ALTERRA
WAGENINGEN UR

Aquatic effect assessment for plant protection products

Dutch proposal that addresses the requirements of the Plant Protection Product Regulation and Water Framework Directive

Alterra Report 2235
ISSN 1566-7197

T.C.M. Brock, G.H.P. Arts, T.E.M. ten Hulscher, F.M.W. de Jong, R. Luttik, E.W.M. Roex, C.E. Smit and P.J.M. van Vliet

ALTERRA WAGENINGEN UR
Rijkswaterstaat *Ministry of Infrastructure and the Environment*
Deltares Enabling Delta Life
ctgb
National Institute for Public Health and the Environment *Ministry of Health, Welfare and Sport*

Aquatic effect assessment for plant protection products

Commissioned by the Dutch Ministries of Economic Affairs, Agriculture & Innovation (project BO-12.07-002-001) and of Infrastructure & Environment (project M/601720)

Aquatic effect assessment for plant protection products

Dutch proposal that addresses the requirements of the Plant Protection Product Regulation and Water Framework Directive

T.C.M. Brock[1], G.H.P. Arts[1], T.E.M. ten Hulscher[2], F.M.W. de Jong[3], R. Luttik[3], E.W.M. Roex[4], C.E. Smit[3] and P.J.M. van Vliet[5]

1 Alterra, part of Wageningen UR
2 Rijkswaterstaat - Ministry of Infrastructure and the Environment
3 National Institute for Public Health and the Environment - Ministry of Health, Welfare and Sport
4 Deltares
5 Ctgb - Board for the Authorisation of Plant Protection Products and Biocides

Alterra Report 2235

Alterra, part of Wageningen UR
Wageningen, 2011

Abstract

Brock, T.C.M., G.H.P. Arts, T.E.M. ten Hulscher, F.M.W. de Jong, R. Luttik R, E.W.M. Roex, C.E. Smit and P.J.M. van Vliet, 2011. *Aquatic effect assessment for plant protection products; Dutch proposal that addresses the requirements of the Plant Protection Product Regulation and Water Framework Directive*. Wageningen, Alterra, Alterra Report 2235. 140 pp.; 10 fig.; 6 decision schemes; 29 tab.; 183 ref.

In this report new proposals for the aquatic effects assessment of plant protection products (pesticides) in the Netherlands are described for edge-of-field surface waters (drainage ditches) falling under the domain of the Plant Protection Product Regulation (pre-registration) and for water bodies falling under the domain of the Water Framework Directive (post-registration). These methods are developed on request of two Dutch ministries (Ministry of Economic Affairs, Agriculture and Innovation; Ministry of Infrastructure and Environment). They are based on specific protection goals proposed by the responsible risk managers of the Dutch ministries, the current European aquatic risk assessment procedures for plant protection products, state-of-the-art knowledge on the ecotoxicology of these chemicals and different aims/claims of the Plant Protection Product Regulation (1107/2009/EC) and the Water Framework Directive (2000/60/EC).

Keywords: Pesticides; Water organisms; Ecological risks; Ecotoxicology; Regulation 1107/2009/EC; Water Framework Directive.

ISSN 1566-7197

The pdf file is free of charge and can be downloaded via the website www.alterra.wur.nl (go to Alterra reports). Alterra does not deliver printed versions of the Alterra reports. Printed versions can be ordered via the external distributor. For ordering have a look at www.rapportbestellen.nl.

Alterra Report 2235
Wageningen, November 2011

Contents

Preface

Chemical monitoring programmes (see www.bestrijdingsmiddelenatlas.nl) revealed that in a large number of surface waters of the Netherlands measured exposure concentrations of certain plant protection products (pesticides) exceeded Dutch water quality standards. This might have been attributed to flaws in the registration procedure or EQS (Environmental Quality Standard) derivation used in the past to assess aquatic risks of plant protection products, but also might have been caused by differences in effect assessment methods used between the registration procedure and the derivation of water quality standards. Responsible risk managers of the Dutch Ministries of Economic Affairs, Agriculture & Innovation and of Infrastructure & Environment requested the authors of this report to update the aquatic effect assessment procedures for plant protection products (PPPs) by taking into account the requirements laid down in European legislation, with reference to PPP registration procedures under Regulation 1107/2009/EC (EC, 2009) and environmental quality standard derivation in line with requirements of Directive 2000/60/EC (Water Framework Directive; EC, 2000).

A project was started to develop decision trees for aquatic organisms to be used in the pre-registration and post-registration environmental risk assessment procedures of PPPs in the Netherlands. In this report we refer to PPP Regulation to indicate both the new Regulation 1107/2009/EC and the Annexes of Directive 91/414/EEC which are still in force.

The core of the approach is that risk assessments are performed at two places in the water system, viz.: (1) in edge-of-field surface water and (2) further downstream in WFD surface water. In smaller edge-of-field surface waters (e.g. drainage ditches) pre-registration criteria of the PPP Regulation apply, whilst in larger water bodies (officially assigned as WFD water bodies) the standards derived according to the WFD methodology apply. Post-registration verification of the exposure concentrations in the WFD water bodies against WFD water quality standards will take place using measurements. If results of chemical monitoring programmes indicate exceeding of EQS values for a specific compound which can be attributed to the current 'GAP' (good agricultural practice), this may have consequences for its authorisation (post-registration risk assessment procedure) and/or adequate mitigation measures have to be implemented.
Within the Dutch project described above four working groups were initiated, viz.:

1. *Exposure assessment working group* to further develop scenarios and exposure models for the pre-registration exposure prediction of PPPs in Dutch drainage ditches (see Tiktak et al., 2012)
2. *Effects assessment working group* to further develop decision trees for (a) the pre-registration effects assessment of predicted exposures of PPPs in Dutch drainage ditches and (b) the derivation of WFD water quality standards for PPPs that will be used in the post-registration risk assessment procedure
3. *Monitoring working group* to provide guidance for the interpretation of chemical monitoring data of PPPs in Dutch surface waters with respect to possible consequences for the authorisation of PPPs (see De Werd and Kruijne, 2011)
4. *Multiple-stress working group* to evaluate whether the risk assessment procedure based on individual PPPs is sufficiently protective for exposure to different PPPs used in crop protection programmes (e.g. for crops like potatoes and fruit)

In this report the decision trees for aquatic organisms in Dutch drainage ditches and WFD water bodies as proposed by the *Effects assessment working group* are presented.

Beleidssamenvatting

Korte samenvatting

Dit rapport presenteert een Nederlands voorstel van een Beslisboom Water voor de effectbeoordeling van Gewasbeschermingsmiddelen in oppervlaktewater in het kader van de pre- en post-registratie beoordeling in Nederland. De Beslisboom Water bestaat uit twee onderdelen: een beslisboom voor de kavelsloot in lijn met de Europese Gewasbeschermingsmiddelenverordening (1107/2009/EC) en de Annexes onder richtlijn 91/414/EEC (toelatingsbeleid), en een beslisboom voor grotere oppervlaktewateren in lijn met de Europese Kaderrichtlijn Water (2000/60/EC), die het bereiken en behouden van een goede chemische en ecologische toestand van Europees oppervlaktewater regelt. In lijn met de Europese toelatingsprocedure voor gewas-beschermingsmiddelen en de toestandsbeoordeling volgens de KRW wordt een onderscheid gemaakt tussen risico's voor korte- en risico's voor lange-termijn blootstelling.

De beslisboom voor de kavelsloot volgt een getrapte benadering, waarbij elke volgende trap gekenmerkt wordt door meer gegevens, meer realisme en minder onzekerheden. In feite is deze beslisboom een stelsel van beoordelingschema's (Figuur A). De beoordelingschema's leiden tot wetenschappelijk onderbouwde RACs (Regulatory Acceptable Concentrations) voor korte- en lange-termijn blootstelling. Deze RACs worden vergeleken met de bijbehorende PECs (Predicted Environmental Concentrations). Dit leidt tot een uitspraak over de acceptatie van het risico (wel/niet acceptabel). Daarbij zijn de opties: (i) geen aantoonbaar ecologisch effect, en (ii) met kortdurend effect gevolgd door herstel, beide uitgewerkt.

In de beslisboom voor de grotere oppervlaktewateren worden op basis van de gegevens uit het dossier en eventuele aanvullende gegevens uit de literatuur de normen afgeleid voor langdurige blootstelling en voor kortdurende piekblootstelling. (respectievelijk de jaargemiddelde milieukwaliteitsnorm, JG-MKN, en de maxi-maal aanvaardbare concentratie, MAC-MKN). Beide normen hebben betrekking op een concentratie waarbij geen effecten optreden. In de post-registratie periode kunnen meetgegevens in KRW-wateren worden gebruikt om te beoordelen of de toegelaten toepassing leidt tot overschrijding van de milieukwaliteitsnormen voor water.

Uitgebreide samenvatting

Bij de implementatie van de Kaderrichtlijn Water in Nederland hebben de toenmalige departementen van LNV, VROM en V&W (nu EL&I en I&M) als uitgangspunt gesteld dat de toelating van gewasbeschermingsmiddelen niet in conflict mag zijn met de doelstellingen van de KRW. Dat betekent dat een toelating van een stof volgens de criteria van de Gewasbeschermingsmiddelenverordening niet mag leiden tot een overschrijding van de normen in KRW-wateren. Als de norm in KRW-wateren wordt overschreden én er een aannemelijk verband is tussen normoverschrijding en de landbouwkundige toepassing van de stof, zou dit gevolgen moeten hebben voor de toelating (herbeoordeling) van deze stof en/of moeten leiden tot het implementeren van adequate mitigerende maatregelen. De zorg hiervoor kwam mede voort uit het feit dat de beschermdoelen in beide kaders niet gelijk zijn. Bij de Europese toelating van bestrijdingsmiddelen mag onder bepaalde voorwaarden een kortdurend effect gevolgd door herstel worden meegenomen in de beoordeling. De Kaderrichtlijn Water heeft als uitgangspunt dat stoffen, en dus ook gewasbeschermingsmiddelen, geen nadelig effect mogen hebben op de structuur en het functioneren van het waterecosysteem. De departementen hebben de werk-

groep verzocht voor de Nederlandse kavelsloot een optie met herstel en een optie zonder herstel uit te werken voor de toelatingsbeoordeling. Voor KRW wateren geldt alleen de optie zonder herstel.

De Beslisboom beschrijft de effectbeoordeling voor de kavelsloot en de effectbeoordeling voor grotere oppervlaktewateren. In de kavelsloot moeten de voorspelde concentraties van een gewasbeschermingsmiddel voldoen aan de wetenschappelijk onderbouwde RACs (Regulatory Acceptable Concentrations) voor korte- en lange-termijn blootstelling. In grotere oppervlaktewateren moeten gemeten concentraties voldoen aan de normen voor langdurige blootstelling en kortdurende piekblootstelling, respectievelijk de jaargemiddelde milieukwaliteitsnorm en de maximaal aanvaardbare concentratie (JG-MKN en MAC-MKN).

Deel 1. Beoordeling voor de kavelsloot

Het eerste deel van dit rapport behandelt de effectbeoordeling voor de kavelsloot. De beslisboom voor de kavelsloot volgt een getrapte benadering, waarbij elke volgende trap gebaseerd is op een grotere beschikbaarheid van gegevens, meer realistisch is en minder conservatief is dan de voorgaande trap(pen).

Eerste trap

De eerste trap bestaat uit drie beoordelingsschema's. Deze schema's volgen de nieuwe Europese Toelatingsprocedure voor Gewasbeschermingsmiddelen (1107/2009/EC) en de voorstellen die daarin worden gedaan. Zo zijn nieuwe standaard testsoorten opgenomen in de beoordelingsschema's (een soort van de kreeftachtigen en additionele waterplanten) en worden de toetsen met algen en waterplanten ook gebruikt voor het bepalen van de normen voor langdurige blootstelling.

Het eerste beoordelingsschema is voor de risico's voor standaard testsoorten als gevolg van directe blootstelling via het water. Voor elke standaard testsoort wordt een RAC (Regulatory Acceptable Concentration) afgeleid door het test eindpunt te delen door een veiligheidsfactor. Dit gebeurt voor zowel kortdurende als langdurige blootstelling. Het tweede beoordelingsschema is voor de risico's voor vis als gevolg van ophoping van de stof in het weefsel van organismen. Het derde beoordelingsschema omvat de risico's voor visetende vogels en zoogdieren als gevolg van doorvergiftiging van de stof via de voedselketen. Met deze laatste twee schema's worden RAC's afgeleid die betrekking hebben op langdurige blootstelling.

Hogere trappen

De beoordelingsschema's in de hogere trappen volgen eveneens de nieuwe Europese Toelatingsprocedure voor Gewasbeschermingsmiddelen (1107/2009/EC) en nemen de nieuwste wetenschappelijke inzichten in beschouwing. Een verschil is dat het rapport een verdere uitwerking beoogt van de hogere trappen. In het rapport wordt voor elke hogere trap aangegeven hoe de RAC wordt afgeleid en welke veiligheidsfactoren worden gehanteerd. De veiligheidsfactoren kunnen verschillen per trap, per organisme of groep van organismen of per ecotoxicologisch eindpunt.

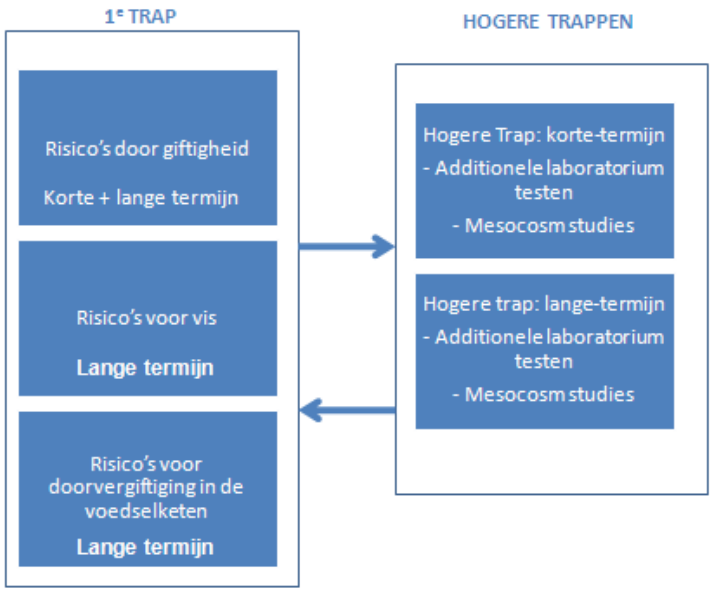
Alvorens te starten met de beoordeling in de hogere trappen wordt het blootstellingsprofiel uit het Nederlandse slootscenario vergeleken met de RAC uit de eerste trap. Het blootstellingsprofiel uit het Nederlandse slootscenario geeft informatie hoe de blootstelling in een experiment in de hogere trappen dient te worden vorm gegeven. Bijvoorbeeld in het geval van risico's van kortdurende blootstelling aan een snel verdwijnende stof kan een herhaalde toediening van de stof in het ecotoxicologisch experiment worden opgenomen indien het voorspelde blootstellingsprofiel hiertoe aanleiding geeft. Het blootstellingsprofiel in het ecotoxicologisch experiment dient realistisch tot conservatief te zijn ten opzichte van het profiel uit het Nederlandse slootscenario.

De beoordelingsschema's voor de hogere trappen vragen verschillende soorten aanvullende informatie (Figuur A). Allereerst kunnen extra gegevens beschikbaar zijn uit laboratoriumtoetsen met andere organismen dan de standaardsoorten. Ook kan informatie worden gebruikt uit laboratoriumexperimenten waarin de test-organismen op een realistischer manier zijn blootgesteld, bijvoorbeeld door meerdere keren kort te doseren, of waarin meer aandacht is voor specifieke ecologische informatie, bijvoorbeeld door meerdere generaties in de tijd te volgen. Tenslotte kunnen semi-veldexperimenten zijn uitgevoerd waarin de blootstelling en ecologische complexiteit zo realistisch mogelijk zijn nagebootst. Voor elk type hogere trap wordt precies aangegeven hoe de RAC wordt afgeleid en welke veiligheidsfactoren worden gehanteerd. De veiligheids-factoren kunnen verschillen, afhankelijk van het soort informatie die aanwezig is.

Als er voldoende ecotoxiciteitsgegevens zijn voor andere organismen dan de standaard toetsorganismen, maar te weinig om een soortgevoeligheidsverdeling (Species Sensitivity Distribution, SSD) te maken, wordt geadviseerd om het geometrisch gemiddelde te nemen van de beschikbare toxiciteitsgegevens van de relevante taxonomische groep. In combinatie met de veiligheidsfactor geeft dit de RAC. Wanneer het aantal toxiciteitsgegevens vijf en meer bedraagt voor vissen of acht en meer voor andere organismen, kan de SSD-methode worden worden toegepast. Dit kan zowel voor chronische als voor acute eindpunten. In de beoor-deling van gewasbeschermingsmiddelen worden deze gevoeligheidsverdelingen toegespitst op de gevoelige organismengroepen. Welke groepen dat zijn, volgt uit de eerste trap beoordeling en uit additionele informatie uit bijvoorbeeld open literatuur, read-across etc. Het rapport geeft een aantal criteria waarmee gevoelige groepen kunnen worden geselecteerd om vervolgens te worden opgenomen in een SSD. Uit deze verdeling wordt de concentratie afgeleid waarbij ten hoogste 5% van de soorten boven het acute of chronische eindpunt wordt blootgesteld. De mediane waarde voor deze 'Hazardous Concentration' (HC_5) leidt in combinatie met een adequate veiligheidsfactor tot de RAC. Bij de risicobeoordeling kan het nodig zijn om voor vis een aparte RAC af te leiden d.m.v. een soortgevoeligheidsverdeling indien de risico's voor planten en evertebraten zijn afgedekt d.m.v. resultaten van micro- of mesocosm experimenten. In deze experimenten worden namelijk meestal geen vissen getest.

Toelatingsdossiers kunnen ook gegevens bevatten van studies met (standaard)testsoorten, waarin zowel de blootstelling als de ecologische opzet realistischer zijn dan in de eerste trap. Hieruit kan met de voorgestelde veiligheidsfactor ook een RAC worden afgeleid.

Als laatste worden in dit deel van het rapport de beoordelingsschema's voor de semi-veldstudies (micro- en mesocosms) besproken. In deze studies worden de effecten van gewasbeschermingsmiddelen op aquatische populaties en levensgemeenschappen gekwantificeerd. Het blootstellingsprofiel dat met het Nederlandse slootscenario is berekend moet als uitgangpunt worden gebruikt voor het blootstellingscenario in het experiment. Uit micro- en mesocosmstudies kunnen verschillende RACs worden afgeleid. De eerste RAC wordt afgeleid op basis van de hoogste concentratie waarbij geen effecten kunnen worden aangetoond op populaties en levensgemeenschappen in de micro-of mesocosm. De tweede RAC is gebaseerd op de laagste concen-tratie waarbij een effect wordt waargenomen, mits dat effect binnen acht weken wordt gevolgd door volledig herstel van de betreffende populatie of levensgemeenschap. Voor de afleiding van beide RACs gelden verschil-lende veiligheidsfactoren. In het rapport wordt aangegeven hoe RACs dienen te worden afgeleid voor kort-durende en voor langdurige blootstelling in micro-/mesocosm studies. Aangezien de meeste micro-/ mesocosm studies geen vis bevatten dient gecontroleerd te worden of de RAC op basis van micro-mesocosm studies tevens beschermend is voor vis.



Figuur A

Weergave van het stelsel van beslisbomen en hun onderlinge relaties (P.M. Simpele weergave in blokken).

Het eerste deel van het rapport eindigt met een beschouwing over de bruikbaarheid van modellen in de risicobeoordeling.

Deel 2. Beoordeling voor grotere oppervlaktewateren

Het tweede deel van dit rapport behandelt de beoordeling van gewasbeschermingsmiddelen voor de grotere oppervlaktewateren, de zogenaamde KRW-waterlichamen. De Kaderrichtlijn Water (KRW) schrijft voor hoe waterkwaliteitsnormen moeten worden afgeleid voor een breed scala aan stoffen. In deel 2 wordt uitgewerkt hoe deze waterkwaliteitsnormen voor gewasbeschermingsmiddelen kunnen worden afgeleid en hoe de normen bij de post-registratiebeoordeling kunnen worden gebruikt. Als uit metingen blijkt dat een toegelaten middel de norm overschrijdt, zal dit bij herregistratie worden meegenomen. Dit kan gevolgen hebben voor de toelating.

KRW kent twee soorten normen: de jaargemiddelde milieukwaliteitsnorm (JG-MKN) die bescherming biedt tegen langdurige blootstelling, en de Maximaal Aanvaardbare Concentratie (MAC-MKN), die geldt voor kort-durende piekblootstelling. In het kader van post-registratie wordt bij toetsing het jaargemiddelde van de gemeten concentraties vergeleken met de JG-MKN. De hoogst gemeten concentratie wordt vergeleken met de MAC-MKN. Zowel aan JG-MKN als aan MAC-MKN moet worden voldaan. Voor gewasbeschermingsmiddelen is de periode van toepassing korter dan een jaar. Middelen van meetgegevens over een periode van een jaar levert hoogstwaarschijnlijk een onderschatting van de werkelijke risico's. Dit rapport beveelt dan ook aan om de JG-MKN te vergelijken met de hoogste tijdgewogen gemiddelde concentratie over een kortere periode gekenmerkt door hogere gemeten blootstellingsconcentraties (bijvoorbeeld drie maanden). Hiervoor is het nodig dat de meetfrequentie tijdens deze periode voldoende hoog is (d.w.z. ten minste twaalf meet-punten bevat).

De wijze van afleiden van beide typen normen staat beschreven in een Europees guidance document, dat begin 2011 is vastgesteld. De methodes bouwen voort op richtsnoeren die eerder in het kader van de KRW zijn opgesteld en vinden hun oorsprong in de methodieken die van toepassing zijn onder REACH. Net als in het eerste deel van dit rapport, is de geldende guidance als uitgangpunt genomen. Het huidige rapport geeft vooral invulling aan die onderwerpen die in de guidance niet (volledig) zijn uitgewerkt, of waarvoor wordt verwezen naar 'expert judgement'.

De normafleiding binnen de KRW kent geen 'getrapte benadering', zoals het geval is onder de Europese toelatingsprocedure voor gewasbeschermingsmiddelen volgens verordening 1107/2009/EC. Afhankelijk van de hoeveelheid en soort gegevens die beschikbaar zijn, kunnen normen op drie verschillende manieren worden afgeleid:

1. Door middel van veiligheidsfactoren op de meest kritische eindpunten van laboratoriumtoetsen.
2. Door het toepassen van statistische extrapolatie op eindpunten van laboratoriumtoetsen (SSD's).
3. Op basis van gegevens van semi-veldstudies (micro- of mesocosmstudies).

Indien alle drie de methoden kunnen worden toegepast, hebben normen op basis van SSD's of micro-/ mesocosms de voorkeur, omdat ze de beschikbare informatie over effecten op waterorganismen/ ecosystemen beter meewegen.

De benadering met veiligheidsfactoren ('assessment factor approach') lijkt in zekere zin op de eerste trap van de beoordeling voor de kavelsloot (zie boven). Een verschil is dat onder verordening 1107/2009/EC per taxonomische groep een veiligheidsfactor geldt, terwijl onder de KRW het aantal en type eindpunten die beschikbaar zijn voor verschillende taxonomische groepen bepalen welke veiligheidsfactor mag worden toegepast. Bovendien moet aannemelijk worden gemaakt dat de gevoelige groepen zijn vertegenwoordigd in de dataset. Wanneer veel gegevens beschikbaar zijn, zoals meestal het geval is in bestrijdingsmiddelen-dossiers, zal voor insecticiden en herbiciden de RAC voor de kavelsloot vergelijkbaar zijn met de norm voor grotere wateren, omdat dezelfde veiligheidsfactoren worden toegepast. Voor fungiciden zal binnen de KRW-systematiek mogelijk een hogere veiligheidsfactor worden toegepast, omdat gegevens over waterschimmels als potentieel gevoelige groep meestal niet beschikbaar zijn. Binnen de KRW is het gebruik van openbare literatuur nadrukkelijk vereist, maar voor veel nieuwe gewasbeschermingsmiddelen is dit niet relevant. De reden hiervoor is dat gepubliceerde gegevens op dat moment nog nauwelijks voorhanden zijn. Het rapport concludeert dan ook dat de 1^{e} trap van de beoordeling voor de kavelsloot en voor de grotere wateren heel vergelijkbaar zijn.

Binnen de KRW gaat men anders om met SSD's dan in de toelatingsprocedure voor de kavelsloot. Het voornaamste verschil bij SSD's is dat de KRW-guidance voorschrijft dat er minimaal tien (liefst vijftien) soorten uit ten minste acht verschillende taxonomische groepen in de SSD vertegenwoordigd moeten zijn. Als is aangetoond dat een specifieke taxonomische groep gevoelig is, kan voor die groep vervolgens een aparte SSD worden gemaakt. Binnen de toelatingsprocedure kan de SSD direct gericht worden op de gevoelige taxonomische groepen. Ook zijn in de toelatingsprocedure minder toxiciteitswaarden nodig (minimaal acht, of vijf in geval van vissen). Dit rapport geeft door middel van concrete voorbeelden een handreiking voor het opstellen van SSD's onder de KRW guidance. Ook geeft het rapport aan hoe bijvoorbeeld informatie uit micro-/ mesocosms kan worden gebruikt wanneer formeel (net) niet wordt voldaan aan de vereisten van een SSD inzake het aantal toxiciteitsgegevens

De KRW-guidance geeft informatie over het gebruik van micro-/mesocosms voor normafleiding, maar een concrete uitwerking voor de praktijk ontbreekt. Dit rapport geeft specifieke aanwijzingen voor het gebruiken van micro-/mesocosms bij het afleiden van chronische en acute normen. Een essentieel punt is dat de KRW niet uitgaat van herstel van effecten. Het uitgangspunt voor de norm is de concentratie waarbij geen ecologische effecten optreden, de zogenaamde ecologische drempelwaarde.

Zowel bij de SSD-methode als voor de micro-/mesocosm-benadering worden bij het afleiden van de KRW-normen andere veiligheidsfactoren gebruikt dan in de toelatingsbeoordeling voor de kavelsloot. Het is goed om te bedenken dat de methodieken van de KRW zijn ontwikkeld voor allerlei soorten wateren en stoffen, dus ook industriële chemicalien, metalen etc., waarvan vaak niet bekend is op welke manier het effect wordt veroorzaakt. De reikwijdte van de KRW is dus breder dan alleen gewasbeschermingsmiddelen. Ook reikt de KRW verder dan alleen landbouwgebieden en betreft de richtlijn juist de algehele waterkwaliteit in de grote watersystemen. Het verdient aanbeveling om bij een volgende herziening van de KRW-guidance speciaal aandacht te geven aan de wetenschappelijke inzichten die in het kader van de toelating van gewasbeschermingsmiddelen zijn ontwikkeld.

Tenslotte

Dit rapport eindigt met het signaleren van een aantal (wetenschappelijke) ontwikkelingen en onderzoeksvragen die aan de orde zouden moeten komen bij toekomstige herziening van internationale guidance documenten voor zowel de toelatingsprocedure als voor de KRW-normafleiding. Deze aandachtspunten betreffen ondermeer:

- Specifieke beschermdoelen.
- Implementatie van ecologische scenario's en effectmodellen.
- Wetenschappelijke onderbouwing van veiligheidsfactoren, met speciale aandacht voor chronische risico's.
- Risico's voor sediment-bewonende organismen.
- Risico's van fungiciden voor waterschimmels.
- Risico's gewasbeschermingsmiddelen met een nieuw werkingsmechanisme.
- Risico van multi-stress en mengseltoxicitiet.
- Mogelijke gevolgen van klimaatverandering voor de beoordelingsmethodiek.

1 Introduction

1.1 Motivation for updating the assessment methodology

In Europe, different legislations (Directives and Regulations) have been developed with different methodologies to assess the aquatic risks/hazards of plant protection products. In particular, these differences are apparent when comparing the authorisation criteria for the compartment water according to the Plant Protection Products (PPPs) Regulation and the water quality standards according to the Water Framework Directive (WFD). These criteria and standards not only are a reflection of knowledge on environmental fate and ecotoxicity of PPPs, but also on different policy decisions about the acceptance of risks in relation to formulated protection goals. More specifically, authorisation criteria for edge-of-field surface waters and generic WFD water quality standards differ in function, usage and the way the effect assessment is linked to the exposure assessment.

Ideally, a common context should be available for the underlying policy decisions and scientific insights. If such a common context is absent, the different aims/claims of the European Directives and Regulations may lead to conflicts between risk assessors (due to different views in technical aspects of the risk assessment) and risk managers (due to different views in protection goals). For example, if the WFD risk assessment procedure is stricter than that of the PPP authorisation procedure, it cannot be excluded that potential risks of PPPs are identified for larger surface waters in agricultural landscapes. This problem came urgently to attention in the Netherlands in the 90s. From measurements, it was shown that the water quality standards for many plant protection products (derived using a methodology which resembles that in the WFD) were seriously failing to be met in larger surface waters (see www.bestrijdingsmiddelenatlas.nl).

Recently, stricter dossier requirements for PPPs have been implemented in Europe by adopting the new Regulation 1107/2009/EC (EC, 2009) and the update of Annex II of Directive 91/414/EEC (EC, 1991). In addition, several scientific opinions of the PPR Panel of the European Food Safety Authority (EFSA) have been published in recent years, which provide new insights in environmental risk assessment procedures for PPPs. Furthermore, a new Technical Guidance Document to derive WFD water quality standards became available (EC, 2011).

An important aim of the present report is to present new effect assessment decision schemes in which these new requirements and developments at EU level are incorporated, while also considering new state-of-the-art knowledge in the field of effect assessment for PPPs.

The proposed effect assessment decision schemes for the Netherlands are based on the following model (also see Figure 1-1). Assessments take place at two points in the water system, each with its own risk assessment procedure:

- Small edge-of-field surface waters (in the Netherlands the drainage ditch): Pre-registration risk assessment procedures for short- and long-term exposure according to the PPP Regulation, as far as possible based on standardised European dossier data and models and (national) exposure scenarios
- WFD water bodies: Generic risk assessment procedures according to the WFD, by comparing the water quality standards for short- (MAC-EQS) and long-term exposure (AA-EQS) with measured (post-registration) exposure concentrations in WFD water bodies.

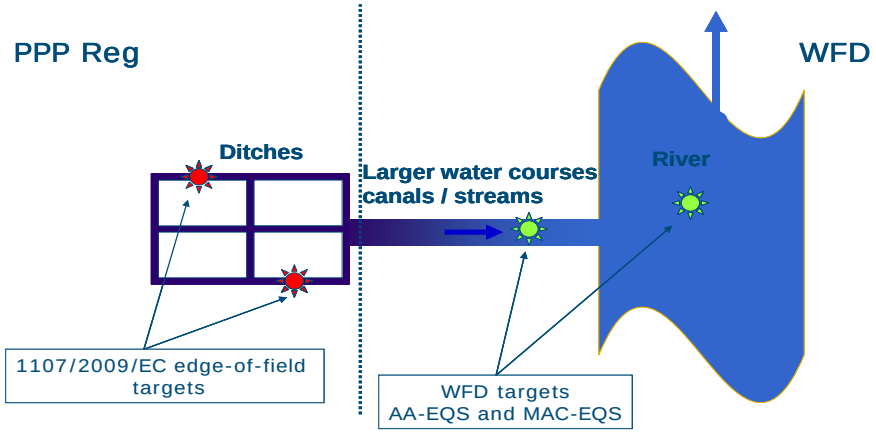


Figure 1-1

Conceptual model for the aquatic risk assessment of PPPs in the Netherlands based on spatial differentiation in compliance to the Plant Protection Product Regulation and the Water Framewerk Directice (WFD).

If in one of the two parts of the water system the specific criteria/standards for PPPs are not met, and this cannot be attributed to misuse (i.e. Good Agricultural Practice (GAP) has been applied), this may have consequences for its authorisation (decisions based on both pre-registration and post-registration risk assessment procedures) and/or additional mitigation measures have to be implemented. This conceptual model does not address the existing disagreement between the ecotoxicological assessments (e.g. by in-/excluding ecological recovery) in the contexts of authorisation and setting water quality standards and acceptance of higher-tier studies. Nevertheless, by using the conceptual model the post-registration assessment is in compliance with the criteria of both the WFD and PPP Regulation.

1.2 Outline of the report

In this report the new proposal for the aquatic effects assessment of plant protection products within the context of the pre- and post-registration procedure will be presented. Before describing the different procedures for effect assessment of PPPs in edge-of-field surface waters (Chapters 5 and 6) and WFD water bodies (Chapters 7 and 8), attention will be paid to the protection goals underlying Regulation 1107/2009/EC and Directive 2000/60/EC (Chapter 2) and to the main features of linking exposure to effects in the aquatic risk/hazard assessment procedure of plant protection products (Chapter 3). An overall description of the proposed decision schemes for risk assessment in Dutch drainage ditches is presented in Chapter 4 and a general description of the water quality standards for WFD water bodies in Chapter 7.

The current guidance documents underlying the PPP Regulation as well as Directive 2000/60/EC leave several decisions to expert judgement. In our report, with a focus on aquatic risk assessment in the Netherlands, we give further guidance on a number of these items and develop tailor-made decision schemes for PPPs. In addition, there are aspects which could be improved on the basis of valid scientific arguments, while the current guidance does not give room to implement these changes. We discuss these issues in Chapter 9. These discussion items may be considered when updating the official guidance documents underlying these directives.

Finally the report presents a glossary of frequently used terms (Chapter 10). To verify the proposed decision schemes (and underlying risk assessment approaches) case studies with selected compounds that differ in fate properties and toxic mode-of-action will be presented in a future report.

2 Protection aims of Regulation 1107/2009/EC and Directive 2000/60/EC

2.1 PPP Regulation (1107/2009/EC)

The PPP Regulation offers a framework for the authorisation of plant protection products (PPPs) on the European market. According to the preamble, it is required that 'plant protection products, when properly applied for the purpose intended, are sufficiently effective and have no unacceptable effect on plants or plant products, no unacceptable influence on the environment in general and, in particular, no harmful effect on human or animal health or on groundwater'. The PPP regulation gives a definition of 'environment': according to Art. 3 (13), 'environment means waters (including ground, surface, transitional, coastal and marine), sediment, soil, air, land, wild species of fauna and flora, and any interrelationship between them, and any relationship with other living organisms'. This definition does, however, not specify the geographical level (local, regional, national or European), nor the level of biological organisation (individual, population, community or ecosystem) which should be considered. More specific information can be found in the Uniform Principles as laid down in Annex VI of Directive 91/414/EEC (EC, 1997). The environmental risk assessment should address the fate and distribution in the environment and the impact on non-target organisms on the acute and long-term time scale.

With respect to the geographical unit of the risk assessment, Annex VI refers to 'the area of envisaged use' (art. 2.5.1.1 and 2.5.1.3). In line with this, the FOCUS-scenarios for estimation of PECs in surface water refer to ditches, ponds or streams next to the treated field (FOCUS, 2001).

Concerning the impact on non-target organisms, Annex VI refers to specific organism groups: birds and mammals, aquatic organisms, honeybees and other beneficial non-target arthropods, earthworms and other non-target soil macro-organisms and soil micro-organisms. Although not explicitly stated, it may be assumed that the underlying reasoning is that if specific organism groups are sufficiently protected, unacceptable effects on the ecosystem level will not occur. The risk assessment for the respective groups is performed at different levels of biological organisation. For birds and mammals, there is a kind of common agreement among risk managers (related to public awareness) that birds and mammals should be protected on an individual level. Although not explicitly stated anywhere, it is not considered acceptable that individual birds or mammals show acute mortality to PPP use, even when this would not affect the population. Bees are considered at population level. For other non-target arthropods and earthworms both population and community studies are performed. The updated Annex II of the PPP regulation mentions 'aquatic organisms' and refers specifically to the fish *Oncorhynchus mykiss*, *Daphnia* (preferably *Daphnia magna*), mysid shrimp *(Americamysis bahia)*, the insect *Chironomus riparius*, green algae (e.g. *Pseudokirchneriella subcapitata*), diatoms (e.g. *Navicula pellicosa*), and the macrophytes *Lemna* sp., *Myriophyllum spicatum/aquaticum* and *Glyceria maxima*. Together, these organisms are considered to represent key-taxa in the aquatic ecosystem. In most European Member States the level of protection for aquatic invertebrates and primary producers is set at the population and/or community level, while there is a tendency towards protection of fish (and other aquatic vertebrates) on an individual to population level. Note that acute and visible mortality of fish due to pesticide application is not considered acceptable. The effects assessment described in this report for the drainage ditch aims to protect fish (and other vertebrates) at the individual level, and plants (including algae)

and invertebrates at the population level. This is in accordance with a recent scientific opinion of the PPR panel of EFSA (EFSA, 2010) and a review paper of Hommen et al. (2010a).

2.2 Water Framework Directive 2000/60/EC

According to the preamble, the Water Framework Directive 2000/60/EC (WFD; EC, 2000) aims at 'maintaining and improving the aquatic environment in the Community'. According to point 27 of the preamble, 'the ultimate aim is to achieve the elimination of priority hazardous substances (...)'. The Directive is focused on water quality, which also includes the control of water quantity.

Member States should aim to achieve the objective of at least a 'good ecological status' and a 'good chemical status' by defining and implementing the necessary measures within integrated programs of measures. Where good water status already exists, it should be maintained. The biological, hydromorphological and physico-chemical parameters that determine the ecological status are presented in Annex V to the Directive. For a good status the WFD requires that Environmental Quality Standards (EQSs) are met, without prejudice to the PPP Regulation (Annex V, Section 1.2). Within the context of the WFD, EQSs are thus one of the instruments to evaluate water quality. They serve as a benchmark to decide whether or not specific measures are required. Two types of EQSs are distinguished to cover both long-term and short-term exposure.

According to the text of the Directive, quality standards should be derived according to the Technical guidance document (TGD) in support of the risk assessment for new and existing substances and biocides (EC, 2003). A more detailed guidance was provided by Lepper (2005). At present, the new and existing substances regulation has been replaced by REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals), but the TGD is still into force for biocides. With respect to the aquatic ecosystem, the risk assessment methodology under REACH (ECHA, 2008) is basically the same as outlined in the TGD. The guidance of Lepper (2005) was revised recently, and an updated TGD for derivation of quality standards under the WFD was published (EC, 2011).

The geographical unit under consideration in the WFD is the river basin, which is defined as 'the area of land from which all surface run-off flows through a sequence of streams, rivers and, possibly, lakes into the sea at a single river mouth, estuary or delta'. Member States must assign the river basins lying within their national territory to 'river basin districts'. For each river basin district - some of which will traverse national frontiers - a 'river basin management plan' will need to be established and updated every six years, and this will provide the context for coordinated measures. The Netherlands belong to four international river basin districts: the rivers Rhine, Meuse, Scheldt and Ems. Within each river basin, the WFD applies to so-called water bodies (see Figure 2-1).

Figure 2-1
Example of WFD water bodies for the river basin Rijndelta in the Netherlands (from: www.kaderrichtlijnwater.nl/publicaties/de_krw_rapportages/?ActItmIdt=19927).

The protection goal of the WFD is human and ecosystem health. The protection of human health obviously refers to the individual level. As for PPPs under the PPP Regulation, protection of birds and mammals is a specific ally addressed and the implicit assumption is that for a good status effects on individual birds and mammals cannot be accepted. The derivation of the QSs for direct ecotoxicity is based on the methodology for establishing Predicted No Effect Concentrations (PNEC) according to the TGD (EC, 2003). This guidance is taken over within the context of the REACH Implementation Project (ECHA, 2008). According to the TGD, it is generally accepted that protection of the most sensitive species should protect structure, and hence function. It is assumed that:

– ecosystem sensitivity depends on the most sensitive species, and
– protecting ecosystem structure protects community function.

The REACH guidance states that ecosystems are expected to be more sensitive than individual organisms in the laboratory. Therefore, the results of tests are not used directly for the risk assessment but used as a basis for extrapolation of the PNEC.
The level of biological organisation as considered for derivation of PNECs (and quality standards) is thus the ecosystem, including its biodiversity. However, as for authorisation of PPPs under the PPP Regulation, this is achieved by using studies with individual species, population or communities.

2.3 Different approaches

From the above, it can be concluded that the protection aims of the PPP-regulation and WFD seem to be very similar, with the exception that the PPR-regulation not excludes that under certain conditions transient, short-term effects on non-vertebrates are acceptable in edge-of-field surface waters. However, the approaches used for defining a 'safe' concentration for the aquatic ecosystem are fundamentally different. Under the PPP-regulation, the aquatic risk assessment is carried out by evaluating the risks for each species group (fish, invertebrates, algae/macrophytes, fish eating birds and mammals) separately in a tiered approach. If, for a certain group the evaluation points at a potential risk, the assessment is further focused on that particular problem. This means that different regulatory acceptable concentrations (RACs) are derived, depending on the species group and time-scale under consideration. This is further outlined in Chapters 4 to 6. Under the WFD, a single chronic and an acute water quality standard is derived for the aquatic ecosystem as a whole, including predatory birds and mammals, and fish eating humans where relevant (see Chapters 7 and 8). Under both frameworks, however, the risk assessment or standard setting will in the end depend on the most critical species group or endpoint, under the assumption that protection of the most sensitive species group will ensure the protection of the ecosystem.

3 Linking exposure to effects in the risk evaluation of plant protection products

3.1 Introduction to linking exposure and effects

The aquatic risk assessment procedure for PPPs, and all other toxic chemicals, consists of two parts:
- Exposure assessment, which is the domain of experts in environmental chemistry, exposure modelling and chemical monitoring, and
- Effects assessment, which is the domain of experts in toxicology, ecotoxicology and ecology (including biological monitoring).

Also within the current project, that aims to scientifically underpin the authorisation policy for PPPs in the Netherlands, different working groups are active that deal with exposure and effects assessment, respectively.

Relevant exposure concentrations in the water courses of concern can be obtained by chemical monitoring, by applying fate models to derive PECs or by a combination of monitoring and modelling. However, in a prospective risk assessment for new PPPs not yet placed on the market, chemical monitoring data are not yet available, and exposure predictions at the landscape level may be characterised by relatively high uncertainty because the scale and intensity of the use of these new PPPs are not yet known. A common, cost-effective approach in the prospective exposure assessment is the development of exposure scenarios. For example, within the European Union, harmonised approaches for conducting aquatic exposure assessments for agricultural pesticides have been developed. These are documented in the 'FOCUS Surface Water Scenarios' report (FOCUS, 2001). Also for the prospective exposure assessment of new PPPs in the Netherlands scenarios are developed by the exposure working group. These scenarios, in combination with models that estimate the emissions to and fate and behaviour of PPPs in surface waters, enable to predict realistic worst-case exposure concentrations in edge-of-field drainage ditches (Tiktak et al., 2012) and possibly in the near future also WFD-water bodies (www.cascade.pesticidemodels.eu).

All prospective approaches to assessing ecological risks at the edge-of-field or watershed level heavily rely on the proper linking of predicted exposure concentrations to ecotoxicological and ecological data. The ecotoxicological data usually concern concentration - response relationships derived from controlled experiments with e.g. standard and additional aquatic test species or micro-/mesocosm tests. The ecological data usually relate to the 'target image' of the aquatic community in the relevant surface waters, including ecological traits of the important aquatic species at risk. An example of an ecological Dutch ditch scenario can be found in Brock et al. (2010b). Uncertainty factors and/or modelling approaches, are used to extrapolate the experimental concentration - response relationships in space and time, e.g. to estimate the threshold concentrations for toxic effects in the field or the potential for recovery of affected populations.

3.2 Ecotoxicologically Relevant Concentration (ERC)

After having collected the relevant data for the exposure and the effects assessment, a crucial step in the risk assessment is the linking of exposure and effects data. Lack of a clear conceptual basis for the interface between the exposure and effect assessment may lead to a low overall scientific quality of the risk/hazard assessment. This interface is defined by EFSA (2005a) and Boesten et al. (2007) as the type of concentration

that gives an appropriate correlation to ecotoxicological effects, and is called the ecotoxicologically relevant concentration (ERC). In the risk/hazard assessment the ERC needs to be consistently applied so that field exposure estimates (PECs) and regulatory acceptable concentrations (RACs; used within the context of the PPP Regulation) or environmental quality standards (MAC-EQS and AA-EQS; used within the context of 2000/60/EC) can be compared as readily as possible. The ecotoxicological considerations determining the ERC may include the following questions:

- In which environmental compartment do the aquatic organisms at risk live (e.g. water or sediment)?
- What is bioavailable for the organism (e.g. for sediment dwellers the fraction in the pore water or the fraction bound to the sediment; for pelagic organisms the fraction in water or the fraction in the food)?
- What is the influence of the time-variable exposure pattern on the effects (e.g. do peak or longer-term concentrations explain the responses)?
- Which information on the 'time to onset-of-effects' is available to determine whether short-term or long-term exposures are relevant?

In ecosystems the ERC may be different for substances that differ in toxic mode-of-action and for different populations of aquatic organisms, life stages of species, and so on. For example, for an aquatic insect living associated with macrophytes in shallow freshwater ecosystems, the ERC could be the maximum concentration over time of the dissolved fraction for a fast-acting insecticide or some time-weighted average (TWA) concentration for a slow acting fungicide (see Section 3.3.2). For sediment dwelling insects that live predominantly in the top centimeters of the sediment, the ERC could be the maximum over time of the pore water concentration of the fast-acting insecticide in the top 2 cm of the sediment. For an aquatic insect that predominantly dwells at the water surface (e.g. water striders) the ERC of a fast acting insecticide may be the water concentration in the top layer of the water column, which may be relevant if initially stratification of the insecticide occurs. After the ERCs for the PPP under evaluation and the aquatic organisms at risk have been determined, the collected exposure data can be linked to the relevant ecotoxicological data. Key is that the type of ERC used to express the 'C' in the PEC estimates should not be in conflict with the ERC used to express the 'C' in the RAC and EQS estimates.

3.3 When to use the peak or a (time weighted) average concentration in the risk/hazard assessment

3.3.1 Current procedures under the PPP Regulation

Generally PEC_{max} values are used in acute risk assessments, whilst in chronic risk assessments, in first instance the PEC_{max}, and under certain conditions, a TWA PEC may be used. The use of the time-weighted average (TWA) concentration approach in the risk assessment of PPPs is based on the observation that effects of PPPs on aquatic organisms may be similar when exposed for a short time to a greater concentration or for a longer time to a smaller concentration, a phenomenon referred to as reciprocity (Giesy and Graney, 1989). Reciprocity relates to Haber's law, which assumes that toxicity depends on the product of concentration and time. For example, an 8-day exposure at 10 µg/L may cause the same effects as a 4-day exposure at 20 µg/L or a 2-day exposure at 40 µg/L, an example of linear reciprocity. Linear reciprocity is the basis of the time-weighted average (TWA) approach where exposure concentration is integrated over time (area under the curve = AUC) and then divided by the duration of the toxicity test. When this approach is applied, different exposure patterns with the same AUC are assumed to have the same effects.

Theoretically, reciprocity should only apply where both uptake and/or elimination of a compound into the test organism (toxicokinetics) and damage and/or repair processes (toxicodynamics) have reached steady state (Rozman and Doull, 2000). In tests with *Gammarus pulex* no reciprocity for chlorpyrifos was observed by Ashauer et al. (2007a) when extrapolating from short- to long-term exposures. These authors found that the

TWA approach based on an acute toxicity test greatly underestimated mortality in longer-term exposure studies, whereas it overestimated mortality caused by pentachlorophenol. In long-term toxicity tests with *Gammarus pulex*, however, Ashauer et al. (2007a) demonstrated that the TWA concentration approach can be used to extrapolate results of a chronic pulse test to other chronic exposures for both chlorpyrifos and pentachlorophenol. This observation supports the use of the TWA concentration approach in chronic risk assessments. In addition, the longer duration of chronic tests implies a greater probability that toxicokinetics and toxicodynamics will approach steady state by the end of the study period.

3.3.2 **Recent developments**

According to the proceedings of the ELINK workshop (Brock et al., 2010a) TWA approaches in the chronic risk assessment have limitations in the following situations, viz.;

- In risk assessments that use RACs derived from effect studies where the exposure is not maintained and loss of the active substance in the test system other than uptake by the test organism is fast.
- When the effect endpoint in the chronic test (used to derive the RAC) is based on a developmental process during a specific sensitive life-cycle stage and when it cannot be excluded that the exposure will occur when the sensitive stage is present.
- When the effect endpoint in the chronic test (used to derive the RAC) is based on mortality occurring early in the test (e.g. in the first 96 h), or if the acute to chronic ratio (acute EC_{50} or LC_{50}/chronic NOEC) based on immobility or mortality is <10.
- If latency of effects (delayed effects) has been demonstrated, or might be expected due to mode of action of the pesticide or by appropriate other data (e.g. in the case of moulting inhibitors and substances suspected of endocrine disruption).

One of the recommendations of the ELINK workshop (no. 3) is that ecotoxicologists must determine, based on knowledge of ecotoxicological data, whether or not the TWA concentration approach is appropriate to use in the chronic risk assessment, and which time window the TWA should be based upon. Following a worst-case approach, the time-window of the TWA PEC should be equal to or smaller than the length of the relevant chronic toxicity test (or life stage of highest ecotoxicological concern) that triggered the risk. Ideally, the selected time-window for the TWA estimate should be justified on a case-by-case basis considering the scientific information available. Note, however, that currently limited concentration - response information is available for pesticides and relevant water organisms on basis of time-variable exposure experiments. This information is key to evaluate the TWA approach in pesticide risk assessment. For pragmatic reasons the ELINK document proposes a default 7-day TWA time window for invertebrates and fish (and possibly also macrophytes), if the TWA concentration approach is deemed appropriate and no further information on the relation between exposure pattern and time to-onset-of the relevant effect is provided. It may be justified to lengthen or shorten the default 7-day TWA period when scientific data are made available that demonstrate that another TWA period is more appropriate. The ELINK workshop recommended further research to scientifically underpin the criteria that can be used to decide whether the TWA approach is appropriate and to set the appropriate TWA time window (Brock et al., 2010a).

If the use of the TWA approach in the chronic risk assessment is appropriate, concentration-response relationships observed in toxicity tests with long-term exposure (which may be variable in time), as well as the derived RAC, can be expressed in terms of TWA concentrations. This RAC value can be compared with the appropriate TWA PEC.

If the TWA concentration approach cannot be used in the chronic risk assessment, the ELINK document describes possibilities for refined exposure studies (single species, population and micro-/mesocosm experiments) in which the organisms are exposed to realistic worst-case long-term exposure concentrations (which may be variable in time). The refined realistic worst-case exposure regime tested should be guided by relevant exposure predictions for the intended agricultural uses (e.g. as deduced from FOCUS surface water scenarios or from national exposure scenarios). More realistic and representative exposure in toxicity tests probably will decrease the need for TWA-calculations.

It appears from the above that the current evaluation under the PPP Regulation follows a risk assessment procedure in which in higher-tier effect assessment procedures the exposure regime is based on realistic worst-case exposure predictions for the field. In addition, in the risk assessment procedure under the PPP Regulation it is recognised that time-variable exposure concentrations of plant protection products are more often the rule than the exception. Consequently, chronic higher-tier effect assessments not necessarily need to be performed by simulating constant chronic exposure regimes.

3.3.3 Toxicological and ecological independence of different pulse exposures

For an appropriate assessment of risks from exposure profiles characterised by repeated pulsed exposures it is in first instance important to determine whether or not the pulses are toxicologically independent or not (EFSA, 2005a). Toxicological dependence of repeated pulses may occur if the life-span of the individuals of the sensitive species is long enough to also experience repeated pulse exposures or when the exposure results in toxicogenetic (trans generation) effects. If, for example, the predicted exposure profile consists of two pulse exposures, the second pulse can be considered toxicologically independent from the first pulse if between the two pulses: (i) the internal exposure concentrations in the individuals of the sensitive species drop below critical threshold levels, and (ii) complete repair of damage occurs. According to the proceedings of the ELINK workshop (Brock et al., 2010a) to demonstrate the toxicological independence of different pulse exposures, either specially designed pulsed exposure tests or toxicokinetic/toxicodynamic models for the relevant organisms and PPP of concern are required. If the toxicological independence of successive pulse exposures can be demonstrated for the species at risk it may be valid to adopt a single pulse exposure regime in higher tier tests to derive a RAC for the threshold level of toxic effects. Note that these approaches should take into account possible latency of toxicological effects if this phenomenon is reported for related compounds (similar toxic mode-of-action).

When evidence can be provided that successive pulse exposures are toxicologically independent, it may be important to also demonstrate their ecological independence, particularly when ecological recovery is taken into account in the effects assessment. Successive pulse exposures may be considered ecologically independent if peak intervals are greater than the relevant recovery time of the sensitive populations of concern. According to the proceedings of the ELINK workshop (Brock et al., 2010a) evaluating the ecological dependence/independence of successive pulse exposures will be important when microcosm and mesocosm tests are used in the risk assessment that aim to derive a NOEAEC (No Observed Ecologically Adverse Effect Concentration). Since only a limited number of ecological recovery scenarios can be investigated in micro-/mesocosm tests, modelling approaches may provide an alternative tool for spatio-temporal extrapolation and to investigate whether successive pulse exposures are ecologically dependent or not. The possible ecological independence of pulse exposures may also be of importance in the risk assessment if the potentially sensitive species, or specific sensitive life stages of these species, are not present in the periods that certain pulse exposures occur (e.g. pulse exposure in winter because of drainage).

3.3.4 **Current procedures under 2000/60/EC**

As pointed out in Section 2.2, two types of quality standards are derived under the umbrella of the Water Framework Directive (WFD), to cover both long-term and short-term exposure to a chemical: the AA-EQS related to annual average concentration to protect against the occurrence of prolonged exposure, and the MAC-EQS to protect against possible effects from short term concentration peaks.

When the chemicals under evaluation are already on the market (currently used or used in the past) comparing chemical monitoring data with the AA-EQS and MAC-EQS is the means by which compliance is assessed (retrospective assessment). Checking compliance with an EQS relies on analysis of discrete chemical monitoring samples. According to the 'Technical Guidance for Deriving Environmental Quality Standards under the Water Framework Directive' EQSs should be linked to an average concentration (AA-EQS) or the maximum of the measured concentrations (MAC-EQS) (EC, 2011). For details on treatment of monitoring data see Rijkwaterstaat (2009).

When linking exposure and effects under the retrospective assessment procedure under the umbrella of 200/60/EC, the AA-EQS is normally compared with the arithmetic mean of concentrations measured in chemical monitoring samples taken at a sampling station over a year. However, for a substance that is used for only a short part of the year, a shorter period may be considered. With a sampling frequency of 1 sample per month (as recommended in Annex V, 1.3.4. of the WFD) the annual average concentration at a sampling station will normally be calculated as arithmetic mean of twelve samples. This implies that for a certain percentage of time the concentration in water may exceed the AA-EQS due to peaks resulting e.g. from intermittent releases of the chemical in question. According to the 'Technical Guidance for Deriving Environmental Quality Standards under the Water Framework Directive' multiple samples must be taken when assessing compliance with a quality standard that is expressed as a long-term average.

A minimum of twelve samplings per year for priority pollutants and minimally four samples per year for other relevant pollutants per year is prescribed. In case of e.g. pesticides, which show peak concentrations within short time periods, enhanced sampling frequency may be necessary in these periods. For example, the best sampling time for detecting concentration peaks of pesticides is after heavy rainfall within or just after the application period. Moreover, failure to comply with good agricultural practice can also cause higher peak concentrations of pesticides than predicted. The results of those measurements should be compared with the MAC-EQS. For the calculation of the annual average concentrations all results are averaged. Collecting composite samples might be another option to detect peak concentrations of seasonally variable compounds (Rijkswaterstaat, 2009).

Without continuous chemical monitoring, it is impossible to know whether or not a MAC-EQS is actually exceeded. Since chemical monitoring usually relies on discrete sampling, a measured peak concentration is actually a *de facto* percentile, depending on the sampling frequency.

For plant protection products, using the annual average exposure concentration is hard to support scientifically because of factors related to specific toxic mode-of-action of many PPPs and their relatively fast time-to-onset-of-effects. An option may be to identify for each PPP its period of frequent agricultural use and to use that period as time-frame for the long-term average PEC, at least if this agricultural use period coincides with the main period of exposure in the WFD water body. A disadvantage of this approach is that the time-frame for the long-term PEC to be linked to the AA-EQS will be substance, crop and region dependent. In Chapter 7.2 this will be discussed in greater detail.

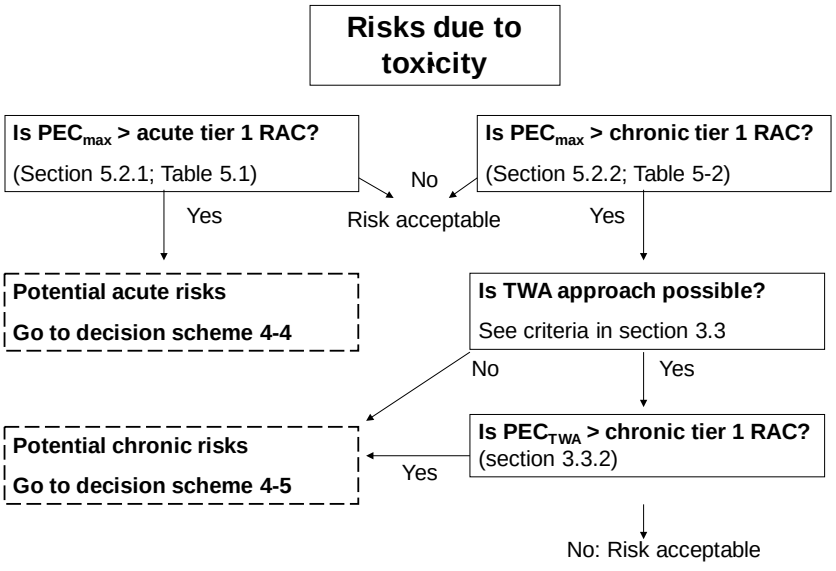
4 Proposed decision schemes for acute and chronic risk assessment for plant protection products in drainage ditches

4.1 Introduction

In this Chapter the proposed decision schemes for the risk assessment of PPPs in Dutch drainage ditches are presented. The effects assessment components of these decision trees are described in detail in the following Chapters (Chapter 5, tier 1 procedure; Chapter 6, higher-tier procedures). The decision schemes also include proposals how to link the exposure and the effects estimates in the risk assessment. In the different boxes of the decision schemes reference is made to the report Chapters/sections and tables in which further information can be found with respect to the scientific underpinning of the effects assessment methodology proposed.

4.2 Decision scheme for tier 1 risk assessment

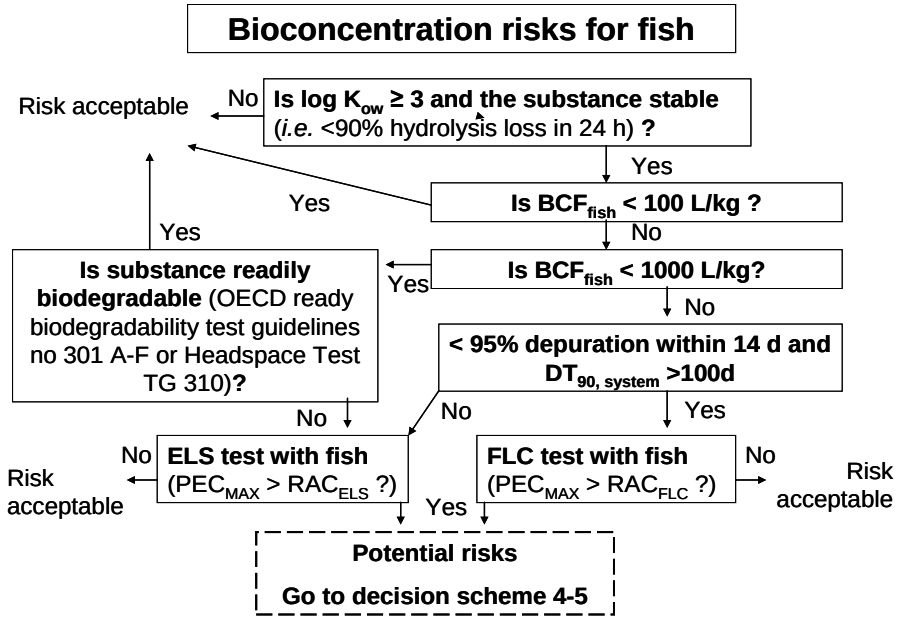
The first tier aquatic risk assessment procedure is based on three types of risks:
1. Risks due to toxicity as assessed with standard test species (see Decision scheme 4-1)
2. Risks due to bioconcentration in fish (see Decision scheme 4-2)
3. Risks due to secondary poisoning (see Decision scheme 4-3)



Decision scheme 4-1

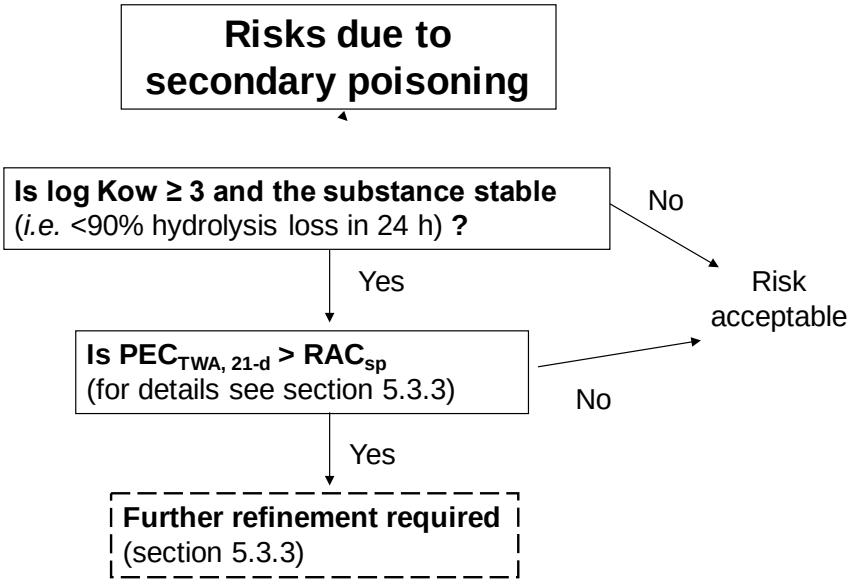
Tier 1 flow chart for acute and chronic risk assessment of pesticide toxicity in edge of field surface waters.

If, as a result of the first-tier assessment, potential risks for one or more aquatic standard test species are identified a higher-tier assessment may be performed that either concerns a refinement of the exposure assessment or a refinement of the effects assessment, or a combination of the two. In this report the approaches for higher-tier effect assessment and the rules how to link exposure to effects estimates are described. The exposure assessment approach for the Dutch ditch scenario is described in Tiktak et al. (2012).



Decision scheme 4-2

Tier 1 flow chart for the risk assessment of pesticide bioconcentration in fish.

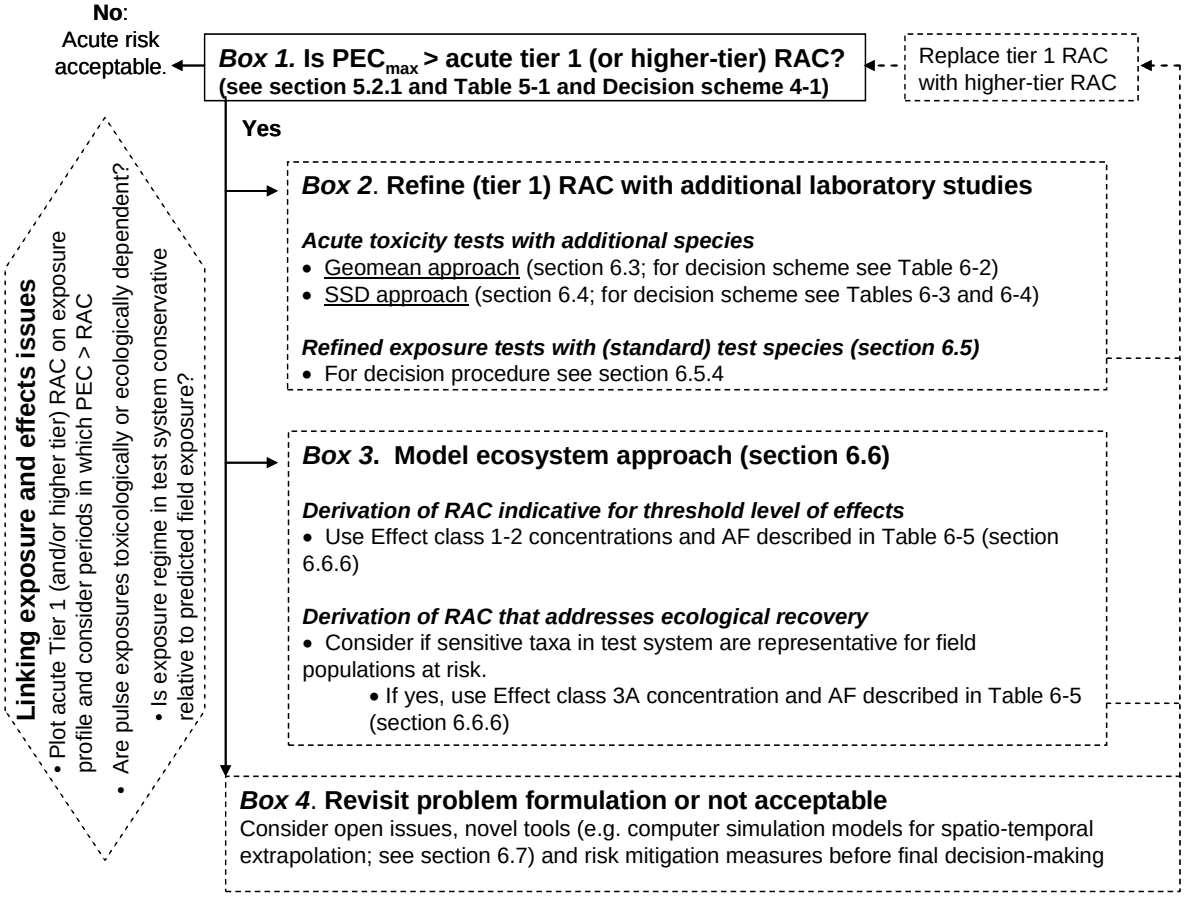


Decision scheme 4-3

Tier 1 flow chart for the assessment of risks of secondary poisoning of PPPs to fish-eating birds and mammals (RAC_{sp} = RAC for secondary poisoning).

4.3 Decision scheme for higher-tier acute risk assessment of toxicity

The basic flow chart for higher-tier acute risk assessment for toxicity of PPPs in Dutch drainage ditches is presented in Decision scheme 4-4. In this scheme the specific report sections are mentioned where detailed guidance for decision making can be found. The scheme also identifies a few linking exposure and effects issues that need to be considered in the higher-tier effect assessment procedures mentioned in Boxes 2 and 3 (see vertical panel on the left in Decision scheme 4-4).



Decision scheme 4-4

Basic flow chart for higher-tier acute risk assessment of PPPs in edge-of-field surface waters in the Netherlands.

Before starting a higher-tier effect assessment the predicted exposure profile for the PPP of concern in the Dutch drainage ditch scenario needs to be compared with the tier 1 RAC. This can best be done by plotting the tier 1 RAC on the predicted exposure profile. As an example the tier 1 acute RAC for the hypothetical insecticide Phantasithrin is plotted on its exposure profile in Figure 4-1A. In this example the exposure profile is characterised by a repeated pulse exposure regime and the peaks of all pulses exceed for short periods the acute tier 1 RAC.

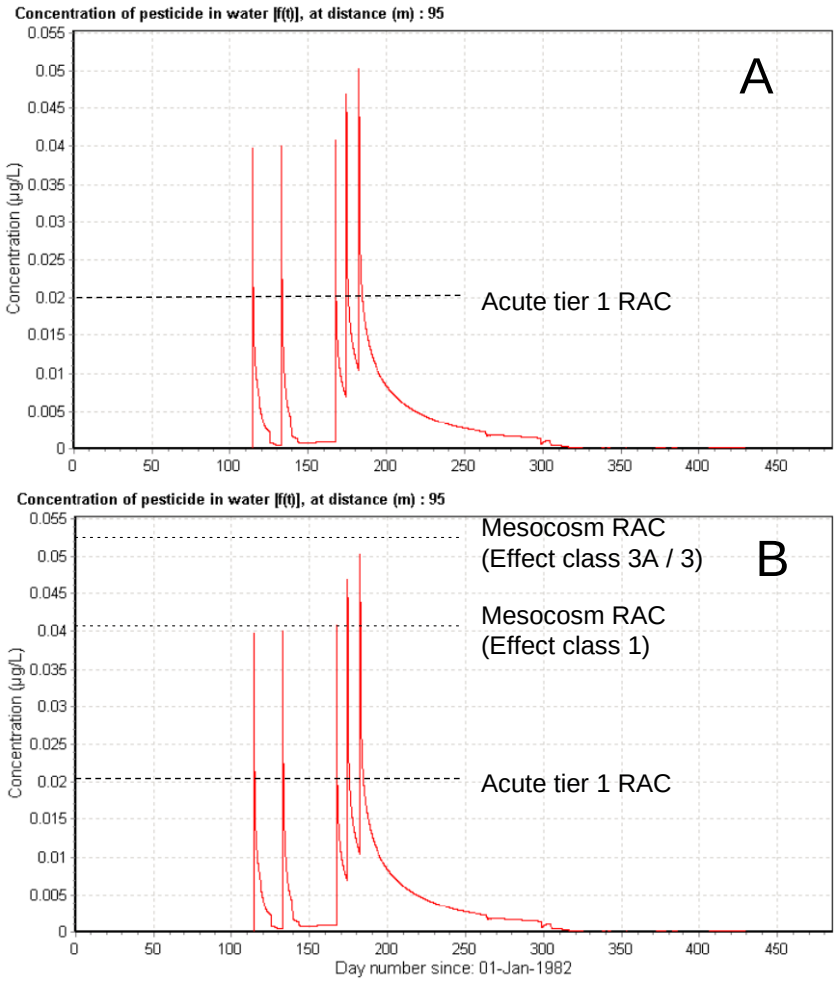


Figure 4-1

Example of an exposure profile for the hypothetical insecticide Phantasithrin in the Dutch drainage ditch scenario on which the tier 1 acute RAC (panel A) and the acute higher-tier RACs derived from a mesocosm experiment (panel B) are plotted.

In first instance it is important to determine whether the pulse exposures that exceed the tier 1 RAC for Phantasithrin are toxicologically and/or ecologically (in)dependent (for details see Section 3.3.3). In the example presented in Figure 4-1A the intervals between pulses are short (particularly the last 3 pulses) relative to the average life-span of individuals of sensitive insects and macro-crustaceans that are at risk. Since no information on the toxicokinetics and toxicodynamics of Phantasithrin for these organisms is available, toxicological dependence of the repeated pulses cannot be excluded. Consequently, in the higher-tier assessment a repeated pulse exposure regime needs to be addressed.

In the Phantasithrin example presented in Figure 4-1 the higher-tier risk assessment is based on results of a mesocosm test in which the insecticide was applied four times at weekly intervals and the overall exposure regime was worst-case relative to the predicted field exposure regime. From this mesocosm test a RAC indicative for the threshold level of effects (based on the highest Effect class 1 concentration) and a RAC that addresses ecological recovery (based on the highest Effect class 3A concentration and the application of an AF of 3) could be derived. In Figure 4-1B these RAC values are plotted on the exposure profile. It appears that the authorisation of Phantasithrin only can be granted if the specific protection goal allows some effects followed by recovery (total effect period <8 weeks). Plotting the RAC indicative for the threshold level of

effects on the predicted field exposure profile, and evaluating the exposure period above this RAC and the time needed for recovery derived from the mesocosm test, provides insight in the total effect-period that might be expected.

Decision scheme 4-4 also contains a 'Revisit problem formulation box (Box 4) to address open issues, novel tools and possible risk mitigation measures. An important open issue always concerns the organisms that are not evaluated in the higher-tier tests performed. For example, for cost-effective reasons the adopted higher-tier approaches usually focus on organisms for which concerns were identified in the first tier. However, if on basis of the higher-tier test a refined RAC is obtained it should always be checked whether this RAC is protective enough for the taxonomic groups not addressed in the higher-tier tests. If not, a lower-tier approach on basis of test species that are representative for the organisms not addressed in the higher-tier test should be leading in the final risk assessment. For the example insecticide Phantasithrin and the data presented in Table 4-1 this means that the RAC derived from the mesocosm test (without fish; 0.053 µg/L) is protective for fish as well (since the tier 1 RAC of 0.260 µg/L and SSD based RAC of 0.580 µg/L for fish are higher). In contrast, for the other example insecticide (Imagiphos) the tier 1 RAC for fish needs to be selected as overall acute RAC, despite the availability of an appropriate mesocosm experiment, since the tier 1 RAC for fish is lower than the mesocosm derived RAC for invertebrates.

Table 4-1

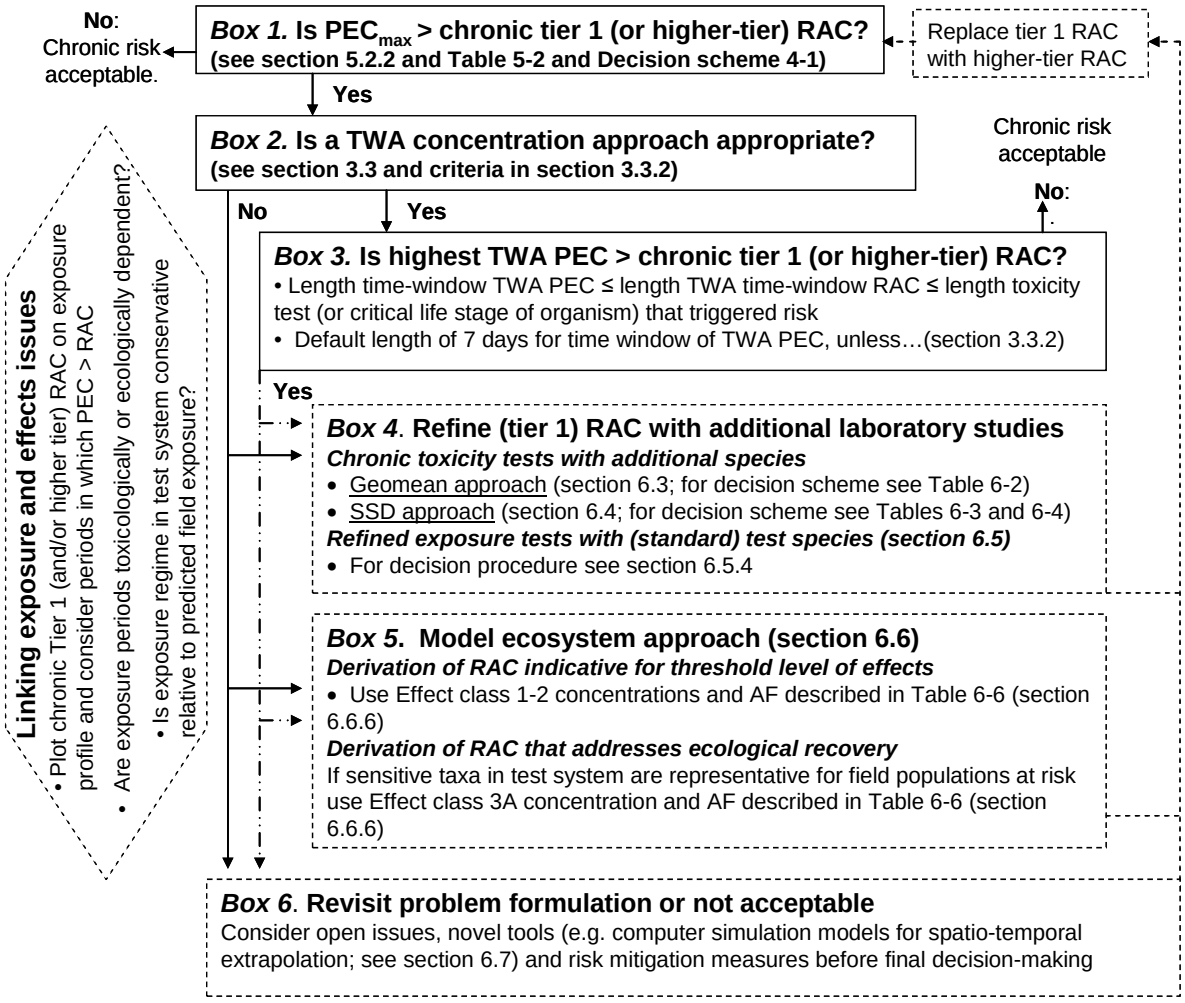
Lower and higher-tier acute RACs for invertebrates and fish and two hypothetical insecticides (Phantasithrin and Imagiphos). For each insecticide the RAC that needs to be selected for the final decision is given in bold.

	$RAC_{Tier\ 1}$	RAC_{SSD}	$RAC_{mesocosm}$
Phantasithrin (invertebrates)	0.020 µg/L	0.038 µg/L	**0.053 µg/L**
Phantasithrin (fish)	0.260 µg/L	0.580 µg/L	-
Imagiphos (invertebrates)	0.14 µg/L	0.62 µg/L	1.89 µg/L
Imagiphos (fish)	**1.51 µg/L**	-	-

4.4 Decision scheme for higher tier chronic risk assessment of toxicity

The basic flow chart for higher-tier chronic risk assessment of PPPs in Dutch drainage ditches is presented in Decision scheme 4-5. Again, in this scheme the specific report sections are mentioned where detailed guidance for decision making can be found. An important difference between the acute and chronic decision schemes is the presence of two boxes in the chronic decision scheme (Box 2 and Box 3 in Decision scheme 4-5) that refer to the Time Weighted Average (TWA) concentration approach.

In chronic risk assessments the PEC_{max} and, under certain conditions described in Section 3.3.2, a PEC_{TWA} may be used. If it is possible to use the TWA approach the ELINK recommendation is proposed (Brock et al., 2010a) by using a PEC_{TWA} with a 7-day time-window as default, at least if no specific information is available on the relation between exposure pattern and time-to-onset of effects for the relevant life stages of the organisms at risk. It may be scientifically justified to lengthen or shorten the default 7-d TWA period of the PEC when appropriate information on time-to-onset of effects is made available for elongated toxicity tests with relevant organisms. According to the ELINK methodology the time-window for the PEC_{TWA} should never be: (1) longer than the duration of the ecotoxicological test that triggered the risk or (2) longer than the duration of the life stage of highest ecotoxicological concern of the test organism (Brock et al., 2010a).



Decision scheme 4-5

Basic flow chart for higher-tier chronic risk assessment of PPPs in edge-of-field surface waters in the Netherlands.

5 Tier 1 risk assessment procedure for Dutch drainage ditches under Regulation 1107/2009/EC

5.1 Tiered approach

Ideally, when many scientifically underpinned methods are available and costs are not a limiting factor, environmental risk assessments can be performed by applying the best available methods. However, in practice environmental risk assessments are not based on an unlimited number of environmental fate and ecotoxicity data but on factors like pragmatism, costs, and efficacy. When both pragmatism and science drive the assessment, one can understand the development of tiered systems (Posthuma et al., 2008).

Tiered approaches are the basis of environmental risk assessment schemes that support the registration of plant protection products under the PPP Regulation (see e.g. Campbell et al., 1999; SANCO, 2002; Boesten et al., 2007). In this context a tier is defined as a complete effect or exposure assessment resulting in an appropriate assessment endpoint, e.g. PEC (Predicted Environmental Concentration) or RAC (Regulatory Acceptable Concentration). The concept of tiered approaches is to start with a simple conservative assessment and to only do additional more complex work if necessary (so it implies a cost-effective procedure both for industry and regulatory agencies). According to Boesten et al. (2007) and Solomon et al. (2008) the general principles of tiered approaches are:

- lower tiers are more conservative than higher tiers
- higher tiers aim at being more realistic than lower tiers
- lower tiers usually require less effort than higher tiers
- in each tier all available relevant scientific information is used
- all tiers aim to assess the same protection goal

In short, the tiered system as a whole needs to be: (i) appropriately protective, (ii) internally consistent, (iii) cost-effective and (iv) address the problem with a higher degree of realism and complexity when going from lower to higher tiers (see Figure 5-1).

An additional practical aspect of the tiered approach is that there has to be some balance between the efforts and the filtering capacity of a tier. For instance, it does not make sense to define a tier that requires 50% of the efforts of the next higher tier but leads in 95% of the cases to the conclusion that this next tier is needed (Boesten et al., 2007).

In pesticide risk assessment under the PPP Regulation the basic data requirements for the first tier risk assessment are strictly defined. Data requirements for the first tier effect assessment in the EU can be found in Annex II of the PPP Regulation (see Section 5.2).

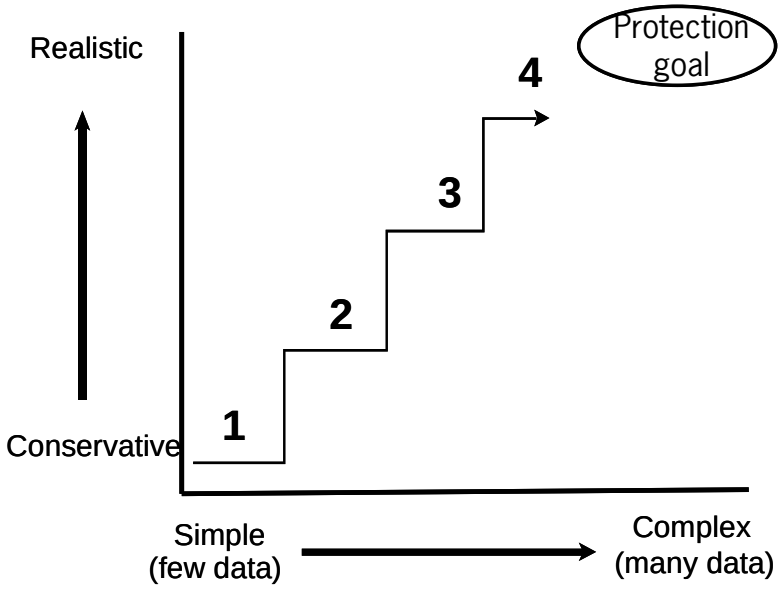


Figure 5-1

Tiers in the risk assessment process, showing the refinement of the process through the acquisition of additional data (redrafted after Solomon et al., 2008).

The 'unless'-clauses described in the Uniform Principles (Annex VI to Directive 91/414/EEC) offer the possibility to perform higher-tier risk assessments (EC, 1997). Procedures for higher-tier testing to evaluate the environmental risks of plant protection products to aquatic organisms before their marketing can be found in the Guidance Document on Aquatic Ecotoxicology (SANCO, 2002), which is currently updated by EFSA (European Food Safety Authority).

A logical consequence of the basic dossier requirements in the process of pesticide registration in the EU is that the risk assessment always starts with the first tier. However, the uncertainties and possible risks indicated by the first tier (and other lower-tiers) inform the risk assessors and risk managers on which organisms and methods to focus in the higher-tier risk assessment. Another logical consequence of the principles of the tiered approach described above is that the highest tier that can be applied in the registration procedure acts as a reference to calibrate for the lower tiers, because the assessment endpoint derived from a higher tier is closer to the actual objectives of the adopted protection goal. In the aquatic effects assessment for pesticides an appropriate mesocosm test (in the sense that it is representative for the target image of the aquatic ecosystem at risk), in combination with an appropriate AF or model for spatio-temporal extrapolation, may be the highest tier when invertebrates or primary producers are at risk. Appropriate intermediate tiers may be refined exposure studies with standard test species and the species sensitivity distribution (SSD) approach based on additional toxicity data with potentially sensitive species.

As explained above, the uncertainties and possible risks indicated by the first tier determine on which organisms and methods to focus in the higher-tier risk assessment. For example, if the first tier effects assessment for an insecticide indicates that the standard test arthropods are at least an order of magnitude more sensitive than the other standard test species (e.g. algae, fish) the higher-tier tests may focus on aquatic arthropods by performing e.g. additional laboratory toxicity tests or microcosm/mesocosm experiments. If these tests lead to a refined RAC for arthropods one has to check whether this refined RAC is still protective for other organisms not at risk in the first tier (e.g. fish). Consequently the tiered approach has to adopt an iterative procedure.

5.2 Tier 1 - Uncertainty Factor approach

In the Tier 1 aquatic risk assessment, information from acute and chronic studies with a selection of standard test species is used to identify potential areas that should be further evaluated. Regulatory Acceptable Concentrations (RACs) are derived for each test species by dividing the test endpoint (e.g. LC_{50}, EC_{50}, NOEC) by the uncertainty factor. For each of the tested species, the RAC should be higher than the PEC (see Section 3.3) to decide that no unacceptable effects are to be expected from the proposed use of the PPP. However, when any of the RACs is lower than the PEC, further information is needed to draw conclusions whether or not authorisation can be granted. The information that is required for the Tier 1 assessment and the derivation of the RACs is outlined below, the accompanying decision scheme is presented in Chapter 4.

In principle, the Tier 1 assessment should be performed for the active substance as well as for any major metabolite (i.e. metabolites for which the concentration in in the water phase in water/sediment studies at any point in time is ≥10% of the added amount). Data on the toxicity of the product should be submitted, if the toxicity of the plant protection product cannot be predicted from the data on the active substance.

5.2.1 Acute (short-term) risk assessment

Acute studies with aquatic organisms include 48 or 96 hours test with fish or invertebrates. Tests with algae are also short-term studies. In view of the generation time of algae, however, the endpoint is considered to refer to chronic effects rather than acute and is included in the chronic dataset (See 5.2.2). Table 5-1 summarises the short-term aquatic ecotoxicity tests that should be submitted for authorisation, and the RACs that are derived from those tests. The tests indicated in bold represent data that should always be submitted. Additional testing is required for specific products or situations. Please note that Table 5-1 describes the situation based on the latest draft Annex II, items that are still under discussion are indicated (see table notes).

Table 5-1

Endpoints available from short-term aquatic toxicity tests; basic dossier data are indicated in bold (based on the updated Annex II).

Taxonomic group	Species/test system	Duration	Endpoint	RAC	Notes
Algae	**green algae (e.g. *Pseudokirchneriella subcapitata*)**	**72 h**	**EC_{50}**	**EC_{50}/10**	1
Algae	Blue green algae / diatom	72 h	EC_{50}	EC_{50}/10	1,2
Macrophyte	*Lemna* sp. or *Myriophyllum* sp. or *Glyceria maxima*	7 d - 14 d	EC_{50}	EC_{50}/10	1,2,3
fish	***Oncorhynchus mykiss***	**96 h**	**LC_{50}**	**LC_{50}/100**	
fish	warm water species	96 h	LC_{50}	LC_{50}/100	4
crustacean	***Daphnia* sp. (*D. magna* preferred)**	**48 h**	**EC_{50}**	**EC_{50}/100**	
crustacean	Additional species, e.g. *Americamysis bahia*	48 h	EC_{50}	EC_{50}/100	5
insect	*Chironomus riparius*	48 h	EC_{50}	EC_{50}/100	6

1 Under 91/414/EC, the RAC for algae and macrophytes is calculated from the EC_{50} with an assessment factor of 10. The endpoint and height of the AF to be used under the new regulation is not yet fixed, it might be the EC_{50} with a factor of 10 or 100, or the NOEC with a factor of 10 (see also Note 7 to Table 5-2). There is also on-going discussion as to whether specific growth rate or biomass should be taken as test parameter. According to OECD 201 (OECD, 2006a), the former is preferred from a scientific point of view. According to OECD 221 (OECD, 2006b) both specific growth rate and yield are required in EC_x calculations for *Lemna.*

2 Required for herbicides, plant growth regulators and fungicides with a herbicidal action.

3 Additional testing may be required on other macrophyte species (*Myriophyllum* sp. or *Glyceria maxima*) depending on the mode of action of the substance, or if clear indications of higher toxicity are apparent to dicotyledonous (e.g. auxin inhibitors, broad leaf herbicides) or other monocotyledonous (e.g. grass herbicides) plant species from efficacy or testing with terrestrial non-target plants.

4 For animal welfare reasons testing of the warm water species is not obligatory anymore in the updated Annex II but this endpoint is often available in the dossier.
5 Required for insecticides or compounds with insecticidal activity; alternatively, other more relevant freshwater non-crustacean species, e.g. *Chironomus* spp. may be used if guidelines or protocols are developed.
6 Only required if guidelines or protocols for the test with this species are developed.

5.2.2 **Chronic (long-term) risk assessment**

Accepted test protocols are available for a number of organism groups. Early life stage (ELS) tests with fish consider the sensitive embryonic and juvenile life stages. The fish Full life cycle (FLC) test is a two-generation test that starts with an adult parental generation and continues until sexual maturation of the F2 generation. The studies with Crustacea and insects are focused on reproductive output of one parental generation. Studies with plants are not so much focused on population-level endpoints, but include sub-lethal endpoints of individual plants. The long-term studies that are required for product authorisation are summarised in Table 5-2, and it is indicated how the RAC is derived. As for the short-term studies, some tests should always be submitted. Additional testing is required in certain cases (see table notes). According to the accepted guidelines, the endpoints obtained from chronic studies are generally expressed as NOECs. For the purpose of deriving the RAC, NOEC and EC_{10} are considered to be interchangeable. Please note that Table 5-2 describes the situation based on the on the latest draft Annex II, items that are still under discussion are indicated (see table notes).

Table 5-2

Endpoints available from long-term aquatic toxicity tests; basic dossier data are indicated in bold (based on the updated Annex II).

Taxonomic group	Species/test system	Duration	Endpoint	RAC	Notes
fish	ELS- test		NOEC (EC_{10})	NOEC/10	1
fish	FLC-test		NOEC (EC_{10})	NOEC/10	2
crustacean	***Daphnia* sp. or additional species**	**21 d**	**NOEC (EC_{10})**	**NOEC/10**	1,3
insect	*Chironomus riparius* (water spiked preferred)	20-28 d	NOEC (EC_{10})	NOEC/10	4
algae	**green algae (e.g. *Pseudokirchneriella subcapitata*)**	**72 h**	**NOEC (EC_{10})**	**NOEC/10**	7
algae	blue green algae/diatom	72 h	NOEC (EC_{10})	NOEC/10	5,7
macrophyte	*Lemna* sp. or *Myriophyllum sp.* or *Glyceria maxima*	7 d - 14 d	NOEC (EC_{10})	NOEC/10	5,6,7

1 Required when exposure of surface water is likely and the compound is deemed stable in water (<90% loss by hydrolysis over 24 h); an ELS test is not necessary when a FLC-test is submitted.
2 Required when BCF_{fish} >1000 L/kg **and** <95% depuration in 14 d **and** $DT_{90,system}$ >100 d, or when other data (e.g. suggesting endocrine disruption) indicate need for FLC-test.
3 Chronic test should be performed with most sensitive species in acute tests if the difference in acute EC_{50} values between *Daphnia* and additional species is larger than an order of magnitude.
4 May be necessary in case of compounds that interfere with moulting (insect growth regulators, IGR) or any other type of compound that has a target specific for insects, or if the compound accumulates in sediment; in the latter case also a sediment spiked test is possible (test should take account of the major route of exposure).
5 Required for herbicides, plant growth regulators and fungicides with a herbicidal action.
6 Additional testing may be required on other macrophyte species (Myriophyllum sp. or Glyceria maxima) depending on the mode of action of the substance, or if clear indications of higher toxicity are apparent to dicotyledonous (e.g. auxin inhibitors, broad leaf herbicides) or other monocotyledonous (e.g. grass herbicides) plant species from efficacy or testing with terrestrial non-target plants.
7 Chronic endpoints and associated assessment factors for algae and macrophytes to be used under 1107/2009/EC are under discussion, see also note 1 to Table 5-1.

5.3 Bioconcentration and secondary poisoning

5.3.1 Bioconcentration in fish

For lipophilic compounds, special attention should be paid to the potential risks resulting from bioconcentration in fish. Determination of the bioconcentration factor (BCF) according to OECD 305 is required for compounds that have a log K_{ow} >3 and that are considered stable (*i.e.* <90% loss of the original substance over 24 hours via hydrolysis). A potential risk is identified when the experimental BCF is >100 L/kg for compounds that are not readily biodegradable, or when the experimental bioconcentration factor (BCF) is >1000 L/kg for compounds that are readily biodegradable. In these cases, authorisation cannot be granted unless results of chronic studies do not indicate an unacceptable risk.

An ELS-test is available for compounds with:
- 100<BCF<1000 or
- BCF >1000 L/kg, and >95% depuration within 14 days or $DT_{90,\ system}$ <100 d.

An FLC-test should only be performed for compounds with:
- BCF >1000 L/kg, and <95% depuration within 14 days and $DT_{90,\ system}$ >100 d.

An unacceptable risk is present when the RAC_{ELS} or RAC_{FLC} is lower than the PEC. Guidance on the choice of the appropriate PEC is given in Section 3.3.

5.3.2 Secondary poisoning

In addition to potential effects on fish special attention should also be paid for potential transfer of lipophilic compounds through the food chain (see Figure 5-2). Bioconcentration (i.e. uptake from water), bioaccumulation (i.e. uptake from water and food) and biomagnification (i.e. increasing concentrations with trophic level) often correlate with lipophilicity. For organic chemicals, a log $K_{ow} \geq 3$ indicates a potential for bioaccumulation. If this condition is met, a risk assessment for secondary poisoning should be carried out. For the aquatic system this risk is assessed for a fish eating bird with a body weight of 1000 g and a fish eating mammal with a body weight of 3000 g.

As bioaccumulation processes often are slow and substances may be persistent, a long-term assessment is appropriate. Relevant metabolites must also be considered. For background information with regard to food chain modelling see Romijn et al. (1993, 1994), Traas et al. (1996), Jongbloed et al. (1996) and Luttik (2003). The stepped approach for assessing the bioaccumulation potential presented below is according to the guidance document for birds and mammals (EFSA, 2008).

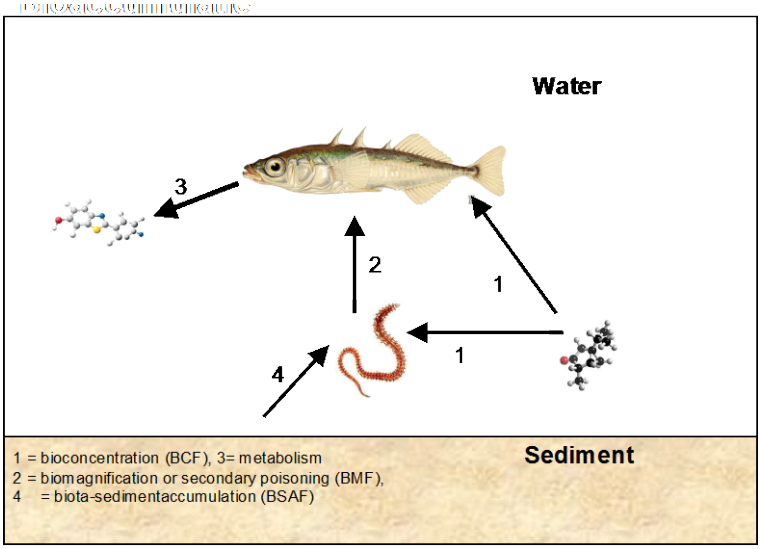


Figure 5-2
The different processes resulting in bioaccumulation of compounds in organisms.

5.3.3 RAC for secondary poisoning (RAC_{sp})

Assuming a food chain from fish to fish-eating birds or mammals, EFSA (2008) proposes a simple worst-case risk assessment in which the exposure of birds and mammals is calculated from the expected residues in fish. To that end, the highest appropriate $PEC_{TWA, 21\text{-}d}$ is selected from the environmental fate section, and multiplied by the whole-body bioconcentration factor of fish to give the Predicted Environmental Concentration in fish:

$$PEC_{\text{fish}} = PEC_{TWA,\,21d} \times BCF_{fish}$$

with
PEC_{fish} = concentration in whole fish [mg/kg]
$PEC_{TWA,\,21\,d}$ = time weighted average PEC in water over 21 days [mg/L]
BCF_{fish} = whole body bioconcentration factor in fish [L/kg]

Note that the default time window of 21 days is chosen unless on basis of scientific reasoning a shorter time window is more appropriate (EFSA, 2008). Then, the PEC_{fish} (in mg/kg) is converted to a daily dose for mammals and birds by multiplying with 0.137 (mammals) and 0.205 (birds) respectively, and compared with the relevant long-term no-adverse-effect-level (NOAEL, in mg/kg bw per day). Multiplications are based on a 3000-g mammal eating 425 g fresh fish per day, and a 1000-g bird eating 159 g per day, according to Smit (2005). The ratio between the relevant NOAEL and the daily dose in fish is denoted as the Toxicity Exposure Ratio (TER), and compared with the appropriate trigger value of 5. For TER ≥5, no further action is required, for TER <5, refinement is needed. Note that the TER-approach with trigger 5 is the reciprocal of an Exposure Toxicity Ratio ('PEC/PNEC') with trigger 0.2.
Refinement options are for instance:
- The use of refined models for calculating exposure concentrations in the surface water,
- The use of measured concentrations either in the surface water or in fish, or
- Modelling of the internal body burden of fish using information on uptake and elimination kinetics in fish as well as information on dissipation kinetics in water, rather than assuming equilibrium and calculating BCF value.

Within the context of this report, instead of calculating the TER, preference is given to the derivation of the Regulatory Acceptable Concentration in water for secondary poisoning (RAC_{sp}), which can be compared with the time weighted average PEC. Using the same input as described above, the following calculations are made:

The relevant long-term no-adverse-effect-level (NOAEL, in mg/kg bw per day) is divided by the assessment factor of 5 to give the 'regulatory acceptable dose', and converted into a concentration in fish by dividing by a factor of 0.205 for birds or 0.137 for mammals. Then, the resulting 'regulatory acceptable concentration in fish' is divided by the BCF_{fish} to yield the corresponding concentration in water. This RAC_{sp} relates to the 21-days TWA concentration in water, unless scientific reasoning indicates otherwise. If the 21-days TWA PEC is higher than the RAC_{sp}, further refinement is necessary. If the 21-days TWA PEC is lower than the RAC_{sp}, no further action is needed. Written in formula, the RAC_{sp} in surface water for fish eating birds and mammals is derived as follows:

$$RAC_{SP} = \frac{NOAEL_{bird}}{5 \times 0.205 \times BCF_{fish}} \text{ or } \frac{NOAEL_{mammal}}{5 \times 0.137 \times BCF_{fish}}$$

with

RAC_{sp}	=	Regulatory Acceptable Concentration in water for secondary poisoning [mg/L]
$NOAEL_{birds}$	=	relevant long-term no-adverse-effect-level [mg/kg bw per d]
BCF_{fish}	=	whole body bioconcentration factor in fish [L/kg]

This RAC_{sp} should be compared with the 21-days TWA PEC in surface water. If RAC_{sp} >21-d TWA PEC_{sw}, no further action is required. If RAC_{sp} <21-d TWA PEC_{sw}, refinement is necessary.

5.3.4 **Biomagnification**

According to the aquatic guidance document (SANCO, 2002), biomagnification has to be taken into account for compounds that meet the trigger for a FLC-test, namely the BCF (whole body) >1000 and the elimination of radioactivity during the 14 day depuration phase in the bioconcentration study is <95% and the substance is stable in water or sediment (DT_{90} >100 days). The guidance document states that if these triggers are met, detailed food chain modelling should be performed, or microcosm/mesocosm studies, which implicitly take into account biomagnification, should be submitted. However, the methodology for food chain modelling as proposed in SANCO (2002) is very complicated and requires a lot of input data. Furthermore, including fish in microcosm/mesocosm experiments can present difficulties and needs to be carefully considered. It is therefore proposed to consider food chain modelling as an option for higher tier assessment. As a first tier, the methodology of the TGD (EC, 2003) and EQS-guidance (EC, 2011) can be adopted, by performing the risk assessment using default biomagnification factors. The TGD proposes the following factors, related to BCF and/or log K_{ow} (Table 5-3).

Table 5-3

Default BMF values for organic substances.

BCF (fish)	log K_{ow} of substance	BMF_1
<2000		1
2000-5000		2
>5000		10
	<4.5	*1*
	4.5-<5	*2*
	5-8	*10*
	>8-9	*3*
	>9	*1*

Note that for compounds with log K_{ow} ≥3 an experimental BCF will always be available, so the selection of BMF based on log K_{ow} is not relevant.

From this table it can be seen that biomagnification may be relevant for compounds with a BCF ≥2000 L/kg. For these compounds, the appropriate BMF will be selected from Table 5-3 and the RAC_{SP} will be derived according to the following formula:

$$RAC_{SP} = \frac{NOAEL_{bird}}{5 \times 0.205 \times BCF_{fish} \times BMF_1} \text{ or } \frac{NOAEL_{mammal}}{5 \times 0.137 \times BCF_{fish} \times BMF_1}$$

with

RAC_{sp} = Regulatory Acceptable Concentration in water for secondary poisoning [mg/L]
$NOAEL_{birds}$ = relevant long-term no-adverse-effect-level [mg/kg bw per d]
BCF_{fish} = whole body bioconcentration factor in fish [L/kg]
BMF_1 = biomagnification factor from Table 5-3 [kg/kg]

This RAC_{sp} should be compared with the 21-days TWA PEC in surface water. If RAC_{sp} >21-d TWA PEC_{sw}, no further action is required. If RAC_{sp} <21-d TWA PEC_{sw}, refinement is necessary. In that case, a higher tier assessment should be carried out and the foodchain modelling approach of the aquatic guidance document (SANCO, 2002) can be followed.

6 Higher tier risk assessment procedures for drainage ditches in line with the PPP Regulation

6.1 Introduction

Additional information is needed in case potential risks are identified in the first tier assessment. The test strategy for the higher tier assessment depends on the areas of concern. A cost-effective option may be to perform tests with additional species (Section 6.3). When the additional dataset is large enough, statistical extrapolation techniques can be applied (Species Sensitivity Distribution; Section 6.4). Performing toxicity tests with refined exposure (Section 6.5) may be an option when the standard laboratory tests with constant exposure do not adequately reflect the predicted exposure under the conditions of use. Model ecosystem studies (Section 6.6) can be seen as the alternative when additional laboratory data do not remove the concern for potential risks. It should be noted, however, that for an adequate higher tier risk assessment according to the PPP regulation jumping from the first tier to the model ecosystem tier may be acceptable, also when no data on additional laboratory toxicity tests are made available. Before going further into the higher tier risk assessment methods with additional species, the next section discusses the subject of dealing with additional data from marine species.

6.2 Dealing with additional data from marine species

For the preparation of the new EQS-guidance, a background document was prepared on the use of ecotoxicity data for freshwater and marine species for derivation of quality standards for fresh- and saltwater ecosystems. This issue is further discussed in Section 7.3.1. Based on the information presented there, the following procedure is proposed for RAC derivation. Where there are sufficient toxicity data for the relevant taxonomic group in both the freshwater and marine datasets a statistical comparison should be made. The null hypothesis is that freshwater and marine organisms of the relevant taxonomic group do not differ in their sensitivity to the compound of interest; i.e. they belong to the same statistical population:

1. Especially for PPPs with a specific mode of action, it is important to identify particularly sensitive taxonomic groups and perform separate statistical analysis for the relevant taxonomic groups. If for the relevant taxonomic group(s) (e.g. crustaceans, arthropods, fish, vertebrates) enough data are available, this may help to determine if there are differences between freshwater and marine species. Note that there are only few marine insects.

2. All freshwater data of the relevant taxonomic group are collected and tabulated (note: this data set contains one toxicity value per species for the ecotoxicologically relevant endpoints[1]). Next, a logarithmic transformation of each of these toxicity values is performed.
3. All marine data of the relevant taxonomic group are collected and tabulated (note: this data set contains one toxicity value per species). Next, a logarithmic transformation of each of these toxicity values is performed.
4. Using an F-test, determine whether the two log-transformed data sets have equal or unequal variances. Perform the test at a significance level (α) of 0.05.
5. A test for differences between the data sets e.g. a two tailed t-test where the data are normally distributed (with or without correction for unequal variances, depending on the results of step 3), is performed. Perform the test at a significance level (α) of 0.05 .

In those cases where there are too few data of the relevant taxonomic group (either freshwater or marine) to perform a meaningful statistical comparison and there are no further indications (spread of the data, read-across, expert judgement) of a difference in sensitivity between freshwater vs. marine organisms of the relevant taxonomic group, the data sets may be combined for RAC derivation. It should be noted that not all species in either freshwater or saltwater have closely related species in the other compartment. For instance, there are not many truly pelagic marine insects (i.e. species that are fully living in seawater), although a number of species are known from intertidal or estuarine ecosystems. For those insecticides for which crustaceans have a low sensitivity, a comparison of sensitivity of freshwater and marine species may probably not be made on the basis of insect data. It can be expected, however, that in this case the potentially most sensitive species group will be covered by the freshwater dataset. On the other hand, there are some exclusively marine taxonomic groups that are sometimes used in toxicity tests. Echinodermata (sea stars, sea urchins) are an example of these. If the lowest endpoint for a PPP is found for an exclusively marine species that has no freshwater relatives, it may be considered not to base the risk assessment for drainage ditches on that endpoint. Similarly, the use of marine mescocosms for risk assessment of PPPs should be carefully considered. In general, it is proposed to use marine data only in addition to freshwater data. In practice, this means that a single marine mesocosm without any equivalent freshwater studies will only be used as supportive evidence, but not as the sole basis for the RAC.

[1] First calculate the geometric mean of multiple comparable toxicity values for the same species and the same endpoint. This can only be done if there are no indications that the difference in toxicity values is caused by differences in e.g. test conditions or life stages. If multiple toxicity values or geometric means for different endpoints are available for one species, the endpoint for which the lowest value is obtained is selected. When, after primary selection, multiple valid toxicity data for one species are left that can not be averaged, the lowest value is selected. Example: There are NOECs or EC_{10} values for three different endpoints, derived from several chronic studies with *Daphnia magna*. The geometric mean of NOECs for reproduction is 0.49 mg/L, the geometric mean of NOECs for mortality is 3.1 mg/L and there is a single EC_{10} value for growth of 0.67 mg/L. The geometric mean value of 0.49 mg/L for reproduction is selected for the aggregated datatable. See EC (2011) and Van Vlaardingen & Verbruggen (2007) for details.

6.3 How to derive a RAC with (a limited number of) additional single species toxicity tests

If additional species (not belonging to the standard test species mentioned in Chapter 5) are tested, it is necessary to consider which toxicity value should be used in the risk assessment, at least if the number of available toxicity data was not high enough to apply the Species Sensitivity Distribution approach. According to the present guidance, 'if a considerable number of additional species was tested in valid studies, then it is possible that the uncertainty factors that are applied to the lowest toxicity value could be lowered by up to an order of magnitude' (SANCO, 2002). It is not further specified how many additional data would be needed to allow for lowering the assessment factor, and to our experience, this option is not often applied in practice. Although more species are tested and thus information on the differences in sensitivity between species is available, the risk assessment is most often still based on the most sensitive species using the default assessment factors. The number of species to be tested according to pesticide legislation effectively sets the first-tier level of protection in the effects assessment. Consequently, when more data are available and the risk assessment is still based on the lowest value without adjusting the trigger value, the average level of protection may exceed the level implied by the provisions of the PPP Regulation.

6.3.1 Approaches considered by EFSA

In 2005, the EFSA PPR Panel published an opinion on the approaches to deal with additional toxicity data, taking into account that the same average level of protection should be maintained.

Option 1
As a first option, the PPR Panel proposed an alternative approach of taking the geometric mean of comparable endpoints within a taxonomic group when more than one species is tested, where the legislation only required one species (EFSA, 2005b). It was shown that this would ensure at least the same average level of protection as implied by the Directive, and avoid most of the increase in conservatism when additional species are tested. This was based on the assumption that toxicity data were normally distributed on a logarithmic scale. Later research (EFSA, 2008) showed that this is true for a wide range of distributions that are symmetric and unimodal (single peak) on a logarithmic scale, and also for asymmetric unimodal distributions where the long tail is to the left. It is also true for asymmetric distributions with long tails to the right[2] and for some examples of bimodal distributions, provided that the standard uncertainty factor includes sufficient allowance for between-species variation in toxicity, which seems likely.

The latter work is mainly based on distributions of acute toxicity data. It remains to be investigated whether the same procedure can be used for chronic toxicity data as well. NOECs may be over/underestimates (e.g. due to wide dose spacing and limited power to detect effects often caused by small sample size). The PPR Panel recommended, however, using the geometric mean for both acute and reproductive toxicity, when multiple species are tested within a taxonomic group. The first tier AF of 10 or 100 should be applied to this geometric mean value of available toxicity data to derive a RAC.

It should be noted that 'taxonomic group' can be interpreted in different ways. For instance, crustaceans and insects represent different taxonomic groups on the phylum level but are sometimes grouped into the

[2] Distributions of acute toxicity data often have long tails to the right on the natural scale, but this is reduced or removed on the logarithmic scale, which is used for the geometric mean.

taxonomic group of arthropods. The default approach should be to treat them as different groups unless scientific arguments can be raised to consider them as one group.

Option 2
For those organism groups where the legislation requires that at least two species are tested, this implies a higher level of protection in the effects assessment in the first tier. In this case, a different procedure is required when additional species are tested. The minimum is then replaced by the i^{th} lowest toxicity value depending on the sample size available, and divided by the current assessment factor (Method 2 described in EFSA, 2005b).

Procedure when more than two species are tested within a taxonomic group:
1. Order the data so that the values are increasing.
2. Choose from the ordered data the i^{th} value where i is determined from Table 6-1 according to the sample size.
3. Divide the obtained data value by the current assessment factor.

This table seems to be less relevant if toxicity data for more than five fish species or more than eight taxa of the sensitive non-vertebrate taxonomic group are available, because in that case the SSD approach seems to be a more scientific solution.

Table 6-1
Sample-size dependence of order statistics to be used with current assessment factors to achieve at least the same level of protection as the current procedure for a sample of size 2.

Sample size	Position in ordered list of data (i)
3-4	1
5-7	2
8-10	3
11-13	4

For deciding whether species belong to the same taxonomic group, again the default approach should be to treat them as different groups unless scientific arguments can be raised to consider them as one group.

6.3.2 Proposal for the derivation of RACs when a limited number of additional single species toxicity tests is available

In some cases additional ecotoxicity data may be available, but their number is too low to apply the SSD approach. For this situation, it is proposed to use the geometric mean of the available toxicity values within a taxonomic group (Option 1 above; Table 6-2).

Table 6-2

Proposal for the derivation of RACs for aquatic organisms when a limited number of additional single species toxicity tests is available. When applying this approach scientific arguments should be given why the selected toxicity data (on which the geomean is based) concern the same taxonomic group relevant for the risk assessment.

Taxonomic group	Number of toxicity data for different taxa of the relevant taxonomic group	RAC	Field exposure concentration
Fish and/or other aquatic vertebrates	< five acute LC_{50}'s	Geomean LC_{50}/100	PEC_{max}
Fish and/or other aquatic vertebrates	< five chronic NOECs (or chronic EC_{10}'s)	Geomean NOEC/10	PEC_{max} or PEC_{TWA}
Invertebrates and/or primary producers	< eight acute EC_{50}'s	Geomean EC_{50}/100	PEC_{max}
Invertebrates and/or primary producers	< eight chronic NOECs (or chronic EC_{10}'s)	Geomean NOEC/10	PEC_{max} or PEC_{TWA}

The benefit of this approach is that all species groups are treated in the same way and that methods do not have to change in the future when more than one standard test species is required for a particular group of species, which implies that in the new situation the level of protection achieved will be different compared to the old situation. The first tier AF of 10 or 100 should be applied to this geometric mean value of available toxicity data to derive a RAC.

6.4 Species Sensitivity Distribution (SSD) Approach

6.4.1 General introduction to the SSD concept

As a result of direct toxicity, species vary markedly in their sensitivity to pesticides. This variation in direct toxicity can be described by constructing a Species Sensitivity Distribution (SSD). The SSD is a statistical distribution estimated from a sample of laboratory toxicity data and visualised as a cumulative distribution function (see Figure 6-1).

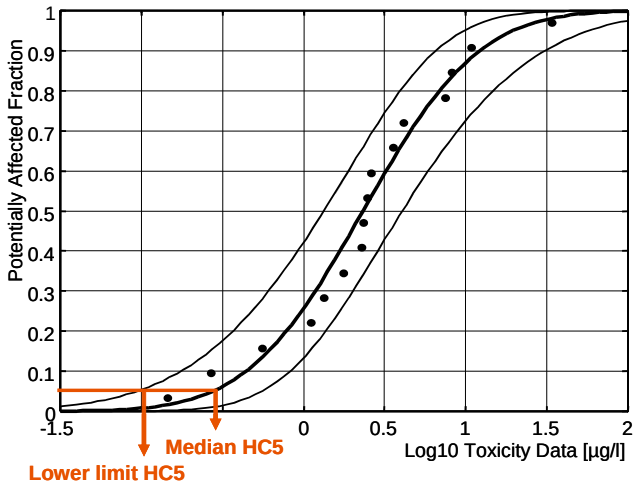


Figure 6-1

Graphical presentation of the Species Sensitivity Distribution curve, its 90% confidence interval, and the derivation of the lower limit and median Hazardous Concentration to 5% of the species (HC_5).

SSDs are used to calculate the concentration at which a specified proportion of species are expected to suffer direct toxic effects. These concentrations, the hazardous concentrations, are expressed as HC_x values and represent the value that affects a specific proportion (x %) of species. When compared with the first-tier effects assessment on the basis of standard test species, SSDs have the advantage of making more use of the available laboratory toxicity data for a larger array of species. They describe the range of sensitivity rather than focusing on a single value, they enable estimates to be made of the proportion of the species affected at different concentrations, and they can be shown together with confidence limits showing the sampling uncertainty due to the limited number of species tested. They can be used in a deterministic risk assessment by taking an appropriate percentile from the SSD, or in a probabilistic risk assessment by using the whole SSD (EFSA, 2006)

The use of the SSD concept in ecological risk assessment is based on several assumptions (Versteeg et al., 1999; Posthuma et al., 2002; Forbes and Calow, 2002a; Van den Brink et al., 2008). These are listed by Brock et al. (2010a):

1. The sample of the species on which the SSD is based is a random selection of the community of concern, and is herewith representative for this community.
   The SSD usually does not represent a known community, but is often interpreted as if it does (Forbes and Calow, 2002a).
2. Interactions among species do not influence the sensitivity distribution.
3. Since most ecotoxicological tests relate to structural endpoints, community structure rather than function is the focus of the SSD.
   This assumption suggests that by protecting the structure of a community, its functions, including energy flows within food webs, are also protected.
4. The laboratory sensitivity of a species approximates its field sensitivity.
   The general conclusion of comparisons between laboratory and (semi-)field sensitivity of species is that, when exposure and developmental stages of the organisms are similar, the laboratory sensitivity of a species to the pesticides evaluated is representative for its field sensitivity.
5. The endpoints measured in the toxicity tests on which the SSD is based are ecotoxicologically relevant.
   This assumption implies that the endpoints measured in the toxicity tests on which the SSD is based must be toxicologically and ecologically relevant.
6. Since in SSDs all species have equal weight it is assumed that all species are equally important for the structure and functioning of the ecosystem of concern.
   For the SSD concept all species are considered to be equal, although we know that some species are more important for the functioning of ecosystems than others, the so-called keystone species.
7. The real distribution of the sensitivity of the community is well modelled by the selected statistical distribution.
   Logistic and lognormal distributions are most often used (Aldenberg and Jaworska, 2000; Aldenberg et al., 2002), because they require less data than distribution-free methods and are relatively easy to fit with standard statistical software (Van den Brink et al., 2008).
8. The number of species data used to fit the distribution is adequate from a statistical, ecological and animal welfare point of view, to describe the real distribution of the sensitivity of the community.
9. The protection of the prescribed fraction of species (e.g. HC_5 or $HC_{1)}$ ensures an 'appropriate' protection of the structure of ecosystems.
10. The validity of the toxicity data used to construct the SSD is assured.

For construction of SSDs, the programme ETX2.0 can be used (Van Vlaardingen et al., 2004). This programme also contains several statistical tools to test the assumptions of normality (see Point 7 above). It should be noted, however, that the performance of these tests strongly depends on the number of data. With a relatively low number of data, a distribution is often accepted as normal, whereas for large datasets deviations from normality will be more easily detected. The outcome of the tests as such should therefore not be used as

a single criterion to decide whether or not the SSD can be applied, or to split datasets to construct specific SSDs for particular taxonomic groups (see 6.4.4). A thorough evaluation of the individual data points and visual inspection of the fit may reveal whether or not violation of the assumptions concerning the distribution is acceptable. For example, violation of the goodness-of-fit test may be acceptable from a regulatory point of view when the fitted distribution in the tail of the SSD is relatively worst case compared to the data points (in the sense that most of the toxicity data around the HC_5 and lower are on the right side of the fitted curve).

6.4.2 Criteria for the selection of acute and chronic toxicity data for aquatic invertebrates and plants

The number of species data used to fit the distribution has to be adequate from a statistical point of view. Suter et al. (2002) concluded that SSDs could be adequate with data points between 3 and 30, dependent on the method used. In the HARAP guidance document Species Sensitivity Distributions are recommended to be based upon a minimum of either eight acute or eight chronic toxicity data for different taxa that are representative for the sensitive taxonomic group, at least if the SSD is not exclusively constructed with toxicity data for fish. An SSD that addresses the sensitivity of fish should be based on a minimum of 5 toxicity data for different fish species (Campbell et al., 1999; see Section 6.4.8).

The endpoints measured in the toxicity tests on which the SSD is based must be toxicologically and ecologically relevant. Acute toxicity data mostly address mortality and immobility as the most frequently studied endpoints for animals, while that is biomass and growth for primary producers. Chronic toxicity data mostly address reproduction, feeding and growth as the most frequently studied endpoints in animals, and again this is biomass and growth for primary producers.

The test duration might be a criterion to be applied for the selection of the toxicity data, however, test duration is taxon and guideline dependent and, as a consequence, a range of test durations for different organisms is often included in the same SSD.

Measurement parameters, from which endpoints are calculated, should preferably be sensitive/responsive in the range of tested concentrations such that SSDs avoid the use of greater or lower than values.

In general, it is not recommended to include unbound values (greater than- or lower than-values) in the SSD. There are situations, however, where ignoring those data would lead to a loss of valuable information. When a <-value is lower than the lowest toxicity endpoint, this means that the other data do not cover the whole range of sensitivities. Leaving out this information might lead to an HC_5 that is underprotective. It is preferred to deal with this issue by adapting the assessment factor to the HC_5. However, to demonstrate the effect of including the information in the SSD, the following procedure can be applied:

- if in a set of available toxicity values for a certain species a greater or lower than value is present, this value will not be used in case the value is inside the range of values but will be used as such (without the < or > sign) in case the value was outside the range;
- if in a set of available toxicity values for a compound for a particular species only a greater or lower than value is present this value will only be used as such (without the < or > sign) in case this value is outside the range of all other values (for other taxa).

If an SSD is used in which unbound values are included, this should always be motivated. Future studies should try to build in test concentrations to avoid greater or lower than values.

In addition, the use of biochemical endpoints or biomarkers in SSDs is not recommended for regulatory purposes due to difficulties in correlating results with tangible ecological effects.

6.4.3 Plant protection products with a specific and non-specific toxic mode of action

SSDs can be based on either acute or chronic toxicity data. According to the HARAP Guidance Document (Campbell et al., 1999), the toxic mode-of-action of a pesticide should be taken into account when constructing SSDs to derive acceptable concentrations. In case of herbicides, vascular plants and/or algae usually comprise the most sensitive groups. For photosynthesis inhibitors the sensitivity distributions between algae and aquatic macrophytes are similar. For insecticides, arthropods (crustaceans and insects) usually are most sensitive. For fungicides, often a range of taxonomic groups are among the sensitive organisms. The next paragraphs give an overview of the sensitive organisms for herbicides, insecticides and fungicides.

6.4.3.1 Herbicides

At environmentally realistic exposure concentrations, herbicides specifically and mainly affect primary producers in aquatic ecosystems, i.e. algae and macrophytes. SSDs are potentially useful tools to determine the relative sensitivity of a range of species to a test substance and, in particular, as a means of comparing the sensitivity of the current Tier 1 standard test species (*Lemna* and standard algal species) with that of other species of primary producers. The AMRAP guidance document (Maltby et al., 2010) gives guidance on the use of the macrophyte data in the SSD approach and defines areas of uncertainty, which are specifically associated with the selection of species and endpoints.

Species selection for SSDs with primary producers

It is not yet clear for which type of herbicides algae and macrophyte data can be combined in the same SSD. Brock et al. (2004) and Van den Brink et al. (2006) show that for photosynthesis inhibitors macrophyte and algae data can be combined in one SSD. Further work of the AMRAP SSD Working Group will generate more recommendations for compounds with other toxic modes-of-action.

For the construction of macrophyte SSDs the AMRAP guidance document (Maltby et al., 2010) recommends that a range of morphologically and taxonomically different macrophytes should be included. Ideally, SSDs should be based on toxicity values for comparable measurement endpoints generated from tests conducted under similar exposure scenarios and exposure durations, preferably using standardised protocols. However, due to the diversity of aquatic plant morphologies and differing test species requirements, this approach is often not practical. The AMRAP guidance document recommends that species included in the SSD ideally should be representative of different growth habits and taxonomic groups whilst also being relevant to the ecological scenarios addressed in the risk assessment. However, for compounds that are known to be selective for a particular group of species, for example submerged species, it may not prove possible to fit a single SSD across a more diverse range of species. Under these circumstances, it may be necessary to focus on a less diverse group of species.

A more or less similar approach as described above for aquatic macrophytes can be followed for algae. Ideally when algae are at risk the SSD should be constructed with a range of taxonomically different groups, e.g. including green algae, diatoms, blue-greens etc., and/or different genera representative for these groups.

Endpoint selection for SSDs with aquatic macrophytes

It appears from the published literature that for aquatic macrophytes a wide array of measurement endpoints is used. This wide array of available measurement endpoints may contribute to the variability in SSDs. The AMRAP guidance document (Maltby et al., 2010) recommends the use of growth rate endpoints for

macrophytes. These growth rate endpoints should be preferably based on biomass or shoot length, as they potentially provide consistency across time and species. From a statistical viewpoint, it is preferable that all endpoints used in development of a SSD are based on common measurement parameters, since each parameter may have a different distribution. An alternative approach is to use the lowest endpoint, no matter what measurement parameter it is based on.

Selection of endpoints should also consider the mode-of-action of the test substance. For example, the effects of auxin-simulating herbicides may lead to distorted growth, but not necessarily a reduction in biomass. In these cases, measurement parameters other than biomass may be more applicable. Note that within and between the scientific and regulatory working groups consensus is not yet reached on the preferred endpoints for macrophyte risk assessment.

In future, the AMRAP SSD Working Group will generate more recommendations on the use of species and endpoints in SSDs with primary producers. They will specifically address species selection for SSDs (what is the sensitivity of *Lemna* species relative to other macrophytes?) and endpoint selection for SSDs.

6.4.3.2 Insecticides

In acute laboratory toxicity tests with invertebrates and insecticides, mortality and immobility are the most frequently studied endpoints. In chronic laboratory toxicity tests with invertebrates and insecticides, endpoints such as reproduction, feeding and inhibition of growth are also studied (Van den Brink et al., 2008).

In case of insecticides, arthropods (crustaceans and insects) usually are most sensitive (Maltby et al., 2005). This implies that the SSD can focus on these taxonomic groups. Note, however, that for some novel types of insecticides (e.g. neonicotinoids) insects may be more sensitive than micro-crustaceans. In that case the SSD should be constructed with the sensitive group within the arthropods (e.g. insects or insects and macro-crustaceans). Specific guidance on the selection of endpoints for arthropods does not exist and in most insecticide SSDs published so far different endpoints for different species were included in the same SSD. However, acute toxicity tests with invertebrates usually address mortality and immobility as endpoints. Consequently, the diversity in endpoints is less than in studies with herbicides and primary producers where most endpoints are sublethal.

Evaluation of the toxicity data of 16 insecticides indicates thatF: (1) arthropods are the preferred taxonomic group to construct acute SSDs, and (2) acute toxicity data for freshwater arthropods from different geographical regions and different freshwater habitats may be combined within a single SSD (Maltby et al., 2005). If necessary, toxicity data for freshwater and saltwater taxa also can be combined in an SSD, but it is important to be aware of differences in taxonomic composition and possible consequences for threshold concentrations that are calculated. SSDs constructed using arthropod species recommended in test guidelines did not differ significantly from those constructed using non-recommended arthropod species (Maltby et al., 2005).

6.4.3.3 Fungicides

For those fungicides that are general biocides, data from all taxonomic groups are recommended to be used to construct SSDs and to assess risk (Maltby et al., 2009; Van Wijngaarden et al., 2010). Note that in these SSDs also toxicity data for fish may be included. The HARAP Guidance Document (Campbell et al., 1999) does not specify the taxonomic groups and level of taxonomic resolution when selecting toxicity data for these generic SSDs.

Of the different groups of pesticides, several fungicides represent the least specific toxic mode-of-action. From this point of view, the generic SSDs as generated for fungicides might resemble the SSDs for biocides. For those fungicides that are general biocides, a default approach could be to include toxicity data from eight different taxonomic groups in the SSD. These data include three to five toxicity data already generated in the 1st tier and five to three additional toxicity data (including fish). The available guidance on pesticides does not yet give further recommendations on which taxa have to be included in SSDs for fungicides.

It should be noted that fungi are not included in the standard dossier dataset as a specific taxon of interest. As a consequence, data on a potentially sensitive species group may be missing. Recent research indicates that aquatic fungi may be particularly sensitive for certain types of fungicides (Dijksterhuis et al., 2009; CBS, 2009; Dijksterhuis et al., 2011). Waterborne fungi species were sampled in the field and isolates of six species were exposed to carbendazim, chlorothalonil, fluazinam, imazalil, epoxiconazole, tebuconazole and azoxystrobine. Effect on fungi growth was most pronounced for the ergosterol inhibitors imazalil, tebuconazole and epoxiconazole, of which the latter two triazoles were most toxic. The results indicate that further research into the potential effects on fungi is urgently needed. It should be noted that the kingdom of fungi is diverse. The selection of relevant species for which standardised ecotoxicity tests may be developed is therefore identified as a further research need.

Note that if fish are included in the SSD for general biocides (non-specific fungicides), the aim is to derive a concentration that is protective at the community level. Since for fish a more stringent protection goal is adopted (see Section 2.1), it should always be checked whether the outcome meets the regulatory lower or higher-tier trigger for fish.

6.4.4 How to generate focused Species Sensitivity Distributions addressing specific groups of organisms

The following criteria have to be considered to decide to which taxonomic groups the SSDs have to be targeted:

1. If the first tier indicates that one standard test species of the basic set is considerably more sensitive (differing by a factor >10) (Campbell et al., 1999) than the others, an SSD should be constructed that is representative for the sensitive taxonomic group. In case of herbicides, vascular plants and/or algae usually comprise the most sensitive groups (see Van den Brink et al., 2006). For photosynthesis inhibitors, algae and aquatic macrophytes can be included in one SSD (Brock et al., 2004). For herbicides with another toxic mode-of-action, this is under study at the moment (Maltby et al., 2010). In case of insecticides, arthropods (crustaceans and insects) usually are most sensitive (Maltby et al., 2005), but within the arthropods a specific sensitive group may exists (e.g. insects for neonicotinoid insecticides). For those fungicides that are general biocides, data from different taxonomic groups are recommended to be used to construct SSDs and to assess risk (Maltby et al., 2009).
2. Data gathered by read-across on related compounds with identical or similar toxic mode-of-action give information on the taxonomic groups which are most likely sensitive for the compound under consideration. This information can be used to decide which taxonomic groups have to be included in an SSD for the compound under consideration.
3. Data in the open literature on the compound may give information on the sensitive taxonomic groups. This information can be used to decide which taxonomic groups have to be included in an SSD for the compound under consideration.
4. Results of micro-/mesocosms tests may shed light on the sensitive taxonomic groups, also when these tests studied the effects of relatively high concentrations not suitable to derive a threshold level of effects.

If for different taxonomic groups different and valid distributions are available, the most sensitive SSD is used in the risk assessment (Van den Brink et al., 2006). The results of higher tier risk assessments based on a specific SSD have to be compared again with the results of the 1st tier to ensure that the RAC based on the specific SSD is protective for taxa not considered in this SSD.

6.4.5 How to generate chronic Species Sensitivity Distributions

In the risk assessment for pesticides, Species Sensitivity Distributions based on chronic data are very scarce. Acute toxicity data are normally more available than chronic data due to experimental and financial constraints (Van den Brink et al., 2006). When chronic data are available, they may be included in the risk assessment for pesticides that cause a chronic (long-term) exposure. Van den Brink et al. (2006) mainly use NOEC values to generate chronic SSDs. EC_{10} effect concentrations also can be included in a chronic Species Sensitivity Distribution. Whereas acute toxicity data relate to a limited number of responses and time scales (e.g., 96-h median lethal concentrations), chronic toxicity data include a wide range of responses and test durations, thereby introducing additional variability into the SSD.

When compiling chronic SSDs, chronic endpoints have to be included for the different species groups, as mentioned before. The test duration has to be of a chronic duration compared to the life cycle characteristics of the species group. More specifically, a chronic toxicity test is defined as a study in which: (1) the species is exposed to the pesticide for at least one full life-cycle, or (2) the species is exposed to the pesticide during one or more critical and sensitive life-stages (see e.g. Holland, 1996; Brock et al., 2010a). Consequently what is considered chronic or acute is very much dependent on the species and endpoint considered.

6.4.6 Calibration of the SSD approach with invertebrate and aquatic primary producer data from micro-/mesocosm studies

Compared to the effects considered in microcosm and mesocosms studies, the SSD approach does not consider recovery nor indirect effects. However, Species Sensitivity Distributions might be very useful in risk assessment as they represent a cost-effective approach for the use of all available laboratory toxicity data for a larger array of species. From this point of view, Hazardous Concentrations derived from Species Sensitivity Distributions for insecticides, herbicides and fungicides were validated with data from mesocosm studies (Van den Brink et al., 2006; Maltby et al., 2005; Brock et al., 2006; Maltby et al., 2009).

For insecticides the lower limit HC_5 of acute SSDs was protective for single and repeated pulse exposures in micro/mesocosm, at least when the effects are expressed in terms of nominal or measured peak concentrations (Maltby et al., 2005). For herbicides the lower limit of the acute HC_5 and the median value of the chronic HC_5 are protective of adverse effects in aquatic microcosms and mesocosms, even under a long-term exposure regime. The median HC_5 estimate based on acute data is protective of adverse ecological effects of herbicides in freshwater ecosystems when a pulsed or short-term exposure regime is used and the effects in the micro/mesocosms are expressed in terms of nominal or measured peak concentrations (Van den Brink et al., 2006). For fungicides, the derived lower limit HC_5 values and the HC_1 values were protective of adverse effects in microcosm and mesocosms studies when effects are expressed in terms of nominal or measured peak concentration (even under more or less long-term exposure regimes). Median HC_5 values were protective for only three of the five fungicides tested (Maltby et al., 2009). Note that these fungicide studies predominantly concerned repeated applications. The latter authors reanalysed the relationships between SSDs constructed with acute toxicity data and threshold concentration derived from microcosm and mesocosm experiments for insecticides (as published by Maltby et al., 2005) and herbicides (as published by Van den Brink et al., 2006) and demonstrated that for these groups of pesticides also the median HC_1 can be used to

derive an appropriate regulatory acceptable concentration (RAC), even for repeated exposure regimes. They conclude that in general the median HC_5 is protective of short-term exposures, the median HC_5 divided by 1.5 is protective of medium-term exposure regimes and the median HC_5 divided by 3 or the HC_1 is protective of repeated longer-term exposure. Note that these acute RAC values need to be compared with the PEC_{max} values of single or repeated pulse exposures.

6.4.7 Proposal for the derivation of RACs for invertebrates and primary producers by means of the SSD approach

Table 6-3 presents a proposal for the derivation of a RAC for Dutch drainage ditches, based on hazardous concentrations derived from Species Sensitivity Distributions with aquatic invertebrates and plants for at least 8 different taxa belonging to the relevant sensitive taxonomic group.

Table 6-3

Proposal for the derivation of a RAC in Dutch drainage ditches, based on hazardous concentrations derived from Species Sensitivity Distributions with aquatic invertebrates and/or plants.

Field exposure regime in drainage ditch scenario	Relevant PEC	Hazardous concentration	AF
Single pulse exposure of short duration (or repeated pulse exposures that are toxicologically independent*) of which the water dissipation DT_{50} in predicted field exposure profile is lower than ten days	PEC_{max}	Median acute HC_5 (based on acute LC_{50} or EC_{50} data)	1
Toxicologically dependent repeated pulse exposures (water dissipation DT_{50} <10d in predicted field exposure profile) or single pulse with a water dissipation DT_{50} in predicted field exposure profile that is larger than ten days	PEC_{max}	Median acute HC_5 (based on acute LC_{50} or EC_{50} data)	3
More or less constant chronic exposure	PEC_{max} or PEC_{TWA}	Median chronic HC_5 (based on chronic NOEC and/or EC_{10} data)	1-2#

* For tests to demonstrate that repeated pulses are toxicologically (in)dependent see Sections 3.3.3.
# The range of 1 to 2 is proposed to address the higher uncertainty in availability of chronic SSDs.

In the past, both the lower limit and the median of the HC_5 have been used as a basis for the RAC. Using the lower limit can be considered as a conservative (safe) approach and may stimulate further research if the confidence interval of the HC_5 is wide. An important drawback of using the lower limit, however, is that the information on the uncertainty associated with the HC_5 may be needed in the future for probabilistic modelling. Therefore, preference is given to the use of the median HC_5 with an appropriate assessment factor.

6.4.8 Proposal for the derivation of RACs for fish by means of the SSD approach

When constructing a Species Sensitivity Distribution with fish, a lower number of toxicity data is accepted. The HARAP Guidance (Campbell et al., 1999) recommends using a minimum of five toxicity data to construct SSDs specific for fish. This lower number of toxicity data is chosen for, among other reasons, animal welfare considerations and because of the overall lower variability in toxicity data when e.g. compared with that of invertebrates. In the risk assessment it is sometimes necessary to consider fish separately and to construct a separate SSD with fish as the most appropriate method to meet this requirement. For example, constructing a separate SSD for fish may be necessary if the risks of a plant protection product to populations of

invertebrates and primary producers have been assessed by means of an appropriate microcosm or mesocosm experiment without fish. In regular mesocosm and microcosm studies fish are recommended not to be included as the effects of fish might interfere with the effects of the compound on the macro-invertebrate community (Giddings et al., 2002). If potential risks to fish cannot be excluded, and fish cannot be part of the ecosystem in a mesocosm or microcosm study, the most appropriate method in risk assessment is to construct a separate SSD for fish.

Acute LC_{10} and acute NOEC values may be used to construct the SSD and to calculate the HC_5 for fish, since a higher protection level is desired for vertebrates than for invertebrates and plants. Another option is to apply an extra AF to the HC_5 based on acute LC_{50} or EC_{50} data.

We propose the following hazardous concentrations and Assessment Factors to derive a RAC for fish and other aquatic vertebrates (Table 6-4). We are aware that the method proposed needs calibration. For the ratio between the acute LC_{50} and chronic NOEC/L(E)C_{10}, usually a factor of 10 is assumed (see e.g. Roex et al., 2000). Taking this into account, assuming a factor of 3 for the ratio between the acute LC_{50} and acute NOEC/LC_{10} for fish seems to be appropriate.

Table 6-4

Proposal for the derivation of a RAC for Dutch drainage ditches, based on hazardous concentrations derived from Species Sensitivity Distributions with fish (and other aquatic vertebrates).

Field exposure regime in drainage ditch scenario	Relevant PEC	Hazardous concentration	AF
Single pulse exposure of short duration (or repeated pulse exposures that are toxicologically independent*) of which the water dissipation DT_{50} in predicted field exposure profile is lower than ten days	PEC_{max}	Median acute HC_5 (based on acute NOEC and/or acute LC_{10} data)	1
		or	
		Median acute HC_5 (based on acute LC_{50} or EC_{50} data)	3
Toxicologically dependent repeated pulse exposures (water dissipation DT_{50} <10d in predicted field exposure profile) or single pulse with a water dissipation DT_{50} in predicted field exposure profile that is larger than ten days	PEC_{max}	Median acute HC_5 (based on acute NOEC and/or acute LC_{10} data)	3
		or	
		Median acute HC_5 (based on acute LC_{50} or EC_{50} data)	5
Chronic exposure: more or less constant chronic exposure	PEC_{max} or PEC_{TWA}	Median chronic HC_5 (based on chronic NOEC and/or EC_{10} data)	1 - 3

* For tests to demonstrate that repeated pulses are toxicologically (in)dependent see Section 3.3.3.

6.5 Refined exposure laboratory toxicity tests

6.5.1 Introduction

The environmental fate properties of a PPP can be an important factor in the mitigation of risk under realistic environmental conditions. For example, if dissipation from water and/or sediment is rapid, risk assessments based on laboratory toxicity tests performed under constant exposure conditions may overestimate potential risks (Campbell et al., 1999).

According to the proceedings from the ELINK workshop (Brock et al., 2010a) in cases where predicted (modelled) field exposure profiles differ considerably from exposure regimes in standard toxicity studies it may be appropriate to design higher-tier laboratory toxicity tests that more closely resemble realistic exposure profiles. Before conducting refined exposure laboratory toxicity tests it is necessary to consider whether the first tier procedure triggers acute or chronic risks. In refined exposure studies supporting acute risk assessments the peak concentration may be used in both the PEC and the RAC estimate, at least if: (1) the height and duration of the pulse exposure in the refined laboratory toxicity test (on which the RAC is based) is relatively worst-case when compared with that of the relevant predicted (modelled) field exposure profile (Figure 6-2), and (2) the predicted repeated pulse exposures in the field are considered to be toxicologically independent (EFSA, 2005a; Boesten et al., 2007; Brock et al., 2010a).

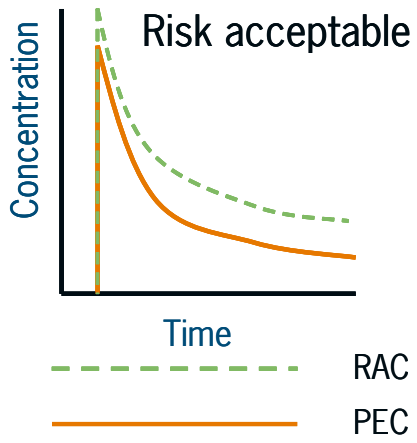


Figure 6-2
The height en duration of the pulse exposure in RACs based on refined exposure laboratory toxicity tests should be worst case relative to the predicted exposure profile (PEC curve).

If the use of the TWA approach in the chronic risk assessment is appropriate, refined exposure laboratory toxicity tests need not to be performed as a higher-tier option, since concentration-response relationships observed in the standard chronic toxicity tests, as well as the derived RACs, can be expressed in terms of TWA concentrations. In the risk assessment the TWA RAC can be compared with the appropriate TWA PEC. According to the proceedings of the ELINK workshop, however, long-term refined ecotoxicological exposure studies, for example simulating repeated pulse exposures, may be a higher-tier option if the TWA approach cannot be used (see Section 3.3.2 for criteria when not to use the TWA approach) (Brock et al., 2010a).

In designing refined exposure laboratory toxicity tests with standard and additional aquatic test species, information on the relevant field exposure predictions should be considered. In order to adopt a realistic worst-case exposure scenario in the toxicity test, the refined exposure regime tested should be deduced from the

relevant field exposure scenarios (e.g. the Dutch ditch exposure scenario) and the relevant intended agricultural use of the PPP. The proceedings of the ELINK workshop recommend to use the most representative generalised exposure profile instead of the one deduced from a very specific scenario and product use, since use of different exposure scenarios (e.g. FOCUS surface water scenarios) and the application of the same PPP in different crops may result in various exposure profiles, all characterised by a specific time-variable field exposure regime. Testing all possible time-variable field exposure profiles simply may not be practical and/or cost-effective, certainly when considering the possible changes over time or location of product use (including mitigation measures).

6.5.2 **Refined exposure tests with standard test species**

Besides information on the relevant field exposure predictions, information on the mode-of-action of the PPP and time-to-onset-of-effects in the Tier 1 acute and chronic toxicity tests should be considered in designing refined exposure laboratory toxicity tests with standard aquatic test species. If, for example, the specific toxic mode-of-action of the PPP results in short time-to-onset-of-effects (e.g. immobility of *Daphnia*, *Americamysis* or *Chironomus* in acute toxicity tests with a certain pyrethroid insecticide) a relatively narrow pulsed exposure in an acute laboratory tests of 48 - 96 h may already yield the maximum effects. Other PPPs may need a longer time to reach the incipient effect level (Figure 6-3) in the standard test species and this time may be dependent on the height and duration of the pulse exposure. Consequently, when designing refined exposure laboratory tests with standard test species it is important to consider the relationship between pulse height and exposure duration and the time needed to express the toxic effects. This information should be used to design the duration of the refined exposure test. This information is also required to appropriately interpret the results of these tests.

According to the proceedings of the ELINK workshop when standard test species are assessed in refined exposure laboratory toxicity tests and these tests are considered appropriate for the risk assessment, a reduction of the AF is not justified when deriving a RAC. However, a higher toxicity value (e.g. acute EC_{50} or chronic NOEC) from the refined exposure study with standard test species and the application of the appropriate AF (e.g. 100 to derive the acute RAC and 10 to derive the chronic RAC) may change the overall risk assessment.

Besides refined exposure tests with standard test species that more or less resemble the design of Tier 1 toxicity studies, long-term refined exposure tests may also be performed at the population-level. Population-level experiments are usually performed with populations of individuals that differ in age and developmental state. In principle, the standard OECD toxicity tests performed with algae (OECD, 2006a) can be considered a population-level test, in contrast to the chronic tests with the other standard test species. In the recent past some experience is gained with the conduct of refined exposure tests with populations of *Daphnia magna* (see e.g. Hanazato and Hirokawa, 2004; Liess et al., 2006).

In case *Daphnia magna* is the most sensitive standard test species in chronic tests with an insecticide, the NOEC of a long-term refined exposure study with a mixed population of *Daphnia*, addressing an insecticide exposure regime guided by relevant field exposure predictions, might be used in the chronic risks assessment by applying the standard AF of 10 to derive the chronic RAC. Since the expertise with these population-level tests still is limited the ELINK document states that this approach warrants 'confirmation/validation' e.g. by comparing the results with the threshold levels for effects derived from appropriate micro-/mesocosm tests that simulated similar exposure regimes.

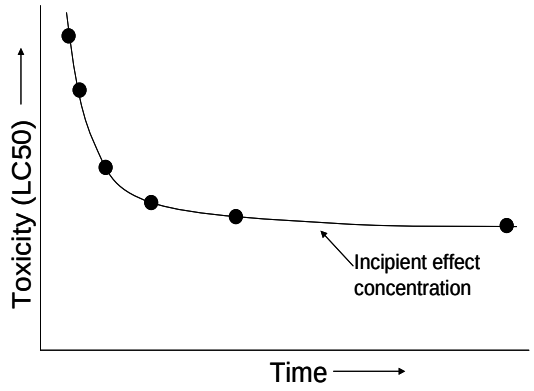


Figure 6-3

Illustration of change in toxicity over time until incipient effect is reached. A refined exposure test should be long enough to reach the incipient effect concentration (after Solomon et al., 2008).

6.5.3 Refined exposure tests with additional test species

In principle, refined exposure toxicity tests can also be performed with additional test species, but again the refined exposure regime tested should be guided by relevant exposure predictions for the intended agricultural uses, and the tests should be performed long enough to allow the expression of the effects.

When enough additional species are tested with a similar refined exposure regime, the results might be used in the effects assessment by applying the Geomean approach (as described in Section 6.3.2) or the SSD approach (e.g. by calculating the HC_5 and using this value as described in Section 6.4). Refined exposure tests with additional test species might also be performed to put the results of other higher-tier tests in perspective. If, for example, a mesocosm study clearly indicates that a certain species in particular is sensitive to the PPP tested, but the lowest concentration tested already resulted in an unacceptable effect, this species might be tested further in the laboratory to address the response of the required exposure regime. Such an approach is more cost-effective than performing a new mesocosm study. Refined exposure tests may also be used to extrapolate the results of an SSD constructed with acute toxicity tests to assess the possible effects of repeated pulse exposures. For example, when addressing the risks of repeated pulse exposures to fish one can select the most sensitive fish species of the acute SSD to perform some further longer-term tests with the realistic pulsed exposure regime. If the repeated pulse exposure tests do not result in a lower EC_x value when compared with that of the standard acute test with these fish species, this is an indication of the toxicological independence of the different pulse exposures. This again may allow the use of the acute HC_5 in the effects assessment for the repeated pulsed exposure regime.

6.5.4 Proposal for the derivation of RACs by means of refined exposure laboratory toxicity tests

For the derivation of an acute RAC by means of refined acute toxicity tests with relevant standard test species it is proposed to apply an AF of 100 to the LC_{50} or EC_{50} (expressed in terms of peak concentration) under the conditions that:

– The pulse exposure in the refined acute laboratory toxicity test is worst-case when compared with the relevant predicted (modelled) field exposure profile.

– The repeated pulse exposures predicted for the field are considered to be toxicologically independent (for explanation see Section 3.3.3); if not the repeated pulses should be addressed in the refined acute toxicity test.
– The duration of the acute test is long enough to reach the incipient effect.
– The refined acute RAC is compared with the $PEC_{max.}$

Long-term refined exposure tests (e.g. simulating repeated pulse exposures) may be a higher-tier option if the TWA approach cannot be used. For the derivation of a chronic RAC by means of refined chronic toxicity tests (at individual or population level) it is proposed to apply and AF of 10 to the NOEC/EC_{10} expressed in terms of nominal (if measured peak exposures do not deviate more than 20% from nominal) or measured peak concentration in the test systems under the conditions that:
– The (repeated pulsed) exposure regime in the refined laboratory toxicity test is worst-case when compared with the relevant predicted (modelled) field exposure profile.
– The duration of the test is long enough to reach the incipient effect.
– The refined chronic RAC is compared with the PECmax.

When refined exposure studies with several additional test species of the relevant taxonomic group are available the derived toxicity values might be used as described in Sections 6.3.2 (geomean method), 6.4.7 (SSD method for invertebrates and primary producers) and 6.4.8 (SSD method for fish), at least when conditions as described above for the derivation of refined RACs are not violated.

Note that our proposal predominantly addresses the uncertainty of the ecotoxicological endpoint. It is assumed that the predicted field exposure profile is sufficiently worst case. Furthermore note that in a refined risk assessment the uncertainty of the exposure estimate can be assessed as well.

6.6 Model Ecosystem Approach

6.6.1 Introduction

Freshwater model ecosystems - usually referred to as microcosms and mesocosms - are bounded systems that are constructed artificially with samples from, or portions of, natural freshwater ecosystems, or that consist of enclosed parts of natural freshwaters. Indoor experimental ecosystems are often referred to as microcosms and outdoor experimental ecosystems as mesocosms, but their difference mainly concerns their size. The most frequently used freshwater model ecosystems in pesticide risk assessment are those that mimic shallow, static freshwater habitats (see Brock and Budde, 1994; Caquet et al., 2000), but ecotoxicological experiments with pesticides in artificial streams are also common (e.g. Schulz et al., 2002; Heckmann and Friberg, 2005; Beketov et al., 2008). In polder landscapes of the Netherlands the communities of edge-of-field surface waters mostly resemble that of shallow, static freshwater habitats. Consequently, microcosms and mesocosms resembling ponds and ditches are most relevant for the risk assessment of PPPs in Dutch edge-of-field surface waters.

Besides the aim of micro- and mesocosm studies to simulate natural conditions and exposing these systems to environmentally realistic pesticide exposure regimes, these studies normally follow experimental designs to demonstrate causality between treatment and effects, and can also identify concentration-effect relationships. Due to confounding factors, causality between pesticide exposure and effects is more difficult to demonstrate in field monitoring studies. The advantage of micro- and mesocosm studies over the other types of experimental higher-tier studies (e.g. additional laboratory toxicity tests to construct SSD's; laboratory population studies) is their ability to integrate more or less realistic exposure regimes with the assessment of

endpoints at higher levels of biological integration, and to study intra- and inter-species interactions and indirect effects. They also allow assessment of latency of effects and population and community recovery.

6.6.2 Selecting the appropriate exposure regime in micro-/mesocosm experiments

In the CLASSIC workshop (Giddings et al., 2002) exposure regime and dosing were already recognized as fundamental issues in the experimental design of micro- and mesocosm studies. The delegates recommended an exposure-response experimental design with preferably five or more concentrations, and at least two replicates per concentration. Preferably, the lowest test concentration should not result in treatment-related responses, while the highest concentration tested should result in pronounced effects on several measurement endpoints. This allows the derivation of threshold concentrations for toxic effects, as well as putting in perspective the possibly more subtle treatment-related responses caused by intermediate concentration levels. This implies that the selected exposure concentrations should always be guided by lower-tier effect information (e.g. single species toxicity tests) and the expected field exposure regime of the substance under evaluation (e.g. risks due to short- or long-term exposure).

Before designing a micro-/mesocosm test for regulatory purposes it is important to evaluate the possible exposure regimes in aquatic ecosystems that may result from normal agricultural use of the PPP of concern, and to identify the relevant exposure regimes that should be addressed in the effect assessment (see ELINK report; Brock et al., 2010a). If the expected and relevant field exposure regime is characterised by a single high pulse (e.g. due to drift application), or by repeated pulses that are both toxicologically and ecologically independent, a single application experimental design is an appropriate exposure regime to study in the micro-/mesocosm experiment. The pulse duration in the micro-/mesocosm experiment should, however, either be equal to or larger than that predicted for the field (see Figure 6-2 in Section 6.5.1), or it should be easy to extrapolate concentration-response relationships for shorter peaks to that for broader peaks (e.g. if the time-to-onset-of-effect is very short for relevant organisms in single species toxicity tests).

If the expected exposure regime in the field triggers concerns of repeated pulse exposures that likely are toxicologically and/or ecologically dependent (see Section 3.3.3), a repeated exposure regime should be adopted in the micro-/mesocosm experiment, or it should be easy to extrapolate population/community level responses due to short-term exposure to that of a longer-term exposure regime. According to the ELINK document, the number of applications has to be considered carefully in relation to the expected biological effects - but should be as low as possible - guided by the responses observed in the toxicity tests that triggered the micro-/mesocosm study and by biological information of the species potentially at risk. In a micro-/mesocosm experiment that aims to derive concentration-response relationships for constant chronic exposure, a realistic worst-case approach is to maintain a more or less constant pesticide concentration for at least the duration of the chronic toxicity test that triggered the micro-/mesocosm test, unless the TWA (Time weighted Average) exposure can be used to express the treatment-related effects (Brock et al., 2010a).

6.6.3 Selecting the appropriate measurement endpoints in micro-/mesocosm experiments

Microcosm and mesocosm experiments are test systems that allow studying treatment-related effects of PPPs at the population and community level. Population responses in micro/mesocosm are usually studied by means of measurement endpoints that provide information on dynamics in population abundance, biomass and/or growth. Measurement endpoints to study community-level responses usually comprise summary parameters like species richness and diversity, but also community metabolism endpoints indicative for

ecosystem processes like dynamics of dissolved oxygen in water and rates of decomposition of particulate organic matter (e.g. in litter bags).

The number of taxa occurring in micro/mesocosms, and consequently the potential measurement endpoints, may be high. Studying all potential measurement endpoints is very expensive. For reasons of cost-effectiveness usually a limited number of measurement endpoints are selected. Available lower-tier studies for the PPP under evaluation (e.g. standard and additional laboratory toxicity tests) and/or results of model ecosystem experiments with related compounds (characterized by a similar toxic mode-of-action) provide insight which structural and functional parameters should be studied intensively. For example, if the PPP under investigation is a selective herbicide and the laboratory toxicity tests indicate that green algae and the macrophytes *Lemna* and/or *Myriophyllum* are at least an order of magnitude more sensitive than the invertebrates *Daphnia* and/or *Chironomus*, the primary focus of the selected measurement endpoints should be on populations of phytoplankton, periphyton and macrophytes (structural endpoints for primary producers) and possibly also on parameters indicative for the functioning of primary producers, such as dissolved oxygen and pH. If the PPP of concern is an insecticide for which standard toxicity tests and model ecosystem experiments with related compounds indicate that crustaceans and insects in particular are sensitive, the focus of the study should be on populations of zooplankton and macro-invertebrates (possibly including emergent insects and effects of shredder populations on the breakdown of particulate organic matter). In contrast, if the differences in toxicity between the standard test organisms is small, as might be the case for fungicides with a biocidal mode-of-action, the selected measurement endpoints should include a variety of taxonomical groups such as populations of primary producers (e.g. algae and vascular plants) and invertebrates (e.g. zooplankton and macro-invertebrates, including non-arthropods).

For detailed guidance on endpoint selection to conduct a proper micro-/mesocosm experiment is referred to workshop documents of SETAC-Europe (Arnold et al., 1991), EWOFFT (Crossland et al., 1993; Hill et al., 1994a), HARAP (Campbell et al., 1999) and CLASSIC (Giddings et al., 2002). If macrophytes comprise the endpoints of concern recommendations for the conduct of specially designed micro-/mesocosm tests can be found in the AMRAP document (Maltby et al., 2010). Studying fish in microcosms and mesocosms can present difficulties and needs to be carefully considered. When the invertebrate community is the principal endpoint of the study, it is recommended not to include free-living fish (Giddings et al., 2002). Note, however, that it may be appropriate to introduce caged fish. In smaller micro-/mesocosms free-living fish usually cannot be introduced at natural biomass levels appropriate to the abundance of their prey, and therefore fish can have an undue influence on other populations inhabiting these confined test systems. However, separate micro/mesocoms may be used to study the individual level effects of a realistic exposure regime on e.g. standard test species of fish (see Section 6.5).

6.6.4 Interpretation of micro-/mesocosm experiments

As explained above, the extensive knowledge and experience gained with model ecosystems and PPPs was used for development and harmonization of micro- and mesocosm studies as a tool for higher-tier effects assessment. The HARAP (Campbell et al., 1999) and CLASSIC (Giddings et al., 2002) workshops also discussed the implementation of the micro-/mesocosm data into the risk assessment. As more experience was gained through the conduct and design of micro- and mesocosm studies, their interpretation was aided by the development of software tools that facilitated multivariate statistical analysis of the data (Van den Brink and Ter Braak, 1999).

In recent years, discussions shifted towards the awareness of inconsistencies in both the way the same mesocosm data are interpreted and the uncertainty (assessment) factors applied by regulatory experts in different EU Member States. The Dutch Platform for Assessment of Higher-tier Studies has produced a

guidance document on how micro-/mesocosm data should be presented and evaluated in a uniform and transparent manner (De Jong et al., 2008). In addition, Brock et al. (2006) and De Jong et al. (2008) also proposed a refinement of the 'Effect classes' used to categorise the results of micro-/mesocosm experiments (see below), which are outlined in the Guidance Document on Aquatic Ecotoxicology in the Context of Directive 91/414/EEC (SANCO, 2002).

Effect class 1 *(No treatment-related effects demonstrated; $NOEC_{micro/mesocosm}$)*
No (statistically and ecologically significant) effects observed as a result of the treatment. Observed differences between treatment and controls show no clear causal relationship.

Effect class 2 *(Slight effects)*
Effects reported as 'slight', 'transient', or other similar descriptions. It concerns a short-term and/or quantitatively restricted response of one or a few sensitive endpoints, usually observed at individual samplings only.

Effect class 3A *(Pronounced short-term effects (< eight weeks, followed by recovery)*
Clear response of sensitive endpoints, but full recovery of affected endpoints within eight weeks after the first[t] application or, in case of delayed responses and repeated applications, the duration of the effect period is less than eight weeks and followed by full recovery. Effects observed at some subsequent sampling instances.

Effect class 3B *(Pronounced effects and recovery within 8 weeks post last application)*
Clear response of sensitive endpoints in micro-/mesocosm experiment repeatedly treated with the test substance and that last longer than eight weeks (responses already start in treatment period), but full recovery of affected endpoints within eight weeks post last application.

Effect class 4 *(Pronounced effect in short-term study)*
Clear effects (e.g. large reductions in densities of sensitive species) observed, but the study is too short to demonstrate complete recovery within 8 weeks after the (last) application.

Effect class 5A *(Pronounced long-term effect followed by recovery)*
Clear response of sensitive endpoints, effect period longer than eight weeks and recovery did not yet occur within eight weeks after the last application, but full recovery is demonstrated to occur in the year of application.

Effect class 5B *(Pronounced long-term effects without recovery)*
Clear response of sensitive endpoints (> eight weeks post last application) and full recovery cannot be demonstrated before termination of the experiment or before the start of the winter period.

6.6.5 How to derive a RAC from the micro-/mesocosm experiment and how to link it to the PEC

Communities and environmental conditions in micro/mesocosms represent only one of the many possible field conditions. Possible variability in exposure-response relationships between different aquatic communities can be evaluated by comparing different micro-/mesocosm experiments performed with the same PPP in e.g. different countries, seasons and/or types of ecosystem (e.g. ponds, streams).

RACs for short-term exposure

The short-term or long-term Regulatory Acceptable Concentration (RAC) representative for the threshold level of effects in the field may be derived by applying an AF (for spatio-temporal extrapolation) to the Effect class 1 - 2 concentration from the micro-/mesocosm experiment. The height of this AF should, amongst others, depend on the relevance of the tested assemblage for the species potentially at risk, the other higher-tier information available (e.g. toxicity data for additional test species and other micro-/mesocosm experiments), and known overlap in Effect class 1 - 2 concentrations with Effect-class 3 - 5 concentrations for related compounds with a similar toxic mode-of-action. If we consider the data presented in Appendix 1 for chlorpyrifos, lambda-cyhalothrin, azinphos-methyl, esfenvalerate and simazine as representative for (short-term) pulsed exposure regimes it seems that the spatio-temporal extrapolation of Effect class 1 and Effect class 2 concentrations is possible with relatively low uncertainty. The comparison of micro-/mesocosm experiments performed with these PPPs suggests that an Effect class 1 and an Effect class 2 concentration of a well-performed micro-/mesocosm study earns confidence as an appropriate indicator of the RAC indicative for the threshold level of toxic effects in Dutch drainage ditches, at least for short-term (single or repeated pulsed) exposures. For the same PPP and a similar exposure regime these Effect class 1 and Effect class 2 concentrations do not overlap with the range of concentrations for higher effect classes. Consequently it may suffice to apply a small AF (1-3) to Effect class 1 - 2 concentrations to derive a RAC indicative for the ecological threshold option.

If an Effect-class 3A concentration for short-term exposures is considered acceptable in Dutch drainage ditches, it appears from the data presented in Appendix 1 that for chlorpyrifos and lambda-cyhalothrin an AF of 3 to 4 may be necessary for spatio-temporal extrapolation to derive a short-term RAC if a single high quality micro-/mesocosm experiment is available. Applying an AF of 3 to the highest Effect class 3A concentration overall avoids the occurrence of unacceptable class 4 - 5 effects caused by pulsed exposures in hydrologically closed systems (lentic micro-/mesocosms or recirculating experimental streams) (Tables A1-1 and A1-2 in Appendix 1). If more appropriate micro-/mesocosm studies are available either the AF may be lowered or the AF may be applied to the highest available Effect class 3A concentration. However, when deriving a RAC on basis of an Effect class 3A concentration, it should be carefully evaluated whether the populations that show recovery in the micro-/mesocosm tests are representative for the populations potentially at risk in the field (e.g. univoltine and semivoltine populations). An example of an ecological Dutch ditch scenario and its typical macro-invertebrates and macrophytes is described in Brock et al. (2010b). In addition, according to ELINK (Brock et al., 2010a) if the derived RAC value from a single- or multiple-application micro-/mesocosm experiment is based on an Effect class 3A concentration (e.g. by application of an AF of 3 for spatio-temporal extrapolation), an appropriate risk assessment can only be performed by also plotting the threshold level for effects (e.g. based on lower tier data or Effect class 1 - 2 concentrations from micro-/mesocosms) on the predicted field exposure profile. If in the appropriate field scenario the pulses are lower than the RAC value based on Effect class 3A concentrations but higher than the threshold level for direct toxic effects, the interval between successive peaks should be carefully considered. If the interval between peaks is smaller than the relevant recovery time of the sensitive populations of concern, these peaks should be considered as ecologically dependent. On the basis of this information, the total period of possible effects can be estimated.

In the assessment of the short-term RAC the Effect class concentrations should be expressed in terms of the nominal or measured/estimated peak concentration in the micro-/mesocosms of concern. For moderately to slow dissipating substances the nominal concentration can be used if the measured exposure concentrations in the integrated water column of the test system do not deviate more than 20% from nominal. Note that the first hours post application a heterogeneous distribution of the test compound in the water column is common which may hamper the proper measurement of peak concentrations. For fast dissipating compounds the proper measurement of the actual peak concentration in the test system may be difficult if not performed shortly after application. An alternative option to estimate the peak concentration in the test systems may be to measure the concentration in the application solutions as well as the amounts of application solution applied

to each test system. In repeated application studies the peak concentration may occur immediately after the last application if the compound does not dissipate completely from the water column between applications. In that case adopting the nominal treatment level to express the effects can be considered a conservative approach.

RACs for long-term exposure

The treatment-related responses caused by a long-term chronic exposure regime to the fungicide carbendazim resulted in similar Effect class 1 concentrations, suggesting little variability in threshold levels for effects between studies (Table A1-6 in Appendix 1). However, long-term exposure studies with the herbicide atrazine (Table A1-5 in Appendix 1) revealed a considerable overlap between Effect class 1 and Effect class 2 concentrations. In addition, an overlap between Effect class 2 and Effect class 3 - 5 concentrations was observed as well for atrazine. As explained in Appendix 1, differences in Effect class 1 - 2 concentrations between studies performed with the photosynthesis inhibiting herbicide atrazine might be explained by differences in light conditions between indoor and outdoor studies presented in Table A1-5. Nevertheless, if we consider the atrazine data representative for chronic exposure regimes of other pesticides, and from a regulatory point of view an Effect class 2 response is acceptable, an AF of 2 to 3 seems to be necessary for spatio-temporal extrapolation from a single high-quality model ecosystem experiment mimicking a chronicexposure regime. Applying an AF of 3 or 4 to the highest Effect class 2 concentrations presented in Table A1-5 (see Appendix 1) will, with a high probability, avoid unacceptable class 3 to 5 effects caused by long-term exposure. If more appropriate micro-/mesocosm studies are available either the AF may be lowered or the AF factor may be applied to the highest available Effect class 2 concentration.

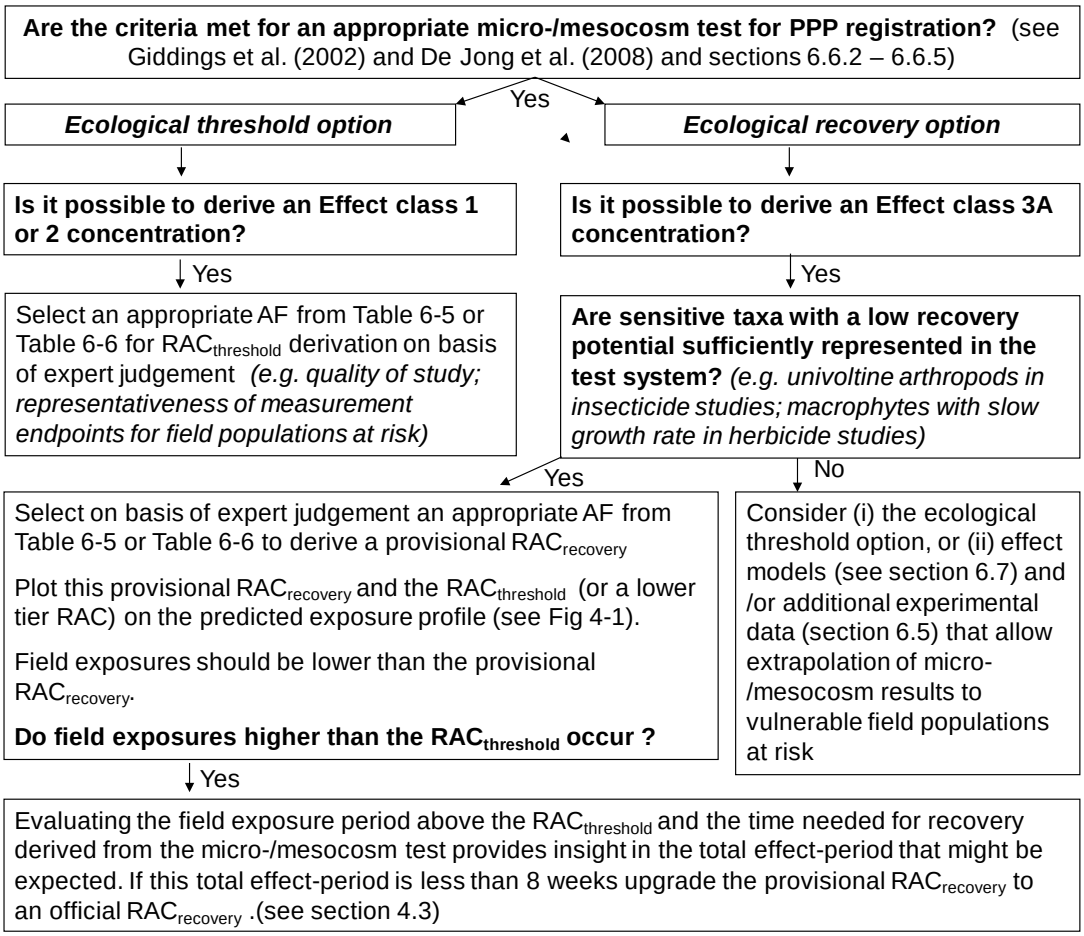
To evaluate risks due to long-term exposure either the peak concentration or a Time Weighted Average (TWA) concentration of the pesticide in the relevant matrix (water, sediment) may be an appropriate PEC. As discussed already the selection of the length of the TWA time-window is based on ecotoxicological considerations (e.g. A/C ratio; time-to-onset-of-effect information; length of the most sensitive life stage of the organisms at risk) and should be guided by the length of the relevant chronic toxicity tests that triggered the micro-/mesocosm experiment. If the TWA approach is considered appropriate (see for criteria Section 3.3.2) participants of the ELINK workshop proposed to adopt a default time-window of 7 days for the TWA estimate of the long-term PEC if no scientific arguments are provided to shorten or lengthen this default time window. Note that for a worst-case approach the time-window for the TWA effect estimate in the micro-/mesocosm study should not be smaller than the selected TWA time-window for the PEC estimate in the field. In addition, the time-window for the TWA effect estimate in the micro-/mesocosm experiment should not be larger than the period in which the exposure remains more or less constant or, in case of a relatively fast dissipating substance, this time-window should not be longer than the application period of the relatively fast dissipating pesticide in the micro-/mesocosm study. The application period is the period in which repeated pulse applications occur. When e.g. a 7-d time-window is adopted for the PEC, the 'Effect class' concentrations derived from a mesocosm experiment characterised by 3 weekly treatments can be expressed in terms of a TWA concentration that is ≥7 days and ≤21 days if in the test systems the pesticide is not very persistent. Note that in repeated application studies, the highest 7-d TWA concentration may be measured later in the application period if the active substance does not completely dissipate between applications.

In case the TWA approach is deemed not to be appropriate in the long-term risk assessment, and consequently the PEC_{max} is used as field exposure estimate, the 'Effect class' concentrations derived from a mesocosm experiment simulating long-term exposure may be expressed in terms of the nominal, peak or average concentration measured/calculated during the application period (or the period in which the exposure remains more or less constant in the micro-/mesocosm test). Adopting the nominal or measured/calculated peak concentration may be realistic if it can be demonstrated that the dissipation from water in the mesocosm experiment overall is less fast, or does not deviate much, from that in the relevant field scenario(s). In that case, and if the concentration builds up due to repeated treatments, adopting the nominal concentration during

the application period can be considered as a more conservative approach than adopting the measured/predicted peak concentration.

6.6.6 Proposal for the derivation of RACs by means of the model ecosystem approach

Decision scheme 6-1 and Table 6-5 and 6-6 present proposals for the derivation of the acute and chronic RAC for Dutch drainage ditches on basis of appropriate micro-/mesocosm experiments.



Decision scheme 6-1

Decision scheme for the derivation of the RAC indicative for the ecological threshold level of effects ($RAC_{threshold}$) or the RAC that consideres ecological recovery ($RAC_{recovery}$) on basis of aquatic micro-/mesocosm tests.

Table 6-5

Proposal for the derivation of the acute RAC (representative for single and repeated pulse exposures) on basis of appropriate micro-/mesocosm experiments. If in the same study several treatments resulted in the same 'Effect class'- response the highest concentration within the same Effect class is selected.

	Assessment factor	Field exposure concentration
Ecological threshold option		
Effect class 1 (based on nominal or measured peak concentration)	1 - 2*	PEC_{max}
Effect class 2 (based on nominal or measured peak concentration)	2 - 3*	PEC_{max}
Ecological recovery option		
Effect class 3A (based on nominal or measured peak concentration)	3 - 4*	PEC_{max}

* The height of the AF is based on expert judgement considering all available lower and higher-tier information. If several adequate micro-/mesocosm studies are available the AF is applied to the highest Effect class 1, 2 or 3 value or a lower AF than reported in the table may be applied.

Table 6-6

Proposal for the derivation of a chronic RAC on basis of appropriate micro-/mesocosm experiments. If in the same study several treatments resulted in the same 'Effect class'- response the highest concentration within the same Effect class is selected.

	Assessment factor	Field exposure concentration
Ecological threshold option		
Effect class 1 (based on time weighted average concentration during the application period#)	1 - 2*	PEC_{max} or PEC_{TWA}
Effect class 1 (based on peak concentration if the pulsed exposure regime is relatively worst-case compared to the predicted field exposure profile)	1 - 2*	PEC_{max}
Effect class 2 (based on time weighted average concentration during the application period#)	2 - 3*	PEC_{max} or PEC_{TWA}
Effect class 2 (based on peak concentration if the pulsed exposure regime is relatively worst-case compared to the predicted field exposure profile)	2 - 3*	PEC_{max}
Ecological recovery option		
Effect class 3A& (based on time weighted average concentration during the application period#)	3 - 4*	PEC_{max} or PEC_{TWA}
Effect class 3A& (based on peak concentration if the pulsed exposure regime is relatively worst-case compared to the predicted field exposure profile)	3 - 4*	PEC_{max}

\# Note that in a long-term micro-/mesocosm test the application period may be shorter than the study duration. The application period is defined as the period in which at regular time intervals the exposure concentrations are adjusted to the required level.

* The height of the AF is based on expert judgement considering all available lower and higher-tier information. If several adequate micro-/mesocosm studies are available the AF is applied to the highest Effect class 1, 2 or 3 value or a lower AF than reported in the table may be applied.

& Note that in micro-/mesocosm experiments that study a long-term exposure regime an 'Effect class 3A' response usually is not observed since a long-term exposure usually results in a long-term treatment-related effect as well. However, theoretically an 'Effect class 3A' response is possible if the selected application period for the chronic or repeated pulsed exposure regime in the test systems is relatively short and worst-case when compared to the predicted field exposure profile.

6.7 Higher-tier modelling approaches

6.7.1.1 Introduction

Where the environmental exposure assessment almost fully relies on modelling approaches, the use of ecological effect models for regulatory purposes is limited. Within the context of this report, ecological effect models are defined as theoretical models that describe or predict effects on different levels of biological organisation in relation to exposure. The main aim of applying models is to extrapolate effect data obtained under standard exposure regimes to realistic time-varying exposure patterns, or to extrapolate population recovery from isolated micro-/mesocosm tests to representative scenarios for surface waters, without having to perform a series of specifically-designed experiment for each possible exposure scenario.

While the parameterisation of (parts of) the models relies on experimental data, the final endpoint is not determined in an experiment, but predicted by mathematical modelling. This is fundamentally different from the model-ecosystem approach (see Section 6.6), in which a (simplified) ecosystem is constructed in an experimental setting. The model for secondary poisoning (see Section 5.3.2) is merely an exposure model, since it describes the distribution from water to prey, while relying on 'classical' measured toxicological endpoints (NOAEL) when considering effects.

The proceedings of the ELINK-workshop (Brock et al., 2010a) provide an overview of the current state of the art with respect to effect modelling, considering toxicokinetic/toxicodynamic modelling, population models, community, food web or ecosystem models, and empirical models. During the LEMTOX-workshop (Thorbek et al., 2010), the potential role of ecological population models for pesticide risk assessment and registration was discussed. Recently application of these models in risk assessment have been discussed (Schmolke et al., 2010; Galic et al., 2010; Hommen et al., 2010a).

6.7.2 Toxicokinetic / toxicodynamic modelling

Toxicokinetic/toxicodynamic (TK/TD) models describe the processes that link exposure to effects in an organism. For aquatic organisms, about seven more or less established TK/TD-models are available, which have been reviewed by Ashauer et al. (2006) and have been discussed in detail during the ELINK-workshop (Brock et al., 2010a). Examples are the Dynamic Energy Budget model (DebTox) by Kooijman and Bedaux (1996) or the Threshold Damage Model (TDM) by Ashauer et al. (2007abc). TK/TD-models consist of two sub-models: a toxicokinetic (TK) model which describes the time course of concentrations within an aquatic organism in relation to concentrations in the external medium, and a toxicodynamic (TD) model to describe the time course of damage and repair to the affected organisms based on specific pattern(s) of exposure to the test compound.

The predicted endpoint of the TK-sub-model is the concentration at the target site. Most TK-models are based on one-compartment first-order kinetics, where the internal (whole body) concentration of the toxicant depends upon the external concentration and uptake and elimination rate constants. More complex models are needed in case distribution over multiple compartments has to be described and/or when significant growth of the organism is expected, e.g. in the case of macrophytes.

The available TD-models differ in their assumptions with respect to the driving parameter for effects. Some models consider the effect to be proportional to the internal concentration, with or without including a certain threshold above which this relationship is present. Others consider that the effect is proportional to damage, which implies that the time course of effects may differ from the time course of the internal concentration.

The approach consists of three major steps: 1) experiments to derive the necessary input parameters for the TK- and TD-sub-models; 2) fitting the model to the experimental data; 3) validation of the model.

All models require a good deal of experimentation to derive the necessary input parameters. For the TK-part, uptake-elimination experiments are performed to establish the uptake and elimination rate constants, or, in case of the DebTox-model, the kinetic parameters are estimated from the development of effects in time. In this way, model input parameters are derived that will allow for the prediction of internal concentrations of chemicals for new situations. For the estimation of uptake and elimination kinetics in fish, an accepted OECD guideline (OECD, 1996) is available which can be used as a starting point for other organisms. While the principle experimental set-up is straightforward and worked out well, the results should be considered with care because variability may be rather large. Especially in case of (bio)degradation of the compound in the exposure medium and/or metabolisation within the organism, it may be hard to establish the 'true' concentrations in water and organisms during the experiment. In addition, if multiple compartments have to be considered, the data demand is extremely high.

The purpose of TD-experiments is to infer the time course of damage and repair to the target organisms based on the time-course of survival in response to specific pattern(s) of exposure to the test compound. In practice, several experiments will have to be carried out with time-varying exposures and frequent (e.g. daily) measurements of exposure concentration and effect endpoint. Although uptake and elimination constants will have been determined independently, measurement of internal concentrations within the organism at less frequent intervals allows checking that toxicokinetics have not deviated markedly from that expected.

Once parameters for both parts have been measured, appropriate modelling software can be used to fit the selected model to the experimental data. None of the TK/TD models has been extensively validated to date. Regulatory use of the models should thus be supported with (a) validation experiment(s) for the particular combination of compound and organism. Ideally, validation experiments should include an exposure profile that contrasts markedly with those used in model calibration (e.g. more/less pulses of shorter/longer duration than previously tested). Longer-term experiments are also useful to demonstrate the ability to extrapolate beyond the precise conditions of the experiments. Consideration of this evaluation phase requires careful definition of validity criteria. Since these criteria are not established, development of guidance is necessary at this point.

As stated in the ELINK-document, a major advantage of TK/TD-models is that the exposure profile is not a limiting factor. Once robust and broadly applicable input parameters have been established, predictions can be easily obtained for a large number of exposure situations. In addition, they allow for the characterisation of the risks from bioaccumulation and can be used to calculate recovery times for individual organisms after single pulses (Ashauer et al., 2007b).

There are, however, several reasons why the applicability of the current models is still limited. First of all, they are developed for simple, small organisms for which single-compartment, first-order toxicokinetics apply. In addition, they are focused on situations in which exposure is shorter than the lifespan of the organism of concern, and growth and changes in lipid content of organisms are negligible. Furthermore, the methodology has generally been applied to survival only. Although there is no reason why the models cannot be adapted, more research is needed to demonstrate the applicability of the approach to different organisms and sub-lethal endpoints (see also Rubach et al., 2010 and Rubach, 2010).

In line with this, the ELINK-workshop identified several areas of research to enable the use of TK/TD-models in regulatory purposes, but concludes that the models which describe lethality of aquatic invertebrates with insignificant growth and reproduction over the course of the exposure period may already be applied, although with care. This latter conclusion is illustrated with a case study in which the TDM-model by Ashauer et al. (2007a) is applied to support authorisation of two applications of a certain pesticide on the basis of results

from a mesocosm experiment with a single peak exposure. Although in the presented figures the TDM-model generally fits reasonably well to the survival data from the TD-experiment, a very large difference between observed and predicted survival is apparent when the calibrated model is used to predict survival in independent validation experiments. Although the model is generally conservative in that it underestimates survival, the conclusion from the ELINK-proceedings that a 'satisfactory fit' is obtained is questionable. The questions at stake are, which criteria to apply in deciding that the model fits the experimental data sufficiently, and whether an important regulatory decision should be based on models that are not yet sufficiently calibrated with experimental data.

In conclusion, TK/TD-models may have potential to be used as supportive evidence in risk assessment. So far, the use will be limited to predicting mortality of single aquatic invertebrate species. However, immediate use in the newly developed Dutch risk assessment scheme is not foreseen, unless validity is adequately demonstrated. Criteria should be developed to evaluate the validity of a model in a certain situation.

6.7.3 Population models

Population models generally aim to describe the dynamics of one population over time. Depending on the properties of the population to be modelled, different types of population models may be distinguished. The ELINK-report discusses several models which differ in their degree of complexity. A summary based on the ELINK-report and additional literature is presented here.

Unstructured models
Relatively simple, so-called unstructured models describe the population with state variables like population abundance or density (N). Most of traditional theoretical population ecology consists of unstructured models, in which it is assumed that individuals can be treated as 'nearly' identical and little differences between individuals are lost by aggregation or averaging over those differences. An example of these kind of models is the simple logistic growth model developed by Barnthouse (2004) to estimate recovery times of different aquatic taxa depending on magnitude of effect (here reduction of abundance) and the intrinsic population growth rate. The model assumes a constant intrinsic growth rate of the population and constant carrying capacity of the environment. Lin et al. (2005) developed a model to combine life-cycle survivorship and fecundity data obtained from individual level responses of medaka exposed to chemicals, into population-level responses defined as reduction of population growth rate (lambda).

Individual-based models
In an individual-based model, the characteristics of each individual within a population are tracked through time. Individual-based models simulate the overall consequences of local interactions of members of a population. These models typically consist of an environment, or framework, in which the interactions occur and a number of individuals defined in terms of their behaviours and characteristics.

Metapopulation models
Metapopulations are sets of local populations connected by migrating individuals. Local populations usually inhabit isolated patches of resources, and the degree of isolation may vary depending on the distance among patches. Metapopulation models consider local populations as individuals. Dynamics of local populations are either not considered at all, or are considered in an abbreviated way. Most metapopulation models are based on colonisation-extinction equilibrium. Elements such as landscape structure, life-history characteristics and the degree to which populations are connected, determine whether effects of toxicants on one or more spatially or temporarily separated populations will lead to extinction or whether recovery is possible. Examples of using metapopulation models in ecological risk assessment can be found in Spromberg et al. (1998) and Angeler and Alvarez-Cobelas (2005). Some metapopulation models are spatially explicit, meaning that they aim to

predict spatial and temporal distribution of populations after pesticide exposure. Van den Brink et al. (2007) developed such a metapopulation model for *Asellus aquaticus* and combined information on concentration-effect relationships and life-history characteristics to model movement patterns of individuals after exposure to pesticides, aiming to predict recolonisation of disturbed populations.

6.7.4 Community, food web or ecosystem models

Community, food web, or ecosystem models aim to describe the dynamics of communities, the fate and transfer of contaminants within food webs or the interaction of communities with their abiotic environment. Food-web models describe the changes in abundance of (groups of) organisms resulting from multiple trophic interactions and the transfer of energy and biomass through a food web. They include top-down and bottom-up processes, trophic cascades, and more complex interactions across multiple trophic levels (Gotelli and Ellison, 2006). Models which describe the transfer of residues through the food chain for compounds with a high BCF can be considered as a specific type of food-web models, although in this case emphasis is put on biomagnification to higher trophic levels and feed-back to lower levels is not applicable.

Traas et al. (2004) developed a food-web model on the basis of a microcosm experiment that addressed the interaction between eutrophication processes and contaminants. The model describes direct and indirect effects of nutrient additions and a single insecticide application on biomass dynamics and recovery of functional groups. Direct toxicant effects on sensitive species could be predicted reasonably well using concentration-response relationships from the laboratory with representative species. The model was extended with recolonisation scenarios, to simulate dose-dependent recovery.

6.7.5 Empirical models

Empirical or 'data mining' models predict effects of pesticides on the basis of correlations with existing data. An example of this is the PERPEST-model by Van den Brink et al., 2002a) in which data from micro-/mesocosm experiments are compiled with respect to e.g. mode of action, exposure pattern (including repeated exposure patterns and mixtures of pesticides), and effect classes. For a given pesticide, predictions are made of effects at population or community level given a certain ecological and exposure scenario.

6.7.6 Use of population models in risk assessment

As stated in Section 2.1, the risk assessment under the PPP Regulation aims at protecting species groups at the population level, implicitly assuming that unacceptable effects on the ecosystem level are prevented in this way. Taking this into account, the use of ecological models to translate the 'traditional' endpoints such as growth or reproduction, to population level parameters like intrinsic rate of population increase could be a way to improve the ecological basis of the current risk assessment procedures.

Forbes and Calow (2002b) argue that population growth rate analysis should be used as a basis for ecological risk assessment. In a review of 41 toxicity studies, which included a total of 28 species and 44 toxicants, they found that in 94 of the 99 cases considered effects on population growth rate were observed at concentrations higher than the most sensitive individual based endpoint. The analysis showed that there is no consistency in which of the measured individual-level parameters was the most sensitive to toxicant exposure, and none of them could be considered to be precise predictors of population growth rate. The most frequently measured parameter, reduction in survival, was not significantly correlated with reduction in population growth rate. The conclusion of the analysis was that in general the most sensitive individual life-cycle parameters are

protective of changes in population growth rate. However, it is not necessarily so that the most sensitive parameter is always measured in ecotoxicity tests.

A second observation from their analyses is that the translation of effects on individual-level parameters to changes in population dynamics can be markedly different for species with different life-cycles. Using a demographic model, Forbes et al. (2001) demonstrated that a reduction of 10% in juvenile survival would result in a 10% decrease in population growth rate λ for a benthic invertebrate, while for algae, fish and daphnids the effect would be 5, 2 and 0.6%, respectively. Although the benthic invertebrate may have a higher LC_{10} value than the daphnid, its population dynamics could thus be more sensitive: a 5% reduction in juvenile survival of the benthic invertebrate would have the same effect on λ as a 80% effect on juvenile survival for daphnids. The other way around, very different responses to individual demographic parameters can lead to comparable changes when population growth rate is considered.

As a third point of concern, Forbes and Calow (2002b) address the fact that the conditions under which ecotoxicity tests are performed (i.e. food, density) are rarely limiting for population growth. It cannot be predicted beforehand whether increased density will enhance or decrease the toxicant effects, experimental studies give mixed results. Model calculations could be used to explore different options.

It should be noted that 'classic' population models (e.g. matrix models or Euler-Lotka equation) are relatively simple, and limited in taking account of natural variability (Galic et al., 2010). The applicability of those models for specific higher tier assessments seems therefore to be rather limited. However, if data from reproduction studies would be reported in such a way that population growth rate can be calculated in addition to the standard test endpoints, this could be used to further underpin the conclusions of a (higher tier) risk assessment, or to identify areas for further research.

A major application of (meta)population models seems to be addressing those cases where recovery is the major concern (Galic et al., 2010). Wogram (2010) gives some examples of situations where models might be used to extrapolate mesocosm findings:

– recovery of populations was demonstrated, but voltinism (i.e. number of generations per year) of the taxa present in the study was not representative for vulnerable species in the field;
– recovery was demonstrated, but the exposure design was not representative of the proposed use (e.g. single application tested but multiple applications applied; early application tested, late application foreseen);
– recovery was not demonstrated within the duration of the experiment (e.g. for univoltine species such as Amphipods).

To date there are almost no cases in which authorisation of a plant protection product was granted based on modelling results. In Germany, several models were submitted to predict recovery from the results from single species toxicity tests or mesocosm experiments (Wogram, 2010). In all cases, the modelled species was the same terrestrial or aquatic invertebrate that turned out to be most sensitive in the toxicity tests. One of the models, an individual model on the phantom midge *Chaoborus chrystallinus*, was used to simulate population dynamics in isolated and connected test systems. The outcome was considered accurate and predictions were considered plausible and reliable. However, none of the models has been accepted by the German competent authority (Federal Environment Agency, UBA) for regulatory decisions, because the species modelled were not considered to be representative of a realistic worst case in agricultural landscapes. That is, while the test species were considered to be representative for field species in terms of toxicological sensitivity, they were not so in terms of ecological traits (e.g. generation time, dispersion). This argument applies, of course, mainly to extrapolation from typical laboratory test species such as Cladocerans, which are selected because of their short generation time. Mesocosms are expected to be representative for agricultural landscape, otherwise the results of the experiment itself, let alone extrapolations, would not be acceptable for

risk assessment. In France, some models have also been submitted as part of regulatory risk assessment, and one model was accepted because species, scenario and ecotoxicological inputs were considered relevant. According to Wogram (2010), an integral part of good modelling practice should be to a proper definition of the regulatory question. The model species should represent a realistic worst case in terms of the combination of ecological and ecotoxicological vulnerability. Together with a proper validation, this might lead to a broader acceptance of models for ecological risk assessment.

From the above, it can be concluded that at present the use of (meta)population models for registration purposes is limited, but that there is scope for their use in the near future. Most models are not developed from a generic risk assessment point of view, but rather to describe a certain (experimental) case. As a result, most risk assessors will question the validity of the model for situations beyond the one used to develop and calibrate the model. During the LEMTOX-workshop (Thorbek et al., 2010), it was noted that the decisions underlying the choice of the model type and structure are not transparent and often seem to be ad-hoc. The lack of validation was identified as a major reason for the limited acceptance of using models for regulatory purposes.

The benefits of ecological modelling can be found in focusing and designing further (higher tier) studies, identifying data gaps and improving the set-up of post-registration monitoring actions. In addition, models can be used to extract more information from complex datasets, increase confidence in safety factors and serve as supportive evidence that the protection goal is achieved. Finally, ecological models can be used as tools to extrapolate to higher levels of biological organisation, to different time scales and different environmental conditions. From the above, however, it appears that especially for this latter purpose further research and validation is needed. As for TK/TD-models, criteria should be developed to evaluate the validity of a model in a certain situation.

Since most 'effect models' developed to date are insufficiently calibrated for a proper regulatory use, and guidance is not yet available on good modelling practice for these simulation tools, assessments based on effect models can only be taken into account on a case-by-case basis and expert judgment. Currently, considerable research efforts take place to address these drawbacks and to further improve modelling approaches in effect assessment procedures for PPPs.

7 Risk/hazard assessment procedure for larger surface waters in line with 2000/60/EC

7.1 Introduction to the derivation of QSs

The aim of the following Chapters is to describe the procedures for deriving water quality standards in line with the Water Framework Directive, using the data that may be included in a regular dossier under the PPP-regulation. As explained in Section 2.2, two types of EQSs are distinguished to cover both long-term and short-term exposure to a chemical. These are called:

– the annual average concentration (AA-EQS), to protect against the occurrence of prolonged exposure, and
– the maximum acceptable concentration (MAC-EQS) to protect against possible effects from short term concentration peaks.

The EQSs should protect freshwater and marine ecosystems from adverse effects as well as human beings from all impacts on health. For the present report, only freshwater standards are taken into consideration, marine and sediment quality standards are not considered. Table 7-1 summarises the different routes that are considered for derivation of water quality standards within the WFD and the temporary standards during derivation.

Table 7-1

Overview of the different types of quality standards for surface water considered in the WFD.

Type of QS	Protection aim	Terminology for temporary standard[1]	Notes	Final selected quality standard
Long-term	Water organisms	$QS_{fw, eco}$	Refers to direct ecotoxicity	**Lowest selected as AA-EQS**
	Predators (secondary poisoning)	$QS_{biota, secpois, fw}$ $QS_{fw, secpois}$	QS expressed as concentration in biota is converted to corresponding concentration in water	
	Human health (consumption of fishery products)	$QS_{biota, hh food}$ $QS_{water, hh food}$	QS expressed as concentration in biota is converted to corresponding concentration in water; valid for fresh and marine waters	
	Human health (surface water for abstraction of drinking water)	$QS_{dw, hh}$	separate standard, not considered in this report	
Short-term	Water organisms	MAC-$QS_{fw, eco}$	Refers to direct ecotoxicity	**MAC-EQS**

[1] Note that the subscript 'fw' refers to the freshwater, subscript 'water' is used for all waters, including marine

For the final selected value for the AA-EQS, direct ecotoxicity, human consumption of fishery products and secondary poisoning of birds or mammals are considered. For each of these routes a quality standard (QS) is derived (when derivation triggers are met). These QSs have a subscripts indicating for which route the value

was derived, for example $QS_{fw, eco}$ is the value derived for freshwater from direct ecotoxicity, $QS_{biota, secpois}$ is the value derived for biota based on the secondary poisoning route. The lowest value is adopted as the overall AA-EQS. The relevance of the latter two routes (human consumption of fishery products or secondary poisoning of birds and mammals) depends on the physico-chemical characteristics and human toxicological information.
The QS for surface waters intended for the abstraction of drinking water is considered as a separate standard, and will not be considered in the present report. The MAC-EQS only relates to direct ecotoxicity. As shown for a series of 23 pesticides, direct ecotoxicity will most often be the critical route, although secondary poisoning and human consumption of fishery products was triggered for about 60% of the compounds (Bodar and Smit, 2008). This is in line with the specific function and design of PPPs.
Before discussing the different derivation procedures in Chapter 8, some general aspects are discussed first.

7.2 Linking exposure to effects

Acceptability of a PPP should be assessed for WFD water bodies by means of a generic risk assessment procedure according to the WFD. For this, MAC-EQS for short-term exposure peaks and the AA-EQS for long-term exposure will be compared with exposure concentrations in WFD water bodies.

For post-registration assessment, the exposure concentrations that result from chemical monitoring programmes will be used. The interpretation of monitoring data for registration purposes is worked out by the *Monitoring working group* (De Werd and Kruijne, 2011). The MAC-EQS will be compared with the highest exposure peak that is available from the selected exposure profiles. As outlined in Section 3.3.4, the AA-EQS is normally compared with the arithmetic mean of chemical monitoring samples taken at a sampling station over a year. However, for a substance that is used for only a short part of the year, a shorter period may be considered. In case of e.g. pesticides, which show peak concentrations within short time periods, enhanced sampling frequency may be necessary in these periods. For example, the best sampling time for detecting concentration peaks of pesticides is after heavy rainfall within or just after the application period. In line with this, it is proposed to compare the AA-EQS with the highest Time Weighted Average concentration over an ecotoxicologically relevant exposure period (e.g. three months), under the condition that the sampling frequency is intensified in this period.

7.3 Specific notes on ecotoxicity data

7.3.1 Dealing with freshwater and marine ecotoxicity data

According to the EQS-guidance the treatment of freshwater and marine toxicity data (*i.e.* species living and tested in water with salinity >0.5 ‰) will be changed. Previously, these datasets were kept separated and the freshwater $QS_{fw, eco}$ was based on freshwater species only. The approach of the EQS-guidance is also adopted for the drainage ditch risk assessment and already briefly presented in Section 6.2.

Additional data can be available for a variety of species, being either freshwater or marine species. The presence of marine data is generally less relevant for PPPs, but the option to include literature data will probably generate more data. Furthermore, studies on marine species are part of the standard dossier for registration in the USA, and will thus sometimes be available in the EU-dossier too. For the purpose of this report, marine species are defined as living and tested in brackish or saltwater (salinity >0.5 ‰). The question how to deal with these data has been subject of discussion within the framework of the WFD, but the considerations made within that context are applicable in general. The following is taken largely from a document that was prepared by the Netherlands as a background document to the WFD-guidance (EC, 2011).

Both the TGD (EC, 2003) and its revision under REACH (ECHA, 2008), recommend that for plant protection products (PPPs) toxicity data for freshwater and saltwater organisms should not be pooled for PNEC derivation. The reasoning in both documents is equal: 'within trophic levels differences larger than a factor of 10 were shown for several metals and pesticides indicating that for these compounds fresh water and saltwater data should not be pooled for hazard assessment and PNEC derivation'. ECETOC (2000) is given here as a single reference to a background document. The methodological choice made in the two guidance documents (TGD, REACH) is not clearly underpinned. The *a priori* separating of aquatic toxicity data for PPPs has been adopted by Lepper (2005). Attempts to retrieve the ECETOC (2000) publication failed, although part of this work was probably published in ECETOC Technical report 82 (2001), and by Hutchinson et al. (1998), Leung et al. (2001) and Wheeler et al. (2002).

Maltby et al. (2005) and Brock et al. (2008) pointed out that for pesticides with a specific mode-of-action, it is rather the taxonomic group than the place in the food chain or food web (trophic level) that determines the sensitivity. In their study, of the ten insecticides of which SSDs were compared based on acute data, no significant differences between HC_5 values for freshwater and saltwater taxa of the same sensitive taxonomic group were found. For two compounds (out of ten; permethrin and chlorpyrifos) differences in HC_5 could be established when arthropods were compared, but this difference could not be demonstrated anymore after selecting crustaceans as sensitive taxonomic group. Maltby et al. (2005) conclude that freshwater and saltwater toxicity data can be combined, but that it is important to be aware of differences in taxonomic position and consequences for threshold concentrations.

Solomon et al. (2001) showed that differences (fresh vs. marine) were observed when comparing acute data for permethrin. For fenvalerate a difference in sensitivity was only observed when data for arthropods (insects and crustaceans) and fish were compared. When comparing complete datasets (including e.g. algae, Mollusca), the 10^{th} or 5^{th} percentile of the freshwater and marine datasets were similar. Note that Solomon et al. (2001) did not make a distinction between crustaceans and insects in comparing marine and freshwater toxicity data for arthropods. Leung et al. (2001) showed a difference in acute sensitivity to chlordane. However, the authors pointed out that 'there is considerable potential for freshwater to saltwater prediction'. They state that differences between the taxonomic compositions of the data sets should be considered. Wheeler et al. (2002) have compared SSDs for pesticides based on acute toxicity data and reported differences in HC_5 values ranging from a factor of 2 to 12 for five (out of seven) of the compounds. They concluded that for pesticides, freshwater data could be used for saltwater risk assessments, but with - possibly - an additional 'modest' safety factor depending on how the sensitive taxonomic groups are represented in the saltwater data set.

A draft report on the SETAC 2006 workshop on quality standards setting (Anonymous, 2007) reports on this topic that: 'Overall the lack of data hampers a sound and definitive comparison, but current scientific opinion is that there is no systematic bias in sensitivity between freshwater and marine species, provided similar tests and endpoints are involved.' Also: 'if there is no indication of differential sensitivity to a particular substance between freshwater and marine organisms, it may be appropriate to combine both datasets in a single SSD, although any resulting quality standard should be regarded as tentative.' Please note that construction of SSDs in quality standard derivation occurs only for very data rich compounds.

Based on the above presented information from the literature, the EQS-guidance states that a statistical evaluation should be performed to test whether or not data from freshwater and marine species should be treated separately. Where there are sufficient toxicity data in both the freshwater and marine datasets to enable a statistical comparison, the following procedure should be followed. The null hypothesis is that freshwater and saltwater organisms do not differ in their sensitivity to the compound of interest; i.e. they belong to the same statistical population:

1. All freshwater data are collected and tabulated (note: this data set contains one toxicity value per species, see Footnote 1 in Section 6.2 for an explanation). Next, a logarithmic transformation of each of these toxicity values is performed.
2. All marine data are collected and tabulated (note: this data set contains one toxicity value per species). Next, a logarithmic transformation of each of these toxicity values is performed.
3. Using an F-test, determine whether the two log-transformed data sets have equal or unequal variances. Perform the test at a significance level (α) of 0.05.
4. A test for differences between the data sets e.g. a two tailed t-test where the data are normally distributed (with or without correction for unequal variances, depending on the results of step 3), is performed. Perform the test at a significance level (α) of 0.05 .
5. Especially for compounds with a specific mode-of-action, it is important to identify particularly sensitive taxonomic groups and perform a separate statistical analysis for this specific group. If enough data are available to make a comparison for individual or related taxonomic groups (e.g. crustaceans, arthropods, fish, vertebrates), this may help to determine if there are differences between saltwater and marine species. Note that there are only few marine insects.

In those cases where there are too few data (either freshwater or marine) to perform a meaningful statistical comparison and there are no further indications (spread of the data, read-across, expert judgement) of a difference in sensitivity between freshwater vs. marine organisms, the data sets may be combined for $QS_{fw, eco}$ derivation. The notes given in Section 6.2 on the use of marine mesocosms also apply to $QS_{fw, eco}$ derivation. In general, it is proposed to use marine mesocosm data only in addition to freshwater data. In practice, this means that a single marine mesocosm without any equivalent freshwater studies will only be used as supportive evidence, but not as the sole basis for the $QS_{fw, eco}$.

7.3.2 **Special considerations on micro-organisms**

According to the EQS guidance (EC, 2011), data for bacteria representing a further taxonomic group may only be used if non-adapted pure cultures were tested. Furthermore, studies with bacteria (e.g. growth tests) are regarded as short-term tests. Consequently, unlike for algae, NOECs or EC_{10} values derived from bacterial studies may not be used in the derivation of the AA-EQS using assessment factors. EC_{50} values from bacterial tests may be used as additional acute data.

The EQS-guidance probably refers to bacteria tests with a short contact time in which a generic parameter such as CO_2 evolution is measured. If, however, a reliable bacteria test is available that is comparable to an algae test in terms of duration and endpoint (i.e. 72 hours and specific growth rate), there is scientific evidence to include the endpoint in the dataset. The same principle applies to toxicity data using protozoans. For the purpose of EQS-derivation for PPPs within the context of the present report, it is therefore proposed to accept NOECs for bacteria and protozoans as chronic endpoints, if obtained in a comparable way as those for algae.

The EQS guidance does not make reference to fungi as a specific taxonomic group. As pointed out previously (see Section 6.4.3.3), data on fungi are considered relevant for fungicide risk assessments and may become available in the (near) future. If growth tests with fungi are present, it is advised for the time being to treat the data similarly to algae, i.e. include the EC50 for the acute dataset and the NOEC in the chronic dataset. It was also noted in Section 6.4.3.3 that the kingdom of fungi is diverse. The selection of relevant species for which standardised ecotoxicity tests may be developed is therefore identified as a further research need. In addition, more research is needed into the life-span and generation time of aquatic fungi, to determine whether or not short-term tests can be used to derive chronic endpoints. These points should be considered when updating the EQS-guidance, and are therefore taken forward to Chapter 9.

7.3.3 **Endocrine disruptors**

When there are indications that a substance may cause adverse effects via disruption of the endocrine system of mammals, birds, aquatic or other wildlife species, the assessor should consider whether the AF that is normally applied for a certain combination of data (see 8.2) would be sufficient to protect against effects caused by such a mode-of-action, or whether a larger AF is needed. Since PPPs with endocrine disrupting properties will not be authorised, this is less relevant for the present report, although the way in which endocrine disruption should be evaluated under the PPP-regulation is still under discussion.

7.3.4 **Use of non-testing methods to reduce uncertainty**

Emphasis is placed on experimental toxicity data for deriving a $QS_{fw, eco}$. However, non-testing methods (e.g. QSARs, read-across methods) are also available which can be used to predict toxicity of certain organic chemicals and endpoints. They should not be used to generate critical data to derive a $QS_{fw, eco}$, but predicted data can play a role in reducing uncertainty and thereby influence the size of AF chosen for extrapolation. In principle, the PPP dossier already contains enough data to derive a $QS_{fw, eco}$ by any of the methods described below. However, in case there is uncertainty as to whether the potentially most sensitive taxonomic group is included in the dataset, or when deciding on the applicability of SSDs, non-testing methods can be considered. Reference to this is made in the following sections where relevant. It should be noted that most QSARs have been derived for those organisms which are already included in the PPP-dossier. Furthermore, care should be taken in the application of QSARs for substances with a specific mode of action.

8 Derivation of the QS_{fw} and MAC-QS

8.1 Introduction

The different assessments that are required for QS-derivation are summarised in Table 8-1 with reference to the section in which they are discussed.
The quality standards based on direct ecotoxicity ($QS_{fw, eco}$ or MAC-$QS_{fw, eco}$) can be derived in three ways, depending on the availability of data:

1. Applying an assessment factor to the lowest credible datum ('AF approach').
2. Species sensitivity distribution modelling ('SSD approach').
3. Using results from mesocosms ('model ecosystem approach').

For the ease of reading, the three methods for derivation of the ecotoxicity-based QSs are discussed in separate sub-chapters: the assessment factor method is discussed in Section 8.2, the SSD-method in Section 8.3 and the mesocosm approach in Section 8.4. For each of these methods, the derivation of the $QS_{fw, eco}$ and MAC-$QS_{fw, eco}$ is discussed in separate sub-sections.
The biota-based assessments for predators and human health are presented in Section 8.5, secondary poisoning of predators is discussed in Section 8.5.1, human health in Section 8.5.2 and conversion to water based standards in Section 8.5.3.
Finally, the selection of the $QS_{fw, eco}$, MAC-$QS_{fw, eco}$ and final AA-EQS and MAC-EQS is dealt with in Section 8.6.

Table 8-1
Overview of WFD assessments relevant for AA-EQS and MAC-EQS derivation in the framework of PPP admission.

Type of quality standard	Relevant route	Terminology and methods	Described in section
AA-EQS	Direct ecotoxicity to water organisms	$QS_{fw, eco}$	
		AF approach	8.2
		SSD approach	8.3
		model ecosystem approach	8.4
		Selection of $QS_{fw, eco}$	8.6
	Predators via fish	$QS_{biota,secpois}$	8.5.1
		Convert to $QS_{fw, secpois}$	8.5.3
	Humans via fish	$QS_{biota,hh\ food}$	8.5.2
		Convert to $QS_{water, hh\ food}$	8.5.3
	All routes	Selection of overall AA-EQS	8.6
MAC-EQS	Direct ecotoxicity to water organisms	MAC-$QS_{fw, eco}$	
		AF approach	8.2.4
		SSD approach	8.3.5
		model ecosystem approach	8.4
		Selection of MAC-EQS	8.6

8.2 Assessment factor approach for derivation of the $QS_{water, eco}$ and MAC-QS

For substances with small datasets that do not meet the requirements of the SSD method (see 8.3), the $QS_{fw, eco}$ and MAC-$QS_{fw, eco}$ are derived by a deterministic approach, i.e. using an assessment factor on the lowest credible datum. The procedures for estimating an $QS_{water, eco}$ are the same as the aquatic effects assessment and the calculation of the PNEC ($\approx QS_{fw, eco}$) described in the guidance prepared for REACH (ECHA, 2008). The derivation of the MAC-$QS_{fw, eco}$ is adapted from the assessment of intermittent releases within REACH.

The quantity and type of data available determines the assessment factors used. The assessment scheme for derivation of the $QS_{fw, eco}$ and MAC-QS is presented in detail in the EQS-guidance (EC, 2011). The schemes have been developed for all types of chemicals, including those for which ecotoxicity data are scarce, and offer the possibility to derive a $QS_{fw, eco}$ and MAC-$QS_{fw, eco}$ in case only acute data for algae, *Daphnia* and fish are available. For the $QS_{water, eco}$ an AF of 1000 is applied in case only L(E)C_{50}-values are available, the factor may be lowered to 10 depending on the amount and nature of additional data. For PPPs, the data that are available from a PPP dossier (see Section 5.2) will in principle allow for lower AFs.

According to the EQS-guidance, acute data for algae, *Daphnia* and fish, and chronic NOECs for three species from three trophic levels, may allow for the use of an AF of 10, provided that the species tested represent one of the more sensitive taxonomic groups. Footnote d to the table with assessment factors in the EQS-guidance states that:

> 'An assessment factor of 10 will normally only be applied when long-term toxicity results (e.g. EC_{10} or NOECs) are available from at least three species across three trophic levels (e.g. fish, *Daphnia*, and algae or a non-standard organism instead of a standard organism). When examining the results of long-term toxicity studies, the $QS_{fw, eco}$ should be calculated from the lowest available long term result. Extrapolation to the ecosystem can be made with much greater confidence, and thus a reduction of the assessment factor to 10 is possible. This is only sufficient, however, if the species tested can be considered to represent one of the more sensitive groups. This would normally only be possible to determine if data were available on at least three species across three trophic levels.'

It is thus very important to notice that an AF of 10 only applies when there is evidence that a potentially sensitive species group is included in the dataset. If this is not the case, a higher AF of 50 or 100 should be considered, according to footnote c:

> 'An assessment factor of 50 applies to the lowest of two long term results (e.g. EC10 or NOECs) covering two trophic levels when such results have been generated covering that level showing the lowest L(E)C50 in the short-term tests. It also applies to the lowest of three long term results (e.g. EC10 or NOECs) covering three trophic levels when such results have not been generated from that trophic level showing the lowest L(E)C50 in the short-term tests. This should however not apply in cases where the acutely most sensitive species has an L(E)C50 value lower than the lowest long term result (e.g. EC10 or NOECs) value. In such cases the QSfw, eco might be derived by using an assessment factor of 100 to the lowest L(E)C50 of the short-term tests.'

This is further explained in the EQS-guidance as follows:

> 'An assessment factor of 10 is applied to the lowest chronic NOEC or EC_{10} if chronic data are available from all three trophic levels of the base set. The trophic levels of NOECs and/or EC_{10}s should include the trophic level of the lowest acute L(E)C_{50}. If acute toxicity data are available for trophic levels not

covered in the chronic toxicity data, and the trophic level of the lowest L(E)C_{50} is not included in that of the NOECs and/or EC_{10}s then:
- an assessment factor of 50 is applied to the lowest NOEC or EC_{10} if the lowest L(E)C_{50} is higher than the lowest NOEC or EC_{10};
- an assessment factor of 100 is applied to the lowest L(E)C_{50} if the lowest L(E)C_{50} is lower than the lowest NOEC or EC_{10}.'

For insecticides, the use of the term 'trophic level' in these citations is complicating because crustaceans and insects may belong to the same trophic level, while for compounds with a certain mode-of-action (e.g. neonicotinoid insecticides) large differences in sensitivity may exist between these taxonomic groups. The choice of the AF is therefore determined by the fact whether or not the potentially most sensitive taxonomic group is represented in the dataset. The focus on taxonomic group rather than trophic level is also applicable for other types of pesticides, like fungicides and herbicides.

The AF-method for the $QS_{fw, eco}$ is outlined below in separate sections for insecticides (8.2.1), herbicides (8.2.2), and fungicides (8.2.3), respectively. Section 8.2.4 deals with the derivation of the MAC-$QS_{fw, eco}$ with the AF-method.

8.2.1 Derivation of the $QS_{water, eco}$ for insecticides

The minimum dossier dataset for insecticides is presented in Table 8-2. Data that are specific for insecticides are indicated by a shaded background.

Table 8-2

Minimum dataset for insecticides obtained from PPP-dossier.

Acute L(E)C_{50}		**Chronic NOEC/EC_{10}**	
Taxon	**Note**	**Taxon**	**Note**
Algae[1]	green, e.g. *Pseudokirchneriella subcapitata*	Algae	green, e.g. *P. subcapitata*
Crustacea	*Daphnia sp.* (*D. magna* preferred)	Crustacea	*D. magna* / *Additional species*[3]
Crustacea[2]	Additional species, e.g. *A. bahia*		
Insecta[2]	freshwater insect, e.g. *Chironomus riparius*	Insecta[4]	*Chironomus riparius* (water spiked preferred)
Pisces	*Oncorhynchus mykiss*	Pisces[5]	ELS/FLC
Pisces[6]	warm water species		

1 The acute EC_{50} for algae usually is derived from the same test as the chronic NOEC.

2 Required for insecticides or compounds with insecticidal activity; alternatively, other more relevant freshwater non-crustacean species, e.g. *Chironomus* spp. may be used if guidelines ore protocols are developed.

3 Chronic test should be performed with most sensitive species in acute tests if the difference in acute EC_{50} values between *Daphnia* and additional species is larger than an order of magnitude.

4 Endpoints from water/sediment systems can only be used if water concentrations during exposure can be accurately described, see text below.

5 It is anticipated that the trigger to conduct a chronic fish test is met for most PPPs; in older dossiers also the 28-d NOECs for fish may be available that can be used

6 For animal welfare reasons a test with warm water fish may not be obliged but information often is available in the dossier.

The presence of acute toxicity data for an additional crustacean (and/or insect) species makes it possible to determine with greater confidence whether the species tested represent one of the more sensitive taxonomic groups. The PPP-dossier for insecticides often contains additional acute studies with insect species, but chronic tests with insects will not necessarily be included in the dossier. Chronic water/sediment tests with *C. riparius* should be submitted for compounds that interfere with moulting (insect growth regulators, IGR) or any other type of compound that has a target specific for insects. Also for the purpose of sediment risk assessment water-spiked water/sediment studies may be available. These studies, however, hardly allow for estimation of the exposure concentration in the water phase over time. In case the acute studies show that insects are more sensitive than crustaceans (incl. *A. bahia*), the chronic studies with *D. magna* or *A. bahia* do not give enough confidence as to whether the chronic data cover the potentially most sensitive species group. It that case, an AF of 10 is no longer appropriate, and the next higher factor of 50 should be considered. It should thus be decided whether a certain taxon is indeed more sensitive than the expected most sensitive species group. Considerable differences may be observed between test results for the same species and endpoint (see e.g. Baird et al., 1990, 1991), a factor of 10 is not uncommon. This would imply that a difference in $L(E)C_{50}$-values is not necessarily related to differences in sensitivity between taxa. However, if for a certain insecticide the LC_{50} for insects is 0.2 mg/L, while the EC_{50} for crustacea is 1 mg/L (i.e. a factor of 5 difference), this will generally be interpreted as an indication that insects may be more sensitive. This means that if insects are not present in the chronic dataset, an assessment factor of 10 is not justified and a higher assessment factor (50 or 100) should be applied. As a pragmatic approach, it is proposed that if the acute endpoint of an insect is less than a factor of 3 lower than that of crustacea, the two taxa are considered to be equally sensitive. In that case, an AF of 10 would be still allowed even when chronic data for insects are absent. It is recognized, however, that the opinions on this subject differ. Therefore, additional relevant information that substantiates the choice of the assessment factor should be considered. Read-across and the use of QSARs (see Section 7.3.4) may also be options to consider in order demonstrating that the potentially most sensitive species group is included in the dataset. Of course, information from additional (higher tier) studies can also be considered. For instance, a 10-days water-only study with *C. riparius* larvae from the open literature does not fit in the data requirements of a PPP dossier, but can give very useful information with respect to the relative sensitivity of insects as compared to crustaceans. However, if such a 10-days test delivers the lowest endpoint, the question should be asked whether this endpoint reflects true chronic exposure and justifies an AF of 10. If from a mesocosm it appears that crustaceans are equally sensitive as insects, this information can be used to underpin a lower AF. On the other hand, if additional studies point at a much more sensitive taxon that is not represented in the laboratory data set, a higher AF should be considered. It should be emphasised that the most sensitive taxon in the acute data set not necessarily needs to be the most sensitive taxon in the chronic data set. In fact comparing the place of specific taxa in species sensitive distributions between acute and chronic SSDs is an important topic for future research (see Chapter 9).

8.2.2 Derivation of the $QS_{fw,\ eco}$ for herbicides

The minimum dossier dataset for herbicides is presented in Table 8-3. Data that are specific for herbicides are indicated by a shaded background.

Table 8-3

Minimum dataset for herbicides obtained from PPP-dossier.

Acute L(E)C_{50}		**Chronic NOEC/EC_{10}**	
Taxon	**Note**	**Taxon**	**Note**
Algae[1]	e.g. *Pseudokirchneriella subcapitata*	Algae	e.g. *P. subcapitata*
Algae[1]	blue green algae/diatom	Algae	blue green algae/diatom
Macrophyta[1]	*Lemna* sp. / *Myriophyllum* sp./*Glyceria maxima*[2]	Macrophyta	*Lemna* sp. / *Myriophyllum* sp./*Glyceria maxima*[2]
Crustacea	*Daphnia sp.* (*D. magna* preferred)	Crustacea	*D. magna*
		Insecta	*Chironomus riparius*[3]
Pisces	*Oncorhynchus mykiss*	Pisces[4]	ELS/FLC
Pisces[5]	warm water species		

1 The acute EC_{50} for algae usually is derived from the same test as the chronic NOEC.

2 Additional testing may be required on other macrophyte species (*Myriophyllum* sp. or *Glyceria maxima*) depending on the mode of action of the substance, or if clear indications of higher toxicity are apparent to dicotyledonous (e.g. auxin inhibitors, broad leaf herbicides) or other monocotyledonous (e.g. grass herbicides) plant species from efficacy or testing with terrestrial non-target plants.

3 If the compound accumulates in sediment; endpoints from water/sediment systems can only be used if water concentrations during exposure can be accurately described.

4 It is anticipated that the trigger for a chronic fish test will apply for most PPPs; in older dossiers also the 28-d NOECs for fish may be available that can be used.

5 For animal welfare reasons a test with warm water fish may not be obliged but information often is available in the dossier.

As stated above, acute data for algae, *Daphnia* and fish, and chronic NOECs for three species from three trophic levels, allow for the use of an AF of 10, provided that the species tested represent one of the more sensitive taxonomic groups. With acute and chronic data present for two algae species and at least one macrophyte, this is normally the case. Of course, it should always be checked whether the data indeed allow for an AF of 10. In theory, additional data could point at an unexpected high acute toxicity for another taxon. If this taxon is not present in the chronic dataset, this should be taken into consideration and a higher assessment factor should be used. As described above for insecticides, the question whether a lower endpoint indeed points at an unexpected sensitive taxon is subject for discussion. As a pragmatic approach, it is proposed that if an unexpected taxon gives an acute endpoint that is less than a factor of 3 lower than that for primary producers, an AF of 10 on the chronic data is considered justified, even if the taxon with the lowest acute endpoint is not represented in the chronic dataset. It is known, however, that the views on this subject differ. Therefore, additional relevant information that substantiates the choice of the assessment factor should be considered.

Blue-green algae should be counted among the primary producers due to their autotrophic nutrition (ECHA, 2008). Thus, cyanobacteria (blue-green algae or Cyanophyta) belong to the trophic level of primary producers. This means that data from (both chronic and acute) tests with cyanobacteria are considered as additional algal data and are treated in the same way (*i.e.* if they represent the lowest endpoint, the AF will be based on cyanobacteria, even when data for green algae are present).

8.2.3 Derivation of the $QS_{fw, eco}$ for fungicides

No specific data requirements are set for fungicides in addition to the basic dossier data. The minimum dossier dataset for fungicides is therefore as follows (Table 8-4).

Table 8-4

Minimum dataset for fungicides obtained from PPP-dossier.

Acute L(E)C_{50}		**Chronic NOEC/EC_{10}**	
Taxon	**Note**	**Taxon**	**Note**
Algae	e.g. *Pseudokirchneriella subcapitata*[1]	Algae	e.g. *P. subcapitata*
Algae	blue green alga/diatom	Algae	blue green alga/diatom[2]
Macrophyte	*Lemna*[2]	Macrophyte	*Lemna*[2]
Crustacea	*Daphnia sp.* (*D. magna* preferred)	Crustacea	*D. magna*
		Insecta	*Chironomus riparius*[3]
Pisces	*Oncorhynchus mykiss*	Pisces[4]	ELS/FLC
Pisces[5]	warm water species		

1 The acute EC50 for algae usually is derived from the same test as the chronic NOEC.
2 For fungicides with a herbicidal mode of action.
3 If the compound accumulates in sediment; endpoints from water/sediment systems can only be used if water concentrations during exposure can be accurately described.
4 It is anticipated that the trigger for a chronic fish test will apply for most PPPs; in older dossiers also the 28-d NOECs for fish may be available that can be used.
5 For animal welfare reasons a test with warm water fish may not be obliged but information often is available in the dossier.

As mentioned in Section 6.4.3.3, a lot of fungicides act as general biocides. For these compounds, it cannot be predicted beforehand which species group is most sensitive, and the variation between species within a taxonomic group may be large. Other fungicides are very toxic for a specific species group. If that appears to be the case, the PPP-dossier will most often contain additional data. However, ecotoxicity data for aquatic fungi will generally not be present in the dossier, which means that a potentially sensitive species group is not represented. Even when one of the 'traditional' species groups is much more sensitive than the other taxa (such as fish in the case of captan), it has to be considered if a $QS_{fw, eco}$ based on that group will also be protective for non-target fungi, *i.e.* that the sensitivity of aquatic fungi is comparable to that of the other aquatic species already included in the dataset.

Maltby et al. (2009) compiled aquatic ecotoxicity data for a series of fungicides. The dataset included acute single-species data for 42 fungicides, semi-field data for twelve fungicides and covered seven modes-of-action and different exposure regimes. SSDs were constructed for separate taxonomic groups (*i.e.* fish, invertebrates, and primary producers) and for all groups together. Based on EC_{50} values, fish were less sensitive for fungicides belonging to the group of ethylene bisdithiocarbamates (EBDC, e.g. mancozeb, maneb, metiram, zineb), and inhibitors of sterol biosynthesis (conazoles, e.g. cyproconazole, tebuconazole), but they were generally more sensitive towards multi-site inhibitors such as captan, that do not belong to the EBDC-compounds. For fungicides that inhibit energy production, such as the quinone inhibitors, no overall significant differences between taxonomic groups were observed. When comparing SSDs for the combined data of different taxonomic groups, there was no significant effect of the mode-of-action on interspecies variation in sensitivity. Maltby et al. (2009) also compared three levels of hazardous concentration (HC_5, LL HC_5 and HC_1) from acute SSDs to NOECs and LOECs from mesocosm studies with fungicides (and separately plus insecticides and herbicides). For three out of nine fungicides, the HC_5 was lower than the NOEC from the mesocosm studies, while the lower limit of the HC_5 and HC_1 were always protective for ecosystem effects. In four studies, leaf decomposition was studied and the LOECs for this parameter was an order of magnitude higher than that for effects on the most sensitive structural parameter. The authors conclude that there is no evidence to suggest that derived threshold values based on hazardous concentrations (HC_p) from acute aquatic SSDs would pose a risk to aquatic hyphomycetes. However, (laboratory) effect data on fungi were not included in the datasets, and none of the semifield studies specifically studied fungi. The authors therefore

also concluded that the underlying data is limited in number and that further research on nontarget fungi should be conducted.

The importance of generating data for aquatic fungi was recently demonstrated in a screening study on the toxicity of fungicides to aquatic fungi (Dijksterhuis et al., 2009; CBS, 2009; Dijksterhuis et al., 2011). Waterborne fungi species were sampled in the field and isolates of six species were exposed to carbendazim, chlorothalonil, fluazinam, imazalil, epoxiconazole, tebuconazole and azoxystrobine. Effect on fungi growth was most pronounced for the ergosterol inhibitors imazalil, tebuconazole and epoxiconazole, of which the latter two triazoles were most toxic. For these compounds, effects were noted at the level of acute HC_5 as assessed by Maltby et al. (2009). This means that there is strong evidence that the current effect assessment, which is based on toxicity data of algae, *Daphnia* and fish, is not protective for non-target fungi in case of fungicides classified as ergosterol inhibitors, particularly triazoles. Ergosterol synthesis is specific for fungi, which can explain the high sensitivity of fungi as compared with other species groups. For other types of fungicides, it cannot be concluded beforehand that the current methodology is protective. There are several fungicides for which the target site is widely conserved across animal, fungal and plant kingdoms, which could be an argument to assume that the differences in sensitivity between species are smaller than for specific acting fungicides. Still, fungi could be more sensitive than other taxa.

If data on aquatic fungi are not available, it cannot be concluded with confidence whether or not the potentially most sensitive species group is represented in the dataset. In view of this, we propose to apply an AF of 50 to the lowest chronic NOEC of fungicides, in case the chronic dataset is complete, but no additional information on (aquatic) fungi is present. This factor may be lowered to 10 if there is supportive information that the available endpoints are also representative for the sensitivity of fungi. Apart from the data of Dijksterhuis et al. (2009, 2011) and CBS (2009), supportive information may be found in the efficacy dossier. Sometimes the results of efficacy tests with fungi are presented in terms of IC_{50} values. These tests generally do not meet the quality criteria for inclusion in the ecotoxicity dataset, but can be used as an indication if sensitivity of fungi is in the same order of magnitude as for the dossier species. Tests with fungi may also be present in the soil ecotoxicity dossier, since soil fungi are more regularly tested. Although effect concentration cannot be converted directly to water concentrations, a comparison with data for earthworms, soil arthropods and plants may give an indication of the relative sensitivity of fungi as compared to other taxa. Finally, the data of Dijksterhuis et al. (2009, 2011) and CBS (2009) may allow for read-across to related compounds.
As already mentioned in previous sections (see 6.4.3.3), more research is needed to determine which aquatic fungi species are most relevant for testing, and which test duration is needed for derivation of chronic endpoints.

8.2.4 Derivation of the MAC-$QS_{fw, eco}$ using the AF-method

The assessment scheme for derivation of the MAC-QS is presented in detail in the EQS-guidance (EC, 2011). Where there are at least for three species short term tests that represent three trophic levels (base set), an AF of 100 is normally applied to the lowest $L(E)C_{50}$ to derive the MAC-$QS_{fw, eco}$. Under some circumstances an AF less than 100 may be justified, e.g.

- For substances which do not have a specific mode of action (e.g. acting by narcosis only), if the available data show that interspecies variations are low (standard deviation of the log transformed $L(E)C_{50}$ values is <0.5) an AF of 10 may be appropriate.
- For substances with a specific mode of action, the most sensitive taxonomic groups can be predicted with confidence. Where representatives of the most sensitive taxonomic groups are present in the acute dataset, an AF of 10 may again be justified.

– Where there is a good understanding of the relationship between acute and chronic toxicity (e.g. acute: chronic ratios for a range of species), the AF used to estimate the MAC-$QS_{fw, eco}$ may be selected to reflect this, or at least to ensure the MAC-$QS_{fw, eco}$ is not lower than the $QS_{fw, eco}$.

In no case should an AF lower than 10 be applied to a short-term L(E)C_{50} value. The accompanying AF-scheme is given below in Table 8-5 (adapted from the EQS-guidance).

Table 8-5

Assessment factors to derive a MAC-$QS_{fw, eco}$.

Toxicity data	Assessment factor	Remark
At least one short-term L(E)C_{50} from each of three trophic levels of the base set (fish, crustaceans and algae)	100	Acute toxicity data for standard test species usually available for PPPs
At least one short-term L(E)C_{50} from each of three trophic levels of the base set (fish, crustaceans and algae)	10[a]	Acute toxicity data for different species do not have a higher standard deviation than a factor of 3 in both directions[b] OR known mode of toxic action and representative species for most sensitive taxonomic group included in data set

[a] Lowest assessment factor to be applied.

[b] To assess the span of the acute toxicity data, all reliable acute toxicity data collected are used, with a minimum of three LC_{50} or EC_{50} values, for species representing each of the base set trophic levels (algae, *Daphnia*, fish). If the standard deviation of the log transformed L(E)C_{50} values is <0.5, an assessment factor of 10 could be applied, otherwise an assessment factor of 100 should be applied.

The considerations listed above for lowering the AF are all applicable to PPPs. The same reasoning applies as discussed above for the $QS_{fw, eco}$ for the different types of PPPs (sections 8.2.1-8.2.3) the question to be answered is if on the basis of the available data it can be concluded with confidence whether or not the potentially most sensitive species group is represented in the dataset. If this is not the case, an AF of 10 cannot be applied.

For insecticides, an AF of 10 will normally be justified when acute insect data are present in the dataset (i.e. 48-hours water-only study with *Chironomus riparius*), or when there is evidence (e.g. from chronic or higher tier studies) that data on crustaceans (incl. *Americamysis bahia*) cover the sensitivity of insects. For herbicides, an AF of 10 will generally be possible, since acute data for algae and macrophytes will be available from the dossier. For fungicides that act as general biocides and for which it may be assumed that aquatic fungi are equally sensitive as compared to other species, an AF of 10 may be considered. This also applied to insecticides, herbicides and fungicides for which the acute data cover the potentially most sensitive species groups. For fungicides, a MAC-$QS_{fw, eco}$ based on the base set might not be protective for non-target fungi, so that a higher AF than 10 (i.e. 100) is required (see Table 8-5).

Some comments should be made on the AF of 10, because this factor does not seem to be consistent with the AF for the chronic assessment. In the chronic assessment, the AF of 10 to the lowest NOEC is meant to cover residual uncertainty, which is related to e.g. variation within the sensitive taxonomic group, and the translation of single species laboratory data to the field situation. This residual uncertainty also applies to the derivation of the MAC-$QS_{fw, eco}$. However, it has also to be taken into account that the MAC-$QS_{fw. eco}$ represents an acute no effect level. The underlying data, however, represent a 50% effect level, and an additional factor is needed to go from the L(E)C_{50} level to the acute NOEC or L(E)C_{10}. It is thus questionable whether an AF of 10 on the lowest L(E)C_{50} is justified on the basis of laboratory data alone. A higher factor would also be more in

line with the treatment of acute data under the PPP-regulations. In the first tier the risk assessment of drainage ditches, a safety factor of 100 is normally applied to acute laboratory data for derivation of the RAC (see Section 5.2.1). This is generally more stringent, although under the umbrella of the PPP Regulation also the geomean method may be applied if additional data on species from the same taxonomic group are present. This point should be considered when updating the EQS-guidance, and is therefore taken forward to Chapter 9.

8.3 Species Sensitivity Distribution method

The SSD-method for the $QS_{fw, eco}$ is outlined below. First, some general aspects are discussed, concerning data requirements (8.3.1), specific options for PPPs (8.3.2) and choice of the distribution (8.3.3). Section 8.3.4 discusses the derivation of the $QS_{fw, eco}$, Section 8.3.5 presents the derivation of the MAC-QS.

8.3.1 Data requirements

Statistical extrapolation in line with the provisions of the REACH guidance (ECHA, 2008), namely the species sensitivity distribution method (SSD), can be used for the derivation of $QS_{fw, eco}$ and MAC-$QS_{fw, eco}$ on the basis of chronic and acute data, respectively. The principles of the methodology are described in detail in Section 6.4. Under the WFD, specific criteria apply with respect to data requirements, which partly differ from those under the PPP directives. An EQS should be protective for the wide range of surface waters and communities that can occur within Europe. Given the broad scope of protection of the WFD, the requirements of the REACH guidance with respect to the number of taxa and species to be included in the dataset (ECHA, 2008) are followed, i.e. the output from an SSD-based $QS_{fw, eco}$ or MAC-$QS_{fw, eco}$ is considered reliable if the database contains preferably more than 15, but at least 10 datapoints, from different species covering at least eight taxonomic groups. The following taxa would normally need to be represented:

- Fish (species frequently tested include salmonids, minnows, bluegill sunfish, channel catfish, etc.).
- A second family in the phylum Chordata (e.g. fish, amphibian, etc.).
- A crustacean (e.g. cladoceran, copepod, ostracod, isopod, amphipod, crayfish etc.).
- An insect (e.g. mayfly, dragonfly, damselfly, stonefly, caddisfly, mosquito, midge, etc.).
- A family in a phylum other than Arthropoda or Chordata (e.g. Rotifera, Annelida, Mollusca, etc.).
- A family in any order of insect or any phylum not already represented
- Algae.
- Higher plants.

From the data requirements for insecticides, herbicides and fungicides as presented above (Sections 8.2.1, 8.2.2 and 8.2.3), it appears that the minimum dossier dataset does not meet the criteria listed above, and additional data have to be submitted in order to construct SSDs. The acute dossier datasets for insecticides and herbicides are relatively large and require little additional testing. For fungicides that act as general biocides, it was already noted in Section 6.4.3.3 that additional data should be gathered in order to cover a broad range of taxonomic groups. The basic chronic dataset from the PPP dossier is small and considerable efforts will have to be made to meet the criteria for using SSDs.

So, on the basis of the standard dossier data for PPPs, the SSD-approach will primarily be applicable to derivation of the MAC-$QS_{fw, eco}$ for insecticides and in some cases for herbicides. Under the new PPP-regulation data from the open literature should be added to the dossier, this offers the possibility to extent the dataset without having to perform additional testing. However, experience with derivation of $QS_{fw, eco}$ for a series of PPPs learned that with a few exceptions, the amount of literature data is scarce (Bodar & Smit, 2008): only one chronic and two acute SSDs could be constructed on a total of 23 PPPs. For authorisation of new PPPs it is expected that the amount of data from the open literature is even lower.

In the EQS-guidance, the data requirements are presented in a strict way, but in specific situations there may be options to deal with the absence of data on a specific taxon. It is not possible to give a straightforward decision scheme on how to deal with different datasets. However, experiences gained with the previously mentioned series of 23 PPPs and other risk limit derivations can be used to illustrate the possibilities. In these cases, deviations from the guidance were accepted by the Dutch Scientific Advisory Group INS, which acts as a peer review group for derivation of risk limits in the Netherlands.

For carbendazim (Dang and Smit, 2008), the available chronic data covered the following taxonomic groups:
1. fish: *Cyprinus carpio* (family Cyprinidae)
2. a second family in the phylum Chordata: *Oncorhynchus mykiss* (family Salmonidae)
3. crustacea: *Gammarus pulex* and *Daphnia magna*
4. insects: *Chironomus riparius*
5. a family in another phylum than Arthropoda or Chordata: *Dugesia lugubris* (Turbellaria, phylum Platyhelminthes)
6. a family in any order of insect or any phylum not already represented: *Stylaria lacustris* (Clitellata, phylum Annelida), and *Bithynia tentaculata* and *Planorbis planorbis* (Gastropoda, phylum Mollusca)
7. algae: *Scenedesmus subspicatus*
8. macrophyta: no data

The dataset did not include macrophytes, but carbendazim was shown not to have a direct toxic effect on macrophytes in a mesocosm study. Therefore, the minimum requirements for performing an SSD were considered to be met. Demonstrating the absence of effects in several taxonomic groups is, however, not a permit to move on directly to an SSD for a specific taxonomic group (see 8.3.2 below).

For lambda-cyhalothrin (Van Leeuwen et al., 2008), the available acute data covered the following taxonomic groups:
1. Fish: *Ictalurus punctatus* (familiy Ictaluridae)
2. A second family in the phylum Chordata: *Oncorhynchus mykiss* (family Salmonidae)
3. Crustacea: *Daphnia magna* and eight other species
4. Insects: nine different species
5. A family in another phylum than Arthropoda or Chordata: **no data**
6. A family in any order of insect or any phylum not already represented: *Hydracarina* (Arachnida)
7. Algae: *Scenedesmus subspicatus*
8. Macrophyta: **no data**

The dataset neither included macrophytes nor a phylum 'other than arthropoda or chordate'. However, lambda-cyhalothrin was shown not to have a direct effect on macrophytes in mesocosm studies nor on molluscs (LOEC value of >8.9 µg/L for *Bithynia tentaculata*) in concentrations below its water solubility. Additionally, a large amount of data was available for the potentially most sensitive taxonomic groups crustacea and insects. Therefore, it was considered justified to perform an SSD.

Based on a similar approach, SSDs were constructed for azinphos-methyl and dimethoate by Moermond et al. (2008ab), although data on macrophytes were absent.

8.3.2 SSD for substances with a specific mode of action

The above listed requirements apply to all substances, including PPPs. The WFD-guidance offers the possibility to construct SSDs on the basis of selected taxonomic groups for substances with a specific mode-of-action. However, where under the PPP directives selection of the sensitive taxonomic groups can be done

beforehand, the WFD-guidance requires that first an SSD is constructed for the entire dataset (i.e. all taxa listed above) so that the relative sensitivities of taxa can be examined. In other words, the minimum requirements to perform an SSD should also be met for a compound with a specific mode-of-action, in order to be able to demonstrate deviations from the expected distribution. As indicated above, there are situations in which the construction of a generic SSD is acceptable, although certain taxa are absent and the criteria are thus not fully met. This will, however, always be a case-by-case decision which depends on the general picture that arises from the available data.

If on the basis of this generic SSD there is clear evidence of a 'break' in the distribution between the sensitive and other species, or poor model fit which can be attributed to specific action, the HC_5 should be estimated using only data from the most sensitive group, provided that the minimum number of 10 datapoints is present. If other evidence is available that indicates there might be a specific sensitive group of species, for example, 'read-across' data from a structurally similar substance, this could also be used.

8.3.3 Choice of the distribution

As already explained in Section 6.4.1, different parametric distributions may be used, but in the EQS guidance preference is given to logistic or log-normal distributions. The choice of a distribution function other than the log-normal or log-logistic distribution should be clearly explained. Whatever the model fitted to a distribution, results should be discussed with regards to the graphical representation of the species distribution and the different p-values (~probability value: the likelihood of wrongly rejecting a statistical hypothesis when it is true) obtained with each test. ($p < 0.05$ means a probability of <5%). If the data do not fit any distribution, the left tail of the distribution (the lowest effect concentrations) should be analysed more carefully. As explained above, if a subgroup of species is particularly sensitive and, if there are sufficient data, an SSD may be constructed using only this subgroup. However, this should be underpinned if possible by some mechanistic explanation e.g. high sensitivity of certain species to this particular chemical. The SSD method should not be used in cases where there is a poor data fit to all available distributions. This can be evaluated by means of goodness-of-fit testing (see also Section 6.4.1 for comments on the use of statistical tests).

8.3.4 Derivation of the $QS_{water, eco}$ using SSDs

For derivation of the $QS_{fw, eco}$ by means of statistical extrapolation, chronic NOEC or EC_{10}-values are used to construct an SSD. The median estimate of the HC_5 is used as the basis of the $QS_{fw, eco}$. According to the requirements set out above, an SSD can only be constructed when enough data are available, but there may still be some residual uncertainty that needs to be accounted for in the final $QS_{fw, eco}$. For this reason, the HC_5 is divided by an additional AF:

$$QS_{fw, eco} = HC_5 / AF$$

An AF of 5 is used by default, but may be reduced where evidence removes residual uncertainty. The exact value of the AF depends on an evaluation of the uncertainties around the derivation of the HC_5. As a minimum, the following five points have to be considered when determining the size of the assessment factor (ECHA, 2008):

1. The overall quality of the database and the endpoints covered, e.g., if all the data are generated from 'true' chronic studies (e.g., covering all sensitive life stages).
2. The diversity and representativity of the taxonomic groups covered by the database, and the extent to which differences in the life forms, feeding strategies and trophic levels of the organisms are represented.

3. Knowledge on presumed mode of action of the chemical (covering also long-term exposure). Details on justification could be referenced from structurally similar substances with established mode of action.
4. Statistical uncertainties around the HC_5 estimate, e.g., reflected in the goodness-of-fit or the size of confidence interval around the 5th percentile, and consideration of different levels of confidence (e.g. by a comparison between the median estimate of the HC_5 with the lower estimate (90% confidence interval) of the HC_5).
5. Comparisons between field and mesocosm studies, where available, and the HC_5 and mesocosm/field studies to evaluate the level of agreement between laboratory and field evidence.

Because all datasets are different, it is hard to present a general decision scheme in which the above listed items are systematically addressed. It is possible, however, to comment on each of the aspects and to give some examples of issues that may be considered.

1. First of all it should be noted that if the intrinsic quality of the data is not adequate, they should not be used in an SSD at all. The question whether a study is a 'true' chronic study generally does not apply to the standard chronic ecotoxicity tests in a dossier, because these tests are performed according to accepted (OECD) guidelines that take account of including sensitive life stages. It merely addresses additional studies with non-standard test species, or 'borderline' cases from the open literature, for example a chronic daphnid study started with adults, or a 10-days study with fish eggs. As a general rule it can be stated that when data are considered adequate for use in the chronic AF-method, they are equally relevant for use in the SSD. Vice-versa, if the quality of the data is questioned when considering the use of an SSD, this should also be taken into account when using the AF-method.
   Another case in which the overall quality was questioned, is a dataset of twelve NOECs in which ten values originated from two studies by (partly) the same authors, who applied the same concentration range for all species. The choice of this single concentration range thus highly determined the NOEC-values and the shape of the distribution. This can be seen as a deficit, and although strictly speaking the requirements to perform statistical extrapolation were met, this was seen as a reason not to lower the AF. This problem can be avoided if EC_{10}-values are available. Although the EC_{10} is more and more reported as an (additional) endpoint, the majority of studies, however, still focuses on the NOEC as chronic endpoint.
2. According to the EQS guidance, constructing an SSD is only allowed when data are available for a wide range of taxa and species with different characteristics. It can thus be argued that the aspects listed under this point are adequately covered. This means that this point is considered less relevant for determination of the AF.
3. Since the information on the mode-of-action is present in the dossier, for this item it should be considered whether the potentially most sensitive species groups are adequately represented in the dataset. The same considerations as discussed above for the AF-method apply. Since the additional data in a PPP dossier will be focussed on the sensitive groups, this will normally not be a problem. However, as explained above, it is possible that information from mesocosm studies points at another sensitive group that is not represented in the laboratory dataset. For fungicides, it should be considered whether or not the dataset used for the SSD adequately represents the potentially most sensitive taxa. It can also be the case that the potentially most sensitive group is represented, but by one datapoint only so that the variability within this group is not accounted for. If there is doubt on whether the sensitive taxa are adequately represented, this is a reason not to lower the AF.
4. With respect to the statistical uncertainties, it has already explained in the last part of Section 6.4.1 that the goodness-of-fit by itself should not be used as a strict criterion because for large datasets deviations from normality will be more easily detected. The confidence interval around the HC_5 is determined by the number of data and the spread in sensitivity. In view of the range of taxa that should be represented in the dataset, it is expected that confidence intervals will be relatively large. When the number of NOECs in the dataset is limited to the minimum of 10, the AF should not be lowered. If additional data are present for

potential sensitive taxa, the confidence interval is expected to decrease. This may be therefore a reason for lowering the AF.

5. If mesocosm studies are available (which for a chronic assessment will not often be the case), it should be checked whether or not at the level of the HC_5/AF effects are seen. This point relates to point 3.

The final choice of the AF depends on the picture that arises when critically reviewing the SSD taking account of the above considerations. Since this is a weight-of-evidence approach, which is very much dependent on the specific characteristics of the dataset, it is not possible to fit this into a single decision scheme.

The choice of 5 as the starting point for the AFs to be applied to an SSD is not further specified in the EQS-guidance and related guidance documents (TGD: EC, 2003; REACH: ECHA, 2008; EQS guidance: EC, 2011). This is different from the proposal for drainage ditches, where an AF of 1-3 is proposed, amongst others based on a comparison with mesocosm data or whether fish or non-vertebrates are the most sensitive species (see Sections 6.4.7 and 6.4.8). It is clear that there is a gap between the two frameworks and it is recognised in the EQS-guidance that PPPs may require specific methods. Chronic mesocosm studies are still scarce, but more data will become available that may allow for the 'validation' of the appropriate AF to apply for the extrapolation of the chronic HC_5.

8.3.5 Derivation of the MAC-$QS_{fw, eco}$ using SSDs

For deriving a MAC-$QS_{fw. eco}$ by statistical extrapolation, acute L(E)C_{50} data are the appropriate input data. Combined acute toxicity data sets for marine and freshwater species may be used, if, after evaluation of the freshwater and saltwater toxicity data, the data can be pooled. Similar to the chronic SSD, an AF is applied to the HC_5 to account for residual uncertainty. Where an AF of 5 is considered sufficient for the chronic $QS_{fw, eco}$, the default AF for the derivation of the MAC-QS is higher. The reason for this is that the MAC-$QS_{fw. eco}$ represents an acute no effect level, while the acute HC_5 refers to a 50% effect concentration for 5% of the species, because the input of the SSD are L(E)C_{50} values. Therefore, an AF of 10 is applied by default as recommended by EC (2011).

In the first part of this report, it is argued that a lower AF of 3 may be protective (see 6.4.7) based on open domain scientific publications that compared acute HC_5 values of herbicides, insecticides and fungicides with threshold levels of effects observed in aquatic micro-/mesocosm studies. On the other hand, there are also indications that the default AF of 10 is appropriate. For lambda-cyhalothrin, an acute SSD was constructed using data for all species (Van Leeuwen et al., 2008). The HC_5 was 4.7 ng/L, and with an AF of 10, the MAC-$QS_{fw. eco}$ was 0.47 ng/L. Acute EC_{10}-values were available for eleven arthropod species. The assumptions of a normal distribution were not fully met, but an SSD was constructed for reasons of comparison. The HC_5 based on these values was 0.65 ng/L. The HC_5 based on L(E)C_{50}-values with an AF of 10 and the HC_5 based on EC_{10}-values (without an AF) differed only by a factor of 1.4. Given the fact that probably an AF would still be needed when using the acute L(E)C_{10}-values, it can be concluded that an AF of 10 on the HC_5 derived from L(E)C_{50}-values is not over-protective.

For MAC-derivation of deltamethrin, De Knecht & Van Herwijnen (2008) refer to the SSD for arthropods as included in the DAR. The HC_5 based on these data was 5.7 ng/L, the default AF of 10 would lead to a value of 0.57 ng/L. Since the dataset included only three taxa, this value could not be used for MAC-derivation, but it is in line with the NOEC of 1 ng/L obtained in a mesocosm study. It should be noted that an AF of 3 was applied to this mesocosm NOEC to derive the MAC-$QS_{fw. eco}$.
The examples given above indicate that for derivation of the MAC-$QS_{fw. eco}$ the default AF 10 on the HC_5 based on EC_{50} values for the full range of species is adequate as a first approach.

8.3.6 **Proposal for a consistent set of assessment factors**

The EQS-guidance only gives default assessment factors for the standard approach, i.e. a generic SSD based on chronic data for the derivation of the $QS_{fw, eco}$ and a generic SSD based on acute $L(E)C_{50}$-values for derivation of the MAC-$QS_{fw. eco}$. It is recognised in the EQS-guidance that PPPs may require specific methods. According to the EQS-guidance, the default AF can be adapted if other lines of evidence suggest that a higher or lower AF is appropriate. Several other situations may be applicable for the $QS_{fw, eco}$ and/or MAC-$QS_{fw. eco}$:

- specific SSD based on chronic NOEC/EC_{10}-values for sensitive taxa
- generic SSD based on acute NOEC/$L(E)C_{10}$-values
- specific SSD based on acute $L(E)C_{50}$-values for sensitive taxa
- specific SSD based on acute NOEC/$L(E)C_{10}$-values for sensitive taxa

In Table 8-6, a proposal is made for assessment factors to be used in these situations, starting from the default factors as presented in the EQS-guidance. The defaults of the EQS-guidance are indicated in grey and are used as a starting point. An explanation is given below.

Table 8-6

Proposal for assessment factors for SSDs based on different types of datasets.

	$QS_{fw, eco}$	MAC-$QS_{fw. eco}$	
	input: chronic NOEC/EC_{10}	input: acute $L(E)C_{50}$	input: acute NOEC/$L(E)C_{10}$
generic SSD	**default 5** **range 5-1**	**default 10 (5 x 2)** **range 10-2**	default 5 range 5-1
specific SSD	default 3 range 3-1	default 6 (3 x 2) range 6-2	default 3 range 3-1

As indicated above, the default assessment factors of 5 and 10 for derivation of the $QS_{fw, eco}$ and MAC-$QS_{fw. eco}$ are considered adequate as a first approach when generic SSDs are used. The residual uncertainty that is the reason for applying an AF in the derivation of the $QS_{fw, eco}$ is also applicable to the MAC-$QS_{fw. eco}$. One could thus argue that the AF of 10 for the generic acute SSD is in fact built up from a factor of 5 to cover the extrapolation from a laboratory dataset to the field ecosystem, and a factor of 2 because the endpoints used in the SSD refer to a 50% effect level whereas the MAC-$QS_{fw. eco}$ represents no effect. This indicates that if enough acute $L(E)C_{10}$- or NOEC-values are available for a generic acute SSD, and thus the factor from 50% to 0% effect is not needed, the MAC-$QS_{fw. eco}$ may be derived starting with a default AF of 5. However, when using acute $L(E)C_{50}$-values for derivation of the MAC-$QS_{fw. eco}$, the AF should never be lower than 2. If enough chronic $L(E)C_{10}$- or NOEC-values are available for a specific SSD, a default AF of 3 is proposed. For the ditch, a default of 1 is proposed in this situation (see Table 6-3), but a higher factor is considered appropriate in view of the broader perspective of the WFD. Analogous to the 2-fold difference between the assessment factor of 5 for the generic $QS_{fw, eco}$ and 10 for the generic MAC-$QS_{fw. eco}$, an AF of 6 is proposed in case a specific MAC-$QS_{fw. eco}$ can be constructed based on acute $L(E)C_{50}$-values. Finally, when enough acute $L(E)C_{10}$- or NOEC-values are present for a specific SSD, a default AF of 3 is proposed.

It is recognized that there is a difference in the assessment factors used under the PPP-regulation and under the WFD. As said before, the perspective of the WFD is broader than PPPs and agricultural areas alone. The system of AFs according to the WFD is based on the methodology that has originally been developed for the risk assessment of industrial chemicals. It is a challenge for the near future to further integrate the experience

gained within the framework of PPP-authorisation into the WFD-methodology. The above cited information on the relationship between acute HC_5 values of herbicides, insecticides and fungicides and threshold levels of effects observed in aquatic micro-/mesocosm studies, can be used to further underpin the AFs to be applied.

8.4 Model Ecosystem Approach

8.4.1 Introduction

Technical guidance for deriving EQSs is published recently (EC, 2011). In this document, the use of micro- or mesocosm studies for deriving a $QS_{fw, eco}$ or a MAC-$QS_{fw, eco}$ for the freshwater compartment are discussed and the following is stated on QS derivation on basis of model ecosystem experiments:

> 'Field studies and simulated ecosystem studies such as microcosm and mesocosm experiments (e.g. ponds and streams) are frequently used to assess the environmental risks posed by pesticides. They can be a valuable tool to assess the impact of a chemical on populations or communities of aquatic ecosystems under more realistic environmental conditions than is achievable with standard single-species laboratory studies. If such studies are available, and they fulfil the criteria regarding reliability and relevance as defined below, they may be used either as the basis of $QS_{fw, eco}$ derivation or, when an SSD is used, to help select the size of AF applied to the HC_5.'

Concerning the exposure concentration to be used to derive the MAC-$QS_{fw, eco}$, the draft EU-guidance states that:

> 'For substances that do not dissipate quickly, the MAC-QSfw, eco values should be based on measured time weighted average (TWA) concentrations, and biological effects determined over a time span that is representative for most acute toxicity studies (i.e. 48-96 h). In the case of a non-persistent compound, measurement of exposure concentrations should take account of both spatial and temporal changes within the mesocosm. Furthermore it is important to determine which part of the exposure profile is most relevant. For example, if the peak concentration causes the effect, the actual initial concentration in the cosms is relevant, as well as the concentration at various time intervals (hours in the case of rapidly-dissipating compounds). An understanding of the exposure phase that is most relevant to any toxic effects (the Ecologically Relevant Concentration, ERC) is important because it: (a) influences how the assessor interprets the mesocosm data and (b) how the resulting EQS should be expressed (e.g. a 24 h or a 1 month peak). Such properties must be drawn to the attention of policy makers because it will affect how compliance is assessed, or indeed whether an EQS for compliance monitoring can be feasibly implemented at all. Such an EQS may still have value for planning purposes.'

With respect to the citation of the EU-guidance given above we like to state that also in short-term pulse exposure micro-/mesocosms experiments a prolonged observation period is necessary to study possible expose-response relationships. Consequently, the biological effects determined over a time span of several weeks may be necessary to gain insight in the effects (including latency) of short-term exposures, even of pulse exposures that do not last longer than 48-96 h. Furthermore, the ERC concept is applicable for both acute and chronic effect assessments, as well as for the exposure assessment to appropriately link exposure to effects (see Chapter 3). In addition, PPPs that show a fast dissipation from water due to sorption may be persistent in other compartments (e.g. sediments, vegetation) and exposure via these compartments may also contribute to the effects observed in micro/mesocosms.

Below, further guidance for the use of micro- and mesocosm studies for the risk assessment in larger surface waters is proposed, taking into account the context of the European WFD guidance (EC, 2011).

For a description of micro- and mesocosm studies see Section 6.6 of this report. The evaluation criteria for these studies are discussed in Section 6.6.4.

8.4.2 Assessment of model ecosystem studies

In the WFD guidance (EC, 2011) it is explicitly stated that only reliable studies should be used for EQS derivation. For the assessment of the reliability of micro-/mesocosm experiments is referred to De Jong et al. (2008). The following criteria should be addressed when assessing microcosm and mesocosm data:

- Is the test system adequate and does the test system represent a relevant freshwater community?
- Is the description of the experimental set-up adequate and unambiguous?
- Is the exposure regime adequately described?
- Are the investigated endpoints sensitive and in accordance with the working mechanisms of the PPP, and with the results of the lower tier studies?
- Is it possible to evaluate the observed effects statistically and ecologically?

8.4.3 Interpretation of micro-/mesocosm experiments

In Section 6.6.4 a classification system is proposed as tool for the interpretation of micro-/mesocosm experiments. Short- and long-term RAC derivation in the context of the PPP Regulation on basis of micro-/ mesocosm experiments usually makes use of Effect Classes that are based on either the threshold level for ecological effects or on ecological recovery after an initial effect period. Micro-/mesocosm studies with PPPs are primarily conducted within the context of the tiered approach in PPP-registration, i.e. to show the absence of unacceptable effects under more realistic conditions, or for validation (e.g. to check the validity of lower tier effect assessments). They generally are interpreted using the classification system mentioned in Section 6.6.4. However, ecological recovery is not considered when deriving EQSs. In the EQS guidance (EC, 2011) it is described that:

> 'The aquatic environment can be affected by chemical pollution both in the short- and long- term, and therefore both acute and chronic effects data should be used as the basis for establishing the EQS. In order to ensure that the aquatic environment and human health are adequately protected, EQS expressed as an annual average value should be established at a level providing protection against long-term exposure, and maximum allowable concentrations should be established to protect against short-term exposure.'

In practice of EQS derivation this means that the $QS_{fw, eco}$ is set at the NOEC/EC_{10} level of the most sensitive, relevant measurement endpoint, i.e. short term or long term exposure should not result in effects at the population-, community- and ecosystem-level. On the most sensitive, relevant measurement endpoint an appropriate assessment or extrapolation factor is applied to derive the $QS_{fw, eco}$.

For this reason, at least Effect Class 3 concentrations and higher are not relevant for EQS derivation, because an initial treatment-related effect on a relevant ecological endpoint is demonstrated. At Effect Class 1 concentrations no consistent and statistically significant treatment-related effects are found. This Effect Class is equal to the NOEC of the most sensitive measurement endpoint in the micro-/mesocosm experiment. About the use of Effect Class 2 concentrations in EQS derivation some dispute is possible. Treatment-related responses are classified as Effect Class 2 responses when:

'treatment-related effects are reported as 'slight', 'transient', or other similar descriptions. It concerns a short-term and/or quantitatively restricted response of one or a few sensitive endpoints, usually observed at individual samplings only.'

In the context of the European WFD guidance (EC, 2011), where only the term NOEC is used in relation to endpoints from micro-/mesocosms, Class 2 effects are not mentioned/considered for EQS derivation. However, application of a larger assessment or extrapolation factor to Effect Class 2 concentrations may ensure appropriate protection and a cost-effective use of micro-/mesocosm experiments.

8.4.4 Selecting the appropriate exposure regime in micro-/mesocosm experiments

In Section 6.6.2 is discussed that an appropriate exposure regime in the micro- or mesocosm experiments should be realistic worst-case and reflect the normal agricultural use of a PPP. However, as stated elsewhere in this report, in the case of risk/hazard assessment procedures for larger surface waters in line with 2000/60/EC, the $QS_{fw, eco}$ should protect against the occurrence of prolonged exposure, and the MAC-$QS_{fw, eco}$ should protect against possible effects from short-term concentration peaks.

The majority of available micro-/mesocosm studies with PPPs were designed to meet the demands of PPP registration. In general, the PPP under investigation is applied to these test systems at one or some repeated moments. The duration of the exposure and the dissipation are determined by height and frequency of application, cosm and substance properties and prevailing weather conditions. Semi-field experiments with PPPs in which water exposure concentrations are kept constant are an exception since in most cases the predicted exposure profiles for edge-of-field surface waters are characterised by time-variable exposures. Also in larger surface waters time-variable exposures of PPPs most likely are the rule rather than the exception, although pulse durations may be larger further downstream.

The fact that an EQS should be derived for both prolonged exposure and short-term concentration peaks implies that it usually will not be possible to derive safe concentrations for both exposure types from a single micro-/mesocosm study.

Below in Section 8.4.4.1 different types of exposure concentration in micro-/mesocosms are discussed. In Section 8.4.4.2 the selection of an appropriate exposure regime for MAC-$QS_{fw, eco}$ derivation is discussed followed by a similar discussion for $QS_{fw, eco}$ derivation in Section 8.4.4.3

8.4.4.1 Types of concentration

Dynamics in PPP concentrations in micro-/mesocosm studies should be measured (and/or calculated by appropriate tools) in order to obtain a reliable estimation of the exposure, both in terms of peak and time-weighted average concentrations (TWA) when relevant.

An important question at stake is how to assess the initial exposure concentration in micro-/mesocosms directly after application. In micro-/mesocosms it will (depending on the way of application and properties of the active substance) generally take some time before the substance applied will be equally dispersed through the water column. For substances that do not dissipate very fast from the water column, experimental data show that the substance will be homogeneously mixed in about 12-24 h, even when efforts are made to mix the substance in the water column immediately after PPP application. Note that the mixing of the test substance in the water column has to be a gentle process to avoid disturbance of the sediment compartment.

In a micro-/mesocosm study the initial peak concentration can be expressed in several types of concentrations:

- theoretical nominal concentration (calculated from amount of active substance (a.s.) in the application solution according to label information (not based on actual measurements of a.s.), volume of application solution applied and volume of the water compartment of the micro/mesocosm)
- checked nominal concentration (on basis of concentration measured in the dosing solution added to the mesocosm, the volume of dosing solution applied and volume of water compartment of the micro/ mesocosm)
- actual concentration (concentration measured in depth-integrated water samples from the micro/ mesocosm soon after application)

In the case of slow dissipating substances, it is expected (and found in practice) that actual, measured concentrations will be close to the checked nominal concentration. For slow dissipating substances it is proposed to use: (i) the actual, measured concentration (when more or less complete mixing can be assumed) when this concentration is not within 20% of the checked nominal concentration, or (ii) when the actual, measured concentration in the mesocosm is within 20% of the checked nominal concentration, the checked nominal concentration may be used.

In the case of fast dissipating PPPs, a problem occurs. Since it takes some time for the substance to mix through the system, measurements in the first hours can show highly variable concentrations, depending on application and sampling methodology, water circulation and dissipation processes of the substance. By the time the substance is homogeneously mixed, however, a considerable part may be dissipated from the water compartment e.g. by sorption to sediments, macrophytes and other types of organic matter. The question here is what concentration is the best measurement for the peak concentration. Measurements of water exposure concentrations shortly after application can be highly variable (in space and time) in the micro/ mesocosm. For example, exposure concentrations may be initially higher at boundary layers between different environmental compartments (e.g. water - air; water - sediment) or lower in dense macrophyte vegetation. This also indicates that the theoretical (calculated) initial concentration might not be fully representative for the initial exposure of organisms in the micro/mesocosm since, dependent on their habits and micro-habitats in which species occur, they may experience lower or higher initial exposure concentrations.

It is proposed to use the checked nominal concentration as representative for the initial peak concentration, and to use both the checked nominal concentration and measurements in the micro-/mesocosm test systems at different time intervals after application to assess the water dissipation rate (dissipation DT_{50} water). This procedure allows calculating mean exposure concentrations over relevant time-intervals (which may be required for proper EQS (particularly MAC-EQS) derivation and makes use of all measurements (in the stock solutions and in the micro/mesocosms). If for the different dosages the checked nominal concentrations do not fit into the time-exposure curve, the TWA can be based on the measured values only. Also note that the fraction that dissipated from the water column may still be present in the test system e.g. sorbed to organic matter, sediments and organisms, and consequently may contribute to the treatment-related effects in micro-/ mesocosm experiments.

8.4.4.2 Use of simulated micro-/mesocosm studies for deriving a MAC-$QS_{fw, eco}$

For determining the MAC-$QS_{fw, eco}$, experiments simulating short-term exposure are most relevant.

Studies with a pulse exposure, in which the PPP more or less rapidly disappears from the water column of the test system, can be used for MAC-$QS_{fw, eco}$ derivation. In studies with a constant exposure over a longer period, it cannot be determined whether the effects are caused by an initial short-term exposure, or by chronic exposure, and for that reason these studies may result in a conservative estimate of the MAC-$QS_{fw, eco}$.

In practice however, most micro-/mesocosm studies delivered for PPP-registration are characterised by single or repeated pulse exposures, in which the overall dissipation time and number of applications determine the exposure pattern.

For MAC-$QS_{fw, eco}$ setting on basis of micro-/mesocosm tests the treatment-related responses should be based on the relevant pulse exposure concentration (see below), and the treatment-related responses should be determined over a time span of at least several weeks after the start of the relevant pulse exposure to allow the detection of possible delayed effects.
Below a procedure for the type of concentration to be used for the MAC-QSfw, eco is proposed:

Test with one application
The MAC-QSfw, eco value derived from single application micro-/mesocosm studies should be expressed in terms of the time-weighted average (TWA) concentration in the water column for the period of the first 48 hours (in case invertebrates trigger the acute risks) or 72 hours (in case non-invertebrates trigger the acute risks) after PPP application, unless:

a) The substance in the water column is below LOD within 48/72 hours post application. In that case the time period for calculating the TWA exposure concentration to express the MAC-$QS_{fw, eco}$ can be lowered to the period in which the PPP can be reliably measured above the LOD, at least if it can be assumed that the dissipation rate in the test system is not substantially faster than under realistic WFD field conditions.

or

b) Experimental/scientific evidence is provided that the treatment-related effects are caused and/or can best be expressed in terms of exposure during a shorter period than 48/72 h (or the initial peak).

Test with repeated applications
In the case repeated exposure is tested in micro-/mesocosm experiments, several situations can occur:
(i) the substance has disappeared by the time of next application (no accumulation in the water compartment) or (ii) a certain amount of the substance is still available by the time of next application so that the second application results in accumulation in the water compartment and (iii) different dosages are applied, on purpose, e.g. simulating different exposure routes, or accidentally due to experimental variation in concentration or volume of dose solution applied (>20% difference from nominal).

It should be noted that for MAC-$QS_{fw, eco}$ derivation a NOEC/EC_{10} value (or an equivalent estimate) for short-term exposure and the most sensitive population or community endpoint needs to be derived from these experiments. This implies that in multiple application studies these endpoints will be a worst case for a situation with a single pulse exposure. For this reason it is proposed in the present report to use TWA exposure concentration based on the highest peak in the treatment without significant effects to express the overall micro-/mesocosm NOEC/EC_{10} and to apply a smaller Assessment Factor for spatio-temporal extrapolation to derive the MAC-$QS_{fw, eco}$. The argument is that when repeated pulses do not result in treatment-related effects, this reflects a worst case situation for a single pulse exposure.

8.4.4.3 Use of simulated micro-/mesocosm studies for deriving a $QS_{fw, eco}$

Below a number of possible exposure situations in micro-/mescosm studies are discussed.

Test with one application

For the derivation of the $QS_{fw, eco}$, micro-/mesocosm tests with a single application can only be used if the substance is relatively persistent. For $QS_{fw, eco}$ derivation, micro-/mesocosm studies with rapidly dissipating compounds (with half-lives of hours) cannot be used unless steps have been taken to replenish the test substance at intervals consistent with the substance's dissipation half-life in the water column.

For slow dissipating compounds, the TWA concentration for a relevant period (immediately post application) is recommended to use as exposure endpoint to express the $QS_{fw, eco}$. In the section concerning edge-of-field assessment it is discussed that the selection of the length of the TWA time-window should be based on ecotoxicological considerations (e.g. time-to-onset-of-effect information; length of the most sensitive life stage of the organisms at risk) and should be guided by the length of the relevant chronic toxicity tests that triggered the micro-/mesocosm experiment. For EQS derivation the chronic exposure should not result in effects on ecosystem structure and function in the exposed ecosystem. In the current report it is proposed to select the length of the chronic toxicity test that triggered the micro-/mesocosm study as time-window for the TWA exposure concentration used to express the $QS_{fw, eco}$.

Test with repeated applications

In the case of repeated treatments of the test system, three situations can occur:

1. The substance is disappeared by the time of next application, resulting in pronounced time-variable exposure, but no accumulation in the water compartment due to repeated applications.
2. The PPP shows a pronounced dissipation from water but a certain amount of the substance is still available by the time of next application so that accumulation in the water column occurs (particularly when applying the same dose at different treatments).
3. The PPP is applied frequently with the aim to obtain a more or less constant exposure concentration.

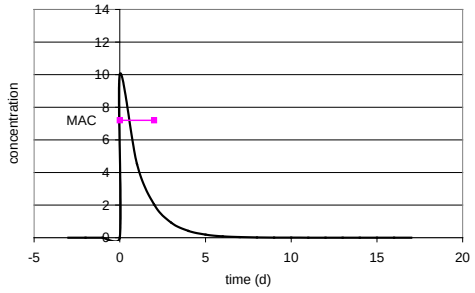
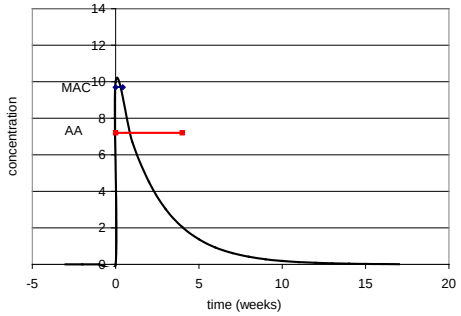
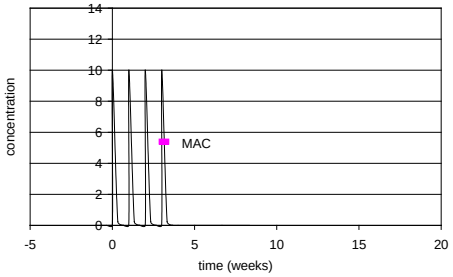
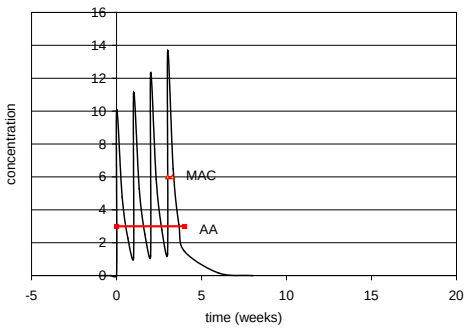
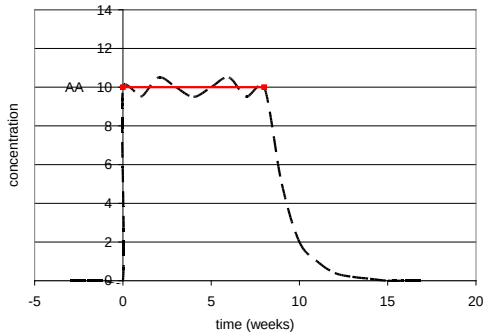
Situation 1) Since the PPP is not permanently present in the water column during the relevant exposure period, it is recommended in the present report not to use such a study for $QS_{fw, eco}$ derivation, unless additional evidence is provided that a TWA concentration approach can be used for intermittent time-variable exposure.

Situation 2) For the $QS_{fw, eco}$, it is proposed in the present report to use the TWA concentration as exposure endpoint if the exposure concentration during the relevant period never drops below 10% of the initial peak concentration, unless additional evidence is provided that a TWA concentration approach can be used for the time-variable exposure regime observed in the test system.

Situation 3) For the $QS_{fw, eco}$, it is proposed in the present report to use the TWA concentration for the relevant period as exposure endpoint.

The different exposure regimes and the concentrations to be used for $QS_{fw, eco}$ or MAC-$QS_{fw, eco}$ are summarised below.

Type of exposure

Exposure in semi-field study / Treatment without effect			Test exposure to link with overall NOEC/EC_{10} to derive $QS_{water, eco}$	Test exposure to link with overall NOEC/EC_{10} to derive MAC-$QS_{water, eco}$
Single pulse exposure	Fast dissipating substances		Not applicable	Initial 48/72 h TWA (or initial peak concentration)
	Slow dissipating substances		TWA (length time-window = duration of chronic toxicity test that triggered the risk)	Initial 48/72 h TWA worst case for peak exposure
Repeated peaks (same dose)	Disappeared before next application		Not applicable, unless scientific data are provided that the TWA approach can be used (length time-window = duration of chronic toxicity test that triggered the risk)	Initial 48/72 h TWA of the highest peak
	Build up of concentrations		TWA (length time-window ≡ duration of chronic toxicity test that triggered the risk)	Initial 48/72 h TWA of the highest peak
Continues (or static renewal) exposure	Chronic exposure concentration		TWA (length time-window = duration of chronic toxicity test that triggered the risk)	Not applicable

8.4.5 Selecting the appropriate measurement endpoints in micro-/mesocosm experiments

For considerations concerning the endpoint from microcosm and mesocosm experiments is referred to Section 6.6.3. For the aim of EQS derivation it is of importance to realize that both acute and chronic exposure should not result in significant effects (larger than Effect Classes 1 - 2) on ecosystem structure and functioning.

8.4.6 Application of an assessment factor to the threshold concentration

The EQS must be protective for all types of surface waters and communities, not just the type covered by a particular mesocosm or field study. We therefore need to assess whether the test system can be considered as representative for the full range of water bodies that might be subject to PPP exposure. Higher-tier (e.g. mesocosm) studies in the context of the PPP risk assessment are normally focused on shallow, mesotrophic, water bodies more or less representative for aquatic ecosystems (ponds, streams, ditches) occurring in agricultural landscapes. An EQS under the WFD, however, must also assure protection of other water bodies that may differ in ecosystem properties (different communities due to different climatic zones, or differences in trophic status). Preferably, the available (semi-)field data should cover this wide range of water types, but in reality this is not the case and therefore the guidance presented here should be considered when deciding on the choice of the AF for spatio-temporal extrapolation (see below).

More relevant NOEC/EC_{10}-values for sensitive populations are likely to arise when the species composition in a mesocosm is representative of that found in the field. This does not mean that the species composition in a micro- or mesocosm experiment should be exactly the same as that in the variable field; it is more important that a sufficient number of representatives of sensitive taxonomic groups are present, especially taxa that are expected to be sensitive given the substance's mode of action (e.g. insect larvae in a study with an insecticide that acts by disrupting moulting). Maltby et al. (2005) showed that taxonomy plays a more important role than habitat and geographical region in predicting the sensitivity of water organisms to PPPs with a specific toxic mode-of-action. Furthermore, the representativeness of the biological traits of the tested species is important.

Usually vertebrates are not incorporated in mesocosm studies. If laboratory data suggest vertebrates belong to the most sensitive group, little weight should be given to a mesocosm study without vertebrates. In general, the more similar the test system is to the potentially sensitive component of the field situation, the higher its relevance for risk assessment and EQS setting. Differences between experimental mesocosms and the field can result in either an over- or underestimation of the response of the field ecosystem. For example, avoidance and drift of organisms are reported in the field (Schulz and Liess, 1999), particularly for organisms that detect and avoid toxic substances by moving to areas with lower concentrations. Sessile organisms cannot avoid exposure. Although avoidance and drift of organisms are relevant endpoints, in general, laboratory and mesocosm studies do not accommodate avoidance reactions.

8.4.6.1 Application of an assessment factor to the threshold concentration from a micro/ mesocosm to derive a MAC-$QS_{fw, eco}$

For substances for which the toxic mode-of-action and/or the most sensitive taxonomic groups are unknown, an assessment factor ranging from 1-5 is proposed by the TGD to extrapolate the lowest threshold concentration from the available micro-/mesocosm study (EC, 2011). For most PPPs the toxic mode of action and the potentially sensitive taxonomic groups are fairly well described.

In determining the size of AF to be applied, the following should be considered:
- What is the overall quality of the micro- or mesocosm study/studies from which the NOEC/EC_{10} values (or equivalent estimates like Effect Classes 1 - 2) have been derived?
- What is the relationship between the mode-of-action of the investigated substance and the species represented in the available micro- or mesocosm studies? Are sensitive species represented?
- As stated above, one micro- or mesocosm study will never be able to be representative for all types of water bodies that should be protected under the Water Framework directive (e.g. lentic-lotic, trophic status, size of the waterbody etc.). For the purpose of $QS_{fw, eco}$ derivation the question at stake is not whether the systems are representative, but whether the endpoint from the micro- or mesocosm study is protective for other systems.

Brock et al. (2006, 2008) compared the outcome of six mesocosm studies that simulated short-term exposure, for each of the insecticides chlorpyrifos (single applications) and lambda-cyhalothrin (repeated applications), by considering Effect Class 1-2 concentrations (expressed in terms of peak concentrations). They looked at the spread (= ratio of the upper and lower limit of the 95% confidence interval) of the threshold concentrations for toxic effects. The spreads were 2.9 for chlorpyrifos and 2.6 for lambda cyhalothrin. Since the calculated spread between Effect Class 1-2 concentrations for micro-/mesocosm studies performed with lambda-cyhalothrin all concern multiple application studies, it should be realised that the studied exposures are worst case for a single pulse exposure on which the MAC-$QS_{fw, eco}$ can be based. In addition, the Effect Class 1 concentrations between single application studies performed with chlorpyrifos and the Effect Class 2 concentrations between repeated application studies performed with lambda-cyhalothrin were remarkably similar (see Appendix 1).

It should be noted that the AF is aimed at quantifying the differences between artificial aquatic ecosystems. Although it is argued before that these systems might be protective (or worst case) for the field situation, differences between these systems and the field situation are present (e.g. presence of vertebrates).

8.4.6.2 Application of an assessment factor to the threshold concentration from a mesocosm to derive a $QS_{fw, eco}$

According to the REACH guidance (ECHA 2008), the AF applied to mesocosm studies or (semi-) field data will need to be reviewed on a case-by-case basis, but no guidance is given with respect to the range of AFs to be applied. In the draft WFD guidance (EC, 2011) it is suggested that:

> 'where there is (a) only a single model ecosystem study, and (b) sensitive taxa are included in the study of a compound with a specific mode of action, an assessment factor of 5 would account for variation in the NOECs'.

In the WFD guidance, however, underlying scientific data to support the height of this AF are hardly provided.

It should be noted that for PPP registration both an Effect Class 1 and an Effect Class 2 concentration may be used for chronic RAC derivation that is representative for the threshold level of effects by applying a small AF (1 to 2 when based on Effect Class 1 concentrations; 2 to 3 when based on Effect Class 2 concentrations) if only one adequate micro-/mesocosm study is available (see Section 6.6.6). For $QS_{fw, eco}$ derivation a similar procedure is proposed in the present report but to apply a higher AF for spatio-temporal extrapolation (see section below). In the context of interpreting Effect Class 2 responses it is worthwhile mentioning that when more measurement endpoints are assessed on several sampling days (which usually is the case in micro-/mesocosm experiments) that the chance of occurrence of Type II statistical errors may increase (demonstrating a statistical difference when there is not a treatment-related effect). For this reason the

evaluator of a mesocosm study could decide that a single Effect Class 2 response could be seen as the NOEC of the study.

Brock et al. (2008) compared micro-/mesocosm experiments for several chemicals in which long-term exposures were simulated. Based on studies with the fungicide carbendazim and the herbicide atrazine, they estimated a geographical extrapolation factor based on the ratio of the upper and lower limit of the 95% confidence interval of NOECs (Effect Classes 1 - 2) for toxic effects. These factors ranged between 1.4 and 5.4. It should be noted that the relatively large variability in overall NOECs between the atrazin studies partly is explained by the indoor microcosm experiments in the data set. In indoor microcosms the quantity of irradiance provided usually is substantially lower than under field conditions and the inhibitory effects of photosynthesis inhibitors may be partly compensated by higher availability of light (Brock et al., 2006; 2008).

8.4.6.3 Proposal for AFs to derive QSs on basis of threshold levels for effects in micro/mesocosms

Based on the information presented in the previous sections, the proposed assessment factors for spatio-temporal extrapolation of a single micro-/mesocosm study for QS derivation is presented below.

	$QS_{fw, eco}$	MAC-$QS_{fw, eco}$
NOEC most sensitive structural endpoint (Effect class 1)	2-4*	1-2* (repeated application studies) 2-3* (single application studies)
Effect class 2 (most sensitive structural endpoint)	4-5*	2-3* (repeated application study) 3-4* (single application study)

* The height of the AF is based on expert judgement considering all available lower and higher-tier information. If several adequate micro-/mesocosm studies are available the AF is applied to the highest Effect class 1or 2 value or a lower AF than reported in the table may be applied.

Note that the proposed AF's concern the situation in which the mode of action of the substance is known, and the sensitive organisms are present in the mesocosm and that the Effect Classes for MAC-$QS_{fw, eco}$ derivation are based on 48-72 h TWA concentrations (while acute RACs are based on Effect Classes expressed in terms of peak concentration). The AFs proposed are based on comparison of micro-/mesocosm studies for a limited number of compounds (see also Appendix 1). Furthermore, potential risks for vertebrates like fish usually are not addressed in model ecosystem experiments. Consequently, if fish are amongst the most sensitive taxa, the risks to fish have to be assessed additionally (e.g. by applying the SSD approach). In addition, only a few micro-/mesocosm studies with PPPs are available that studied the effects of a more or less constant exposure regime (see Appendix 1).

If more than one appropriate micro-/mesocosm test is available for the same PPP the AFs mentioned in the table above might be applied to the highest Effect Class 1 or Effect Class 2 concentration observed, or it may be decided to lower the AF when it is decided to use the lowest Effect Class 1-2 concentration.

8.5 Quality Standards based on biota

8.5.1 Secondary poisoning

The standard approach from the TGD (EC, 2003; ECHA, 2008) uses the concentration in the diet of the toxicity test as the basis for the quality standard in biota. Mammalian or avian toxicity data may be expressed as NOECs relating to concentration in food ($NOEC_{oral}$, expressed in units of mg/kg_{food}) or as no observed adverse effect levels relating to dose ($NOAEL_{oral}$, expressed in units of $mg/kg_{bw}/d$). For the standard derivation of QSs for secondary poisoning, the results need to be expressed in terms of the concentration in food because this is the basis of the adopted risk model. The general rule for the conversion is that the concentration in food is equal to the daily dose multiplied by the body weight (bw) divided by the daily food intake (DFI), or

$$NOEC_{oral} = NOAEL_{oral} \frac{bw}{DFI}$$

where:

- $NOEC_{oral}$ = no observed effect concentration (mg/kg_{food})
- $NOAEL_{oral}$ = no observed adverse effect level [$mg/kg_{bw}/d$]
- DFI = daily food intake (g food/d) and
- bw = body weight (g)

Table 8-7, which is copied from the REACH- and EQS-guidance, presents a guide with a standard set of conversion factors that can be used to promote internal consistency when converting concentrations from dose into diet for mammals[3]. The guide should be used only in the absence of more specific data from the study itself or other sources. For example, a rabbit *(Oryctolagus cuniculus)* typically consumes around one third of its body weight per day, and so the conversion factor in this case would be 33.3 $kg_{bw} \cdot d^{-1} kg^{-1}$ food. It should be noted that the conversion factors for young birds and mammals might differ from those for adults. For avian reproduction studies, a default factor of 10 can be used as a conversion factor (i.e. bw/DFI = 10) (see Appendix 6 of EFSA, 2008). For this conversion to be valid, no food avoidance should have occurred in the study. According to the EQS-guidance, recommendations from EFSA (2008) should be considered as indicative and the REACH guidance (ECHA, 2008) should be followed rather than EFSA (2008). For PPPs, however, it is considered justified to rely on the information of EFSA (2008) which is based on an extensive analysis of data on energy demand of birds and mammals and caloric content of different food types. NOECs derived from NOAELs in this way are assumed to be equivalent to directly measured NOECs.

[3] Please note that the original table the chicken *Gallus domesticus* is included too; however, this species is nowadays hardly tested anymore.

Table 8-7

Conversion factors for converting NOAELs (dose) from mammalian toxicity studies into NOECs (concentration) (copied from ECHA, 2008).

Species	Age/study	Conversion Factor (bw/DFI) (ECHA, 2008; EC, 2003)	Conversation Factor (bw/DFI) (EFSA, 2008)
Rat *(Rattus norvegicus)*	>6 weeks	20	
Rat	<6 weeks	10	
Rat	28 and 90days		10
Rat	Two generation study first mating[a]		12.5
Rat	Two generation study overall (females)[a]		8.33
Mouse *(Mus musculus)*	28 and 90days	8.3	5.0
Vole (*Microtus* spp)		8.3	
Rabbit *(Oryctolagus cuniculus)*		33.3	
Dog *(Canis domesticus)*	Adult/all	40	40
Monkey (*Macaca* spp)		20	

[a] The first mating value for a two-generation study should be used for assessment when effects (general or on reproduction) are seen to relate to the pre-mating phase of the first mating, or effects are seen only in male F0 parents at any time. For all other aspects of a two-generation study, the overall conversion figure should be used

The quality standard that describes the threshold concentration of a substance in the food of a predator, $QS_{biota,\ secpois,\ fw}$ ($\approx PNEC_{oral}$, in mg/kg_{food}), is derived by applying appropriate assessment factors (AF_{oral}; see Table 8-8) to the selected NOEC oral for each species. Data from two different toxicological studies should only be merged if they have been conducted according to a similar guideline, use the same species and test conditions and report the same key endpoints. There may be more than one chronic study for the same species. Under these circumstances, preference is given to the study with the longest duration. Usually, taking account of the appropriate assessment factor, this will yield the most critical endpoint. It may be that a test with a shorter exposure duration reports a lower endpoint than the test with longest exposure duration. In such a case, the assessment factor corresponding to the longest exposure time may be applied to the most sensitive endpoint. For example, if a 90-days test gives a $NOEC_{oral}$ of 0.1 mg/kg food, and a multi-generation study gives a $NOEC_{oral}$ of 0.2 mg/kg food, it may be considered to use 0.1 mg/kg food with an AF of 30, instead of 90.

Table 8-8

Assessment factors for the extrapolation of mammalian and bird toxicity data into $QS_{biota,\ secpois,\ fw}$ (EC, 2003).

TOX_{oral}	Duration of test	AF_{oral}
$NOEC_{oral,\ birds}$	chronic	30
$NOEC_{oral,mammals}$	28 days	300
	90 days[a]	90
	chronic	30

[a] For consideration of reproduction studies.

$$QS_{biota,\,sec\,pois,\,fw} = \frac{TOX_{oral}}{AF_{oral}}$$

The final value for the $QS_{biota,\ secpois,\ fw}$ is selected by comparison of the different values for the tested species and choosing the lowest resulting values (EC, 2003; Lepper, 2005). If sufficient data are available, there is no reason why a probabilistic approach to extrapolation (i.e. an SSD approach) should not be used. However it should be noted that in the applied assessment factor the factor of 10 to extrapolate from the lowest chronic NOEC values to the $QS_{biota,\ secpois,\ fw}$ is already included and that when applying a statistical extrapolation, the NOECs need only to be converted from subacute (28d; factor 10) and subchronic (90d; factor 3) to chronic and from laboratory diet to fish or mussels (all data; factor 3). For the application of a species sensitivity distribution (SSD), data should be available for a minimum of 10 species The dataset should include both birds and mammals and should also include wildlife-relevant predatory species of both birds and mammals. It should be noted that these requirement are very conservative and in practice will never be met. Relevant predatory wild-life species (e.g. otters, cormorants, herons) are never tested under standard conditions. It is also noted that the requirements for fish are less stringent (minimum of five species). The width of the sensitivity distribution of birds and mammals is even smaller than for fish. Therefore, the same or even smaller sample size would be sufficient for birds as well as for mammals.

8.5.1.1 Refined approach using key species

The EQS-guidance offers a refined approach, based on EFSA (2008). In this approach, which is described in Section 5.3.3, the dose rather than the diet concentration, is used as a starting point. This helps to minimise bias relating to different food intake rates between laboratory and field situations. Information on body weight, dietary composition and feeding rate by predators are used to select those species most likely to experience the highest exposures to contaminants through the aquatic food web. A group of key species should represent all the organisms at risk from secondary poisoning. By definition, if these are protected (and the assumptions are correct) other species will also be protected. For the edge of field ditch, the otter and great crested grebe ('fuut') or kingfisher ('ijsvogel') could be selected. These species are also relevant for larger surface water, and could be used for EQS-derivation as well. Therefore, an approach similar to that described in Section 5.3.3 can be used to derive the $QS_{biota,\ secpois,\ fw}$.

The key species is defined as the most susceptible species on the basis of its ratio of body and daily food intake and its position in the trophic chain (the latter only of the substance is subject to significant biomagnification). The NOEC for the key indicator wildlife species can then be calculated from the lowest reliable NOAEL from laboratory studies using information on body weight (bw) and daily food intake (DFI) for these species as indicated below:

$NOEC_{wildlife} = NOAEL_{laboratory} * (bw_{wildlife}/DFI_{wildlife})$

Only the mammals NOAEL is used to extrapolate to mammalian wildlife species. Similarly, only the avian NOAEL is used to extrapolate to avian wildlife species.
The DFI for a key indicator wildlife species can be calculated with the information provided in Appendix G of the guidance document for birds and mammals of the EFSA (EFSA, 2008). Then the $QS_{biota,\ secpois,\ fw}$ is derived from the $NOEC_{wildlife}$ in this case using the assessment factors from Table 8-9. In this table the extra factor of three for the difference in caloric content between laboratory food and a diet based on fish and/or mussels is omitted.

Table 8-9

Assessment factors for the extrapolation of mammalian and bird toxicity data into $QS_{biota,\ secpois,\ fw}$ in a refined assessment (based on Table 8-7[a]).

TOX_{oral}	Duration of test	AF_{oral}
$NOEC_{oral,\ birds}$	chronic	10
$NOEC_{oral,mammals}$	28 days	100
	90 days[a]	30
	chronic	10

[a] The AF of 3 accounting for extrapolation from laboratory to field is omitted because the method already takes the dietrary intake differences between laboratory and field into account

The resulting AF should allow for interspecies variation in sensitivity to account for differences in toxicity. A factor of 10 accounting for interspecies variation is appropriate for this purpose. An additional AF of 3 to 10 is applied when exposure periods are not truly chronic (i.e. subchronic to chronic extrapolation).
The same considerations as in the standard approach may be applied with regard to the use of acute avian data and data treatment for the same species. For application of the SSD method the same considerations as in the standard approach are valid with the exception that in this case the input data should be based on dose and not diet concentrations.

8.5.2 $QS_{biota,\ hh\ food}$ based on human exposure via fish

For compounds that meet certain triggers concerning classification with respect to human toxicology and bioaccumulation, a quality standard should be derived that addresses the potential risks for humans from consumption of fishery products (i.e. fish or shell fish). For humans, the derivation of a biota standard is triggered solely on the basis of the hazardous properties of the chemical of interest. The available mammalian and bird toxicity data is used to give an indication of possible risks to top wildlife predators as well as humans since there is usually standard mammalian toxicity data available for well-studied chemicals. Effects on reproduction, fertility and development are of particular concern since these are long-term effects which could impact on populations of organisms. Specific triggers are as follows:

- a known or suspected carcinogen (Cat. I-III, R-phrases R45 or R40) or
- a known or suspected mutagen (Cat. I-III, R-phrases R46 or R40) or
- a substance known or suspected to affect reproduction (Cat. I-III, R-phrases R60, R61, R62, R63 or R64) or
- possible risk of irreversible effects (R68) or
- the potential to bioaccumulate (see protection of top predators) plus danger of serious damage to health by prolonged exposure (R48) or harmful/toxic/fatal when swallowed (R22/R25/R28).

The H-statements will soon replace the R-phrases in EU chemicals legislation via the Classification, Labelling and Packaging Regulation (1272/2008/EC). The conversion between H and R phrases is provided below:

- R22 H302: Harmful if swallowed
- R25 H301: Toxic if swallowed
- R28 H300: Fatal if swallowed
- R40 H351: Suspected of causing cancer
- R45 H350: May cause cancer
- R46 H340: May cause genetic effects
- R48 H373: May cause damage to organs through prolonged or repeated exposure
- R60 H360: May damage fertility or the unborn child

– R61 H360: May damage fertility or the unborn child
– R62 H361: Suspected of damaging fertility or the unborn child
– R63 H361: Suspected of damaging fertility or the unborn child
– R64 H362: May cause harm to breast-fed children
– R68 H341: Suspected of causing genetic effects

When the derivation of a QS based on human fish consumption is triggered, a maximum permissible concentration in (shell) fish ($QS_{biota,\ hh\ food}$) is calculated using the method of Lepper (2005). It assumes that the uptake of a substance from fishery products does not exceed 10% of the relevant threshold level (TL), estimated from experimental data and expressed in µg/kg_{bw}/d for humans. For practical purposes, the acceptable daily intake (ADI), tolerable daily intake (TDI) or NO(A)EL_{oral} (the latter divided by an assessment factor) provides such an estimate. The $QS_{biota,\ hh\ food}$ (expressed as µg/kg) is calculated using defaults for human bw (70 kg) and for the consumption of fishery products (0.115 kg/d) as follows:

$$QS_{biota,hhfood} = \frac{0.1 \times TL \times 70}{0.115}$$

This approach does not specifically consider possible sensitive groups, such as the developing foetus or subpopulations that consume more fishery products than the European average. However, the assumption that fishery products make up no more than 10% of the threshold level value (0.1·TL) at the European average level of compound uptake provides a margin of safety.

8.5.3 Conversion of QS_{biota} to QS for water

The assessment of secondary poisoning (secpois in the subscript below), and human fish consumption (hh food in the subscript below) both lead to QS in biota. These biota standards have to be converted into $QS_{water,\ hh\ food}$ and a $QS_{fw,\ secpois}$. Therefore, experimental BCF and BMF data, or a field derived BAF, are required. In general the water concentration value is calculated from the biota concentration as follows:

$$QS_{fw} = \frac{QS_{biota}}{BAF}$$

The term bioaccumulation refers to transfer mechanisms of hydrophobic contaminants by both bio-concentration (accumulation via media) and biomagnifications (accumulation via food). Normally, the combined effects of each step are combined in a multiplicative approach. Therefore, the BAF may be calculated as:

$$BAF = BCF \cdot \prod_{i=1}^{n} BMF_i$$

where the number of BMFs depends on the trophic level or position of the organism in the food web. According to REACH Guidance (ECHA, 2008), for freshwater a simple food web is assumed that consists of water -BCF→ fish/mussel -BMF_1→ fish-eating predator.

$$QS\,(\mu g/L) = \frac{QS_{biota}\,(\mu g/kg)}{BCF(L/kg) \times BMF_1}$$

Ideally, field BAF values for the correct trophic level should be used and BMFs should be based on measured data. In general, the most reliable data on biomagnification originate from trophic magnification studies.

In such studies, the levels of contaminants in several species in an ecosystem are measured and expressed as a function of the trophic level. The trophic level is mostly derived from stable nitrogen isotope ratios and a regression is made between contaminant concentration and trophic level. The contaminant values should preferably be normalised to the fraction in the organisms that contains the substance, e.g. lipids.

The availability of reliable field bioaccumulation and biomagnification data is, however, limited. Therefore, the default BMF values given in Table 8-10 (EC, 2003) may be necessary. A reliable experimental BCF value is always preferable to the log K_{ow} to estimate the BMF value because it takes the metabolism of the substance into account, which is an important parameter in food web accumulation.

Table 8-10

Default BMF values for organic substances (table adapted from EQS-guidance).

BCF (fish)	log K_{ow} of substance	BMF_1
<2000		1
2000-5000		2
>5000		10
	<4.5	*1*
	4.5-<5	*2*
	5-8	*10*
	>8-9	*3*
	>9	*1*

From this table it can be seen that biomagnification is relevant for compounds with a BCF ≥2000 L/kg. For these compounds, the appropriate BMF will be selected from in Table 8-10. Note that for PPPs with log K_{ow} ≥3 an experimental BCF will always be available, so the selection of BMF based on log K_{ow} is not relevant. Furthermore, for PPPs, a log K_{ow} >8-9 is not relevant, so the default BMF will be either 1, 2 or 10. Generally, substances with a BCF of 500 L/kg or less can be converted into an equivalent water concentration with reasonable confidence. Where it is necessary to convert a biota QS into an equivalent water-column QS, the uncertainties involved in making the extrapolation may be taken into account by performing the conversion for extreme BAF values as well as using the typical BAF value. If the QS for water lies within the range of possible extrapolated values of the QS for biota, when considering the uncertainties of the extrapolation, it is not possible to determine with high confidence which is the 'critical' QS. Bioconcentration data will most often refer to fish, but if relevant data on e.g. mussels are available, these should be considered as well.

When this route is critical for the final AA-EQS, i.e. is lower than the routes secondary poisoning and direct ecotoxicity (see Section 8.5.1), there is no further option for refinement because the human toxicological threshold is a fixed value, as are the defaults for bodyweight, and fish consumption. As already indicated in Section 7.1, it is not likely that this is the case.

8.6 Selection of the appropriate $QS_{fw, eco}$, final AA-EQS and MAC-EQS

In the three methods described above (AF-, SSD- and mesocosm approach), remaining uncertainty is taken into account by applying an assessment factor. The derived $QS_{fw, eco}$, will, however, differ between the methods. According to the EQS-guidance, the $QS_{fw, eco}$ and MAC-$QS_{fw, eco}$ should preferably be based on the results from the SSD-method or the model ecosystem-studies if all methods can be performed. The reason for this is that the latter include a more scientific approach towards ecosystem effects. The final choice between the SSD-

or mesocosm-based $QS_{fw, eco}$ and MAC-$QS_{fw, eco}$ should be based on expert judgement. It is further stated that possible remaining uncertainties involved with the SSD and model ecosystem studies used to derive the $QS_{fw, eco}$ need to be tabulated to allow a transparent decision which method should prevail. An explanation of possible discrepancies in the results and the reason for choosing the final $QS_{fw, eco}$ and MAC-EQS should be provided. Once the $QS_{fw, eco}$ is selected, it should be compared with the $QS_{fw, secpois}$ and $QS_{water, hh food}$. The lowest of these values is selected as the AA-EQS. It is expected that direct ecotoxicity, and thus the $QS_{fw, eco}$, will most often determine the final AA-EQS because of the specific function and design of PPPs.

As already argued in Section 8.3.1, it is expected that the construction of SSDs will only be possible in a limited number of cases. This implies that mesocosm studies will be the main option next to the AF-method, at least for the MAC-EQS. In 2008, RIVM derived ERLs for 23 pesticides, based on data present submitted within the context of pesticide authorisation under Directive 91/414/EEC. For nine compounds, reliable mesocosm data were available that could be considered for the MAC-EQS, while for two compounds SSD could be applied. SSD was also applied for the MAC-derivation for 2,4,6-trichlorophenol (Bodar and Smit, 2008; Smit, 2009). Since only very few chronic mesocosm studies are available, the experience with derivation of the $QS_{fw, eco}$, for PPPs on the basis of such studies is limited. Kresoxim-methyl and carbendazim are two examples in which the $QS_{fw, eco}$ has been based on mesocosm data (Smit and Dang, 2008; Van Leeuwen and Vonk, 2008).

An overview of the outcome of the MAC-EQS derivations presented by Smit (2009) shows that if mesocosm data were available or SSDs could be constructed, the latter were used, except in one case, where the AF-method and mesocosm resulted in virtually the same value (Table 8-11).

Table 8-11

Summary of MAC-QSfw, eco derived by different methods. Bold values represent the final MAC-EQS values as reported in the original reference; the AF that was applied to the critical endpoints is presented as well.

Compound	**MAC-$QS_{fw, eco}$ [µg/L] derived by**						**Reference**
	AF-approach	AF	**SSD-approach**	AF	**mesocosm approach**	AF	
abamectin	**0.018**	10			0.016	1	Scheepmaker, 2008a
deltamethrin	3.1 x 10^{-5}	10	5.7 x 10^{-4}[b]	10	**3.0 x 10^{-4}**	3	De Knecht and Van Herwijnen, 2008
dodine	0.0069[a]	1000			**2**	3	Smit and Van der Veen, 2008
esfenvalerate	**0.00085**[a]	100			NOEC <0.01		Van Vlaardingen et al., 2008
fenoxycarb	5.2[a]	100			**0.026**	1	Smit and Vonk, 2008
imidacloprid	0.1	10			**0.2**	3	Posthuma-Doodeman, 2008
lambda-cyhalothrin	0.00023		**0.00047**[c] 0.00065[d]	10 1	NOEC <0.002		Van Leeuwen et al., 2008
teflubenzuron	0.05	10			**0.0017**	3	Scheepmaker, 2008b
pyriproxyfen	**0.026**	10			NOEC = 5 (no insects)		Moermond, 2008
2,4,6-trichlorophenol	3.6[a]	100	**32**	10			Moermond et al., 2009
azinphosmethyl	0.014	10	HC5 0.06[e]				Moermond et al., 2008b

[a] For these compounds, an additional AF of 10 was applied to account for bioaccumulation; this factor will no longer be used in the new EQS-guidance; resulting in a 10-times higher MAC according to the AF-method.

[b] Based on HC_5 of $L(E)C_{50}$ of arthropods with AF 10; requirements for SSD not met (see 8.3.5).

[c] Based on HC_5 of $L(E)C_{50}$ with AF 10.

[d] Based on HC_5 of $L(E)C_{10}$ of arthropods; no AF; requirements for SSD not fully met.

[e] HC_5 of $L(E)C_{50}$ for crustaceans and insects; fit not good, AF 4-5 proposed.

9 Scientific developments and research needs to consider when updating guidance documents

9.1 Aspects specifically related to the PPP regulation

9.1.1 Extrapolation of effect assessment scheme to other ecosystems.

The effect assessment decision schemes proposed in this report have been developed for Dutch drainage ditches on the one hand and for larger WFD water bodies in the Netherlands on the other hand. In Dutch agricultural landscapes edge-of-field surface waters are mainly represented by ditches. Note that streams in the Netherlands are all assigned to WFD water bodies in contrast to drainage ditches. The proposed effect assessment schemes for Dutch drainage ditches may however also be representative for edge-of-field ponds and streams. The principles of the effect assessment scheme are applicable to other ecosystems, assuming that sensitive and vulnerable taxonomic groups are represented in test systems used to derive the RAC because the ecological threshold level is expected to lie in the same range for these ecosystems. The variability in sensitivity/vulnerability of water organisms between different types of surface waters likely is covered by the application of the assessment factors proposed. An important difference between different types of edge-of-field surface waters however is the exposure regime. Streams, and to a lesser extent ditches, are hydrologically open systems characterized by inflow of water and organisms thereby influencing ecological effects and recovery. Ponds are hydrologically isolated from other surface waters. The exposure regimes and recovery potential in ditches, ponds and streams may therefore substantially differ between different types of edge-of-field surface waters and this will determine the outcome of the risk assessments. Ecological scenarios for ditches, streams and ponds may be helpful to interpret the ecological relevance of higher-tier tests, particularly micro- or mesocosm tests and effect models.

9.1.2 Specific protection goals

The effects assessment procedures described in this report for Dutch drainage ditches aim to protect vascular plants, algae and invertebrates at the population-level and fish and other aquatic vertebrates at the individual-level. This is in accordance with a recent EFSA opinion (EFSA, 2010) in which the following groups of aquatic organisms are identified as most important key drivers for specific protection goals: microbes (bacteria and fungi), algae, vascular plants, invertebrates and aquatic vertebrates (including fish and amphibians). The related specific protection goals usually concern the maintenance of biodiversity in the (agricultural) landscape/watershed by allowing temporary effects on local field or edge-of-field populations only. For most aquatic plants and invertebrates the ecological entities to be protected are (meta)populations. However, according to the EFSA opinion this may be individuals when it concerns aquatic vertebrates, and functional groups when it concerns aquatic microbes in edge-of-field surface waters.

Although in the EFSA opinion on protection goals (EFSA, 2010) microbes were identified as important key drivers, the tiered risk assessment procedures and decision schemes presented in this report do not specifically address aquatic bacteria and fungi. Although we assume that microbial processes will be protected when applying the decision schemes developed for aquatic flora and fauna, this needs to be verified. Particularly research on the effects of fungicides on aquatic fungi may warrant further attention.

9.1.3 Ecological modelling

Individual-level effects of pesticides may depend on factors such as toxicokinetics and toxicodynamics, exposure history and adaptation, the developmental stage of the organism and avoidance behaviour. On the population-level, effects of pesticides not only depend on exposure and toxicity profiles, but also on factors such as biological traits (e.g. life history characteristics), demographic structure of the populations of concern, food web interactions, ecological infrastructure (e.g. connectivity of waterways), spatio-temporal aspects of multi-stress and the presence of refuges in space and time. Since it is practically not feasible to perform experiments that address all these factors, computer simulation models may be the appropriate tools to integrate the results of focussed ecotoxicological experiments.

As discussed already in Section 6.7.2 promising individual-level models in the future risk assessment to extrapolate time-variable exposure regimes of pesticides comprise toxicokinetic/toxicodynamic (TK/TD) models. These models, however, have been developed for a limited number of aquatic species only so that it primarily concerns a research activity up till now. Whether the parameters and model concepts derived with TK/TD models for these focal species can be easily extrapolated to other aquatic species is an important research activity (Rubach, 2010). So far TK/TD models do not consider distribution and metabolism of the toxicant within the organism. Thus, the description of the TK is usually restricted to the process of uptake and elimination only, and the models differ mainly in their assumptions on the TD. The TD concepts differ in the range of toxic mechanisms for which they are valid. Consequently, another important research activity is to further develop TK/TD models for pesticides that differ in toxic mode-of-action (Hommen et al., 2010b).

To date, a broad range of ecological models to predict population and community responses is available in the scientific literature (see sections 6.7.3 to 6.7.6). However, ecological models in support of the regulatory risk assessment for pesticides not often have been used because of lack of understanding of model assumptions, uncertainties about model inputs and outputs, and lack of validation and good modelling practise (Schmolke et al., 2010). Nevertheless, currently considerable research efforts take place to address these drawbacks and to further improve modelling approaches in the effect assessment procedures for pesticides (Grimm et al., 2009).

9.1.4 Ecological scenarios for Dutch drainage ditches

To facilitate the future use of ecological models to assess the risks of pesticide exposure in Dutch drainage ditches the life history and ecotoxicological profiles of the populations/communities modelled should match the populations and communities that are typical for Dutch drainage ditches. The diversity of aquatic vascular plants and macro-invertebrates in Dutch drainage ditches is reported to be potentially high and enough information on these taxa is available to construct a target image typical for Dutch drainage ditches (Brock et al., 2010b). However, more information is needed on the species composition and densities of typical zooplankton and phytoplankton assemblages. Furthermore, a systematic overview of typical fish species and their densities in Dutch drainage ditches is lacking.

The normal operating range of environmental and ecological conditions in ditches can be characterised as dynamic in space and time due to the mechanical cleaning and dredging regime. This management regime favours plant and animal species with pioneering properties. Invertebrates with two or more generations per year are widely distributed in Dutch drainage ditches of agricultural landscapes, but taxa with one generation per year are also common (approximately 33% of the taxa and 19% of the individuals). For insects that differ in number of generations per year no significant difference in overall sensitivity to insecticides could be demonstrated. This indicates that the RAC as derived by means of the SSD approach is not dependent on the voltinism of the species incorporated in the SSD curve. In addition, this also suggests that a mesocosm-RAC indicative for the threshold level of effects (e.g. based on Effect class 1-2 concentrations) and derived from test systems primarily populated with species that have short life cycles can be used as well for communities characterised by species with a larger variation in length of life cycle. When considering recovery in the risk assessment, however, species-specific properties such as generation time and dispersal abilities within water courses of the agricultural landscape cannot be ignored when extrapolating results of micro-/mesocosm tests to the field (Brock et al., 2010b).

Mesocosm experiments (Brock et al., 2009; 2010c) and metapopulation studies (Van den Brink et al., 2007) suggest that in drainage ditches the ecological impact of pesticides may be smaller in exposed sections of ditches if only part of the system suffers pesticide-stress. However, an important question at stake is:

At the landscape level, how large should be the surface area, and distance to refuges to allow an overall acceptable impact (including recovery and effects at a distance) of pesticide-stress on populations and communities?

In order to answer this question ecological scenarios and model exercises at the landscape level are needed that not only take into account biological traits of potentially sensitive populations but also the possible effects of multiple stress caused by spatio-temporally variable exposures to different pesticides.

9.1.5 Verification of chronic risk assessment procedures

Higher-tier risk assessment procedures predominantly have been calibrated/validated for acute toxicity and only sparsely for chronic effect assessments. For example, in the open literature for a few pesticides only (predominantly herbicides) enough chronic toxicity data for additional species can be found, hampering the application of the SSD approach to derive a chronic RAC. In the present report it is assumed that the regularities in species sensitivity distributions found for acute toxicity data and pesticides with a similar toxic mode-of-action also apply for chronic toxicity data. Although there are no clear indications that this assumption is wrong its verification needs further attention from a scientific and regulatory point of view. Furthermore, since micro-/mesocosm experiments that studied chronic, more or less constant exposure regimes are scarce as well, not enough data are available to demonstrate the protective value of RACs derived from chronic lower-tier tests (including standard test species, SSD and refined exposure test approaches). Consequently, the calibration/validation of the chronic effect assessment procedure requires a larger number of chronic micro-/mesocosm tests with pesticides that differ in toxic mode-of-action.

In this report we follow the criteria proposed by the ELINK report to decide whether the TWA approach can be used in the chronic risk assessment. The use of the TWA approach has been predominantly tested in (elongated) laboratory single species toxicity tests. However, the criteria proposed in the ELINK document and described in Section 3.3.2 need to be scientifically underpinned for different levels of biological organisation. In addition, more attention should be paid in future research on the time required to express the effects in chronic toxicity tests. Information on the time-to-onset-of-effects for different species and pesticides that differ

in toxic mode-of-action may allow developing scientifically underpinned criteria to set the appropriate time-window for the TWA PEC in chronic risk assessments.

9.1.6 Ecological consequences of exposure regimes that vary in space and time

In edge-of-field surface waters time-variable exposure concentrations are more often the rule rather than the exception. The largest ecological impact of a fast acting pesticide on a specific site usually occurs during or immediately after periods of high pulse exposures. However, for sensitive taxonomic groups these pulse exposures most likely will be more detrimental for more or less sessile organisms than for more mobile organisms with a larger territory, certainly when these mobile organisms are characterised by avoidance behaviour. Differences in mobility and territory (or home range) between species in Dutch drainage ditches may be important when defining the spatial unit of the PEC estimates. For example, the relevant spatial unit for the PEC estimate may be several hundreds of metres of ditch length for mobile fish species, while that may be less than ten metres ditch length for sessile species. This example illustrates the importance of interaction between exposure and effects experts when developing risk assessment schemes. Ecotoxicologists need to inform the exposure experts what is the ecotoxicologically relevant concentration (ERC) including the relevant spatial and temporal units that should be considered for the PEC estimates. This ERC may be different for different types of organisms.

Pulse exposures may locally be more detrimental when organisms are metabolically active and populations are in their growing phase and/or when the exposure occurs when specific sensitive life stages are present. For this reason most laboratory toxicity tests are performed with young life stages of test species. For this reason also the majority of micro-/mesocosm experiments with pesticides usually are initiated in spring and early summer. However, for certain types of PPPs (particularly pre-emergence herbicides subject to leaching) high pulse exposures in drainage ditches may occur in periods when the potentially sensitive organisms hibernate (late autumn to early spring). To date, hardly any experiments have been performed to study the ecological consequences of pulse exposures in the period late autumn to early spring. Consequently, it needs to be verified that exposure in periods when sensitive species are metabolically less active (or hibernate) will not result in unacceptable latent effects. This type of information is of importance to determine whether different pulse exposures that may occur in different periods of the year should be evaluated individually or in combination.

9.2 Aspects specifically related to the WFD

9.2.1 Data requirements for the Species Sensitivity Distribution-approach

One of the most prominent differences between the edge-of-field assessment and the WFD-framework, is the way statistical extrapolation is used for derivation of RACs or standards. According to the WFD-guidance, SSDs can only be performed when at least ten (preferably fifteen) endpoints are available for at least eight taxonomic groups. If the data show that a taxonomic group is particularly sensitive, a separate SSD may be constructed for this group if at least ten endpoints are available. Within the authorisation procedure, the SSD can be focused directly on sensitive taxa, and a smaller number of data is required as minimum input (five for fish, eight for other taxa). In practice, it will sometimes be the case that there will be enough data to construct an SSD focused on the potentially sensitive species groups, but additional testing will have to be done to cover the taxonomic diversity of the generic SSD as required under the WFD- and REACH-guidance. Since a focused SSD generally leads to an HC_5 with less uncertainty (smaller confidence interval) than a generic one, the added value of the generic SSD may be questioned. On the other hand, there are examples from PPPs which are marketed for a specific use (i.e. herbicide), whereas the sensitivity of presumed 'sensitive' target taxa is

similar to that of the 'non-sensitive' ones. Citeria have to be developed to decide when a focused SSD may be constructed without first having to make a generic SSD. For the edge-of-field assessment (Section 6.4.4) we propose to construct a focused SSD if the first tier indicates that one standard test species of the basic set is considerably more sensitive (differing by a factor >10) than the others. The focused SSD should be constructed with toxicity data for species representative for the sensitive taxonomic group.

9.2.2 Assessment factors: scientific basis and consistency

Relationship between acute and chronic data

In general, the choice of the assessment factors in the WFD guidance is historically determined. For drafting the WFD-guidance, the European Commission required that the former TGD and current REACH-guidance should be followed as close as possible. In this way, the choices that have been made in the past have been carried over to the new guidance. It has to be recognised that decisions on assessment factors are not always scientifically underpinned. More often, it is a combination of science, policy and the wish to have coherent schemes with round figures that are easy to apply.

There are several implicit assumptions in the assessment factor schemes that should be underpinned with data. One of these items is the assumption that sensitivity on the acute and long term time scale is related. According to the assessment factor scheme (see 8.2), low assessment factors can only be applied to long term results, when such results have been generated covering the level showing the lowest $L(E)C_{50}$ in the short-term tests. The question is whether this connection is present in reality. Comparing the position of specific taxa in species sensitivity distributions between acute and chronic SSDs is an important topic for future research.

Assessment factors for SSDs

For derivation of the $QS_{fw, eco}$, the AF to be applied on the results of a chronic SSD is 5 by default, and may be lowered to 1 depending on several criteria. In this report, some guidance is presented on the choice of the factor, based on case studies. However, a step-by-step decision scheme is still to be developed. Similar to the chronic SSD, an AF is applied to an acute SSD when deriving the MAC-$QS_{fw, eco}$ by statistical extrapolation. Where an AF of 5 is considered sufficient for the chronic $QS_{fw, eco}$, the default AF for the derivation of the MAC-$QS_{fw, eco}$ is set to 10. The reason for this is that the MAC-$QS_{fw, eco}$ represents an acute no effect level, while the acute HC_5 refers to a 50% effect concentration. For PPPs it may be relevant to construct SSDs based on the potentially sensitive taxonomic group. The EQS guidance does not specify the AFs to be applied on the result of such a specific chronic or acute SSD to derive a QS/MAC-QS, neither for the AF to be applied when an SSD can be constructed on the basis of acute no-effect data. Therefore, a proposal for a consistent set of assessment factors has been made in Section 8.3.6, starting from the default assessment factors of the EQS-guidance. As pointed out in Section 8.3.4, data from chronic mesocosm studies may allow for the 'validation' of the appropriate AF to apply for the extrapolation of the chronic HC_5. However, these studies are still scarce. Similarly, the open domain scientific publications that compared acute HC_5 values of herbicides, insecticides and fungicides with threshold levels of effects observed in aquatic micro-/mesocosm studies should be used to further underpin the default AFs that can be applied when acute SSDs are constructed based on acute $L(E)C_{50}$-values or $NOEC/EC_{10}$-values for either a broad range of species or for sensitive taxonomic groups.

Assessment factor for the MAC-QSfw, eco using the base set data

As pointed out in Section 8.2.4, the AF of 10 that is used for derivation of the MAC-$QS_{fw, eco}$ does not seem to be consistent with the AFs for the chronic assessment. In the latter, an AF of 10 is used on the lowest NOEC to cover residual uncertainty, which is related to e.g. variation within the sensitive taxonomic group, and the translation of single species laboratory data to the field situation. This residual uncertainty also applies to the derivation of the MAC-$QS_{fw, eco}$. In addition, the MAC-$QS_{fw, eco}$ represents an acute no effect level, while the underlying data represent a 50% effect level. Since an additional factor is needed to go from the $L(E)C_{50}$ level

to the acute NOEC or L(E)C_{10}, it is questionable whether an AF of 10 on the lowest L(E)C_{50} is justified on the basis of the minimum laboratory data set alone. Furthermore, an AF of 10 is also applied if the MAC-$QS_{fw, eco}$ is derived using an SSD on acute data, while in that case much more information is available. If this latter factor is considered justified, a higher value should be applied when the data do not allow for construction of an SSD.

9.3 General issues

9.3.1 Risks to sediment-dwelling organisms

This report and the decision schemes presented in Chapter 4 have their focus on risk assessment procedures for aquatic organisms that (largely) occur in the water column and less so on typical benthic organisms that predominantly dwell in the sediment compartment. Developing risk assessment procedures and decision schemes for sediment-dwelling organisms, however, may be important in the near future for the following reasons:

- Many of the modern pesticides dissipate relatively fast from the water column by sorption to sediment particles (e.g. benzoyl urea and pyrethroid insecticides)
- Sediment-associated pesticides are reported to be toxic to benthic organisms and to hamper ecological recovery processes (e.g. Weston et al., 2008; Domagalski et al., 2010; Brock et al., 2010c)
- The protective value of the sediment RAC on basis of the standard *Chironomus riparius* test for other benthic invertebrates needs to be verified.

9.3.2 Risks of fungicides to aquatic fungi

Almost no information is available concerning the potential risks of fungicides (or PPPs in general) to aquatic fungi. Maltby et al. (2009) compiled aquatic ecotoxicity data for a series of fungicides. The dataset included acute single-species data for 42 fungicides, semi-field data for 12 fungicides and covered seven modes of action and different exposure regimes. SSDs were constructed for separate taxonomic groups (*i.e.* fish, invertebrates, and primary producers) and for all groups together. They conclude that there is no evidence to suggest that derived threshold values based on hazardous concentrations (HC_p) from acute aquatic SSDs would pose a risk to aquatic hyphomycetes. However, laboratory toxicity data on fungi were not included in the datasets, since they were not available. In the micro-/mesocosm studies reviewed, only functional responses of micro-organisms in the form of litter decomposition received attention. None of the semifield studies specifically studied structural endpoints of fungi. Maltby et al. (2009) therefore also concluded that the underlying data is limited in number and that further research on nontarget fungi should be conducted. The relevance of further research into the sensitivity of aquatic fungi was demonstrated recently in screening studies by Dijksterhuis et al. (2009, 2011) and CBS (2009). Their data indicate that HC_5 concentrations derived by Maltby et al. (2009) for ergosterol inhibitors may show an effect on aquatic fungi. Further research is needed to address the relevance of aquatic fungi as additional non-target groups in the risk assessment of PPPs. Special attention should be paid to the selection of appropriate test species, given the enormous diversity within the kingdom of fungi. When these data are collated, it will be a risk manager decision to set the specific protection goal for aquatic fungi (e.g. structure and/or function).

9.3.3 Pesticides with a novel toxic mode-of-action

As described in Section 5.1, a tiered risk assessment scheme needs to be appropriately protective, internally consistent, cost-effective and address the problem with a higher accuracy and precision when going from lower to higher tiers. This also means that the predictive value of lower tiers can be calibrated/validated by

means of higher-tier experiments. In this report we assume that the regularities observed between results of different effect tiers can be extrapolated between compounds with a similar toxic mode-of-action. This also implies that for compounds with a novel toxic mode-of-action the internal consistency of the tiered approach and the protective value of the lower tiers need to be verified.

9.3.4 Multiple stress and mixture toxicity

In many crops during the growing season more than one compound will be used. In some crops this can ad up to more than 50 applications and some of these compounds will be applied together, e.g. an herbicide together with an insecticide and/or fungicide. Sometimes even two or three herbicides or two or three fungicides or two insecticides may be applied simultaneously, up to 5 or 6 compounds at the same time. When these combinations (e.g. tank mixes) are not sold as a formulation the legislative process does not take account for the potential combined effects of the use of these tank mixes. Neither does the legislative process take into account that different compounds of the same group (e.g. insecticides) or of different groups (e.g. insecticides, herbicides, fungicides) are used over time in the same growing season.

When a compound is allowed on the market this decision is sometimes based on the potential of recovery. Whether under different crop scenarios the recovery option is appropriate to use in the derivation of the RAC needs to be evaluated from an ecological point of view, since during the growing season drainage ditches may be affected multiple times by the use of plant protection products. Research on multiple stress of pesticides on aquatic communities representative for Dutch drainage ditches, and how to deal with mixture toxicity of pesticides, has already been initiated in the past (Hartgers et al., 1998; Deneer, 2000; De Zwart, 2005; Van Wijngaarden et al., 2004; Arts et al., 2006; Van den Brink et al., 2002b and 2009). In 2009 a literature research was started to update the knowledge on mixture toxicity (Verbruggen and Van den Brink, 2010). In addition, a working group has been installed to look into the problem of multiple stress caused by pesticides in Dutch drainage ditches. This group will analyse some of the more realistic worst cases of pesticide use in crops (e.g. potatoes and fruit). As the multistress topic is currently being addressed in another working group, this topic will not be included and elaborated in the current report.

For WFD-water bodies, the situation may be more complex. It can be expected that the extremes in time-variable exposures are less pronounced. However, the exposure patterns may be less easy to assess because of the larger scale of the water bodies (relative to edge-of-field ditches) and the variety in land uses in the region connected to them. In addition, other chemicals resulting from industrial use or diffuse sources will be present.

9.3.5 Possible consequences of climate change

Climate change is a factor that influences water quantity and quality. De Nijs et al. (2008) mention a number of aspects related to this. Effects will occur on abiotic as well on biotic processes. Increasing temperature will lead to enhanced degradation of substances. Increased water loads in rivers can lead to release of contaminated sludge, which can be deposited in the river delta. On the other hand, a decrease in rainfall during summer, and thereby a decrease in dilution, can lead to an increase in concentrations. The capacity to complexate may decrease due to a decrease in dissolved organic matter. It is expected that the sensitivity of organisms towards PPPs and chemicals in general will be affected too. It has been demonstrated that the sensitivity of taxonomically related species is comparable between temperate regions and tropical regions (Daam et al., 2009b), although differences exist for some compounds (Kwok et al., 2007). However, exposing species to chemicals under conditions that they are not (yet) adapted to, leads to different effects. It has been demonstrated in several studies that an increase in temperature leads to enhanced toxicity

(e.g. Ferrando et al., 1987; Lydy et al., 1999; Prato et al., 2006). It has also been demonstrated that subtle changes in environmental conditions and changes in abundance of key species can cause a shift towards a different stable state of the ecosystem (Scheffer, 2009).

9.3.6 **Exposure modelling in WFD water bodies**

In surface waters, time-variable exposure regimes of plant protection products are the rule rather than the exception. Consequently, in WFD water bodies the actual peak concentration usually will not be measured at the monitoring frequency adopted. Furthermore, it is questionable whether the annual average concentration on basis of e.g. twelve measurements can be considered as the ecotoxicologically relevant concentration (ERC) for plant protection products. To assist the conduct and interpretation of chemical monitoring programmes, catchment-level exposure scenarios and exposure models might be used, but still have to be developed/improved for interconnected water bodies in the Netherlands. These catchment exposure tools not only may help to identify PPPs of potential concern in larger WFD water courses so that cost-effective post-registration monitoring programmes can be established, but they also may be useful for the interpretation of monitoring data in relation to PPP use in certain crops and, if deemed necessary by risk managers, in prospective risk assessment procedures.

10 Glossary

AA-EQS	annual average environmental quality standard = environmental quality standard protecting aquatic ecosystems and organisms depending on it from long-term exposure
ADI	acceptable daily intake
AMRAP	SETAC Europe workshop Aquatic Macrophyte Risk Assessment for Plant Protection Products (Maltby et al., 2010)
AUC	area under the curve
AF	assessment factor
BAF	bioaccumulation factor
BBW	beslisboom water
BCF	bioconcentration factor
BMF	biomagnification factor
DT_{50}	half life, or time to 50% degradation
$DT_{90,system}$	time to 90% degradation
EC_{10}	effective concentration to 10% of the test organisms
EC_{50}	effective concentration to 50% of the test organisms
EEC/EC	European Commission
EFSA	European Food Safety Authority
ELINK	SETAC Europe workshop on linking aquatic exposure and effects for pesticides (Brock et al., 2010b)
ELS-test	early life stage test
EQS	environmental quality standard
ERC	ecotoxicologically relevant concentration
ETR	exposure toxicity ratio
EU	European Union
FLC-test	full life cycle test
FOCUS	forum for the co-ordination of Pesticide Fate Models and Their Use
HARAP	higher-tier Aquatic Risk Assessment for Pesticides
HC_5	hazardous concentration for 5% of the species, predicted from an SSD curve
IGR	insect growth regulator
LC_{50}	lethal concentration to 50% of the test organisms
Log Kow	log of octanol water partition coefficient
MAC-EQS	final maximum acceptable concentration - environmental quality standard = environmental quality standard protecting aquatic ecosystems from short-term concentration peaks
$MAC\text{-}QS_{fw,\ eco}$	maximum acceptable concentration - quality standard for freshwater ecosystems, several $MAC\text{-}QS_{fw,\ eco}$ can be derived using different methods
NOEAEC	no observed ecologically adverse effect concentration
NOAEL	no observed adverse effect level
NOEC	no observed effect concentration
OECD	Organisation for Economic Cooperation and Development

PEC	predicted environmental concentration
PEC_{max}	maximum predicted environmental concentration
PEC_{fish}	predicted environmental concentration in fish
PEC_{water}	predicted environmental concentration in water
PNEC	predicted no effect concentration
PPP	plant protection product(s)
QS	Quality Standard
$QS_{fw, eco}$	long term quality standard for aquatic species, expressed as concentration in freshwater
$QS_{fw, secpois}$	long term quality standard for predatory birds and mammals, expressed as concentration in freshwater
$QS_{water, hh food}$	long term quality standard for fish-eating humans, expressed as concentration in water
$QS_{biota, secpois, fw}$	long term quality standard for predatory birds and mammals, expressed as concentration in freshwater biota (e.g. fish, mussels)
$QS_{biota, hh food}$	long term quality standard for fish-eating humans, expressed as concentration in aquatic biota (e.g. fish, shellfish)
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RAC	regulatory acceptable concentration
RAC_{ELS}	RAC for the early life stage test with fish
RAC_{FLC}	RAC for full life cycle test with fish
$RAC_{recovery}$	RAC derived from a micro-/mesocosm test by using the ecological recovery option (e.g. Effect class 3A)
RAC_{SP}	RAC for secondary poisoning
$RAC_{threshold}$	RAC derived from a micro-/mesocosm test by using the ecological threshold option (e.g. Effect class 1-2)
SSD	species sensitivity distribution
TDI	tolerable daily intake
TGD	technical guidance document
TK/TD models	toxicokinetic/toxicodynamic models
TL	threshold limit
TWA	time weighted average
WFD	Water Framework Directive

11 Literature

Aldenberg, T. and J.S. Jaworska, 2000. Uncertainty of hazardous concentrations and fraction affected for normal species sensitivity distributions. *Ecotoxicol Environ Saf* 46:1-18.

Aldenberg, T., J.S. Jaworska and T.P. Traas, 2002. Normal species sensitivity distributions in probabilistic ecological risk assessment. In: Posthuma L, Traas TP, Suter GW (eds). Species Sensitivity Distributions in Risk Assessment. Boca Raton (FL), CRC Press, pp. 49-102.

Angeler, D.G. and M. Alvarez-Cobelas, 2005. Island biogeography and landscape structure: Integrating ecological concepts in a landscape perspective of anthropogenic impacts in temporary wetlands. *Environ Pollut* 138:420-424.

Anonymous, 2007. Draft report on 'SETAC Technical Workshop on the Derivation and Use of Environmental Quality and Human Health Standards for Chemical Substances in Water and Soil'. Faringdon, Oxfordshire, UK; 16-19 October, 2006.

Arnold, D., I.R. Hill, P. Matthiessen and R. Stephenson (eds.), 1991. Guidance Document on testing procedures for pesticides in freshwater mesocosms. SETAC-Europe publication, Brussels. ISBN 90-5607-002-9.

Arts, G.H.P., L.L. Buijse-Bogdan, J.D.M. Belgers, C.H. van Rhenen-Kersten, R.P.A. van Wijngaarden, I. Roessink, S.J. Maund, P.J. van den Brink and T.C.M. Brock, 2006. Ecological impact in ditch mesocosms of simulated spray drift from a crop protection programme for potatoes. *Integr Environ Assess Manag* 2:105-125.

Ashauer, R., A.B.A. Boxall and C.D. Brown, 2006. Predicting effects on aquatic organisms from fluctuating or pulsed exposure to pesticides. *Environ Toxicol Chem* 25: 1899-1912.

Ashauer, R., A.B.A. Boxall and C.D. Brown, 2007a. New ecotoxicological model to simulate survival of aquatic invertebrates after exposure to fluctuating and sequential pulses of pesticides. *Environ Sci Technol* 41(4): 1480-1486.

Ashauer., R, Boxall ABA, Brown CD (2007b). Simulating toxicity of carbaryl to Gammarus pulex after sequential pulsed exposure. *Environ Sci Technol* 41(15): 5528-5534.

Ashauer, R, A.B.A. Boxall and C.D. Brown, 2007c. Modelling combined effects of pulsed exposure to carbaryl and chlorpyrifos on Gammarus pulex. *Environ Sci Technol* 41: 5535-5541.

Baird, D.J., I. Barber and P. Calow, 1990. Clonal variation in general responses of Daphnia magna Straus to toxic stress. I. Chronic life-history effects. *Funct Ecol* 4: 399-407.

Baird, D.J., I. Barber, M. Bradley, A.M.V.M. Soares and P. Calow, 1991. A comparative study of genotype sensitivity to acute toxic stress using clones of Daphnia magna straus. *Ecotox Environ Saf* 21:257-265.

Barnthouse, L.W., 2004. Quantifying population recovery rates for ecological risk assessment. *Environ Toxicol Chem* 23(2):500-508.

Beketov, M.A., R.B. Schäfer, A. Marwitz, A. Paschke and M. Liess, 2008. Long-term stream invertebrate community alterations induced by the insecticide thiacloprid: Effects concentrations and recovery dynamics. *Sci Total Environ* 405:96-108.

Biever, R.C., J.M. Giddings, M. Kiamos, MF. Annunziato, R. Meyerhoff and K. Racke, 1994. Effects of chlorpyrifos on aquatic microcosms over a range of off-target drift exposure levels. Brighton Crop Protection Conference - Pests and Deseases - 1994. pp. 1367 - 1372.

Bodar, C. and C.E. Smit, 2008. Nieuwe normen voor bestrijdingsmiddelen: schoner water? H_2O 24: 53-55.

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

Brock, T.C.M. and B.J. Budde, 1994. On the choice of structural parameters to indicate responses of freshwater ecosystems to pesticide stress. In: Hill IA, Heimbach F, Leeuwangh P, Matthiesen P (Eds), Freshwater Field Tests for Hazard Assessment of Chemicals Lewis Publishers, Michigan, pp. 19-56.

Brock, T.C.M., S.J.H. Crum, J.W. Deneer, F. Heimbach, R.M.M. Roijackers and J.A. Sinkeldam, 2004. Comparing aquatic risk assessment methods for the photosynthesis-inhibiting herbicides metribuzin and metamitron. *Environ Poll* 130: 403-426.

Brock, T.C.M., G.H.P. Arts, L. Maltby and P.J. van den Brink, 2006. Aquatic Risks of Pesticides, Ecological Protection Goals and Common Aims in European Union Legislation. *Integr Environ Assess Manag* 2(4): E20-E46.

Brock, T.C.M., L. Maltby, C.W. Hickey. J. Chapman and K. Solomon, 2008. Spatial extrapolation in ecological effect assessment of chemicals. In: Solomon et al. eds. Extrapolation practice for ecotoxicological effect characterization of chemicals, Chapter 7. Boca Raton, FL, USA: CRC Press (SETAC books).

Brock, T.C.M., Roessink I, Belgers JDM, Bransen F, Maund SJ. 2009. Impact of a benzoyl urea insecticide on aquatic macroinvertebrates in ditch mesocosms with and without nonsprayed sections. *Environ Toxicol Chem* 28:2191-2205.

Brock, T.C.M., A. Alix, C.D. Brown, E. Capri, B.F.F. Gottesbüren, F. Heimbach, C.M. Lythgo, R. Schulz and M. Streloke, 2010a. Linking Aquatic Exposure and Effects: Risk Assessment of Pesticides. SETAC Press & CRC Press, Taylor & Francis Group, Boca Raton, FL, 398 pp.

Brock, T., G. Arts, D. Belgers and C. van Rhenen-Kersten, 2010b. Ecological characterization of drainage ditches in the Netherlands to evaluate pesticide stress. In: Brock, T.C.M., A. Alix, C.D. Brown, E. Capri, B.F.F. Gottesbüren, F. Heimbach, C.M. Lythgo, R. Schulz and M. Streloke, 2010. Linking Aquatic Exposure and Effects: Risk Assessment of Pesticides. SETAC Press & CRC Press, Taylor & Francis Group, Boca Raton, FL, pp. 269-287.

Brock, T.C.M., J.D.M. Belgers, I. Roessink, J.G.M. Cuppen and S.J. Maund, 2010c. Macroinvertebrate responses to insecticide application between sprayed and adjacent non-sprayed ditch sections of different sizes. *Environ Toxicol Chem* 29:1994-2008.

Brockway, D.L., P.D. Smith, F.E. and Stancil, 1984. Fate and effects of atrazine in small aquatic microcosms. *Bull Environ Contam toxicol* 32: 345-353.

Bryfogle, B.M. and W.F. McDiffet, 1979. Algal succession in laboratory microcosms as affected by an herbicide stress. *Am Midl Nat* 10(2): 344-354.

Campbell, P.J., D.J.S. Arnold, T.C.M. Brock, N.J. Grandy, W. Heger, F. Heimbach, S.J. Maund and M. Streloke, 1999. Guidance document on higher-tier aquatic risk assessment for pesticides (HARAP). Brussels (BE): SETAC-Europe, 179 p. ISBN 90-5607-011-8.

Caquet, T., L. Lagadic and S.R. Sheffield, 2000. Mesocosms in ecotoxicology. 1. Outdoor aquatic systems. *Rev Environ Contam Toxicol* 165:1-38.

CBS, 2009. De gevoeligheid van schimmels in het oppervlaktewater voor fungiciden die worden toegepast in de landbouw. Centraal Bureau voor Scimmelcultures, Utrecht, Weesp, 22 pp.

Crossland, N.O., F. Heimbach, I.R. Hill, A. Boudou, P. Leeuwangh, P. Matthiessen and G. Persoone 1993. European Workshop on Freshwater Field tests (EWOFFT), Summary and recommendations. Workshop held in Potsdam, Germany, June 25 -26, 1992.

Cuppen, J.G.M., P.J. van den Brink, E. Camps, K.F. Uil and T.C.M. Brock, 2000. Impact of the fungicide carbendazim in freshwater microcosms. I. Water quality, breakdown of particulate organic matter and responses of macroinvertebrates. *Aquat Toxicol* 48: 233-250.

Daam, M.A., S.J.H. Crum, P.J. van den Brink and A.J.A. Nogueira, 2008. Fate and effects of the insecticide chlorpyrifos in outdoor plankton-dominated microcosms in Thailand. *Environ Toxicol Chem* 27:2530-2538.

Daam, M.A., K. Satapornvanit, P.J. van den Brinkand A.J.A. Nogueira, 2009a. Sensitivity of macroinvertebrates to carbendazim under semi-field conditions in Thailand: Implications for the use of temperate toxicity data in a tropical risk assessment of fungicides. *Chemosphere* 74: 1187-1194.

Daam, M.A., P.J. van den Brink and A.J.A. Nogueira, 2009b. Comparison of fate and ecological effects of the herbicide linuron in freshwater model ecosystems between tropical and temperate regions. *Ecotoxicol Environ Saf* 72: 424-433.

Dang, Z. and C.E. Smit, 2008. Environmental risk limits for carbendazim. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). Letter report no. 601716014.

De Jong, F.M.W., T.C.M. Brock, E.M. Foekema and P. Leeuwangh, 2008. Guidance for summarizing and evaluating aquatic micro- and mesocosm studies. RIVM Report 601506009, Bilthoven, the Netherlands, 59 p.

De Knecht, J.A .and R. van Herwijnen, 2008. Environmental risk limits for deltamethrin. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716015, 56 p.

Deneer, J.W. 2000. Toxicity of mixtures of pesticides in aquatic systems. *Pest Manag Sci* 56:516-520.

DeNoyelles, Jr. F., S.L. Dewey, D.G. Huggins and W.D. Kettle, 1994. Aquatic mesocosms in ecological effects testing: detecting direct and indirect effects of pesticides. In: Graney, R,L,, J.H. Kennedy and J.H. (jr) Rodgers (eds), Aquatic mesocosm studies in ecological risk assessment. Lewis Publishers, Boca Raton, FL, pp. 577-603

De Nijs, A.C.M., A. Driesprong, H.A. Den Hollander, L.R.M. De Poorter, W.H.J. Verweij, J.A. WHJ, J.A. Vonk and D. de Zwart, 2008. Risico's van toxische stoffen in de Nederlandse oppervlaktewateren. Bilthoven, Nederland: RIVM. Rapport nr. 607340001.

Detenbeck, N.E., R. Hermanutz, K. Allen and M.C. Swift, 1996. Fate and effects of the herbicide atrazine in flow-through wetland mesocosms. *Environ Toxicol Chem* 15(6): 937-946.

De Werd, H.A.E. and R. Kruijne, (eds), 2011. Interpretation of surface water monitoring results in the authorisation procedure of plant protection products in the Netherlands. Applied Plant Research Report nr. 2011-02, DLO Foundation, Wageningen, 87 pp.

De Zwart, D., 2005. Ecological effects of pesticide use in the Netherlands: Modeled and observed effects in the field ditch. *Integr Environ Assess Manag* 1:123-134.

Domagalski, J.L., D.P. Weston, M. Zhang and M. Hjadik, 2010. Pyrethroid insecticide concentrations and toxicity in streambed sediments and loads in surface waters of the San Joaquin Valley, California, USA. *Environ Toxicol Chem* 29:813-823

Dijksterhuis, J., T. van Doorn and J. Postma, 2009. De gevoeligheid van oppervlaktewaterschimmels blootgesteld aan azolen en strobilurines die worden toegepast in de landbouw. Centraal Bureau voor Schimmelcultures / Ecofide, Utrecht, Weesp, 32 pp.

Dijksterhuis, J., T. van Doorn and J. Postma, 2011. Effects of seven fungicides on non-target aquatic fungi. *Wat Air Soil Pollut* DOI 10.1007/s11270-011-0836-3

EC, 1991. Council Directive 91/414/EEC of 15 July 1991 concerning the placing of plant protection products on the market. Official Journal of the European Union L 230:1-32. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31991L0414:EN:NOT

EC, 1997. Council Directive 97/57/EC of 22 September 1997 establishing Annex VI to Directive 91/414/EEC concerning the placing of plant protection products on the market. Official Journal of the European Union C 240: 1-23. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31997L0057:EN:NOT

EC, 2000. Directive 2000/60/EC of the European parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Union L 327: 1-72. http://ec.europa.eu/environment/water/water-framework/index_en.html

EC, 2003. Technical Guidance Document on Risk Assessment in Support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances and Commission Regulation (EC) No 1488/94 on Risk Assessment for Existing Substances and Directive 98/8/EC of the European Parliament and the Council Concerning the placing of biocidal products on the market. European Commission, Joint Research Centre.

EC, 2009. Regulation (EC) No 1107/2009 of the European parliament and the council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal of the European Union L 309:1-50.

EC, 2011. Technical Guidance Document for deriving Environmental Quality Standards. Draft version 5.0 (29 January 2010).

ECETOC, 2001. Risk assessment in marine environments. Brussels, Belgium: European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC). Technical Report No. 82.

ECHA, 2008. Guidance on information requirements and chemical safety assessment. Chapter R.10: Characterisation of dose [concentration]-response for environment. Helsinki, Finland: European Chemicals Agency. Guidance for the implementation of REACH. May 2008. http://guidance.echa.europa.eu/docs/guidance_document/information_requirments_r10_en.pdf?vers=20_08_08

EFSA, 2005a. Opinion of the Scientific Panel on Plant health, Plant Protection Products and their Residues on a request from the EFSA related to the evaluation of dimoxystrobin. *The EFSA Journal* 178, 1-45.

EFSA, 2005b. Opinion of the Scientific Panel on Plant health, Plant Protection Products and their Residues on a request from the EFSA related to the assessment of the acute and chronic risk to aquatic organisms with regard to the possibility of lowering the assessment factor if additional species were tested. *The EFSA Journal* 301, 1-45.

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

EFSA, 2008. Risk Assessment for Birds and Mammals. The EFSA Journal 734:1-181.

EFSA, 2009. Risk Assessment for Birds and Mammals. *The EFSA Journal* 7(12):1438 [358 pp.].

EFSA, 2010. Scientific opinion on the development of specific protection goal options for environmental risk assessment of pesticides, in particular in relation to the revision of the Guidance Documents on Aquatic and Terrestrial Ecotoxicology (SANCO/3268/2001 and SANCO/10329/2002). *The EFSA Journal* 2010; 8(10):1821 [55 pp.].

Fairchild, J.F., T.W. La Point and T.R. Schwartz, 1994. Effects of herbicide and insecticide mixture in aquatic mesocosms. *Arch Environ Contam Toxicol* 27: 527-533.

Farmer, D., I.R. Hill and S.J. Maund, 1995. A comparison of the fate and effects of two pyrethroid insecticides (lambda-cyhalothrin and cypermethrin) in pond mesocosms. *Ecotoxicology* 4: 219-244.

Ferrando, M.D., E. Andreu-Moliner, M.M. Almar, C. Cebrian and A. Nuñez 1987. Acute toxicity of organochlorined pesticides to the European eel, *Anguilla* anguilla: the dependency on exposure time and temperature. *Bull Environ Contam Toxicol* 39: 365-369.

FOCUS, 2001. FOCUS Surface Water Scenarios in the EU Evaluation Process under 91/414/EEC'. Report of the FOCUS Working Group on Surface Water Scenarios, EC Document Reference SANCO/4802/2001-rev.2. 245 pp.

Forbes, V.E., P. Calow and R.M. Sibly, 2001. Are current species extrapolation models a good basis for ecological risk assessment? Environ Toxicol Chem 20: 442-447.

Forbes, V.E. and P. Calow, 2002a. Species Sensitivity Distributions revisited: a critical appraisal. *Human Ecol Risk Assess* 8:473-492.

Forbes, V.E. and P. Calow, 2002b. Population growth rate as a basis for ecological risk assessment of toxic chemicals. *Phil Trans R Soc Lond B* 357: 1299-1306.

Galic, N., U. Hommen, H. Baveco and P.J. van den Brink, 2010. Potential application of population models in the European ecological risk assessment of chemicals II: Review of models and their potential to address environmental protection aims. *Integr Environ Assess Manag* 6:338-360.

Giddings, J.M., R.C. Biever, R.L. Helm, G.L. Howick and F.J. deNoyell, 1994. The fate and effects of Guthion (azinphos methyl) in mesocosms. In: Aquatic mesocosm studies in ecological risk assessment. R.L. Graney, J.H. Kennedy, J.H. Rogers (Eds.). Special Publication of the Society of Environmental Toxicity and Chemistry, Michigan: Lewis Publishers, pp. 469 - 495.

Giddings, J.M., T.C.M. Brock, H. Heger, F. Heimbach, S.J. Maund, S. Norman, H.T. Ratte, C. Schäfers and M. Streloke (eds), 2002. Community-level aquatic system studies-interpretation criteria. (CLASSIC) Pensacola (FL): SETAC 44 p.

Giesy, J.P. and R.L. Graney, 1989. Recent developments in and intercomparisons of acute and chronic bioassaysand bioindicators. *Hydrobiologia* 188/189:21-60

Goldborough, L.G. and G.G.C. Robinson, 1983. The effect of two triazine herbicides on the productivity of freshwater marsh periphyton. *Aquat Toxicol* 4: 95-112.

Goldborough, L.G. and G.G.C. Robinson, 1986. Changes in periphytical algal community structure as a consequence of short herbicide exposures. *Hydrobiologia* 139: 177-192.

Gotelli, N.J. and A. Ellison, 2006. Food-Web models predict species abundances in response to habitat change. *PloS Biol* 4(10).

Grimm, V., R. Ashauer, V. Forbes, U. Hommen, T.G. Preuss, A. Schmidt, P.J. van den Brink, J. Wogram and P. Thorbek, 2009. CREAM: A European project on mechanistic effect models for ecological risk assessment of chemicals. *Environ Sci Pollut Res* 16:614-617.

Gruessner, B. and M.C. Watzin, 1996. Response of aquatic communities from a Vermont stream to environmentally realistic atrazine exposure in laboratory microcosms. *Environ Toxicol Chem* 15: 410-419.

Guasch, H. and S. Sabater, 1998. Light history influences the sensitivity to atrazine in periphytic algae. *J Phycol* 34:233-241.

Hanazaro, T. and H. Hirokawa, 2004. Changes in vulnerability of *Daphnia* to an insecticide application depending on the population phase. *Freshw Biol* 49:402-409.

Hartgers, E.M., G.H. Aalderink, P. van den Brink, R. Gylstra, J.W.F. Wiegman and T.C.M. Brock, 1998. Ecotoxicological threshold levels of a mixture of herbicides (atrazine, diuron and metolachlor) in freshwater microcosms. *Aquatic Ecology* 32: 135-152.

Heckman, L.H. and N. Friberg, 2005. Macroinvertebrate community response to pulse exposure with the insecticide lambda-cyhalothrin using in-streaam mesocosms. *Environ Toxicol Chem* 24:582-590.

Hill, I.R., F. Heimbach, P. Leeuwangh and P. Matthiessen (Eds). 1994a. Freshwater Field Tests for Hazard Assessment of Chemicals. Boca Raton (FL), Lewis Publishers, 561.

Hill, I.R., J.K. Runnalls, J.H. Kennedy and P. Ekoniak, 1994b. Lambda-cyhalothrin: A mesocosm study of its effects on aquatic organisms. In: Aquatic mesocosm studies in ecological risk assessment. R.L. Graney, J.H. Kennedy, J.H. Rogers (Eds.). Special Publication of the Society of Environmental Toxicity and Chemistry, Michigan: Lewis Publishers, pp. 403-467.

Holland, P.T., 1996. Glossary of terms relating to pesticides (IUPAC Recommendations 1996). *Pure Appl Chem* 68:1167-1193.

Hommen, U., J.M. Baveco, N. Galic and P. van den Brink, 2010a. Potential application of ecological models in the European environmental risk assessment of chemicals. I. Review of protection goals in EU directives. *Integr Environ Assess Manag* 6:325-337.

Hommen, U., R. Ashauer, P. van den Brink, T. Caquet, V. Ducrot, L. Lagadic and T. Ratte, 2010b. Extrapolation methods in aquatic effect assessment of time-variable exposures to pesticides. In: Brock, T.C.M., A. Alix, C.D. Brown, E. Capri, B.F.F. Gottesbüren, F. Heimbach, C.M. Lythgo, R. Schulz and Streloke, (Eds), 2010. Linking Aquatic Exposure and Effects: Risk Assessment of Pesticides. SETAC Press & CRC Press, Taylor & Francis Group, Boca Raton, FL, pp. 211-242.

Hutchinson, T., N. Scholz and W. Guhl, 1998. Analysis of the ECETOC aquatic toxicity (EAT) database IV - Comparative toxicity of chemical substances to freshwater versus saltwater organisms. Chemosphere 36 (1): 143-153.

Jenkins, D.O. and A.L. Buikema Jr., 1990. Response of a winter plankton food web to simazine. *Environ Toxicol Chem* 9: 693-705.

Johnson, B.T., 1986. Potential impact of selected agricultural chemical contaminants on a northern prairie wetland: a microcosm evaluation. *Environ Toxicol Chem* 5: 473-485.

Jongbloed, R.H., T.P. Traas and R. Luttik, 1996. A probabilistic model for deriving soil quality criteria based on secondary poisoning of top predators. II. Calculations for dichlorodiphenyltrichloroethane (DDT) and cadmium. *Ecotox Environ Saf* 34: 279-306.

Jüttner, .I, A. Peither, J.P. Lay, A. Kettrup and S.J. Ormerod, 1995. An outdoor mesocosm study to assess ecotoxicological effects of atrazine on a natural plankton community. *Arch Environ Contam Toxicol* 29: 435-441.

Knuth, M.L., L.E. Anderson, L.J. Heinis, R.E. Hicks, A.W. Jarvinen, S.J. Lozano, S.L. O'Halloran, J.A. Patzer, R.W. Seelke, R.E. Siefert, M.E. Sierzen, F.S. Stay and D.K. Tanner, 1992. Effects, persistence and distribution of azinphos-methyl in littoral enclosures. U.S. EPA, Environmental Research Laboratory-Duluth, Ecosystem Response Branch, Duluth, MN 55804.

Kooijman, S.A.L.M. and J.J.M. Bedaux, 1996. The Analysis of Aquatic Toxicity Data. VU University Press, Amsterdam, Netherlands.

Kosinski, R.J. and M.G. Merkle, 1984. The effect of four terrestrial herbicides on the productivity of artificial stream algal communities. *J Environ Qual* 13: 75-82.

Kosinski, R.J., 1984. The effect of terrestrial herbicides on the community structure of stream periphyton. *Environ Pollut Ser A* 36: 165-189.

Krieger, K.A., D.B. Baker and J.W. Kramer, 1988. Effects of herbicides on stream Aufwuchs productivity and nutrient uptake. *Arch Environ Contam Toxicol* 17: 299-306.

Kwok, K.W.H., K.M.Y. Leung, G.S.G. Lui, V.K.H. Chu, P.K.S. Lam, I.I.D. Morritt, L. Matlby, T.C.M. Brock, P.J. van den Brink, M.S.J. Warne and M. Crane, 2007. Comparison of tropical and temperate freshwater animal species' acute sensitivities to chemicals: implications for deriving safe extrapolation factors. *Integr Environ Assess Manag* 3: 49-67.

Lepper, P., 2005. Manual on the methodological framework to derive environmental quality standards for priority substances in accordance with Article 16 of the Water Framework Directive (2000/60/EC). Schmallenberg, Germany: Fraunhofer-Institute for Molecular Biology and Applied Ecology.

Leung, K.M.Y., D. Morritt, J.R. Wheeler, P. Whitehouse, N. Sorokin, R. Toy, M. Holt and M. Crane, 2001. Can saltwater toxicity data be predicted from freshwater data? *Mar Pollut Bull* 42(11): 1007-1013.

Liess, M., B.J. Pieters and S. Duquesne, 2006. Long-term signal of population disturbance after pulse exposure to an insecticide: Rapid recovery of abundance, persistent alteration of structure. *Environ Toxicol Chem* 25: 1326-1331.

Lin, B.L., A. Tokai and J. Nakanishi, 2005. Approaches for establishing predicted-no-effect concentrations for population-level ecological risk assessment in the context of chemical substances management. *Environ Sci Technol* 39(13):4833-4840.

López-Mancisidor, P., G. Carbonell, C. Fernández and J.V. Tarazona, 2008. Ecological impact of repeated applications of chlorpyrifos on zooplankton community in mesocosms under Mediterranean conditions. *Ecotoxicology* 17:811-825.

López-Mancisidor, P., G. Carbonell, A. Marina, C. Fernández and J.V. Tarazona, 2007. Zooplankton community responses to chlorpyrifos in mesocosms under Mediterranean conditions. *Ecotoxicol Environ Saf* 71: 16-25.

Lozano, S.J., S.L. O'Halloran, K.W. Sargent and J.C. Brazner, 1992. Effects of esfenvalerate on aquatic organisms in littoral enclosures. *Environ Toxicol Chem* 11:35-47.

Luttik, R., 2003. Risk assessment scheme for the impact of plant protection products on birds and mammals. Thesis, Leiden University.

Lydy, M.J., J.B. Belden and M.A. Ternes, 1999. Effects of temperature on the toxicity of m-parathion, chlorpyrifos, and pentachlorobenzene to *Chironomus tentans*. *Arch Environ Contam Toxicol* 37: 542-547.

Maltby, L., N. Blake, T.C.M. Brock, P.J. van Den Brink, 2005. Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems. *Environ Toxicol Chem* 24: 379-388.

Maltby, L., T.C.M. Brock and P.J. van den Brink, 2009. Fungicide risk assessment for aquatic ecosystems: importance of interspecific variation, toxic mode of action and exposure regime. *Environ Sci Technol* 43: 7556-7563.

Maltby, L., D. Arnold, G. Arts, J. Davies, F. Heimbach, C. Pickl and V. Poulsen, (eds.), 2010. Aquatic Macrophyte Risk Assessment for Pesticides. SETAC Press & CRC Press, Taylor & Francis Group, Boca Raton, FL, 140 pp.

Moermond, C.T.A., 2008. Environmental risk limits for pyriproxyfen. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716022, 36 p.

Moermond, C.T.A., P.L.A. van Vlaardingen, J.H. Vo and E.M.J. Verbruggen, 2008a. Environmental risk limits for dimethoate. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). Report no. 601714001.

Moermond, C.T.A., J.H. Vos and E.M.J. Verbruggen, 2008b. Environmental risk limits for organophosphorous pesticides. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). Report no. 601714004 (with erratum).

Moermond, C.T.A. and E.H.W. Heugens, 2009. Environmental risk limits for trichlorophenols. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). Report no. 601714005, 95 p.

Muňos, I., M. Real, H. Guasch, E. Navarro and S. Sabater, 2001. Effects of atrazine on periphyton under grazing pressure. *Aquatic Toxicology* 55: 239-249.

Nystrőm, B., M. Paulsson, K. Almgren and H. Blank, 2000. Evaluation of the capacity for development of atrazine tolerance in periphyton from a Swedish freshwater site as determined by inhibition of photosynthesis and sulfolipid synthesis. *Environ. Toxicol. Chem.* 19: 1324-1331.

OECD, 1996. OECD Guidelines for testing of chemicals. Section 3 Degradation and accumulation: 305. Bioconcentration - flow through fish test.

OECD, 2006a. OECD Guidelines for the testing of chemicals. Section 2 Effects on biotic systems: 201. Freshwater alga and cyanobacteria, growth inhibition test.

OECD, 2006b. OECD Guidelines for the testing of chemicals. Section 2 Effects on biotic systems: 221. *Lemna* sp. growth inhibition test.

Posthuma, L., G.W.I. Suter and T.P. Traas (eds), 2002. Species-Sensitivity Distributions in Ecotoxicology. CRC Press, Lewis, Boca Raton, FL, 587 pp.

Posthuma, L., D. de Zwart, K.R. Solomon and T.C.M. Brock, 2008. Guidance on the application of extrapolation methods in ecological exposure and effects characterization of chemicals. In: Solomon, K.R., T.C.M. Brock, D. de Zwart, S.D. Dyer, L. Posthuma, S.M. Richards, H. Sanderson, P.K. Sibley and P.J. van den Brink (eds), 2008 Extrapolation Practice for Ecotoxicological Effect Characterization of Chemicals, SETAC Press & CRC Press, Boca Raton, Fl, USA, pp. 281-322.

Posthuma-Doodeman, C.J.A.M., 2008. Environmental risk limits for imidacloprid. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716018, 25 p.

Prato, E., C. Scardicchio and F. Biandolino, 2006. Effects of temperature on the acute toxicity of cadmium to Corophium Insidiosum. *Environ Monit Assess* 136: 161-166.

Pratt, J.R., N.J. Bowers, B.R. Niederlehner and J. Cairns Jr., 1988. Effects of atrazine on freshwater microbial communities. *Arch Environ Contam Toxicol* 17: 449-457.

Pusey, B.J., A.H. Arthington and J. McClean, 1994. The effects of a pulsed application of chlorpyrifos on macroinvertebrate communities in an outdoor artificial stream system. *Ecotoxicol. Environ. Saf.*, 27: 221 - 250.

Roessink, I., G.H.P. Arts, J.D.M. Belgers, E. Bransen, S.J. Maund and T.C.M. Brock, 2005. Effects of lambda-cyhalothrin in two ditch microcosm systems of different trophic status. *Environ Toxicol Chem* 24: 1684-1696.

Roex, E.W.M., C.A.M. van Gestel, A.P. van Wezel en N.M. van Straalen, 2000. Ratios between acute aquatic toxicity and effects on population growth rates in relation to toxicant mode of action. *Environ Toxicol Chem* 19:685-693.

Romijn, C.A.F.M., R. Luttik, D. van der Meent, W. Slooff and J.H. Canton, 1993. Presentation of a general algorithm to include effect assessment on secondary poisoning in the derivation of environmental quality criteria. Part I. Aquatic food chains. *Ecotox Environ Saf* 26: 61-85.

Romijn, C.A.F.M., R. Luttik, W. Slooff and J.H. Canton, 1994. Presentation of a general algorithm to include effect assessment on secondary poisoning in the derivation of environmental quality criteria. II. Terrestrial food chains. *Ecotox Environ Saf* 27: 107-127.

Rozman, K.K. and J. Doull, 2000. Dose and time as variables of toxicity. *Toxicology* 144:169-178.

Rubach, M., 2010. Predicting the response of aquatic invertebrates to stress using species traits and stressor mode of action. PhD Thesis, Wageningen University, 194 pp.

Rubach, M.N., R. Ashauer, S.J. Maund, D.J. Baird and P.J. van den Brink, 2010. Toxicokinetic variation in 15 freshwater arthropod species exposed to the insecticide chlorpyrifos. *Environ Toxicol Chem* 29: 2225-2234.

Rijkswaterstaat, Ministerie van Verkeer en Waterstaat, Instructie. Richtlijn Monitoring Oppervlaktewateren en Protocol Toetsen & Beoordelen. 30 maart 2009.

SANCO, 2002. Guidance document on aquatic ecotoxicology in the context of the Directive 91/414/EEC. European Commission, Health & Consumer Protection Directorate-General, SANCO/3268/2001 rev, 4 (final), Brussels (BE).

Scheepmaker, J.W.A., 2008a. Environmental risk limits for abamectin. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716003, 32 p.

Scheepmaker, J.W.A., 2008b. Environmental risk limits for teflubenzuron. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716023, 33 p.

Scheffer, M., 2009. Critical Transitions in Nature and Society. Princeton University Press, Princeton, 384 p.

Schmolke, A., P. Thorbek, P. Chapman and V. Grimm, 2010. Ecological models and pesticide risk assessment: current modelling practice. *Environ Toxicol Chem* 29:1006-1012.

Schulz, R. and M. Liess, 1999. A field study of the effects of agriculturally derived insecticide input onstream macroinvertebrate dynamics. *Aquat Toxicol* 46:155-176.

Schulz, R., G. Thiere and J.M. Dabrowski, 2002. A combined microcosm and field approach to evaluate the aquatic toxicity of azinphosmethyl to stream invertebrates. *Environ Toxicol Chem* 21: 2172-2178.

Seguin, F., C. Leboulanger, F. Rimet, J.C. Druart and A. Berard, 2001. Effects of atrazine and nicosulfuron on phytoplankton in systems of increasing complexity. *Arch Environ Contam Toxicol* 40: 198-208.

Siefert, R.E., S.J. Lozano, J.C. Brazner and M.L. Knuth, 1989. Littoral enclosures for aquatic field testing of pesticides: effects of chlorpyrifos on a natural system. *Miscellaneous Publication Series Entomological Society of America,* 75: 57-73.

Slijkerman, D.M.E., D.J. Baird, A. Conrad, R.G. Jak and N.M. van Straalen, 2004. Assessing structural and functional plankton responses to carbendazim toxicity. *Environ Toxicol Chem* 23: 455-462.

Smit, C.E., 2005. Energy and moisture content and assimilation efficiency of bird and mammal food. Chapter 4 in: Factsheets for the (eco)toxicological risk assessment strategy of the National Institute for Public Health and the Environment - Part V (J.W.A. Scheepmaker, C.E. Smit and M.T.M. van Raaij, eds.). Bilthoven, The Netherlands: RIVM. Report 601516013.

Smit, C.E. and Z.C. Dang, 2008. Environmental risk limits for carbendazim. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716014, 33 p.

Smit, C.E. and M. van der Veen, 2008. Environmental risk limits for dodine. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716007, 30 p.

Smit, C.E. and J.U.W. Vonk, 2008. Environmental risk limits for fenoxycarb. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716008, 36 p.

Smit, C.E., 2009. Maximum Acceptable Concentrations within the context of the Water Framework Directive. An initial evaluation. Bilthoven, The Netherlands: RIVM. Letter Report 601714012, 39 p.

Solomon, K.R., J.M. Giddings and S.J. Maund, 2001. Probabilistic risk assessment of cotton pyrethroids: I. Distributional analyses of laboratory aquatic toxicity data. *Environ Toxicol Chem* 20 (3): 652-659.

Solomon, K.R., T.C.M. Brock, D. de Zwart, S.D. Dyer, L. Posthuma, S.M. Richards, H. Sanderson, P.K. Sibley and P.J. van den Brink (Eds), 2008. Extrapolation Practice for Ecotoxicological Effect Characterization of Chemicals, SETAC Press & CRC Press, Boca Raton, FI, USA, 380 pp.

Spromberg, J.A., B.M. John en W.G. Landis, 1998. Metapopulation dynamics: Indirect effects and multiple distinct outcomes in ecological risk assessment. *Environ Toxicol Chem* 17(8):1640-1649.

Stay, F.S., A. Katko, C.M. Rohm, M.A. Fix and D.P. Larsen, 1989. The effects of atrazine on microcosms developed from four natural plankton communities. *Arch Environ Contam Toxicol* 18: 866-875.

Suter, G.W., T.P. Traas and L. Posthuma, 2002. Issues and practises in the derivation and use of species sensitivity distributions. In: Posthuma L, Traas TP, Suter GW (eds), Species Sensitivity Distributions in Risk Assessment. Boca Raton (FL), CRC Press, pp. 473-474.

Tanner, D.K. and M.L. Knuth, 1995. Effects of Azinphos-methyl on the reproductive success of the Bluegill Sunfish, *Lepomis macrochirus*, in littoral enclosures. *Ecotoxicol Environ Saf*, 32: 184 - 193.

Thorbek, P., V.E. Forbes, F. Heimbach, U. Hommen, H.H. Thulke, P.J. van den Brink. J. Wogram and V. Grimm (Eds). 2010. Ecological models for regulatory risk assessments of pesticides. Developing a strategy for the future. Pensacola/Boca Raton USA: SETAC Press/CRC Press.

Tiktak, A., P.I. Adriaanse, J.J.T.I. Boesten, J. Delsman, C. van Griethuysen, M.M.S. ter Horst, B.H.J. Linders, A.M.A. van der Linden and J.C. van de Zande, 2012. Scenarios for exposure of water organisms in the Netherlands. RIVM Report 607407002/2011

Traas, T.P., R. Luttik and R.H. Jongbloed, 1996. A probabilistic model for deriving soil quality criteria based on secondary poisoning of top predators. I. Model description and uncertainty analysis. *Ecotox Environ Saf* 34: 264-278.

Traas, T.P., J.H. Janse, P.J. van den Brink, T.C.M. Brock and T. Aldenberg, 2004. A freshwater food web model for the combined effects of nutrients and insecticide stress and subsequent recovery. *Environ Toxicol Chem* 23(2): 521-529.

Van den Brink, P.J., E. van Donk, R. Gylstra, S.J.H. Crum and T.C.M. Brock, 1995. Effects of chronic low concentrations of the pesticides chlorpyrifos and atrazine in indoor freshwater microcosms. *Chemosphere* 31: 3181-3200

Van den Brink, P.J., R.P.A. van Wijngaarden, W.G.H. Lucassen, T.C.M. Brock and P. Leeuwangh, 1996. Effects of the insecticide Durban 4®E (a.i. chlorpyrifos) in outdoor experimental ditches: II. Community responses and recovery. *Environ Toxicol Chem* 15: 1143 - 1153.

Van den Brink, P.J., Ter Braak CJF.1999. Principal response curves: Analysis of time-dependent multivariate responses of biological communities to stress. *Environ Toxicol Chem* 18:138-148.

Van den Brink, P.J., J. Hattink, F. Bransen, E. van Donk and T.C.M. Brock, 2000. Impact of the fungicide carbendazim in freshwater microcosms. II. Zooplankton, primary producers and final conclusions. *Aquatic Toxicology* 48:251-264.

Van den Brink, P.J., J. Roelsma, E.H. van Nes, M. Scheffer and T.C.M. Brock,2002a. Perpest model, a case-based reasoning approach to predict ecological risks of pesticides. *Environ Toxicol Chem* 21(11): 2500-2506.

Van den Brink, P.J., E.M. Hartgers, R. Gylstra, F. Bransen and T.C.M. Brock, 2002b. The effects of a mixture of two insecticides on freshwater microcosms. II. Water quality, responses of zooplankton, phytoplankton and periphyton and ecological risk assessment. *Ecotoxicology* 11, 181-197.

Van den Brink, P.J., N. Blake, T.C.M. Brock and L. Maltby, 2006. Predictive value of species sensitivity distributions for effects of herbicides in freshwater ecosystems *Human Ecol Risk Assess* 12, 645-674.

Van den Brink, P.J., J.M. Baveco, J. Verboom and F. Heimbach, 2007. An individual-based approach to model spatial population dynamics of invertebrates in aquatic ecosystems after pesticide contamination. *Environ Toxicol Chem* 26(10):2226-2236.

Van den Brink, P.J., P.K. Sibley, H.T. Ratte, D.J. Baird, J.V. Nabholz and H. Sanderson, 2008. Extrapolation of effect measures across levels of biological organization in ecological risk assessment. In: Solomon, K.R., T.C.M. Brock, D. de Zwart, S.D. Dye, L. Posthuma, S.M. Richards, H. Sanderson, D. Sibley and P.P.K. van den Brink (Eds.), Extrapolation Practice for Ecotoxicological Effect Characterization of Chemicals, SETAC Press & CRC Press, Boca Raton, Fl, USA, pp. 105-133.

Van den Brink, P.J., S.J.H. Crum, R. Gylstra, F. Bransen, J.G.M. Cuppen and T.C.M. Brock, 2009. Effects of a herbicide-insecticide mixture in freshwater microcosms: Risk assessment and ecological effect chain. *Environmental Pollution* 157:237-249.

Van Leeuwen, L.C., J.H. Vos and B.J.W.G. Mensink, 2008. Environmental risk limits for lambda-cyhalothrin. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716001, 43 p.

Van Leeuwen, L.C. and J.W. Vonk. 2008. Environmental risk limits for kresoxim-methyl. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716019, 26 p.

Van Vlaardingen, P.L.A., T.P. Traas, A.M. Wintersen and T. Aldenberg, ETX 2.0. 2004. A programme to calculate Hazardous Concentrations and Fraction Affected, based on normally distributed toxicity data. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). Report 601501028. www.rivm.nl/bibliotheek/rapporten/601501028.pdf. Software available via www.rivm.nl/rvs/risbeoor/Modellen/ETX.jsp

Van Vlaardingen, P.L.A. and E.M.J. Verbruggen, 2007. Guidance for the derivation of environmental risk limits within the framework of 'International and national environmental quality standards for substances in the Netherlands' (INS). Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Report 601782001/2007.

Van Vlaardingen, P.L.A., J.W. Vonk and F.M.W. de Jong, 2008. Environmental risk limits for esfenvalerate. Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report no. 601716017, 42 p.

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

Van Wijngaarden, R.P.A., T.C.M. Brock and M.T. Douglas, 2005. Effects of chlorpyrifos in freshwater model ecosystems: the influence of experimental conditions on ecotoxicological thresholds. *Pest Management Science* 61: 923 - 935.

Van Wijngaarden, R.P.A., T.C.M. Brock, P.J. van den Brink, R. Gylstra and S.J. Maund, 2006. Ecological effects of spring and late summer applications of lambda-cyhalothrin in freshwater microcosms. *Arch Environ Contam Toxicol* 50: 220 - 239.

Van Wijngaarden, R.P.A., G.H.P. Arts, J.D.M. Belgers, H. Boonstra, I. Roessink, A.F.W. Schroer and T.C.M. Brock, 2010. The Species Sensitivity Distribution approach compared to a microcosm study: A case study with the fungicide fluazinam. *Ecotoxicol Environ Saf* 73:109-122.

Verbruggen, E.M.J. and P.J. van den Brink, 2010. Review of recent literature concerning mixture toxicity of pesticides. RIVM report 601400xxx/2010.

Versteeg, D.J., S.E. Belanger and G.J. Carr, 1999. Understanding single-species and model ecosystem sensitivity: data-based comparison. *Environ Toxicol Chem* 18:1329-1346.

Webber, E.C., W.G. Deutch, D.R. Bayne and W.C. Seesock, 1992. Ecosystem-level testing of synthetic pyrethroid insecticide in aquatic mesocosms. *Environ Toxicol Chem*, 11: 87-105.

Weston, D.P., J. You, E.L. Amweg and M.J. Lydy, 2008. Sediment toxicity in agricultural areas of California and the role of hydrophobic pesticides. In: Gan, J., F. Spurlock, P. Hendley and D. Weston (Eds). Synthetic Pyrethroids: Occurrence and Behavior in Aquatic Environments. ACS Book Series 991. American Chemical Society, Washington DC, pp. 26-54.

Wheeler, J.R., K.M.Y. Leung, D. Morritt, N. Sorokin, H. Rogers, R. Toy, M. Holt, P. Whitehouse and M. Crane, 2002. Freshwater to saltwater toxicity extrapolation using species sensitivity distributions. *Environ Toxicol Chem* 21 (11): 2459-2467.

Willis, K.J., P.J. van den Brink and J.G. Green, 2004. Seasonal variation in plankton community responses of mesocosms dosed with pentachlorophenol. *Ecotoxicology* 13:707-720.

Wogram, J., 2010. Regulatory challenges for the potential use of ecological models in risk assessment of plant protection products. Chapter 4 in: Thorbek, P., V.E. Forbes, F. Heimbach, U. Hommen, H.H. Thulke, P.J. van den Brink, J. Wogram and V. Grimm (Eds.), Ecological models for regulatory risk assessments of pesticides. Developing a strategy for the future. Pensacola/Boca Raton USA: SETAC Press/CRC Press.

Appendix 1 Variability in exposure-response relationships between micro-/mesocosm experiments performed with the same PPP

Given the natural variability in the structure and function of freshwater communities, it is reasonable to question the spatio-temporal extrapolation of results of model ecosystem experiments with pesticides. Within the context of the risk assessment for pesticides one of the questions at stake is how unique such test systems are with respect to the concentration-response relationships observed. However, for a few pesticides only, more than two micro-/mesocosm experiments have been performed. In addition, when more than one micro-/mesocosm experiments for a certain pesticide are available, they often vary in exposure regime (single pulse, repeated pulse, chronic exposure). Fortunately, for a few PPPs extensive datasets exists that allow the evaluation of concentration-response relationships caused by similar exposure regimes. The information available for the organophophorous insecticide chlorpyrifos in particular allows the evaluation of effects of a single pulse exposure regime (Table A1-1).

Table A1-1

Effect class concentrations (in µg/L) of the most sensitive measurement endpoint in micro-/mesocosm experiments that studied the impact of single pulse, repeated pulse and chronic exposures of the insecticide chlorpyrifos. The Effect classes are expressed in terms of nominal concentrations. These nominal concentrations generally were within 20% of the exposure concentrations on basis of measurements in the application solutions or in the water column of the test systems.

Exposure Regime	Effect class 1	Effect class 2	Effect class 3A	Effect class 4-5	Type of test system	Reference; Country
Single pulse (peak)	0.1	0.3	1.0	3.0	Outdoor lentic microcosm	Biever et al., 1994; USA
Single pulse (peak)	0.1	-	-	0.9	Outdoor lentic mesocosms	Van den Brink et al., 1996; NL
Single pulse (peak)	0.1	-	-	1.0	Outdoor lentic mesocosms	Lopez-Mancisidor et al., 2007; Spain
Single pulse (peak)	0.1	-	-	1.0	Outdoor lentic mesocosm	Daam et al., 2008; Thailand
Single pulse (peak)	0.1	-	(5*)	-	Outdoor lotic mesocosm	Pusey et al., 1994; Australia
Single pulse (peak)	-	-	0.5	6.3	Outdoor lentic mesocosm	Siefert et al., 1989; USA
Single pulse (peak)	0.1	-	1.0	10	Indoor lentic cosm; 16 °C, mesotrophic	Van Wijngaarden et al., 2005; NL
Single pulse (peak)	0.1	-	1.0	-	Indoor lentic cosm; 26 °C, mesotrophic	Van Wijngaarden et al., 2005; NL
Single pulse (peak)	0.1	-	-	1.0	Indoor lentic cosm; 26 °C, eutrophic	Van Wijngaarden et al., 2005; NL
Repeated pulse (4x)	0.033		0.1	1	Outdoor lentic mesocosms	Lopez-Mancisidor et al., 2008; Spain
Constant chronic (28 d)	-	-	-	0.1	Indoor lentic microcosm	Van den Brink et al., 1995
Constant chronic (28 d)	-	0.01**	-	0.1**	Indoor lentic microcosm	Cuppen et al., 2002

* Recovery is fast because of constant input of propagules in experimental stream after pulse exposure.

** Exposure to a mixture of chlorpyrifos and lindane; all treatment-related effects were assigned to chlorpyrofos.

It appears from the model ecosystem experiments performed with the non-persistent insecticide chlorpyrifos that the range in exposure concentrations resulting in 'Effect class 1' responses is remarkably small for the single pulsed exposure regime. For eight aquatic micro-/mesocosm experiments, performed in different parts of the world and/or under different experimental conditions, an Effect class 1 response was observed at a peak concentration of 0.1 µg chlorpyrifos/L. Note that this is partly due to the fact that similar exposure concentrations were selected by the different experimenters. In addition, the similarity between Effect class 1 responses between different studies can be explained by the fact that both crustaceans and insects are sensitive to this insecticide and that the communities of the micro-/mesocosm test systems used all contained a reasonably high diversity of these arthropods, while also several arthropod populations belonging to zooplankton or macro-invertebrates occurred in high enough densities.

It appears from the data presented in Table A1-1 that for a single pulse exposure regime of chlorpyrifos information on Effect class 2 concentrations for a single pulse exposure is scarce (only one study available). More information is available for Effect class 3A concentrations due to single pulse exposures, and it appears that differences in these concentrations between studies are relatively large. Note, however, that from a regulatory point of view it is fair to make a distinction in recovery of sensitive arthropods between hydrologically isolated test systems (lentic micro-/mesocosms: Effect class 3A concentrations $\leq$1.0 µg/L) and the outdoor stream in which a more or less constant inflow of sensitive stream invertebrates was possible (resulting in an Effect class 3A concentration of 5 µg/L). It also appears from the chlorpyrifos data presented in Table A1-1 that the threshold concentration (Effect class 1) of the repeated (4x) pulse exposure study is a factor of approximately 3 lower than that of the single exposure studies. Treatment-related effects due to a constant chronic exposure probably occur at concentrations equal to higher than 0.01 µg chlorpyrifos/L.

For the pyrethroid insecticide lambda-cyhalothrin the majority of micro-/mesocosm experiments available concerns repeated application studies (Table A1-2). Again it appears that the variability in Effect class 1 (n=2) and Effect class 2 (n=4) responses between different studies is remarkably low, while that for Effect class 3A (n=3) responses is somewhat higher.

In accordance with the data for chlorpyrifos and lambda-cyhalothrin described above more or less similar Effect class 1 - 2 and Effect class 3A concentrations were observed for different model ecosystems treated once or repeatedly with the insecticides azinphos-methyl and esfenvalerate (Table A1-3).

Furthermore, microcosm and mesocosm experiments studying the impact of a single application of the relatively persistent herbicide simazine on biomass and densities of primary producers indicate that Effect class 1 - 2 concentrations (based on peak exposure) varied a factor of 2 only (Table A1-4). In addition, lake enclosure studies exploring effects of a single application of pentachlorophenol to plankton communities in spring, summer, autumn and winter indicated that threshold levels for effects (Effect class 1 concentrations based on peak exposure) varies by approximately a factor of 2 (Willis et al., 2004).

Table A1-2

Effect class concentrations (in ng/L) of the most sensitive measurement endpoint in micro-/mesocosm experiments that studied the impact of pulsed exposures of the insecticide lambda-cyhalothrin. The Effect classes are expressed in terms of nominal peak concentrations. In most studies the nominal concentrations were in accordance with measurements of the test substance in the application solutions.

Exposure Regime	Effect class 1	Effect class 2	Effect class 3A	Effect class 4-5	Type of test system	Reference; Country
Single pulse	-	-	50	-	Outdoor lotic mesocosms	Heckmann and Friberg, 2005; Denmark
Repeated pulse (12x)	2.7*	-	-	27*	Outdoor lentic mesocosms	Hill et al., 1994b; USA
Repeated pulse (2x)	4.0**	-	16**	85**	Outdoor lentic mesocosms	Arts et al., 2006; NL
Repeated pulse (5x)	-	10**	-	25**	Indoor lentic microcosms	Van Wijngaarden et al., 2004; NL
Repeated pulse (3x)	-	10	-	25	Outdoor lentic microcosm	Roessink et al., 2005; NL
Repeated pulse (3x)	-	10	50	-	Outdoor lentic microcosm	Roessink et al., 2005; NL
Repeated pulse (3x)	-	10	25	50	Outdoor lentic microcosms	Van Wijngaarden et al., 2006; NL
Repeated pulse (3x)	-	-	-	17	Outdoor lentic mesocosms	Farmer et al., 1995; UK

* Experiment was characterized by both spray-drift (nominal 1.6 µg/L) and run-off applications (nominal 4.7 µg/L). As exposure concentration the median value for the spray drift and run-off application was used.

** Exposure to a realistic package of different pesticides used in a specific crop including lambda-cyhalothrin; all treatment-related effects were assigned to lambda-cyhalothri.

Table A1-3

Effect class concentrations (in µg/L) of the most sensitive measurement endpoint in micro-/mesocosm experiments that studied the impact of (short-term) pulsed exposures of the insecticides azinphos-methyl and esfenvalerate. The Effect classes are expressed in terms of nominal peak concentrations. In most studies the nominal concentrations were in accordance with measurements of the test substance in the application solutions.

Exposure Regime	Effect class 1	Effect class 2	Effect class 3A	Effect class 3B-5	Type of test system	Reference; Country
Single pulse **Azinphos-methyl**	-	0.2	1.0 - 4.0	20	Littoral enclosures	Knuth et al., 1992; USA
Singe pulse **Azinphos-methyl**	-	-	1.0 - 4.0		Littoral enclosures	Tanner en Knuth, 1995; USA
Repeated pulse (8x) **Azinphos-methyl**	0.22	-	-	0.95	Outdoor mesocosms	Giddings et al., 1994; USA
Repeated pulse (2x) **Esfenvalerate**	-	0.01	-	0.08	Littoral enclosures	Lozano et al., 1992; USA
Repeated pulse (10x) **Esfenvalerate**	0.01	-	-	0.25	Outdoor mesocosms	Webber et al., 1992; USA

Table A1-4

Effect class concentrations (in µg/L) of the most sensitive structural measurement endpoint in micro-/mesocosm experiments (fish not present) that studied a single application of the herbicide simazine (field dissipation DT50 in water approximately 20 days).

Exposure Regime	Effect class 1	Effect class 2	Effect class 3A	Effect class 4	Type of test system	Reference; Country
Single application	-	100	-	1000	Experimental swamp	Goldsborough and Robinson, 1983, 1986; Canada
Single application	100	-	-	500	Pond microcosms	Jenkins and Buikema, 1990; USA
Single application	50	-	150	-	Laboratory microcosms	Bryfogle and McDiffett, 1979

In general there is a lack of different micro-/mesocosm data dealing with the effects of long-term, more or less constant, exposure regimes of the same pesticide. Only for the persistent herbicide atrazine a large data set is available (Table A1-5).

Table A1-5

Effect class concentrations (in µg/L) of the most sensitive measurement endpoint in micro-/mesocosm experiments that studied the impact of more or less constant long-term exposure of the herbicide atrazine.

Exposure Regime	Effect class 1	Effect class 2	Effect class 3B	Effect class 4-5	Type of test system	Reference
Long-term	-	2	-	30	Outdoor lentic mesocosms	Seguin et al., 2001
Long-term	5	-	-	-	Indoor lentic microcosms	Van den Brink et al., 1995
Long-term	-	5	-	-	Indoor lotic microcosms	Gruessner and Watzin, 1996
Long-term	5	10	-	22	Outdoor lentic microcosms	Jüttner et al., 1995
Long-term	-	10	-	100	Indoor lentic microcosm	Johnson, 1986
Long-term	5	-	50	100	Indoor lentic microcosm	Brockway et al., 1984
Long-term	10	-	-	32	Indoor lentic microcosms	Pratt et al., 1988
Long-term	-	-	-	10	Indoor lotic microcosms	Kosinsky, 1984, Kosinsky and Merkle, 1984
Long-term	14	25	-	80	Indoor lotic microcosms	Nyström et al., 2000
Long-term	-	-	-	14	Indoor lotic microcosms	Muňos et al., 2001
Long-term	-	-	-	15	Experimental swamp	Detenbeck et al., 1996
Long-term	-	-	-	20	Outdoor lentic mesocosms	DeNoyelles et al., 1994 (and literature cited)
Long-term	-	20	-	100	Indoor lentic microcosms	Stay et al., 1989
Long-term	-	-	-	24	Indoor lotic microcosms	Krieger et al., 1988
Long-term	-	-	-	50	Outdoor lentic mesocosms	Fairchild et al., 1994

Data available for atrazine suggest a larger variability in class 1 and class 2 effect concentrations between chronic exposure experiments; however, also a larger number of studies is available. Effect class 1 concentrations could be derived from five different atrazine studies, and Effect class 2 concentrations for six studies (Table A1-5).

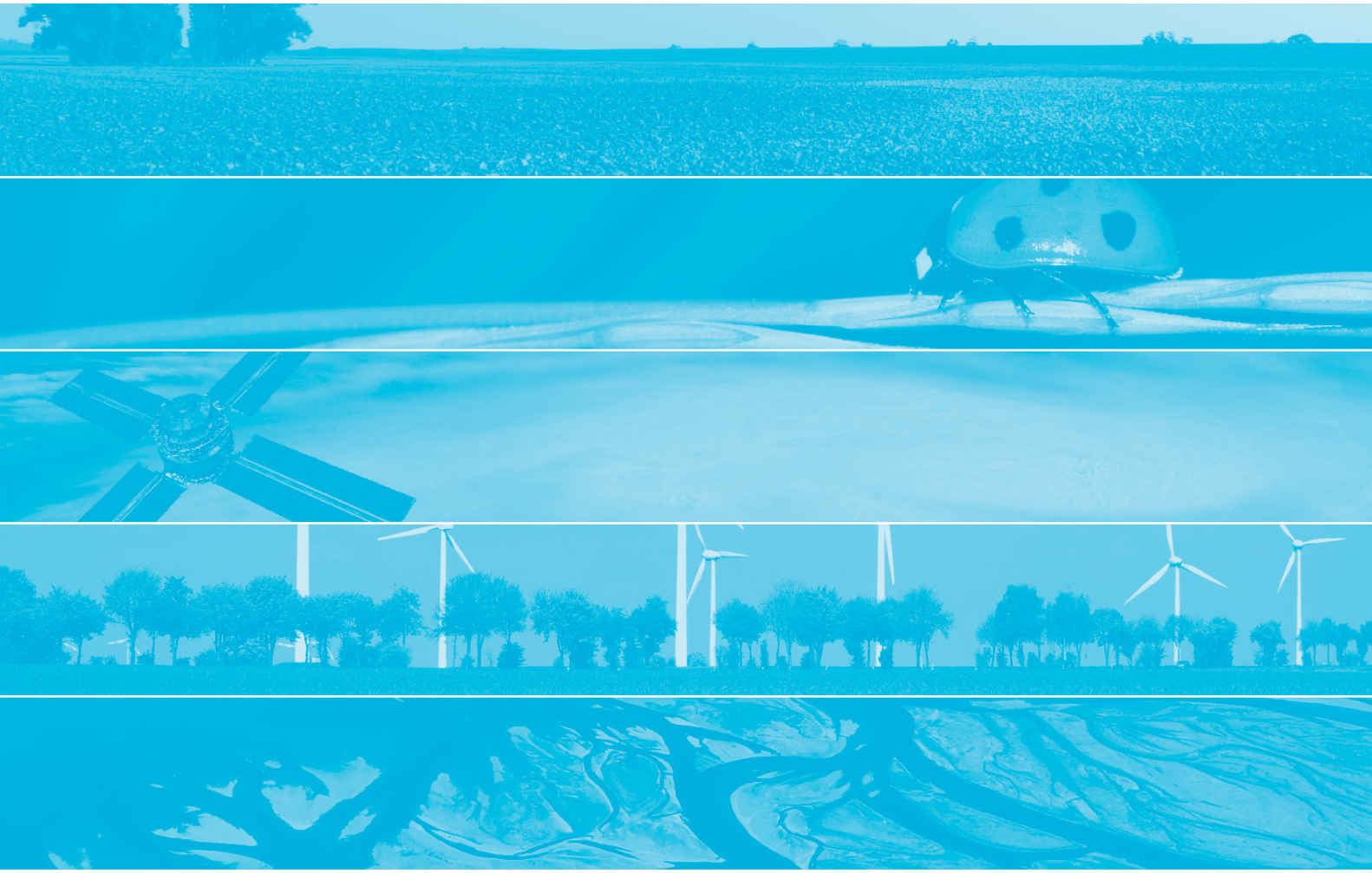
The relatively high variability in Effect class 1 - 2 concentrations for chronic studies with atrazine when compared with those with pulsed exposures to chlorpyrifos and lambda-cyhalothrin might be explained by differences in toxic mode-of-action between these substances. Atrazine is a photosystem II inhibitor. According to Guasch and Sabater (1998), inhibition of photosynthesis by atrazine is influenced by ambient light conditions, which most probably considerably varied between the different micro-/mesocosm studies reported in Table A1-5. Note that the difference in Effect class 1 concentrations observed for the photosystem II inhibitor simazine (Table A1-4) might also be caused by different light conditions between the indoor and outdoor test systems. Consequently, a question at stake is whether the results from the chronic micro-/ mesocosm studies with atrazine are representative for pesticides with another toxic mode-of-action.

There appears to be limited other information on other pesticide-treated model ecosystems comparing Effect class 1 or Effect class 2 concentrations for direct toxic effects as a result of more or less constant chronic exposure. The limited microcosm/mesocosm information available for the persistent fungicide carbendazim suggests little variation in Effect class 1 concentrations between experiments as a result of a long-term chronic exposure regime (Table A1-6).

Table A1-6

Effect class concentrations (in µg/L) of the most sensitive measurement endpoint in micro-/mesocosm experiments (fish not present) that studied the impact of more or less constant exposure of the fungicide carbendazim.

Exposure Regime	Effect class 1	Effect class 2	Effect class 3	Effect class 4	Type of test system	Reference; Country
Long term	2.6	-	-	26.4	Outdoor microcosms	Daam et al., 2009a; Thailand
Long term	2.2	-	-	20.7	Outdoor mesocosms	Slijkerman et al., 2004; Netherlands
Long term	3.3	-	-	33.0	Indoor microcosms	Cuppen et al., 2000, Van den Brink et al., 2000; Netherlands



Alterra is part of the international expertise organisation Wageningen UR (University & Research centre). Our mission is 'To explore the potential of nature to improve the quality of life'. Within Wageningen UR, nine research institutes – both specialised and applied – have joined forces with Wageningen University and Van Hall Larenstein University of Applied Sciences to help answer the most important questions in the domain of healthy food and living environment. With approximately 40 locations (in the Netherlands, Brazil and China), 6,500 members of staff and 10,000 students, Wageningen UR is one of the leading organisations in its domain worldwide. The integral approach to problems and the cooperation between the exact sciences and the technological and social disciplines are at the heart of the Wageningen Approach.

Alterra is the research institute for our green living environment. We offer a combination of practical and scientific research in a multitude of disciplines related to the green world around us and the sustainable use of our living environment, such as flora and fauna, soil, water, the environment, geo-information and remote sensing, landscape and spatial planning, man and society.

More information: www.alterra.wur.nl/uk