeze, et al. 1992	The toxicity data obtained from this study of spaced plants could be used to assess short-term response to herbicide drift.	Effects from competition between plants in the environment are not considered in the design. All plants were treated on the shoot apex, but drifting droplets of pesticide in field situations will land on other surfaces

Ranunculus acris L., *Stachys officinalis* L., *Torilis japonica*, *Trifolium pretense* L.	given doses of herbicide. Results were used to indicate the risk to each species from drift damage.			of the plant, and may cause less damage for the same dose (or greater toxicity may result from a widely distributed dose). These issues decrease realism of methodology.
Controlled versus realistic environmental conditions				
Plants: *Ranunculus repens, Thlaspi arvense, R. crispus, P. rhoeas, Elymus repens, Festuca rubra, Poa trivalis,* among others	This design utilizes adjacent ditch-bank vegetation in an agricultural field to investigate the effects of spraying crop edges. The presence and abundance of plant species in adjacent ditch-bank vegetation were compared along sprayed and unsprayed crop edges in the same fields.	de Snoo and van der Poll 1999	Methodology may generate useful data and statistical conclusions due to the consistent use of pairwise comparison of ditch banks of the same field.	High variability inherent in natural populations may still confound results.

4.0 References

Abrantes, N., R. Pereira, et al. (2008). "Evaluation of the Ecotoxicological Impact of the Pesticide Lasso (R) on Non-target Freshwater Species, Through Leaching from Nearby Agricultural Fields, Using Terrestrial Model Ecosystems." Water, air and soil pollution **192**(1-4): 211-220.

Addison, J. A. and S. B. Holmes (1995). "Comparison of forest soil microcosm and acute toxicity studies for determining effects of fenitrothion on earthworms." Ecotoxicology and Environmental Safety **30**(2): 127-133.

ASTM (1991). Standard guide for conducting a terrestrial soil-core microcosm test <Document Title> Annual book of ASTM standards, section ll: Water and environmental technology. Vol. 11.04: Pesticides; resource recovery; hazardous substances and soil spill responses; waste disposal; biological effects, ASTM <Publisher Location> Philadelphia.

Barrett, K. L., N. Grandy, et al. (1994). Guidance Document on Regulatory Testing procedures for pesticides and Non-Target Arthropods Wageningen, The Netherlands, SETAC.

Barry, M. J. and D. C. Logan (1998). "The use of temporary pond microcosms for aquatic toxicity testing: direct and indirect effects of endosulfan on community structure. ." Aquat.Toxicol. **41**(1-2): 101-24.

Baveco, J. M. and A. M. de Roos (1996). "Assessing the impact of pesticides on lumbricid populations: an individual-based modelling approach. ." J.Appl.Ecol. **33**(6): 1451-1468.

BBA (1993). Guidelines for testing plant protection products in the authorisation procedure. 25-1: Testing of baits, granules and treated seeds for hazards to birds - acceptance tests. . , Biologische Bundesanstalt für Land- und Forstwirtschaft, Berlin, Germany.

BBA (1994). Guidelines for testing of Plant Protetion products within Registration Part IV. Department for plant protection products and application techniques. Braunschweig, Germany.

Beketov, M. A., R. B. Schafer, et al. (2008). "Long-term stream invertebrate community alterations induced by the insecticide thiacloprid: Effect concentrations and recovery dynamics." Science of the Total Environment **405**: 96-108.

Bluemel, S., S. Aldershof, et al., Eds. (2000). Guidance document to detect side effects of plant protection products on predatory mites (Acari: Phytoseiidae) under field conditions: vineyards and orchards. Guidelines to evaluate side-effects of plant protection products to non-target arthropods, IOBC/WPRS.

Bluemel, S., F. M. Bakker, et al. (2000). Laboratory residual contact test with the predatory mite *Typhlodromus pyri* Scheuten (Acari; Phytoseiidae) for regulatory testing of plant protection products. Guidelines to evaluate side effects of plant protection products to non-target arthropods. . M. P. Candolfi, S. Blümel, R. Forsteret al. Dijon, IOBC.

Blus, L. J. and C. J. Henny (1997). "Field studies on pesticides and birds: unexpected and unique relations." Ecological Applications 7(4): 1125-1132.

Bogomolov, D. M., S. K. Chen, et al. (1996). "An ecosystem approach to soil toxicity testing: a study of copper contamination in laboratory soil microcosms." Applied soil ecology : a section of Agriculture, Ecosystems & Environment **4**(2): 95-105.

Bouldin, J. L., J. L. Farris, et al. (2005). "Evaluated fate and effects of atrazine and lambda-cyhalothrin in vegetated and unvegetated microcosms." Environmental toxicology **20**(5): 487-98.

Boutin, C., N. Elmegaard, et al. (2004). "Toxicity testing of fifteen non-crop plant species with six herbicides in a greenhouse experiment : implications for risk assessment." Ecotoxicology **13**(4): 349-369.

Boutin, C., H.-B. Lee, et al. (2000). "Effects of the sulfonylurea herbicide metsulfuron methyl on growth and reproduction of five wetland and terrestrial plant species." Environmental Toxicology and Chemistry **19**(10): 2532 - 2541.

Boxall, A. B. A., C. D. Brown, et al. (2002). "Higher-tier laboratory methods for assessing the aquatic toxicity of pesticides. ." Pest Manage.Sci. **58**(7): 637-648.

Brasse, D., W. Muehlen, et al. (2003). "Honey bee brood ring-test 2002: Method for the assessment of side effects of plant protection products on the honey bee (Ap s mellifera) brood under semi-field conditions (tunnel test)
<Original> Ringversuch 2002: Methode zur Einschaetzung der Auswirkungen von Pflanzenschutzmitteln auf Bienenbrut (Apis mellifera) unter Halbfreilandbedingungen (Zelt-Versuch); Test de l'anneau 2002 : m thode pour valuer les effets des produits phytosanitaires sur le couvain d'abeilles (Apis mellifera) en condition de semi-plein champ (essai en tunnel)." Apidologie **34**(5): 481-482.

Breeze, V., G. Thomas, et al. (1992). "Use of a Model and Toxicity Data to Predict the Risks to Some Wild Plant Species from Drift of Four Herbicides. ." Ann.Appl.Biol. **121**(3): 669-677.

Breneman, D. H. and K. W. Pontasch (1992). "Stream microcosm toxicity tests: predicting the effects of fenvalerate on riffle insect communities. ." Environ.Toxicol.Chem. **13**: 59.

Brown, K. (1998). Field studies with pesticides and non-target arthropods in orchards
Ecotoxicology: pesticides and beneficial organisms. Ecotoxicology: pesticides and beneficial organisms, Dordrecht London Kluwer Academic 1998**:** 137-149.

Brown, K. and H. Vogt (2004). Non-target arthropod field studies : asking the right questions for their purpose. Bulletin OILB/SROP, International Organization for Biological and Integrated Control of Noxious Animals and Plants (OIBC/OILB), West Palaearctic Regional Section (WPRS/SROP <Publisher Location> Dijon**:** 107.
Brown, K. C. and M. Miles (2006). "How much precision does a regulatory field study need?" Pesticides and Beneficial Organisms IOBC/wprs Bulletin **29**(10): 43-52.

Brown, K. C., M. J. Miles, et al. (2008). An ecological study of the use of vineyards by birds in Southern France. SETAC Europe: 18th Annual Meeting. Warsaw.

Brown, R. A., L. C. McMullin, et al. (1990). Beneficial arthropod toxicity assessments with three insecticides in laboratory, semi-field and field studies Proceedings of the Brighton Crop Protection Conference, Pests and Diseases Volume 2, British Crop Protection Council Publications<Publisher Location> Thornton Heath.

Burdett, A. S., M. M. Stevens, et al. (2001). "Laboratory and field studies on the effect of molinate, clomazone, and thiobencarb on nontarget aquatic invertebrates." Environmental toxicology and chemistry / SETAC **20**(10): 2229-2236.

Burrows, L. A. and C. A. Edwards (2002). "The use of integrated soil microcosms to predict effects of pesticides on soil ecosystems." European Journal of Soil Biology **38**(3-4): 245-249.

Campbell, P. J., D. J. S. Arnold, et al. (1999). Guidance Document: Higher-tier Aquatic Risk Assessment for Pesticides (HARAP). h. a. L. O. From the SETAC-Europe/OECD/EC Workshop, France, 19-22 April 1998.

Campbell, P. J., K. C. Brown, et al. (2000). "A hazard quotient approach for assessing the risk to non-target arthropods from plant protection products under 91/414: hazard quotient trigger value proposal and validation ment." J. Pest Science **73**(5): 117-124.

Candolfi, M., K. L. Barrett, et al. (2001). Guidance Document on Regulatory Testing and Risk Asessment Procedures for Plant Protection Products with Non-target Arthropods. ESCORT 2 (European standard characteristics of non-target arthropod regulatory testing), Wageningen, Netherlands, SETAC.

Candolfi, M., F. Bigler, et al. (2000). "Principles for regulatory testing and interpretation of semi - field and field studies with non-target arthropods." Anzeiger fur Schadlingskunde **73**(6): 141-147.

Candolfi, M. P., S. Blümel, et al. (2000). Guidelines to evaluate side effects of plant protection products to non-target arthropods., IOBC/OILB.

Caquet, T., L. Lagadic, et al. (1996). "Outdoor experimental ponds (Mesocosms) designed for long-term ecotoxicological studies in aquatic environment." Ecotoxicology and environmental safety **34**(2): 125-133.

Carter, N. (1993). "Field trials to study the within-season effects of pesticides on beneficial arthropods in cereals in summer." Bulletin OEPP **23**(4): 709-712.

Chang, K. H., M. Sakamoto, et al. (2005). "Impact of pesticide application on zooplankton communities with different densities of invertebrate predators: an experimental analysis using small-scale mesocosms." Aquatic Toxicology **72**(4): 373-82.

Choo, L. P. D. and G. H. Baker (1998). "Influence of four commonly used pesticides on the survival, growth, and reproduction of the earthworm Aporrectodea trapezoides (Lumbricidae)." Australian Journal of Agricultural Research **49**(8): 1297-1303.

Cilgi, T. (1994). "Selecting arthropod "indicator species" for environmental impact assessment of pesticides in field studies." Aspects of applied biology 1994(**37**): 131-140.

Cole, J. F. H. and L. Canning (1993). Rationale for the choice of species in the regulatory testing of the effects of pesticides on terrestrial non-target plants. British Crop Protection

Conference, Brighton, British Crop Protection Council (BCPC) Farnham (United Kingdom).
Colin, M. E., J. M. Bonmatin, et al. (2004). "A Method to Quantify and Analyze the Foraging Activity of Honey Bees: Relevance to the Sublethal Effects Induced by Systemic Insecticides." Archives of Environmental Contamination and Toxicology 47(3): 387-395.
Coors, A., J. Kuckelkorn, et al. (2006). "Application of in-situ bioassays with macrophytes in aquatic mesocosm studies." Ecotoxicology **15**(7): 583-591.
Crane, M. (1997). "Research needs for predictive multispecies tests in aquatic toxicology." Hydrobiologia 346(<ASFA> CS9722816): 149-155.
Crone, E. E., M. Marler, et al. (2009). "Non-target effects of broadleaf herbicide on a native perennial forb: a demographic framework for assessing and minimizing impacts." Journal of Applied Ecology **46**: 673 - 682.
Daam, M. A. and P. J. v. d. Brink (2003). Effects of three pesticides that differ in mode of action on the ecology of small indoor aquatic microcosms : an evaluation of the effects of the insecticide chlorpyrifos, the herbicide atrazine and the fungicide carbendazim, Alterra, Green World Research Wageningen (Netherlands).
De Jong, F. M. W., F. Bakker, et al. (2009). Guidance for summarising and evaluating non-target arthropod field studies. SETAC Europe: 19th Annual Meeting, Gothenburg, Sweden.
De Jong, F. M. W., M. H. M. M. Montforts, et al. (2009). Guidance for summarising and evaluating eathworm field studies. SETAC Europe, Gothenburg, SETAC.
de Snoo, G. R. and R. J. van der Poll (1999). "Effect of herbicide drift on adjacent boundary vegetation." Agriculture, Ecosystems & Environment **73**: 1 - 6.
Delabie, J. (1984). Study under controlled conditions of the effects of the synthetic pyrethroid, PP321, on the bee. Bures sur Yvettte, France, INRA-CNRS, SOPRA Internal Report.
Dixon, F. L., D. V. Clay, et al. (2005). "The tolerance of young trees to applications of clopyralid alone and in mixture with foliar-acting herbicides." Forestry **78**(4): 353 - 364.
Duso, C. (1994). "Comparison between field and laboratory testing methods to evaluate the pesticide side-effects on the predatory mites Amblyseius andersoni and Typhlodromus pyri <Original> [Comparaison entre des methodes de tests en champ et au laboratoire pour evaluer les effets secondaires des pesticides sur les acariens predateurs Amblyseius andersoni et Typhlodromus pyri] Working group "Pesticides and beneficial organisms". Side-effects of pesticides on beneficial organisms: comparison of laboratory, semi-field and field results <Original> Groupe de travail "Pesticides et organismes utiles". [Effets secondaires des pesticides sur les auxiliaires: comparaison des resultats au laboratoire, en semi-plein champ et en plein champ]." Bulletin OILB SROP (France) **17**(10): 7-19.
Edwards, C. A., T. Knacker, et al. (1998). The prediction of the fate and effects of pesticides in the environment using tiered laboratory soil microcosms <Document Title> Brighton Crop Protection Conference: Pests & Diseases - 1998: Volume 1: Proceedings of an International Conference, Brighton, UK, 16-19 November 1998. Brighton Crop Protection Conference: Pests & Diseases - 1998: Volume 1: Proceedings of an

International Conference, Brighton, UK, 16-19 November 1998, British Crop Protection Council <Publisher Location> Farnham**:** 267-272.

EFSA (2005). "Opinion of the Scientific Panel on Plant Health, Plant Protection products and their Residues on a request from EFSA related to the assessment of the acute and chronic risk to aquatic organisms with regard to the possibility of lowering the uncertainty factor if additional species were tested. ." The EFSA journal **301**: 1-45.

Eisenbeiss, G., R. Lenz, et al. (1999). "Organic residue decomposition: the minicontainer system: a multifunctional tool in decomposition studies." Environ. Scoi. Pollut. Res. **6**: 220-224.

ELINK (2008). "Draft guidance from the Workshop on Linking Aquatic Exposure and Effects in the Registration Procedure of Plant Protection Products (ELINK)".

Fairbrother, A. (1993). Ecotoxicological Principles for Avian Field Studies Using Radiotelemetry or Remote Sensing <NOTE> Symposium paper. Pellston Workshop: Radiotelemetry for avian field studies**:** 10.

Fairchild, J. F., T. W. La Point, et al. (1992). Population-, community- and ecosystem-level responses of aquatic mesocosms to pulsed doses of a pyrethroid insecticide.

Forster, B., C. A. M. Van Gestel, et al. (2004). "Ring-testing and Field-validation of a Terrestrial Model Ecosystem (TME) - An Instrument for Testing Potentially Harmful Substances: Effects of Carbendazim on Organic Matter Breakdown and Soil Fauna Feeding Activity." Ecotoxicology **13**(1-2): 129-141.

Forsyth, D. J., P. A. Martin, et al. (1997). "Effects of herbicides on two submerged aquatic macrophytes, *Potamogeton Pectinatus* L. and *Myriophyllum sibiricum* Komarov, in a prairie wetland." Environmental Pollution **95**(2): 259-268.

Frampton, G. K. (1994). "Sampling to detect effects of pesticides on epigeal Collembola (springtails)." Aspects of Applied Biology **37**: 121-130.

Freier, B., S. Kuehne, et al. (2001). "Field study on effects of insecticide applications in wheat on the arthropod community of field boundaries. ." Mitt.Biol.Bundesanst.LandForstwirtsch. **383**: 82-87.

Fryday, S. L., S. A. Chandler-Morris, et al. (2001). "Robust tests for the acceptance of bait and treated seed by birds." Central Science Laboratory Contract **PN0909**.

Fryday, S. L., S. A. Chandler-Morris, et al. (1999). "Validation of methods for testing the avoidance of treated seeds by birds." Central Science Laboratory Contract **PN0914**.

Ganio, L. M. (1994). Considerations for Planning Terrestrial Field Studies. Workshop on radio-telemetry in support of avian/pesticide field studies. Covallis**:** 22.

Giddings, J. M., T. C. M. Brock, et al. (2002). Community-level aquatic system studies-interpretation criteria (CLASSIC). Pensacola (FL): SETAC. 44p. SETAC.

Giddings, J. M., K. R. Solomon, et al. (2001). "Probabilistic risk assessment of cotton pyrethroids: II. Aquatic mesocosm and field studies." Environmental Toxicology and Chemistry **20**(3): 660-668.

Grade, R., J. Gonzalez-Valero, et al. (2000). "(A "Higher Tier" flow-through system in a steady-state for the assessment of toxicity of herbicides to the green algae Selenastrum

capricornutum) <Orig> Ein "Higher Tier" Durchfluss-System im Fliessgleichgewicht zur Bestimmung der Toxizitaet von Pflanzenschutzmitteln gegenueber der Gruenalge Selenastrum capricornutum." Sci. Total Environ. **247**(2-3): 355-361.

Graney, R. K., JH; Rodgers, JH (eds)| (1994). Aquatic mesocosm studies in ecological risk assessment, CRC PRESS/LEWIS PUBLISHERS, BOCA RATON, FL 33431 (USA), 1994,.

Gregg Morris, R. K. J. H. J. P. C. H. F. E. and R. L. K. J. H. R. J. H. Graney (1994). Pyrethroid Insecticide Effects on Bluegill Sunfish in Microcosms and Mesocosms and Bluegill Impact on Microcosm Fauna Aquatic mesocosm studies in ecological risk assessment. Setac Special Publications Series. Arlington, VA, Boca Raton Lewis Publishers 1994.

Gretenkord, C. and W. Drechser (1997). Laborzucht der Dunklen Erdhummel *Bombus terrestris* und toxicologische Untersuchungen unter Labor- und Halbfreiland- versuchen, Rheinischen Friedrich-Wilhelms- Universitat Bonn, Germany. Dissertation.

Gyorffyne Molnar, J. and A. Polgar (1994). "Effects of pesticides on the predatory mite Typhlodromus pyri Scheuten. A comparison of field and laboratory results <Original> [Effet des pesticides sur l'acarien predateur Typhlodromus pyri Scheuten. Comparaison des resultats au champ et au laboratoire] Working group "Pesticides and beneficial organisms". Side-effects of pesticides on beneficial organisms: comparison of laboratory, semi-field and field results <Original> Groupe de travail "Pesticides et organismes utiles". [Effets secondaires des pesticides sur les auxiliaires: comparaison des resultats au laboratoire, en semi-plein champ et en plein champ]." Bulletin OILB SROP (France) **17**(10): 21-26.

Hanson, M. L., P. K. Sibley, et al. (2003). "Field level evaluation and risk assessment of the toxicity of dichloroacetic acid to the aquatic macrophytes Lemna gibba, Myriophyllum spicatum, and Myriophyllum sibiricum." Ecotoxicology and Environmental Safety **55**(1): 46-63.

Hantschel, R. E., H. Flessa, et al. (1994). "An automated microscosm system for studying soil ecological processes." Soil Science Society of America Journal **58**(2): 401-404.

Harrison, A. F., P. M. Latter, et al. (1988). Cotton strip assay: An index of decomposition in soils. Grange over Sands, Cumbria, U.K., Institute of Terrestrial Ecology, ITE, **:** 176p.

Hart, A. D. M. and H. M. Thompson (2005). "Improved approaches to assessing long-term risk to birds and mammals." Ecotoxicology 14.

Harte, J. (1984). An evaluation of freshwater aquatic microcosms.

Hassan, S. A. (1992). "Guidelines for testing the effects of pesticides on beneficial organisms: description of test methods " Bulletin IOBC/WPRS **15**(3).

Heckmann, L.-H. and N. Friberg (2005). "Macroinvertebrate community response to pulse exposure with the insecticide lambda-cyhalothrin using in-stream mesocosms." Environmental toxicology and chemistry / SETAC **24**(3): 582-590.

Heimbach, F. (1992). Correlation between data from laboratory and field tests for investigating the toxicity of pesticides to earthworms. International Society of Soil Science, Avignon, France.

Heimbach, F. (1993). Comparison of results from laboratory and field studies for a hazard assessment of earthworms. 14th Annu. meetingof SETAC - Ecological Risk Assesssment: lessons Learned?, Houston.

Heimbach, F., J. Berndt, et al. (1994). Fate and Biological Effects of an Herbicide on Two Artificial Pond Ecosystems of Different Size Aquatic mesocosm studies in ecological risk assessment. Setac Special Publications Series. Arlington VA, Boca Raton Lewis Publishers 1994.

Heimbach, U., W. Buechs, et al. (1992). "A semi-field method close to field conditions to test effects of pesticides on Poecilus cupreus L. (Coleoptera, Carabidae) <Original> [Une methode au champ ("semi-field") proche des conditions de champ pour evaluer les effets secondaires des pesticides sur Poecilus cupreus L. (Coleoptera, Carabidae)] Guidelines for testing the effects of pesticides on beneficial organisms: description of test methods. Meeting of the working group "Pesticides and beneficial organisms", University of Southampton, UK, September 1991 <Original> [Directives pour l'evaluation des effets des pesticides sur les organismes utiles: description des methodes de test. Reunion du groupe de travail "Pesticides et organismes utiles", Universite de Southampton, Royaume-Uni, septembre 1991]

Hickey, G. L., P. S. Craig, et al. (2009). Species non-exchangeability for ecotoxicological risk assessment. SETAC Europe: 19th Annual meeting 21 May - 4 June 2009. Goteborg, Sweden.

Hose, G. C. and P. J. Van den Brink (2004). "Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm , and field data." Archives of environmental contamination and toxicology **47**: (4).

Houx, N. W. H., A. Dekker, et al. (1996). "Acute toxicity test for terrestrial hazard assessment with exposure of Folsomia candida to pesticides in an aqueous medium." Archives of Environmental Contamination and Toxicology **30**(1).

Howick, G. L. D. F. G. J. M. G. R. L. (1994). Earthen Ponds vs. Fiberglass Tanks as Venues For Assessing the Impact of Pesticides on Aquatic Environments: A Parallel Study With Sulprofos Aquatic mesocosm studies in ecological risk assessment. Setac Special Publications Series. Arllington VA, Boca Raton Lewis Publishers 1994.

Inglesfield, C. (1989). A long-term field study to investigate the effects of alpha-cypermethrin on predatory and parasitic arthropods. Mededelingen van de Faculteit Landbouwwetenschappen, Rijksuniversiteit Gent. Gent. **54:** 895-904.

Inglesfield, C. (1990). Demonstrating the safety of alphacypermethrin to bees: results of a large-scale field monitoring study in the UK <Document Title> Proceedings of the Fourth International Symposium on the Harmonization of Methods for Testing the Toxicity of Pesticides to Bees, May 15-18, 1990, Rez near Prague, Czechoslovakia. Proceedings of

the Fourth International Symposium on the Harmonization of Methods for Testing the Toxicity of Pesticides to Bees, May 15-18, 1990, Rez near Prague, Czechoslovakia, Research Institute of Apiculture <Publisher Location> Dol**:** 46-47.

INRA (1990). Methodologie d'acceptation de nourriture ou de semences traitees par une substance repulsive, par des oiseaux en conditions de captivite. , Laboratoire de phytopharmacie unite d-eco-etho-toxicologie

ISO (1998). "ISO 11268-2. Soil quality - Effects of pollutants on Earthworms (Eisenia fetida) - Part 2: Determination of effects on reproduction." International Standard.

ISO (1999). ISO 11268-3. Soil Quality - Effects of Pollutants on Earthworms - Part 3: Guidance on the determination of effects in field situations. Switzerland, International Organization for Standardization (ISO).

Jagers op Akkerhuis, G. A. J. M., C. Kjaer, et al. (1999). "Temperature-Dependent, time-dose-effect model for pesticide effects on growing, herbivorouos arthropods: bioassays with dimethoate and cypermethrin." Environmental Toxicology and Chemistry **18**(10): 2370-2378.

Jansch, S., G. K. Frampton, et al. (2006). "Effects of pesticides on soil invertebrates in model ecosystem and field studies : a review and comparison with laboratory toxicity data." Environmental Toxicology and Chemistry **25**(9): 2490-2501.

Jansen, J. P. (2000). "A three-year field study on the short-term effects of insecticides used to control cereal aphids on plant-dwelling aphid predators in winter wheat." Pest Management Science **56**(6): 533-539.

Jepson, P. and M. Mead-Briggs (1992). "A discussion of methods used in semi-field studies to evaluate pesticide toxicity to beneficial invertebrates." Bulletin OILB/SROP 15(3): 4-17.

Johnson, P. C., Kennedy J. H., Gregg Morris R., Hambleton F. E. (1994). Fate and Effects of Cyfluthrin (Pyrethroid Insecticide) in Pond Mesocosms and Concrete Microcosms Aquatic mesocosm studies in ecological risk assessment. Setac Special Publications Series. Arlington VA, Boca Raton Lewis Publishers 1994.

Johnston, G., A. Dawson, et al. (1996). "Effects of prochloraz and malathion on the red-legged partridge : a semi-natural field study." Environmental pollution : (1987) **91**(2): 217-225.

Juettner, I., A. Peither, et al. (1995). "An outdoor mesocosm study to assess ecotoxicological effects of atrazine on a natural plankton community. ." Arch.Environ.Contam.Toxicol. **29**(4): 435-441.

Kalsch, W., T. Junker, et al. (2006). "A chronic plant test for the assessment of contaminated soils. Part 1: method development." JSS - J Soils & Sediments 6(1): 37 - 45.

Kjaer, C., M. Strandberg, et al. (2006). "Effects on hawthorn the year after simulated spray drift." Chemosphere **63**: 853 - 859.

Knacker, T., C. A. M. Van Gestel, et al. (2004). "Ring testing and field-validation of a terrestrial model Ecosystem 9TME)- an istrument fro testing potenytialy harmful substances: Conceptual approach and study design." Ecotoxicology **13**: 9-27.

Koolhaas, J. E., C. A. M. Van Gestel, et al. (2004). "Ring-Testing and Field-validation of a Terrestrial Model Ecosystem (TME) - An Instrument for Testing Potentially Harmful Substances: Effects of Carbendazim on Soil Microarthropod Communities." Ecotoxicology **13**(1-2): 75-88.

Kula, K. and S. Guske (2001). "Minutes of a meeting on the requirement of data according to Annex III, point 10.6.2 (November 27/28 2000) " Braunschweig, BBA. 9 p Unpublished report.

Landner, L., H. Blanck, et al. (1989). Community testing , microcosm and mesocosm experiments : Ecotoxicological tools with high ecological realism. Chemicals in the aquatic environment: advanced hazard assessment., . L. Landner, Springer**:** Pg 216-254

Langhof, M., A. Gathmann, et al. (2003). "Field study on the effects of insecticide drift on non-target arthropods : residual toxicity, effect on populations and recolonisation processes." Bulletin OILB/SROP **26**(5): 73-82.

Leeuwangh, P., T. C. Brock, et al. (1994). "An evaluation of four types of freshwater model ecosystem for assessing the hazard of pesticides." Human & experimental toxicology **13**(12): 888-99.

Lewis, G. B., H. J. Gough, et al. (1990). "A Comparison of a Laboratory Residual Toxicity Test with a Semi-Field 'Tunnel' Trial to Assess the Effects of Pesticide Residues on Honey Bees. ." Proc.Br.Crop Prot.Conf.Pests Dis. **3**: 981-986.

Leymann, B., W. Muehlen, et al. (2000). "A semi field test to evaluate the side effects of pesticides on brood in honeybee colonies (Apis mellifera L.). ." Bulletin IOBC/WPRS **23**(3): 3.

Liebig, M., G. Schmidt, et al. (2008). "Direct and indirect effects of pollutants on algae and algivorous ciliates in an aquatic indoor microcosm." Aquatic Toxicology **88**(2): 102-110.

Liess, M. and P. C. Von Der Ohe (2005). "Analyzing effects of pesticides on invertebrate communities in streams." Environmental Toxicology and Chemistry **24**(4): 954-965.

Maltby, L., N. Blake, et al. (2005). "Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems." Environmental Toxicology and Chemistry 24(2): 379-388.

Matthiessen, P. (1994). Guidelines for pesticide hazard assessment with freshwater field tests at the international level, <Full> LEWIS PUBLISHERS, BOCA RATON, FL (USA).

Maycock, D. S., M. M. Prenner, et al. (2003). "Incorporation of in situ and biomarker assays in higher-tier assessment of the aquatic toxicity of insecticides." Water Research (Oxford) **37**(17): 4180-4190.

Mead-Briggs, M., K. Brown, et al. (2000). A laboratory test for evaluating the effects of plant protection products on the parasitic wasp, *Aphidius rhopalosiphi* (DeStephani-Perez) (Hymenoptera: Braconidae). Guidelines to evaluate side effects of plant protection products to non-target arthropods.. M. P. Candolfi, S. Blümel, R. Forsteret al. Dijon, IOBC.

Mead-Briggs, M., M. Moll, et al. (2009). "An extended laboratory test for evaluating the effects of plant protection products on the parasitic wasp, *Aphidius rhopalosophi* (Se Stefani-Perez) (Hymenoptera, Braconidae)." Biocontrol in press.

Meissle, M. and L. Andreas| (2005). "Comparing methods to evaluate the effects of Bt maize and insecticide on spider assemblages." Agriculture, ecosystems & environment **107**(4): 359-370.

Miles, M. and R. Dutton (2003). Testing the effects of spinosad to predatory mites in laboratory, extended laboratory, semi - field and field studies. Bulletin OILB/SROP, International Organization for Biological and Integrated Control of Noxious Animals and Plants (OIBC/OILB), West Palaearctic Regional Section (WPRS/SROP) <Publisher Location> Dijon**:** 9-20.

Mineau, P. (2002). "Estimating the probability of bird mortality from pesticide sprays on the basis of the field study record. ." Environ.Toxicol.Chem. **21**(7): 1497-1506.

Moser, T., H. J. Schallnass, et al. (2004). "Ring-testing and Field-validation of a Terrestrial Model Ecosystem (TME) - An Instrument for Testing Potentially Harmful Substances: Effects of Carbendazim on Nematodes." Ecotoxicology **13**(1-2): 61-74.

Mothes-Wagner, U., H. K. Reitze, et al. (1992). "Terrestrial multispecies toxicity testing [of pesticides]. 1. Description of the multispecies assemblage." Chemosphere **24**(11): 1653-1667.

Muhlen, W. (1996). Implication of the IGR Alsystin WP 25 on the development of honeybee colonies underfield and semi-field conditions. Procedings of 6th ICP-BR International Symposium on Hazards of Pesticides, Braunschweig, Germany.

Nagel, R. F., H. Fromm, et al. (1995). "The influence of earthworms and soil mesofauna on the C and N mineralization in agricultural soils: a microcosm study." Acta Zool Fennica 196: 22-26.

Naranjo, S. E., G. Head, et al. (2005). "Field Studies Assessing Arthropod Nontarget Effects in Bt Transgenic Crops: Introduction Comparing methods to evaluate the effects of Bt maize and insecticide on spider assemblages." Environmental Entomology 34(5): 1178-1180.

OECD (1984). "OECD Guideline for Testing of Chemicals "Earthworm, Acute Toxicity Tests"."

OECD (2006). "OECD Guideline for Testing of Chemicals. Terrestrial Plant Test: Vegetative Vigour Test."

OECD (2006). "OECD Guidelines for the Testing of Chemicals. Terrestrial Plant Test: Seedling Emergence and Seedling Growth Test."

OECD (2006). "OECD Series on Testing and Assessment Number 53. Guidance document on simulated freshwater lentic field tests (outdoor microcosms and mesocosms). ." ENV/JM/MONO(2006)17.

OEPP/EPPO (1992). "Decision making scheme for the environmental risk assessment of plant protection products, Honeybees." EPPO/OEPP Bulletin 10: 151-159.

Oomen, P. A., A. d. Ruijter, et al. (1992). "Method for honeybee brood feeding tests with insect growth-regulating insecticides." Bulletin OEPP 22(4): 613-616.

Peither, A., I. Juttner, et al. (1996). "A pond mesocosm study to determine direct and indirect effects of lindane on a natural zooplankton community." Environmental Pollution 93(1): 49-56.

Poche, R. M., F. D. L., et al. (1993). Use of radiotelemetry to monitor survival of song-birds exposed to an insecticide on golf courses. Avian radiotelemetry in support of pesticide field studies. Pacific Grove CA, Pensacola, FL Society of Environmental Toxicology and Chemistry 1998.

Pontasch, K. W. and J. Cairns, Jr. (1989). "Establishing and maintaining laboratory-based microcosms of riffle insect communities: their potential for multispecies toxicity tests." Hydrobiologia **175**(1): 49-60.

Pratt, J. R. and R. Barreiro (1998). "Influence of trophic status on the toxic effects of a herbicide: a microcosm study. ." Arch.Environ.Contam.Toxicol **35**(3): 404-411.

PSD and HSE (1995). Guideline to study the within-season effects of insecticides on non-target terrestrial arthropods in cereals in summer, HSE.

Pullen, A. J., P. C. Jepson, et al. (1992). "Terrestrial non-target invertebrates and the autumn application of synthetic pyrethroids: experimental methodology and the trade off between replication and plot size. ." Archives of Environmental Contamination and Toxicology, **23**: 246-258.

Rand, G. M. (2004). "Fate and effects of the insecticide - miticide chlorfenapyr in outdoor aquatic microcosms." Ecotoxicology and environmental safety **58**(1): 50-60.

Reboulet, J. N. (1994). Impact des produits phytosanitaires sur la faune auxiliare; methodologie d'experimentation en verger. Paris, Association de Coordination Technique Agricole.

Reet, K., Viik Eneli, et al. (2007). "Impact of alpha-cypermethrin on honey bees foraging on spring oilseed rape (Brassica napus) flowers in field conditions." Pest Manage.Sci. **63**(11): 1085-1089.

Reinecke, S. A. and A. J. Reinecke (2007). "Biomarker response and biomass change of earthworms exposed to chlorpyrifos in microcosms." Ecotoxicology and environmental safety : (Print) **66**(1): 92-101.

Roessink, I., G. H. P. Arts, et al. (2005). "Effects of lambda-cyhalothrin in two ditch microcosm systems of different trophic status." Environmental Toxicology and Chemistry **24**(7): 1684-1696.

Rombke, J., F. Heimbach, et al. (2002). Effects of Plant Protection Products on Functional Endpoints in Soil. EPFES. SETAC. Lisboa.

Rombke, J., C. A. M. Van Gestel, et al. (2004). "Ring-Testing and Field-Validation of a Terrestrial Model Ecosystem (TME) - An Instrument for Testing Potentially Harmful Substances: Effects of Carbendazim on Earthworms." Ecotoxicology **13**(1-2): 105-118.

SANCO (2000). Guidance document on Risk Assessment for Birds and Mammals Under Council Directive 91/414/EEC, European Commission - Health and Consumer Protection Directorate General.

SANCO (2002). Guidance Document on Terrestrial Ecotoxicology Under Council Directive 91/414/EEC. SANCO/10329/2002 rev 2 final 17 October 2002, European Commission.

SANCO (2002). Working document: Guidance document on Aquatic Ecotoxicology in the context of the Directive 91/414/EEC (Sanco/3268/2001 rev.4 (final).

Schmidt, H. W., D. Brasse, et al. (2003). "Introduction of indices for the evaluation of tent tests and field tests with honey bees." Bulletin of Insectology **56**(1): 111-117.

Schulz, R., G. Thiere, et al. (2002). "A combined microcosm and field approach to evaluate the aquatic toxicity of azinphosmethyl to stream communities." Environmental Toxicology and Chemistry 21(10): 2172-2178.

SETAC (1992). Guidance document on testing procedures for pesticides in freshwater mesocosms. Society of Environmental toxicology and Chemistry - Europe, 46pp.

SETAC (1992). Proceedings of a workshop on aquatic microcosms for ecological assessment of pesticides. (Wintergreen, Virginia, US. October 1991). SETAC Foundation for Environmental Education and the RESOLVE Program of the World Wildlife Fund, 56 pp.

Shaw, J. L. and J. P. Manning (1996). "Evaluating macroinvertebrate population and community level effects in outdoor microcosms: Use of in situ bioassays and multivariate analysis." Environmental Toxicology and Chemistry **15**(5): 608-617.

Sherratt, T. N. and P. C. Jepson (1993). "A Metapopulation Approach To Modelling the Long-Term Impact of Pesticides on Invertebrates. ." J.Appl.Ecol. **30**(4): 696-705.

Shillabeer, N., D. V. Smyth, et al. (2000). Higher tier risk assessment of agrochemicals, incorporating sediment into algal test systems. The BCPC Conference: Pests and diseases, Volume 1 Proceedings of an international conference held at the Brighton Hilton Metropole Hotel, Brighton, UK, 13-16 November 2000, British Crop Protection Council <Publisher Location> Farnham**:** 359-364.

Sibly, R. M., H. R. Akcakaya, et al. (2005). "Population -level assessment of risks of pesticides to birds and mammals in the UK." Ecotoxicology **14**(8): 863-876.

Sousa, J. P., J. M. L. Rodrigues, et al. (2004). "Ring-Testing and Field-validation of a Terrestrial Model Ecosystem (TME) - An Instrument for Testing Potentially Harmful Substances: Effects of Carbendazim on Soil Microbial Parameters." Ecotoxicology **13**(1-2): 43-60.

Stark, J. D., P. C. Jepson, et al. (1995). "Limitations to use of topical toxicity data for predictions of pesticide side effects in the field." Journal of Economic Entomology **88**(5): 1081-1088.

Steen, J. M., van der and A. Dinter (2007). "A monitoring study to assess the acute mortality effects of indoxacarb on honey bees (Apis mellifera L.) in flowering apple orchards." Pest Management Science **63**: 1095-1099.

Steen, J. v. d., C. Gretenkord, et al. (1996). Methods to determine the acute oral LD50 and acute contact LD50 of pesticides for bumble beees (Bombus terrestris L.). 6th Int. Symp. on the hazard of pesticides to bees (ICPBR), , Braunschweig, Germany.

Steen, J. v. d. and A. d. Ruijter (1990). Test methods to determine the hazards of IGRs (insect growth regulators) on honey bee brood, using phenoxycarb as an example Proceedings of the Fourth International Symposium on the Harmonization of Methods for Testing the Toxicity of Pesticides to Bees, May 15-18, 1990, Rez near Prague, Czechoslovakia, Rez, Near Prague, Czechoslovakia, Research Institute of Apiculture_<Publisher Location> Dol.

Sterk, G., P. Creemers, et al. (1994). "Testing the side effects of pesticides on the predatory mite Typhlodromus pyri (Acari, Phytoseiidae) in field trials <Original> [Tests sur les effets secondaires des pesticides sur l'acarien predateur Typhlodromus pyri (Acari, Phytoseiidae), essais en champ] Working group "Pesticides and beneficial organisms". Side-effects of pesticides on beneficial organisms: comparison of laboratory, semi-field and field results <Original> Groupe de travail "Pesticides et organismes utiles". [Effets secondaires des pesticides sur les auxiliaires: comparaison des resultats au laboratoire, en semi-plein champ et en plein champ]." Bulletin OILB SROP (France) **17**(10): 27-40.

Sugiura, K. (1992). "A multispecies laboratory microcosm for screening ecotoxicological impacts of chemicals." Environmental Toxicology and Chemistry **11**(9): 1217-1226.

Thacker, J. R. M. and P. C. J. P (1993). "Pesticide Risk Assessment and Non-Target Invertebrates: Integrating Population Depletion, Population Recovery, and Experimental Design. ." Bull.Environ.Contam.Toxicol. **51**(4): 523-531.

Tiebout, H. M. and K. E. Brugger (1995). "Ecological risk assessment of pesticides for terrestrial vertebrates: evaluation and application of the US environmental protection agency's quotient model." Conservation Biology **9**(6): 1605-1618.

Topping, C. J. and P. Odderskaer (2004). "Modeling the Influence of Temporal and Spatial Factors on the Assessment of Impacts of Pesticides on Skylarks." Environmental Toxicology and Chemistry **23**(2): 509-521.

Topping, C. J., R. M. Sibly, et al. (2005). "Risk assessment of UK skylark populations using life-history and individual-based landscape models." Ecotoxicology **14**(8): 925-936.

Torne, E. v. (1990). "Assessiing feeding activities of soil living animals." Pedobiologia **34**: 89-101.

Tornier, I. and W. Drescher (1990). Development of a semifield testing method for the evaluation of pesticides with hoverflies (Dipt., Syrphidae)<Document Title> Proceedings of the Fourth International Symposium on the Harmonization of Methods for Testing the Toxicity of Pesticides to Bees, May 15-18, 1990, Rez near Prague, Czechoslovakia. Proceedings of the Fourth International Symposium on the Harmonization of Methods for Testing the Toxicity of Pesticides to Bees, May 15-18, 1990, Rez near Prague, Czechoslovakia, Research Institute of Apiculture<Publisher Location> Dol**:** 131-135.

Tornier, I. and W. Drescher (1992). "A semi-field testing method for the evaluation of pesticides with the hoverfly Episyrphus balteatus (Dipt., Syrphidae) <Original> [Une methode de test au champ ("semi-field") pour l'evaluation des effets secondaires des pesticides sur Episyrphus balteatus (Dipt., Syrphidae)] Guidelines for testing the effects of pesticides on beneficial organisms: description of test methods. Meeting of the working group "Pesticides and beneficial organisms", University of Southampton, UK, September 1991 <Original> [Directives pour l'evaluation des effets des pesticides sur les organismes utiles: description des methodes de test. Reunion du groupe de travail "Pesticides et organismes utiles", Universite de Southampton, Royaume-Uni, septembre 1991]." Bulletin OILB SROP (France) **15**(3): 166-170.

Touart, L. W. (1988). Aquatic mesocosm tests to support pesticide registrations. Hazard Evaluation Division Technical Guidance Document. U.S. Environmental Protection Agency, Washington, DC. Report no. EPA/540/09-88-035. U. EPA.

Traas, T. P., J. H. Janse, et al. (2004). "A freshwater food web model for the combined effects of nutrients and insecticide stress and subsequent recovery." Environmental toxicology and chemistry **23**(2): 521-529.

Traunspurger, W., H. Schaefer, et al. (1996). "Comparative investigation on the effect of a herbicide on aquatic organisms in single species tests and aquatic microcosms. Chemosphere **33**(6): 1129-1141.

van den Brink, P., N. Blake, et al. (2006). "Predictive value of Species Sensitivity Distributions for Effects of Herbicides in Freshwater Ecosytems." Human and Ecological Risk Assessment **12**: 645-674.

Van Der Steen Jozef J M (2001). "Review of the methods to determine the hazard and toxicity of pesticides to bumblebees." Apidologie **32**(5): 399-406.

Van Der Steen Jozef J M and Dinter Axel (2007). "A monitoring study to assess the acute mortality effects of indoxacarb on honey bees (Apis mellifera L.) in flowering apple orchards." Pest Management Science **63**: 1095-1099.

Van Gestel, C. A. M., J. E. Koolhaas, et al. (2004). "Ring-Testing and Field-Validation of a Terrestrial Model Ecosystem (TME) - An Instrument for Testing Potentially Harmful Substances: Effects of Carbendazim on Nutrient Cycling." Ecotoxicology **13**(1-2): 119-128.

Van Wijngaarden R P A, Van Den Brink P J, et al. (1996). "Effects of the insecticide Dursban (R) 4E (active ingredient chlorpyrifos) in outdoor experimental ditches: I. Comparison of short-term toxicity between the laboratory and the field." Environmental toxicology and chemistry **15**(7): 1133-1142.

Van Wijngaarden, R. P., T. Brock, et al. (2005). "Effects of chlorpyrifos in freshwater model ecosystems: the influence of experimental conditions on ecotoxicological thresholds." Pest Management Science **61**(10): 923-935.

Van Wijngaarden, R. P. A., J. G. M. Cuppen, et al. (2004). "Aquatic risk assessment of a realistic exposure to pesticides used in bulb crops: a microcosm study." Environmental Toxicology and Chemistry **23**(6): 1479-1498.

Vogt, H., S. Rumpf, et al. (1992). "A field method for testing effects of pesticides on larvae of the green lacewing Chrysoperla carnea Steph. (Neuroptera, Chrysopidae) <Original> [Methode en champ pour tester les effets des pesticides sur la chrysope Chrysoperla carnea Steph. (Neuroptera, Chrysopidae)] Guidelines for testing the effects of pesticides on beneficial organisms: description of test methods. Meeting of the working group "Pesticides and beneficial organisms", University of Southampton, UK, September 1991 <Original> [Directives pour l'evaluation des effets des pesticides sur les organismes utiles: description des methodes de test. Reunion du groupe de travail "Pesticides et organismes utiles", Universite de Southampton, Royaume-Uni, septembre 1991]." Bulletiin OILB SROP (France) **15**(3): 176-182.

Webber, C. E., W. G. Deutsch, et al. (1992). Ecosystem-level testing of a synthetic pyrethroid insecticide in aquatic mesocosms Aquatic mesocosms in ecotoxicology, Toronto, ON, Canada, October 28-November 2, 1989.

Weber, D., M. Dorgerloh, et al. (2009). Time-variable exposure effects on green algae under flow-through conditions: Modelling and experimental results. SETAC Europe: 19th Annual meeting 31 May - 4 June 2009 Goteborg, Sweden.

Weyers, A., B. Sokull-Kluettgen, et al. (2004). "Use of Terrestrial Model Ecosystem Data in Environmental Risk Assessment for Industrial Chemicals , Biocides and Plant Protection Products in the EU." Ecotoxicology **13**(1-2): 163-176.

White, A. L. and C. Boutin (2007). "Herbicidal effects on nontarget vegetation: investigating the limitations of current pesticide registration guidelines." Environmental Toxicology and Chemistry **26**(12): 2634 - 2643.

White, J. S., C. J. Everett, et al. (1990). Assessing the effects of lambda-cyhalothrin on natural enemies using laboratory, small plot and field studies <Original> [Etude en laboratoire, en petite parcelle et en champ, des effets de la lambda-cyhalothrine sur les ennemis naturels] [International colloquium: Relations between phytosanitary treatments and animal reproduction, 25-26 April 1990, Paris] <Original> Colloque international: Relations entre les traitements phytosanitaires et la reproduction des animaux, 25-26 Avril 1990, Paris A long-term field study to investigate the effects of alpha- cypermethrin on predatory and parasitic arthropods. Relations entre les traitments phytosanitaires et la reproduction des animaux Published in Annales ANPP (France). Paris, ANPP Paris (France). **2:** p 53-60.

Wick, M. and B. Freier (2000). "Long-term effects of an insecticide application on non-target arthropods in winter wheat - a field study over 2 seasons. ." Anz.Schaedlingskd.Pflanz.Umweltschutz **73**(3): 61-69.

Wiles, J. A. and G. K. Frampton (1996). "A field bioassay approach to assess the toxicity of insecticide residues on soil to Collembola." Pesticide science **47**(3): 273-285.

Williams, R. R., B. Dehdashti, et al. (1992). "Performance of an aquatic multispecies system in evaluating the effects of a model microbial pest control agent on nontarget organisms." Journal of Toxicology and Environmental Health **37**(2): 247-264.

Wolf, C., M. Riffel, et al. (2009). Chlorpyrifos, a widely used OP-insecticide: A multiple field study program assessing the risk to wild birds. SETAC Europe: 19th Annual Meeting. Goteborg.

Appendix 1: Literature search protocols and number of items retrieved for the DIALOG databases

Set	Items	Description
S1	8550333	(TIER??? OR FIELD OR SEMI-FIELD? OR SEMIFIELD? OR SEMI()FIELD? OR CAGE? ? OR TUNNEL? ? OR MESO-COSM? OR MESOCOSM? OR MODEL? OR POPULATION? ? OR MICROCOSM? OR MICRO-COSM?)/TI,DE,ID,AB
S2	1684778	S1 AND (AQUATIC? OR BEE OR BEES OR APIS OR MELLIFER? OR MELIFER? OR BOMBUS OR EARTHWORM? OR EARTH-WORM? OR LUMBRICID? OR ARTHROPOD? OR BIRD OR BIRDS OR AVE OR AVES OR MAMMAL? OR FISH OR FISHES OR PISCES OR INVERTEBRAT? OR ALGA? OR SEDIMENT? OR SOIL?)/TI,DE,ID,AB
S3	262041	S2 AND (PLANT? ?(4N)PROTECT? OR PESTICID? OR HERBICID? OR ACARI? OR INSECT? OR FUNGICID? OR MITICID? OR PLANT? ?(4N)GROW?)/TI,DE,ID,AB
S4	101776	S3 AND (PLANT? ?(4N)PROTECT? OR PESTICID? OR HERBICID? OR ACARICID? OR INSECTICID? OR FUNGICID? OR MITICID? OR PLANT? ?(4N)GROW?)/TI,DE,ID
S5	47436	S4 AND (TIER??? OR FIELD OR SEMI-FIELD? OR SEMIFIELD? OR SEMI()FIELD? OR CAGE? ? OR TUNNEL? ? OR MESO-COSM? OR MESOCOSM? OR MODEL? OR POPULATION? OR MICROCOSM? OR MICRO-COSM?)/TI,DE,ID
S6	40850	S5 AND (AQUATIC? OR BEE OR BEES OR APIS OR MELLIFER? OR MELIFER? OR BOMBUS OR EARTHWORM? OR EARTH-WORM? OR LUMBRICID? OR ARTHROPOD? OR BIRD OR BIRDS OR AVE OR AVES OR MAMMAL? OR FISH OR FISHES OR PISCES OR INVERTEBRAT? OR ALGA? OR SEDIMENT? OR SOIL?)/TI,DE,ID
S7	19865	S6 AND (TEST? ? OR TESTING OR STUDY? OR STUDIES OR STUDIED OR APPROACH? OR MODEL? OR TIER???)/TI,DE,ID
S8	2208	S7 AND (HIGHER(2W)TIER? OR FIELD?(2W)(STUDY? OR STUDIE?) OR SEMI-FIELD? OR SEMIFIELD?)/TI,DE,ID
S9	1757	S8/ENG
S10	1377	RD (unique items)
S11	**1377**	**Sort S10/ALL/PY,D**
S12	855	S7 AND (CAGE? ? OR TUNNEL? OR MESO-COSM? OR MESOCOSM? OR MICRO-COSM? OR MICROCOSM?)/TI,DE,ID
S13	797	S12 NOT S8
S14	750	S13/ENG
S15	356	RD (unique items)
S16	**356**	**Sort S15/ALL/PY,D**
S17	9193	S6 AND (RESEARCH OR PROCEDURE? OR METHOD? OR PREDICT? OR ASSESS?)/TI,DE,ID
S18	5916	S7 AND (RESEARCH OR PROCEDURE? OR METHOD? OR PREDICT? OR ASSESS?)/TI,DE,ID
S19	4290	S18 AND (RESEARCH OR PROCEDURE? OR METHOD? OR PREDICT? OR ASSESS?)/DE,ID
S20	1971	S19 AND (ECOTOX? OR TOXIC? OR ENVIRONMENT? OR RISK?)/TI,DE,ID
S21	1678	S20 NOT (S8 OR S12)
S22	1423	S21/ENG
S23	1060	RD (unique items)

S24	**1060**	**Sort S23/ALL/PY,D**
S25	2753037	(NON-STANDARD OR NONSTANDARD OR SPECIES OR MULTI-SPECIES OR MULTISPECIES OR MULTIPLE(2W)SPECIES OR TIME(2W)EVENT? ? OR SENSITIVE?(4N)LIFE?)/TI,DE,ID,AB
S26	1891097	((ARTIFICIAL? OR MANMADE? OR MAN-MADE OR CONSTRUCTED)(2W) STREAM? ? OR DITCH? OR ECOSYSTEM? OR ECO-SYSTEM? OR SENSITIVIT? OR MESO-COSM? OR MESOCOSM? OR MICRO-COSM? OR MICROCOSM? OR RECOVERY OR SEMI-FIELD? OR SEMIFIELD? OR SEMI()FIELD? OR FIELD?(2W)(STUDY? OR STUDIE?) OR HIGHER(2W)TIER?)/TI,DE,ID,AB
S27	659324	(S25 OR S26) AND (ECOTOX? OR TOXIC? OR ENVIRONMENT? OR RISK?)/TI,DE,ID
S28	147331	(S25 OR S26)(6N)(ECOTOX? OR TOXIC? OR ENVIRONMENT? OR RISK?) /TI,DE,ID
S29	50300	S28 AND (SPECIES OR TIER? OR MULTI-SPECIES OR MULTISPECIES OR ENSITIV? OR STREAM? OR DITCH? OR ECOSYSTEM? OR ECO-SYSTEM? OR MESOCOSM? OR MESO-COSM? OR MICROCOSM? OR MICRO-COSM? OR RECOVERY OR SEMI-FIELD? OR SEMIFIELD? OR FIELD?)/TI
S30	23082	S29 AND (ECOTOX? OR TOXIC? OR ENVIRONMENT? OR RISK? OR PLANT? ?)/TI
S31	3121	S30 AND (EXPERIMENT? OR MODEL? OR TEST? ? OR TESTING OR APPROACH?)/TI
S32	2995	S31 NOT (S8 OR S12 OR S20)
S33	2820	S32/ENG
S34	1316	RD (unique items)
S35	399	S34 AND (PLANT? ? OR PROTECT? OR CHEMICAL? ? OR PESTICID? OR HERBICID? OR ACARI? OR INSECT? OR FUNGICID? OR MITICID?)/TI,DE,ID
S36	**399**	**Sort S35/ALL/PY,D**

Appendix 2: Literature search protocols and number of items retrieved for the STN databases

Results for Search Question:

non-standard or nonstandard or species or multi-species or multispecies or time or event or stream* or ditch* or ecosystem* or eco-system* or sensitivit* or meso-cosm* or mesocosm* or micro-cosm* or microcosm* or recovery or semi-field* or semifield* or "Semi field" or field stud* or higher tier* or tier* AND eco-tox* or ecotox* or environ* or risk* or toxic* AND title: eco-tox* or ecotox* or environ* or risk* or toxic* AND title: non-standard or nonstandard or species or multi-species or multispecies or time or event or stream* or ditch* or ecosystem* or eco-system* or sensitivit* or meso-cosm* or mesocosm* or micro-cosm* or microcosm* or recovery or semi-field* or semifield* or "Semi field" or field stud* or higher tier* or tier* AND language: english AND title: test* or study* or studie* or approach* or experiment* or model* AND pubyear: 1990-current

30 answers in CAplus (Food & Agriculture focus)

70 answers in CROPU

100 total hits

Results for Search Question:

tier* or higher or field or semi-field* or semifield or "semi field" or cage* or tunnel* or meso-cosm* or mesocosm* or model* or population* or micro-cosm* or microcosm* AND aquatic* or bee or bees or honeybee* or honey-bee* or "honey bee" or apis or mellifera* or melifera* or bombus or bumble-bee* or bumblebee* or earth-worm* or earthworm* or lumbricid* or arthropod* or bird or birds or aves or ave or avian or mammal* or fish or fishes or pisces or invertebrat* or alga* or sediment* or soil* AND (plant* and protect*) or pesticid* or herbic* or acari* or insect* or fungic* or (plant* and grow*) or miticid* AND title: tier* or higher or field or semi-field* or semifield or "semi field" or cage* or tunnel* or meso-cosm* or mesocosm* or model* or population* or micro-cosm* or microcosm* AND title: (plant* and protect*) or pesticid* or herbic* or acari* or insect* or fungic* or (plant* and grow*) or miticid* AND title: aquatic* or bee or bees or honeybee* or honey-bee* or "honey bee" or apis or mellifera* or melifera* or bombus or bumble-bee* or bumblebee* or earth-worm* or earthworm* or

lumbricid* or arthropod* or bird or birds or aves or ave or avian or mammal* or fish or fishes or pisces or invertebrat* or alga* or sediment* or soil* AND language: english AND pubyear: 1990-current AND title: test* or study* or studie* or model* or higher or trial*

141 answers in CAplus (Food & Agriculture focus)

69 answers in CROPU

210 total hits

Appendix 3. Databases searched

The databases used in the search and their details are summarised below.

DIALOG DATABASES:

File 50:CAB Abstracts 1972-2009/Jan W3
(c) 2009 CAB International
File 10:AGRICOLA 70-2009/Jan
(c) format only 2009 Dialog
File 203:AGRIS 1974-2009/Dec
Dist by NAL, Intl Copr. All rights reserved
File 6:NTIS 1964-2009/Feb W1
(c) 2009 NTIS, Intl Cpyrght All Rights Res
File 66:GPO Mon. Cat. 1978-2008/Dec
(c) format only 2008 Dialog
File 156:ToxFile 1965-2008/Nov W2
(c) format only 2008 Dialog
File 65:Inside Conferences 1993-2009/Jan 26
(c) 2009 BLDSC all rts. reserv.
File 144:Pascal 1973-2009/Jan W2
(c) 2009 INIST/CNRS
File 143:Biol. & Agric. Index 1983-2009/Dec
(c) 2009 The HW Wilson Co
File 24:CSA Life Sciences Abstracts 1966-2009/Mar
(c) 2009 CSA.
File 40:Enviroline(R) 1975-2008/May
(c) 2008 Congressional Information Service
File 76:Environmental Sciences 1966-2009/Mar
(c) 2009 CSA.
File 44:Aquatic Science & Fisheries Abstracts 1966-2009/Feb
(c) 2009 CSA.

STN DATABASES:

CAPLUS [Chemical Abstracts; 1907-present]

CROPU [Derwent Crop Protection File]

1. DIALOG
(online database aggregator and service provider)

2. STN
(online database aggregator and service provider)

3. TOXNET
(series of databases that are Internet-searchable for free and maintained by the U.S. National Library of Medicine)

SELECTED DIALOG DATABASES THAT WILL USED:

CAB ABSTRACTS is a comprehensive file of applied life science information containing all records in the 44 abstract journals published by CAB International (CABI), plus many more records which appear online only.

CABI has long been recognized as a leading scientific information service in agriculture and related sciences. Of particular note are sections in the database comprehensively covering literature in the fields of veterinary medicine, human nutrition, horticulture, forestry, leisure, recreation, and tourism.

More than 9,000 serial journals in more than 50 languages are scanned, as well as books, reports, and other publications. About 225,000 items per year are selected for inclusion in CAB Abstracts.

CAB ABSTRACTS covers every branch of the applied life sciences, including:

- Agricultural biotechnology
- Agricultural economics & rural sociology
- Agricultural engineering
- Animal health & veterinary medicine
- Animal production & genetics
- Biodeterioration & biodegradation
- Crop production
- Crop protection
- Dairy science
- Environmental degradation, conservation, & amelioration
- Forestry
- Genetic resources
- Horticulture
- Human nutrition & diet-related disorders
- Human parasitic diseases
- Leisure, recreation, & tourism
- Plant breeding & genetics
- Postharvest science
- Soil science
- Sugar industry
- Rural development

As one of the most comprehensive sources of U.S. agricultural and life sciences information, the **AGRICOLA** (AGRICultural OnLine Access) database serves as the catalog and index to the collections of the National Agricultural Library (NAL) and the research of the U.S. Department of Agriculture (USDA). **AGRICOLA** has been available online since 1970 and contains more than 4.1 million citations to journal articles, book chapters, monographs, theses, patents, software, audiovisual materials, and technical reports related to agriculture. The database contains thousands of records with links to online full-text documents.

AGRICOLA encompasses all aspects of agriculture and allied disciplines, including animal and veterinary sciences, entomology, plant sciences, forestry, aquaculture and fisheries, farming and farming systems, agricultural economics, extension and education, food and human nutrition,

agricultural engineering and technology, and earth and environmental sciences. The NAL Agricultural Thesaurus (NALT) and Library of Congress Subject Headings (LCSH) serve as the controlled vocabularies for indexing and cataloging records. This extensive database has been maintained since 1970 to provide selective worldwide coverage of primary information sources in agriculture and related fields.

AGRIS International is the international information system for agricultural sciences and technology. The AGRIS International database serves as a comprehensive inventory of worldwide agricultural literature which reflects research results, food production, and rural development to help users identify problems involved in all aspects of world food supply. Emphasis in AGRIS International is non-U.S. This file corresponds in part to the printed publication, *Agrindex*, published monthly by the Food and Agriculture Organization (FAO) of the United Nations.

AGRIS is a cooperative, decentralized system in which over 100 national and multinational centers take part. It collects and makes available current information on the agricultural literature of the world appearing in journals, books, reports, and conference papers. Each country which participates in AGRIS does so by submitting information about documents published within its own territories. All contributing sources are of non-U.S. origin. FAO acts as a coordinating agency within this global information system, facilitating the exchange of agricultural information to its member countries.

AGRIS International includes coverage of the following main subject groups:

- Administration and Legislation
- Animal Production
- Aquatic Sciences and Fisheries
- Economics, Development, and Rural Sociology

- Education, Extension, and Advisory Work
- Food Science
- Forestry
- General Agriculture
- Geography and History
- Home Economics
- Human Nutrition
- Machinery and Buildings
- Natural Resources
- Pollution
- Plant Production
- Protection of Plants and Stored Products
- Transgenics

The **NTIS: National Technical Information Service** database comprises summaries of U.S. government-sponsored research, development, and engineering, plus analyses prepared by federal agencies, their contractors, or grantees. It is the means through which unclassified, publicly available, unlimited distribution reports are made available for sale from agencies such as NASA, DOD, DOE, HUD, DOT, Department of Commerce, and some 240 other agencies. Additionally, some state and local government agencies contribute summaries of their reports to the database. NTIS also provides access to the results of government-sponsored research and development from countries outside the U.S.

The **GPO Monthly Catalog** (Government Printing Office) is the machine-readable equivalent of the printed *Monthly Catalog of United States Government Publications*. It contains records of reports, studies, fact sheets, maps, handbooks, conference proceedings, etc., issued by all U.S. federal government agencies, including the U.S. Congress. Also included in this database are records of all of the Senate and House hearings on private and public bills and laws. The GPO Monthly Catalog contains a wealth of information on a wide range of topics including

agriculture, economics, energy research, public policy, tax reform, business law, health, and many other subjects.

All agencies of the U.S. Federal Government are sources of government publications. Congressional publications include bills, hearings, reports, documents from the committees and subcommittees, and laws from Congress acting as a body. Executive-branch publications include Presidential statements as well as agency annual reports, general informational and operational reports, long-established statistical series, technical reports, maps, administrative regulations, treaties, and periodicals. Other sources are the independently established agencies and government corporations and the various boards and committees whose functions are not strictly limited to the internal operation of a parent department or agency, e.g., the Water Resources Council.

ToxFile covers the toxicological, pharmacological, biochemical, and physiological effects of drugs and other chemicals: adverse drug reactions, chemically induced diseases, carcinogenesis, mutagenesis, teratogenesis, environmental pollution, waste disposal, radiation, and food contamination are typical areas of coverage. ToxFile includes toxicology records derived from MEDLINE (). These are journal citations related to toxicology, also called TOXBIB (or TOXLINE Core) records by the National Library of Medicine (NLM).

Inside Conferences is produced by the *British Library*. The database contains details of all papers given at every congress, symposium, conference, exposition, workshop, and meeting received at the British Library Document Supply Centre (BLDSC) since October 1993. Each year over 16,000 proceedings are indexed, covering a wide range of subjects published as serials or monographs. Over 500,000 bibliographic citations for individual conference papers will be

added annually. Most records are in English, with many languages represented in the source documents.

PASCAL is produced by the Institut de l'Information Scientifique et Technique (INIST) of the French National Research Council (CNRS). It provides access to the world's scientific and technical literature and includes about 450,000 new citations per year. Available in machine-readable form since 1973, PASCAL corresponds to the print publication *Bibliographie internationale* (previously *Bulletin signaletique*).

Each citation includes the article's original title, and, in most cases, a French translated title; for material since 1973, an English translated title is also provided. Most abstracts are in French. Analyzed documents come from all over the world, in 100 different languages.

PASCAL is multidisciplinary, covering the core of the world's scientific and technical literature. The principal subject areas covered are: the fundamental disciplines of physics and chemistry; life sciences (including biology, medicine, and psychology); applied sciences and technology; earth sciences; and information sciences.
In addition, a number of fields are covered exhaustively, often in cooperation with a variety of specialized research organizations. These fields are: energy; metals and metallurgy; building and public works; earth sciences; biotechnology; fundamental and applied zoology of invertebrates; agricultural sciences (specifically plant production); tropical medicine; and information science documentation.

Biological & Agricultural Index provides thorough, reliable indexing of 258 periodicals common to most libraries. Periodical coverage includes a wide range of scientific journals, from

popular to professional, that pertain to biology and agriculture. About 45% of the focus is on agriculture. Types of materials indexed include feature articles, biographical sketches, reports of symposia and conferences, review articles, abstracts and summaries of papers, selected letters to the editor, special issues or monographic supplements, and book reviews.

Indexed by a staff of librarians and subject specialists with expertise in the field, **Biological & Agricultural Index** covers a broad range of subjects, including:

- Agricultural Chemicals
- Agriculture
- Animal Husbandry
- Biochemistry
- Biology
- Biotechnology
- Botany
- Cytology
- Ecology
- Entomology
- Environmental Science
- Fishery Sciences
- Food Science
- Forestry
- Genetics
- Horticulture
- Limnology
- Microbiology
- Neuroscience
- Nutrition
- Physiology
- Plant Pathology
- Soil Science
- Veterinary Medicine
- Zoology

CSA Life Sciences Abstracts contains abstracts and bibliographic citations from recent worldwide research literature in major areas of biology, medicine, biochemistry, biotechnology, genetics, immunology, ecology, and microbiology, and some aspects of agriculture and veterinary science. **CSA Life Sciences Abstracts** is produced by CSA and corresponds to print series of more than 20 abstracting journals.
Informative abstracts are included for about 90% of the records.

Enviroline® covers the world's environmental related information. It provides indexing and abstracting coverage of more than 1,000 international primary and secondary publications reporting on all aspects of the environment. These publications highlight such fields as management, technology, planning, law, political science, economics, geology, biology, and chemistry as they relate to environmental issues.

Environmental Sciences contains abstracts and bibliographic citations providing comprehensive coverage of the environmental sciences. The research areas range from agricultural biotechnology and air quality to waste management and water resource issues. Abstracts and citations are drawn from over 6,000 serials including scientific journals, conference proceedings, reports, monographs, books and government publications.

Overwhelmingly cited by a majority of aquatic science librarians as their primary database, the **ASFA (Aquatic Sciences and Fisheries Abstracts)** series is the premier reference in the field of aquatic resources. Input to ASFA is provided by a growing international network of information centers monitoring more than 5,000 serial publications, books, reports, conference proceedings, translations, and limited distribution literature. ASFA is a component of the Aquatic Sciences and Fisheries Information System (ASFIS), formed by four United Nations agency sponsors of ASFA and a network of international and national partners.

SELECTED STN DATABASES THAT WERE USED:

Chemical Abstracts Plus is the most current and most comprehensive chemistry bibliographic database available from Chemical Abstracts Service (CAS). CAplus covers international journals, patents, patent families, technical disclosures, technical reports, books, conference proceedings, dissertations, electronic-only journals, and web preprints from all areas of chemistry, biochemistry, chemical engineering, and related sciences from 1967 to the present.

Bibliographic information and available abstracts for the articles from nearly 1,500 key chemical journals are added within one week of journal receipt. CAplus is updated daily with new bibliographic records and weekly with indexing.

CROPU, the Derwent Crop Protection File, provides references to the worldwide literature on all aspects of pesticides, including their use in crop protection and pest control. The database offers a unique combination of biological and chemical information. Besides bibliographic data, CROPU provides full English-language abstracts detailing the latest developments and applications in crop protection, as well as precise indexing and coding, developed specifically for the crop protection industry.
CROPU covers 1985 to 2003.

SUBJECT COVERAGE

All aspects of pesticides: - Analysis - Biochemistry - Chemistry - Toxicology - Insecticides	- Herbicides - Fungicides - Molluscides - Rodenticides - Biological pest control