Working document on Risk Assessment of Plant Protection Products in the Central Zone

Ecotoxicology

Version 2.0, August 2023

Editing log - Working document on Risk Assessment of Plant Protection Products in the Central Zone

Date	Revision	Issues	Responsible	Implementation date
05-02-2021	Draft	Discussion draft Dec 2020 in CZSC (05-02-2021) after commenting round CircaBC forum July 2020	Ctgb	At publication date of version 1.0
10-05-2021	1.0	Finalisation after discussion in CZSC (editorial comments GE (UBA), comments SL and comments AT).	Ctgb	25th May 2021
August 2023	2.0	Update with: - General bullet Ecotox in the core and nat add (adopted by the CZSC 2020) - bullet points from the CZHW Nov 2019 (adopted by the CZSC April 2022) - bullet point availability aquatoxtool (and NTTP) (adopted by the CZSC March 2021) - bullet points from the CZHW June 2022 (adopted by the CZSC May/June 2023)	Ctgb	1st of September 2023

Table of contents

1. Legal status

This document does not intend to produce legally binding effects and by its nature does neither prejudice any measure taken by a Member State/country within the Regulation (EC) No 1107/2009 or previous implementation prerogatives under Annex II, III and VI of Council Directive 91/414/EEC, nor prejudice any case law developed with regard to these provisions. This document also does not preclude the possibility that the European Court of Justice may give one or another provision direct effect in Member States.

2. Introduction

This document describes procedures and evaluation criteria for the ecotoxicological assessment of applications for authorisation, re-authorisation and amendments of plant protection products following approval of an active substance under Regulation (EC) No 1107/2009 in the Central zone and thereof an inclusion in regulation (EU) No 540/2011.

The Central Zone Working document has been agreed by the responsible competent authorities in Austria, Belgium, Czech Republic, Germany, Hungary, The Netherlands, Poland, Ireland, Romania, Slovakia and Slovenia.

Disclaimers:

This guidance is solely intended for assembling a core assessment document and does not fully cover the various national requirements for risk assessments.

3. Ecotoxicology

The present guidance for the ecotoxicological risk assessment regarding applications for approval of plant protection products in the Central Zone highlights parts for which MS in the Central Zone wants a better clarification or deviates from available EU and EFSA Guidance Documents.

In principle only points agreed by the Central Zone Steering Committee (CZSC) are included in this document (bullet points). However, there are several points which give valuable information and which were only taken note by the CZSC, but not officially agreed in a bullet point. These points are also included.

Furthermore relevant points for the zonal assessments from the EFSA Pesticides Peer Review Meetings on general recurring issues in ecotoxicology, which are not already part of the bullet points, are included.

3.1 General issues ecotoxicology

3.1.1 General aspects

3.1.1.1 General aspects on the ecotoxicological assessment in the core dRR and national addenda (CZSC, June 2020)

- All MS agree that the full ecotoxicological assessment, including all parts that may be relevant for the concerned Member States (cMS) must be done in the core dRR.
- More specifically, the summary of all submitted studies (also when aimed at national circumstances), the evaluation of higher tier refinements and of risk mitigation measures must be included in the core.
- National specific elements such as national exposure models or considerations about relevance of species in the specific Member State may be included in the national addenda.
- Any national addenda evaluated by the zonal rapporteur (zRMS) should be made available to the cMS.

3.1.1.2 Decision-making (CZHW Brno, November 2019; CZSC, April 2022)

To further facilitate the harmonisation of approaches for the aspect ecotoxicology in the Core assessment, a „majority decision“ (or “majority of MS”) is considered if not more than 1/3 of MSs disagree. The CZSC agreed that the harmonised approach of majority decisions must be used in the core assessment.

3.1.1.3 Data-Matching (CZHW, Ede (NL), June 2022; CZSC May/June 2023)

Products: If applicants wish to use data generated with a different product, the question to be answered is whether the products are similar or not. Thus the zRMS should check if the applicant has shown that formulations are sufficiently similar for

the data from the lead formulation to be considered relevant for the product in question.

If applicants provide alternative studies with similar tests/same species the “rule of 3” (i.e., endpoints which differ by >3x indicate significantly different (more critical) toxicity and the studies should therefore be considered more critcal) should be applied unless a test shows another species to be more sensitive.

3.1.2 General ecotoxicological aspects

3.1.2.1 Mixture toxicity

The following points related to combination or formulation toxicity have been discussed in harmonisation workshops and agreed upon by the Central Zone Steering Committee (CZSC). Note that some of the issues are not just ecotoxicology related, while others are specific for a certain area of the risk assessment. Below, the entire list from the CZSC is presented:

CZSC March 2014: National addendum - safeners
- The assessment of safeners is by most MS addressed in the national addendum until data-requirements are set; after that moment the assessment should be included in the core dossier. For work sharing purposes, DE will always include data on safeners in the core dossier.

CZSC January and April 2016:
- Long-term combitox for birds and mammals should be assessed for applications submitted from 1st of June 2016:
− In the (draft) Registration Report, a calculation of the long-term combitox risk according to the concentration addition (CA) model should be presented for tier 1.
− Refinement options and possible consequences are not clear yet, however:
− when the CA combitox assessment indicates no acceptable risk, applicants may present information to demonstrate that adverse effects of the actives are not similar.
- Industry will be asked to cover combitox assessment (birds and mammals, aquatic) in DRR for Article 43 applications.

CZSC May 2016: combitox and art. 43 applications: for PPP containing 2 or more active substances: where the renewal of the second active substance is more than 12 month apart from the renewal of the first one, the applications are to be dealt with separately.
The full combitox for all active substances in PPP should be addressed by the applicant in the core dRR (OPEX (Operator Exposure), chronic birds and mammals, aquatic). There was no full agreement among member states and there will be differences between member states in the approach to combitox. Therefore, when combitox was not assessed by the ZRMS, combitox will be assessed by the individual MS in the corresponding national addenda. Applicants are advised to go to particular MS to be informed about their individual national approaches. Please note, the combitox assessment for birds and mammals (chronic) is nevertheless to be considered for applications by 1st of June 2016.

CZSC November 2017: Regarding the assessment of ecotoxicology in connection with Article 43, agreement has thus far been reached on the following points (please also refer to "2016-07 Bullet points CZSC May 2016"):
− As agreed in May 2016, the full combitox for all active substances in PPP should be addressed by the applicant in the core dRR.
− If the assessment is performed at renewal of the first a.s., new endpoints for the first a.s. and old endpoints for the others are applied.
− For the Tier 1 combitox assessment, MS rely on the respective guidance documents (and where applicable also on already existing agreements at zonal level).
− For higher Tier refinements, there are various approaches by the MS, most of whom would rely on a WoE approach if no agreed methods/ guidance are available; some MS would exhaust single a.s. refinements as a first step for the refined combitox assessment.

Long-term combitox for birds and mammals should also be assessed for Article 29/33 applications (please refer to "2016-05 Bullet points CZSC January-April 2016").

3.1.2.2 How to consider the formulation within the evaluation of the active substance (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

When a PPP appears to be more toxic, i.e. its toxicity endpoint is three times lower than the equivalent endpoint of the active substance, according to the data requirement the lower endpoint should be used for the risk assessment or risk assessments for both the active substance and PPP could be provided.

Background information

The purpose of this discussion point was to achieve a better understanding and enhance the harmonisation between Member States on how to consider the toxicity of the formulation relative to the toxicity of the active substance and how to deal with the risk assessment of the PPP within the peer review of the active substances. The discussion concerned those situations in which some data on both the active substance and formulation are available in the EU dossier (usually only for acute toxicity). In particular, EFSA proposed for discussion two main points for the different groups of non-target organisms:

1. In which situations should a formulation be considered as being more toxic than the substance under assessment?
2. What is the best approach to take when a formulation is more toxic and a comprehensive risk assessment has not been performed?

In relation to 'when a formulation should be considered more toxic than the active substance', the proposal was to account for a difference of a factor of three, as recommended in the guidance from the Directorate-General for Health and Food Safety (SANCO/10597/2003 rev. 10.1) (European Commission, 2012) on the equivalence of batches and in the aquatic guidance (EFSA PPR Panel, 2013). This means that when the endpoint of the PPP (expressed in terms of the active substance) is at least three times lower than the equivalent endpoint for the active substance, it should be considered to be more toxic. This factor was agreed by the majority of the experts, to be applied consistently to Tier 1 studies for all groups of non-target organisms.

For birds and mammals, the data on mammals from the mammalian toxicology section should be considered first. If, based on the comparison of data on mammals, it is clear that the formulation is more toxic, it was agreed that the risk assessment should be performed based on the formulation endpoint, expressed in terms of the active substance, as reported in Regulation (EU) 284/2013. However, before asking for further vertebrate studies (e.g. on birds), other elements should be considered, such as the margin of safety in the risk assessment for mammals or factors which may have an impact on the overall toxicity of the formulation (e.g. carriers, dose spacing, method of dosing).

In the case that multiple studies are available that give contradictory information in terms of the comparison of toxicity between active substance and formulation, it was recommended that all the available data should be considered and a decision made on a case-by-case basis; for example, by considering the sensitivity of the tested species.

For aquatic organisms, if the formulation is more toxic than the active substance, the majority of the experts considered that separate risk assessments for the active substance and for the formulation with their respective endpoints could be provided. In the absence of a comprehensive exposure characterisation for the formulation, the predicted environmental concentrations in surface water (PECSW) values generated for the active substance accounting for all the routes of exposure should be used in combination with the formulation endpoint expressed as active substance.

For bees and soil organisms, if the formulation is more toxic than the active substance, the majority of the experts agreed to follow the same approach as described above for the aquatics, i.e. to perform separate risk assessments: one with the active substance and the other with the endpoint for the formulation expressed as active substance.
Some experts expressed the concern that when more than one substance is included in the formulation, the approach of assuming that the toxicity is entirely due to the substance under evaluation may result in a too conservative risk assessment. This is because the entire toxicity of the formulation will be attributed to the substance under evaluation. However, the approach agreed at the meeting is in line with Regulation (EU) 284/2013 and will only be used when an applicant does not provide a comprehensive formulation risk assessment.
There was no discussion on this point for NTAs and non-target terrestrial plants, since only data on formulation are usually available for these organisms. Where data on the active substance and on the formulation are available, a separate risk assessment should be performed as for the other organism groups.

3.1.2.3 Purity of the test item (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

The experts at the meeting agreed that for substances with less than 90 % purity, when the endpoints are expressed in terms of nominal concentrations, these should be corrected for the purity of the technical material. It must be noted that in such situations the tested item is to be considered a mixture. Expressing the endpoint in terms of pure active ingredient content may overestimate the toxicity of the active substance, but it would ensure consistency when the toxicological endpoint is compared with the exposure estimates in the risk assessment.

3.1.2.4 Use of EC10 values in environmental risk assessments (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

In the first general ecotoxicology meeting (Pesticides Peer Review Meeting 133) the evaluation of the reliability of EC10 calculations were discussed and some guidance was developed, as reported in the technical report of the meeting (EFSA, 2015). A follow-up discussion was proposed for the second general meeting, in order to consolidate the previous agreement.
The experts at the meeting concluded that an update of the guidance given in Appendix F of the technical report (EFSA, 2015) was needed. The update is included in Appendix E of the report of the EFSA PPR Meeting in 2019, which gives a synthesis of the whole process and the agreed approach.

3.2 Birds and mammals

The risk assessment methodology for birds and mammals in EU context has been elaborated in the EFSA GD (2009).

The following points related to birds and mammals have been discussed in harmonisation workshops and agreed upon by the Central Zone Steering Committee (CZSC). The respective harmonisation workshops (CZHW) and CZSC meeting in which the different points were discussed and decided upon are mentioned at each point.
Relevant points from the EFSA Pesticides Peer Review Meetings on general recurring issues in ecotoxicology, which are not already part of the bullet points, are also included.

3.2.1 Voles (CZHW Wageningen (NL), 2014; CZSC July 2014)

Relevance

The small herbivorous mammal needs to be included in the core when relevant in EFSA GD, 2009 (crop groups in Annex I Table 1.2). Generic refinements should be discussed in the core, MS-specific refinements (related to ecological and agricultural circumstances) in the national addendum.

Level of protection

The risk for voles on population level could be lower than for other mammalian species at the same calculated TER. Population modelling is expected to be a promising way forward to resolve this issue on a scientific basis. Nevertheless, the majority of MS are not willing to change trigger values for voles.

3.2.2 Proportion of time spent in the area (PT) (CZHW Wageningen (NL), 2014; CZSC July 2014)

- **Percentage**

Every study should be well described in the core, including presenting both mean and 90th percentile PT values.

- **Time period**

PT data should be relevant (or worst-case) for the part of the application period which is associated with the highest risk.

- **Group**

In principle, use consumer only for calculating PT, but additional data can be used in a weight of evidence.

- **Long Term**

PT (i.e. other numerical parameters) only for long term assessment, unless requirements of EFSA GD 6.1.1 are met.

Background information, EFSA GD 6.1.1

> For example, in the past, one of the most common refinements has been to reduce the value used for PT (e.g. based on radio tracking data) on the grounds that most individuals have PT less than 1. However, the birds that were present in the field studies used to evaluate the level of protection (LoP) for acute assessments of sprayed pesticides also had values of PT less than 1. Therefore the effect of lower values of PT in reducing acute risk is already reflected in the outcomes of the field studies. Consequently, the evaluated level of protection for Tier 1 (Appendix C) already takes account of lower PT values, so replacing PT = 1 with lower values in a refined TER will double-count their effect[62]. The same logic applies to other common refinements including changes to PD and using pesticide-specific residue data, or arguments based on avoidance and/or metabolism: the same factors would also have been operating in the field studies (to varying extents) and will therefore be double-counted (to varying extents) if a refined acute TER is compared to the Tier 1 trigger value. This does not mean that refined TER calculations should not be done. Specifically, if there is evidence that one (or more) of the inputs to the TER calculation for a particular pesticide consistently differs from the range of values expected for the pesticides in the original field studies, in a way that reduces the risk, then the refinement can be supported. This might be the case if, for example, it could be shown that the distribution of PT (and particularly its upper tail, which is most relevant for acute risk), is lower than most of the distributions that would have been expected in the original field studies; or if, the pesticide is more strongly avoided in field conditions than most of the pesticides in the original field studies (organophosphates and carbamates). It is clear, then, that the level of protection provided by refined acute assessments must be re-evaluated case by case, including careful comparison with the field study calibration (Appendix C).

3.2.3 Focal bird species in Central Zone (CZHW Wageningen (NL), 2014; CZSC July 2014)

It is agreed that a 'living' central zone focal species list is useful for refinement. Species on the list will be automatically accepted as focal species without further data once the list is finalized and agreed. Other species can be used provided that adequate supportive information is submitted by applicants. Endangered species and chick diets are currently not explicitly considered in the risk assessment – this was identified as a research need.

[Note NL: to date the central zone focal species list has not been finalised.]

Follow-up (CZHW Vienna (AU), 2015; CZSC October 2016)

In 2016 EFSA launched a project to generate a database including ecological and residue data evaluated in a harmonized way. Though currently it is not clear how this database will and can be used in future, it is not likely to be meaningful to start the work on a comparable

database. Before further actions are initiated the final report of the EFSA project will be awaited.

3.2.4 MAF*twa (CZHW Wageningen (NL), 2014; CZSC July 2014)

The majority is in favor of keeping the first tier as it is (however if the applicant already uses the moving time window in the first tier it could also be accepted, because it is worst-case); in the higher tier the moving time window approach should be used. As a default, the 21 day period should be used, unless another period is mentioned in the DAR of the active substance. Interception is only taken into account at later growth stages with high vegetation coverage (as described in EFSA GD appendix E, table 2). Starting from those stages the FOCUS groundwater interception values can be used for refinement. Note NL: please also refer to the point 3.2.16 below,

3.2.5 Pore-water approach (CZHW Wageningen (NL), 2014; CZSC July 2014)

For now the calculation based on bulk soil concentrations will be used. A calculation based on pore water concentrations would only become meaningful when adequate PEC pore water measurements or calculations are available.

3.2.6 LD50/10 (CZHW Wageningen (NL), 2014; CZSC July 2014)

All MS consider the LD50/10 in the reproductive risk assessment for birds. The lowest NOAEL or LD50/10 is used in the risk assessment.

3.2.7 Averaging interval for calculation of the twa, MAF x twa or moving time window: discussion of the available tools (CZHW Vienna (AU), 2015; CZSC October 2016)

BE and DE have developed tools for calculation of the moving time window. Both tools give similar results and thus any of them can be used in the risk assessment. The application of the time moving window approach was taken note by the Central Zone Steering Committee and hence should be applied as explained above (N.B. see the September 2014 agreements) for the second tier risk assessment.

3.2.8 Exposure outside the breeding season (CZHW Vienna (AU), 2015; CZSC October 2016)

Should exposure outside the breeding season be considered relevant and without the option for waiving the risk, based on delayed effects and/or effects on pair formation and nest building etc.?

Exposure outside the breeding seasons as a waiver for risk assessment was not taken into account in a consistent way in core assessments, due to concerns related to e.g. delayed effects but also as breeding time was considered to depend on various parameters.

The majority of the participants was in favor of including a risk assessment in all cases (even outside the breeding season) in the core assessments. Where refinement due to exposure outside the breeding season was considered on national level, a respective remark could be included in the core assessment.

The point was taken note by the Central Zone Steering Committee. Exposure outside the breeding season can be considered on national level, a respective remark could be included in the core assessment.

3.2.9 Refinement parameters for seed treatment: avoidance factor, dehusking factor (CZHW Vienna (AU), 2015; CZSC October 2016)

UK provided a proposal how to proceed with studies on de-husking, which was generally agreed by the meeting. Currently, no sufficient information to have a standardised factor for de-husking is available.

Avoidance might be considered by a weight of evidence approach, but not by number in a quantitative way for long-term assessment. It shall also not be used in a quantitative way for acute assessments.

The Central Zone Steering Committee did not see the necessity to officially confirm this approach. The incorporation of de-husking and avoidance factors should therefore be followed as discussed above.

3.2.10 PEC to be used in the risk assessment for earthworm-eating birds and mammals (Circabc forum discussion; CZSC January and April 2016)

Based on ecotoxicology forum discussions in the Central Zone, the following decisions were made by the CZSC in January and April 2016 regarding secondary poisoning by earthworms:

For the ecotoxicological risk assessment, secondary poisoning of birds and mammals through earthworms, the dry soil approach, Step 3a of the EFSA GD (2009):
The residue in earthworms (i.e. PEC earthworm) should be estimated by multiplying the appropriate PEC soil (see below) with the BCF earthworm:
− For Non-persistent substances: take PECsoil,twa,21 days (or PECsoil,max as a worst-case)
− For Persistent substances: take PECsoil,twa,21 days + PECsoil,plateau (or PECsoil,max as a worst-case)

Background information

The EFSA guidance on risk assessment for birds and mammals (2009) (EFSA GD (2009)) presents the possibility of assessing the bioaccumulation potential of lipophilic organic substances (i.e. log Pow ≥ 3). In case of the uptake of the substance via the food chain "earthworm to earthworm eating birds and mammals", two options for assessment are presented, the dry soil and the pore water approach. For the purpose of this document, only the dry soil approach will be discussed.
Regarding the dry soil approach, the EFSA GD (2009), states that the PECsoil with an appropriate TWA according to the reproductive assessment should be used, however the EFSA GD (2009) does not distinguish between persistent and non-persistent substances. Therefore Conclusion 1, provides two possibilities for the assessment. Please note that the environmental fate section should assess if a substance is persistent or not and accordingly calculate the appropriate PEC values. For the ecotoxicological risk assessment, secondary poisoning of birds and mammals through earthworms, the dry soil approach, Step 3a of the EFSA GD (2009) the residue in earthworms (i.e. PEC earthworm) should be estimated by multiplying the appropriate PECsoil as presented under Conclusion 1 by the BCFearthworm.

3.2.11 PEC to be used in the risk assessment for fish-eating birds and mammals (CIRCABC forum discussion; CZSC January and April 2016)

Based on ecotoxicology forum discussions in the Central Zone, the following decisions were made by the CZSC in January and April 2016 regarding secondary poisoning by fish:

For the secondary poisoning of fish eating birds and mammals, the EFSA GD (2009) recommends to calculate the PEC fish by multiplying the highest PEC water based on the RAC with an appropriate TWA according to the reproductive risk assessment.
However it is not clear what the EFSA GD (2009) means with 'the relevant PECwater': a

max PEC or the PEC21days? It is also unclear which TWA should be used in the secondary poisoning, and whether this is incorporated in the PEC21 days.
The CZSC therefore decided that in a screening step (i.e. before first tier) the lowest acceptable surface water concentration for aquatic organisms should be used:
− All substances: take RAC-aqua[1] as a screening step, provided that the aquatic risk assessment did not use a twa PEC for one or more groups (and so has a maximum PEC that exceeds the RAC)

Explanation:
Not the actual PECsw, but the critical endpoint with assessment factors must be used for secondary poisoning (e.g. if the PECsw is 0.04 mg/L and the lowest endpoint is 50 mg/L with an assessment factor of 100, than use 0.5 mg/L). However, this is a screening step, and if the risk is not acceptable then the risk assessment can be conducted with TWA concentration. Of course, one can always skip the screening step, and then nothing is different from the old way of risk assessment.

3.2.12 Errors in the EFSA bird and mammal guidance document (CZHW Liverpool (UK), 2017)

Whilst using the EFSA Guidance Document of Risk Assessment for Birds and Mammals several errors in the text have been noticed. These have been collected in a document by FERA:
http://www.hse.gov.uk/pesticides/resources/E/Ecotox_BirdMammal_errors_clarification.pdf

3.2.13 Seedling scenario for treated seeds (section 5.2.1 from the EFSA Guidance) (CZHW Liverpool (UK), 2017)

In determining the DDD for herbivorous birds and mammals following the germination of treated seeds, the approach outlined in section 5.2.1 of the Guidance Document needs to be followed. As stated in the guidance, the seed treatment most closely resembles the “newly sown grassland” or “early post-emergence uses on cereals” scenarios and therefore the relevant focal species are small omnivorous birds and large herbivorous birds and mammals. It is noted however that large herbivorous birds and mammals are mentioned in the text of section 5.2.1, but do not appear in table 19 in the guidance. Therefore when

[1] Regulatory acceptable concentration for aquatic organisms. The use must be safe using the RAC-aqua to use it in the screening step.

performing a risk assessment, these focal species should be added using for birds a FIR/bw of 0.3 and for mammals a FIR/bw of 0.4 (i.e goose and rabbit eating cereal shoots resp., as suggested in the guidance).

3.2.14 Refinement of RUD and DT50 (CZHW Dessau (DE), 2018; CZSC February 2020)

The MS agreed on the following handling of refinements of RUD and DT_{50} in the risk assessment for birds and mammals:

(i) No refinement of RUD and DT50 without detailed justification (by the applicant) would be accepted, and refinements will not be based on one single trial.

(ii) Field trials need replications to be acceptable.

(iii) The trials for refinement need to be close to the GAP, and that deviations need to be justified on a case-by-case basis. If the crop is not eaten, then deviation from the GAP may be accepted.

(iv) All aspects of the study (formulation used, weather conditions, circumstances of application) should be documented in the study report as a comprehensive description of the trials and should also be included in the RR study summary.

(v) The evaluation of residue dissipation studies for refinement of DT50 should be done in collaboration with fate experts.

(vi) If a refined DT50 value for residues is accepted, the time weighted average concentrations should be determined according to EFSA GD, Appendix H (moving time window approach)

(vii) The zRMS should provide the description and an evaluation on the validity of the submitted studies in the core assessment in order to support a harmonized authorisation on the national level.

(viii) Detailed recommendations are now also available in the EFSA PRAS 185 report and should be followed in addition to the agreements of the 4th CZHW. Generally, the PRAS 185 report overrules the 4th CZHW discussions for contradictory conclusions.

3.2.15 Refinement of interception values (CZHW Dessau (DE), 2018; CZSC February 2020)

The MS agreed to use the new interception values according to EFSA guidance on $DegT_{50}$ values (2014) in the tier 2 assessment, but not yet at tier 1 in zonal assessments.

3.2.16 Extrapolation of studies between different agroclimatic conditions (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

In relation to the higher tier studies for birds and mammals, the experts considered that the recommendations given by EFSA (2009) are sufficient for spray applications, i.e. any refinements of the risk based on identification of specific focal species and definition of related ecological data should be representative of the area of use of the active substance. This means, for example, to extrapolate a focal species from one zone to another requires consideration of whether the criteria for selecting the focal species are still met. However, the experts noted that higher tier studies for seed treatment uses would need further attention, in order to take into account specific agronomic practices (e.g. sowing rates) and conditions. The experts suggested that any issue related to the agronomic practices may be addressed in the European Commission's guidance document on seed treatments which is under development and can be considered in the context of the revision of the EFSA Guidance (EFSA, 2009).

3.2.17 Trials for residue decline (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

Kinetic assessment: General principles for the kinetic assessment were agreed.
Extrapolation: Rules for extrapolation within and between defined item groups were agreed.
Plant material: It was agreed that in order to refine the default value for residue decline, residue trials should be performed in at least four sites per item and regulatory zone. However, it was also agreed that in some cases there may be a possibility to extrapolate between areas (e.g. northern France).
Invertebrates: no agreement regarding the minimum number of trials or sites was reached and this should be resolved by the ongoing working group for the revision of the EFSA Guidance (2009).

3.3 Aquatic organisms

In the core assessment, in principle a risk assessment in accordance with Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters (EFSA Journal 2013; 11(7):3290) (abbreviated as EFSA (2013) in this CZ Working Document) should be presented.

The points below concern issues which needed a more detailed elaboration compared with how it is described in the existing EFSA (2013) guidance or which are not sufficiently addressed in the guidance, and which are agreed upon by the Central Zone MSs and also established by the Central Zone Steering Committee (CZSC). The goal is to have clearer guidance available for the Central Zone Core Assessments and also to raise concerns where necessary.
Relevant points from the EFSA Pesticides Peer Review Meetings on general recurring issues in ecotoxicology, which are not already part of the bullet points, are also included.

3.3.1 Use of refined exposure studies in Aquatic Risk Assessment (CZHW Liverpool (UK), 2017; CZSC May 2017)

a) Refined exposure studies should be evaluated and the evaluations should be put into the core of the registration report.

b) In such evaluations, the number and spacing of exposure events should be specified when reporting endpoints from such studies.

c) For refined exposure aquatic studies incorporating a sediment phase, analytical verification of concentrations in the water phase is mandatory.

d) For refined exposure aquatic studies incorporating a sediment phase, in which the test concentrations decline over time:

_ in Tier 1 studies the results should be presented against the mean of the measured concentrations.

_ In Tier 2 studies the results should be presented against both the mean of the measured concentrations and also the peak/initial measured concentration.

Follow-up: Use of refined exposure studies in aquatic Risk Assessment (CZHW Dessau (DE), 2018; CZSC February 2020)

The MS agreed on the following two pre-requisites:

- the GAP must be covered in terms of exposure pattern, and

- if a refined exposure toxicity is delivered by the applicant, all information must be provided in order to facilitate its evaluation and potential implementation in the RA.

Although no final agreement was reached, most MS consider:

- that the Tier 2C approach should generally not be supported at zonal level, considering that implementation in ERA is complex and linked to high uncertainties

- if a conclusion of low risk based on a lower tier approach with RMM is possible this should be favoured over a conclusion based on a Tier 2C approach, considering the uncertainties related to such a Tier 2C approach
- if applicants still decide to deliver a refined exposure toxicity test (Tier 2C option), a lower tier (e.g. Tier 1) risk assessment should always be also presented up to FOCUS step 4 with an agreed level of Risk Mitigation Measures (RMM).

Background information

In the position paper regarding this point some further explanations are presented which are useful to mention here:
It is about the i) non delivery of all information by the applicant, i.e. if the information requested is not made available by the applicant, the RA will not be based on Tier 2C data, but on lower tier data. However the test should be evaluated (with derivation of endpoint(s)), and this evaluation included in Appendix 2 of the core assessment as agreed at the CZSC Feb17 or in or in the DAR/RAR, and ii) if applicants deliver a refined exposure toxicity test (Tier2C option), they must provide all the information requested in order to facilitate its evaluation and use (implementation) in the RA. This includes a) in all cases: all exposure data, profiles, PECmax, and envelope curve comparisons, and b) whenever relevant, information related to the (in)dependence of pulses to justify the simplification of FOCUS scenario(s) tested whenever relevant (e.g. lowering the number of or pooling peaks).

There are strong concerns and uncertainties supporting the position of MSs not wanting these tests. In the discussion about this topic, in addition to issues previously raised (i.e. uncertainties about the representativity of the i) selected Focus profiles tested and ii) life-stage and strategy of the tested species with regard to the most sensitive species occurring in the field), some important points were discussed:
For the toxicological (in)dependence, it is clearly noted that by default peaks are considered as dependent. For long-living species toxicological independence is hard to show. But when attempting to demonstrate independence, the aspect was raised that a short development time species e.g. Daphnia should be representing invertebrates with longer development time. Therefore, peaks that are independent for Daphnia will become dependent for other invertebrates with longer life-cycles. This is an issue if the relevant endpoint is e.g. an EC50.
Furthermore, it was pointed out that influences on the organisms might be transferred to the next generation (voting for stronger dependency of peaks). However, effects on next generation are not assessed in Tier 1 either (to be considered in risk assessment/uncertainties).
Overall, the implementation of this approach in the RA is perceived as complex and linked to high uncertainties.

3.3.2 ETO vs ERO – Use of data from mesocosm (CZHW Liverpool (UK), 2017; CZSC May 2017)

The Central Zone (CZ) Ecotoxicology Harmonisation Group are of the view that an "ecological threshold option" (ETO) should be determined when assessing a mesocosm study. Furthermore, the CZ Ecotoxicology Harmonisation Group considers that an ETO should be used to set the regulatory acceptable concentration (RAC). In light of this, when a new mesocosm study is considered as part of an application for product approval, an ETO should be determined and used, along with an appropriate assessment factor of 2-3 (see Table 8 and 9 of EFSA (2013)) to generate the RAC.

(According to EFSA (2013), the Applicant *may* be able to demonstrate that "all relevant processes that determine population viability and the propagation of effects to the community-, ecosystem- and landscape-level" (Section 5.5 of EFSA (2013)) have been considered. If the Applicant has addressed all the issues regarding recovery, then it may be feasible to determine an "ecological recovery option" (ERO) and along with an appropriate assessment factor of 3-4. Both endpoints – ETO and ERO – as well as the corresponding RAC should be quoted in the core. MS may wish to use the ERO as the justification of using this endpoint may be MS specific (e.g. minor use in a specific area etc.). However, the over-riding view of the CZ Ecotoxicology Harmonisation Group is to use the ETO approach.

As regards what endpoint to use "if the ETO is not report", it is assumed that this is related to where an active substance has been reviewed and as part of that assessment a mesocosm study has been considered and an endpoint agreed. The terms ERO and ETO are new terms and will only be relevant to those active substances considered after EFSA (2013) was noted (i.e. for those dossiers received after 1st January 2015 – see SANCO/10605/2014 – rev. 0 (11 July 2014) Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters) and it is assumed that both endpoints – ERO and ETO will be presented in the LoEP of the EFSA conclusion. The terms ERO and ETO are unlikely to appear in EFSA conclusions prior to this date. As for those active substances that were considered prior to the implementation of EFSA (2013), it is likely that there will be endpoints based either on recovery, minimal or "no effects" as well as a range of associated assessment factors. It is important to have a consistent way in which these previously agreed endpoints are interpreted and used, especially when it is considered that EFSA (2013) should be used for the assessment of products considered after 1st January 2015 – see SANCO/10605/2014. With this in mind, outlined below is a proposal in which the variety of endpoints and assessment factors could be dealt with:

Where an ERO or ETO has not been defined in the list of endpoints, it is assumed that an ETO is broadly equivalent to a NOEC whilst an ERO is equivalent to a recovery based endpoint. It should be noted that in previous assessments the period for recovery may have been longer than specified in EFSA (2013).

1. If the endpoint presented is a NOEC, then this could be assumed to be equivalent to an ETO, hence an assessment factor of 2 could be applied to this endpoint to derive an ETO RAC.
2. If the endpoint presented in the LoEP is based on a recovery endpoint and hence may be quoted as a NOAEAC, then the original assessment in the DAR should be considered, and if a NOEC has been determined as part of the study evaluation[2], then this, along with an assessment factor of 2 should be used.
3. If the endpoint presented is a RAC or some other endpoint where the effects endpoint and the assessment endpoint have been combined and it is unclear from the LoEP what the exact effects endpoint is, then the original assessment in the DAR should be considered and if a NOEC has been determined, then this, along with an assessment factor of 2 should be used. If a NOEC has not been determined, then one should be determined from the study summary in the original DAR if possible.

For points (2) and (3) above, if a NOEC has not been determined, the following course of action is proposed:

1. Can one be determined on the basis of the evaluation in the DAR? If so, then use that along with an assessment factor of 2
2. If a NOEC was not quoted in the DAR and cannot be determined on the basis of the study evaluation, revisit the original study; it is not proposed to re-evaluate it, but to see if a NOEC was determined. If it was, then it is proposed to use that, providing that it is lower than the NOAEAC quoted in the LoEP.
3. If a NOEC cannot be determined due to effects at the lowest concentration then there needs to be a consideration of how many species and what the level of effects were. If there was an impact on two species and the effect

[2] It is not proposed to revisit the study but to work from the original assessment. Whilst the NOEC may not have been subject to detailed discussion during the peer review stage it is assumed that the study will have been and hence endpoints other than the previously agreed endpoint can be considered reliable.

deemed to be a Class 2[3] effect[4], then it may be feasible to use this endpoint along with an assessment factor of 3. If there is uncertainty regarding the relevance of the effects at the lowest concentration then it is proposed to go back to the Applicant and for further information[5].

If the Applicant has represented the mesocosm study for product registration purposes, possibly along with a consideration of minimum detectable difference (MDD), then this should be considered along with any previous comments made during the peer review process regarding the robustness of the mesocosm study(ies) and a ETO (and possibly an ERO) derived.

Whilst the above outlines a proposal regarding the use mesocosm studies, it is proposed that EFSA (2013) should be used to derive other higher tier endpoints, for example those associated with the use of multispecies data (e.g. SSD).

See for extrapolation of studies between different agroclimatic conditions point 3.3.9.

3.3.3 Need for aquatic risk assessment of metabolites to be considered in product authorisation (CZHW Liverpool (UK), 2017; CZSC May 2017)

a). There were proposed positions for 5 different situations, positions

for situations 1 and 2 were agreed at the meeting. These are presented here. [There is further active consideration underway in relation to situations 3, 4 and 5].

	Situation	Proposal
1	Data considered at EU review of active and metabolite not identified as ecotoxicologically relevant	No further consideration required at product approval stage
2	Data considered at EU review of active and metabolite identified as ecotoxicologically relevant	Risk assessment required at product approval stage

[3] See for example Section 2.1.6 of EFSA (2013).
[4] There needs to be a consideration of any additional information in the DAR that could put the effects in to perspective. There also needs to a consideration of whether the impacted species are key and the only relevant ones. For example, if the compound effects moulting and there are only two species that go through a moult in the mesocosm study, then this is of greater concern, compared to say a broad spectrum toxicant where there is an impact on one species in the lowest concentration.
[5] Further information could be in the form of a minimum detectable difference analysis to try to provide some indication as to the robustness of the effects observed on key species.

3.3.4 Consideration of metabolites in data-matching; use of (Quantitative) Structure-Activity Relationship [(Q)SAR] arguments (CZHW Liverpool (UK), 2017; CZSC May 2017)

a) For data-matching purposes, new vertebrate data on metabolites should not be generated; an applicant should seek access to the existing studies.
b) There is ongoing consultation on proposals for when (Q)SAR arguments should and should not be accepted as satisfying information requirements for metabolites in datamatching.]

3.3.5 The need to conduct aquatic ecotoxicity studies on formulations containing multiple active substances; and the derivation of endpoints from such studies where not all active substances had their aqueous concentrations measured (CZHW Liverpool (UK), 2017; CZSC May 2017)

a) As a specific illustration of a general principle, if each active substance is known to be at least 10 times more toxic in algae than in fish, it is not necessary to conduct a study on the toxicity of the formulation in fish.
b) If the concentration of the least stable active substance was measured and was maintained for the course of the study, then ecotoxicological endpoints can be established based on the nominal concentration of each active substance. If the measured concentration of the least stable active substance declined during the study, then ecotoxicological endpoints can be established by adjusting the nominal concentration of each active substance in line with the percentage recovery of the least stable active substance.
See for the latest developments regarding this issue point 3.3.6.

3.3.6 Derivation of endpoints for aquatic tests with instable substances (CZHW Dessau (DE), 2018; CZSC February 2020)

A paper has been agreed on by the MS of the Central Zone (“Expressing endpoints from Tier 1 tests and formulation tests (with one or more active substances) for unstable substances”). This is now published as Appendix J of the EFSA Report on general recurring issues (EFSA Supporting publication 2019:EN-1673).

Background information

At Tier 1, laboratory standard tests must be performed under standard (i.e. mostly worst case) exposure. Therefore, OECD guidelines recommend that the concentrations should be maintained and must be > 80 % and < 120 % of nominal at the end of the exposure period (or at the end of the renewal period for semi-static design).
If the concentration cannot be maintained (i.e. if the substance is dissipating 'fast'), the validity of the study should be questioned and the test may be rejected as highlighted during the EFSA peer review meeting on general recurring issues in ecotoxicology (EFSA, 2015).

During this EFSA peer review meeting, Member States agreed that in principle:
1) **Nominal concentrations** can be used to express the toxicity from any kind of test if the test concentrations were maintained at ± 20 % of the nominal at all times throughout the test including the study end sampling. Mean measured is also an option for this situation.
2) **Initial measured concentrations** can be used to express the toxicity from any kind of test if the initial test concentrations were below 80 % of the nominal and this concentration was maintained throughout the test (within ± 20 % of the initial) including the final sampling. Mean measured is also an option for this situation.
3) **Mean measured concentrations** must be used to express the toxicity from any kind of test when the test concentrations were not maintained within the range of ± 20 % of the nominal or initial measured, but significant concentrations of the test item were still present at the end of the exposure period (or at the end of the renewal period for semi-static design).
4) When the test concentrations were not maintained and significant residues were not present at the end of the exposure period (or at the end of the renewal period for semi-static design), the **validity of the study should be questioned**.

In practice (and not due to a causal relation), however, semi-static and/or flow-through design is rarely used for tests with:
- algae for which semi-static tests are very uncommon and flow-through tests not established in the regulatory context, due to the technical complexity when conducting the test
- formulated products with one or more active substance, especially for tests with algae.

This proposal addresses these issues. It especially considers the cases where the recovery of an active substance at the end of a test is < 80 % (i.e. the test substance is dissipating fast) and where requesting a new semi-static or flow-through test (as required by EFSA, 2015) may not be feasible or desirable (i.e. algae tests and vertebrate tests).
An adequate expression of the endpoint from formulated product tests is needed:
- for the purposes of classification and labelling, and

- as the basis for mixture toxicity assessment since it should enable an assessment of potential synergism or additive toxicity due to one or more co-formulants or additional active substances.

The proposed approach aims to serve both purposes.
Until a revision of the (EFSA PPR Panel, 2013) , this position paper is intended to fill the gap as an interim solution, i.e. for such cases where above-cited requirements 3 and 4 cannot be easily fulfilled and performing tests under semi-static or flow-through conditions are an issue.

The proposal was in a draft phase at the time of the CZHW in Dessau (Sep. 2019) and was therefore not discussed. After the meeting, an updated version of the document was sent to the member states of the Central Zone and an agreement was reached. Thereafter the document was shared with EFSA and is included as Appendix J of the EFSA report “outcome of pesticides peer review meeting on recurring issues on ecotoxicology (June 2019)”.

3.3.7 Use of ErC50 or EbC50 values for algae and aquatic plants (CZHW Liverpool (UK), 2017; CZSC September 2017)

Until Revision of Aquatic GD (EFSA, 2013) it is recommended for product evaluation to include the following updated version (see attached document EbC50 and ErC50 action point update from 18th of November 2015) blocktext in the Core assessment:
“*The endpoint ErC50 is selected in this Core Assessment but there are some uncertainties regarding the level of protection reached for primary producers. This is indicated for macrophytes in the aquatic Guidance Document (EFSA Journal 2013;11(7):3290) that recommends: “... a proper calibration between different tiers (higher and lower tier data) for macrophytes should be performed in the future”. Such calibration should be extended to algae. Until available relevant information on the level of protection reached is considered at EU level, it is recommended to address this uncertainty at Member State level in the National Addendum if considered necessary, although it would be highly appreciated to have a harmonised approach in the central zone.”*

3.3.8 Refinement of the exposure by different risk mitigation measures (RMMs) (CZHW Dessau (DE), 2018; CZSC February 2020)

The MS agreed that RMM up to 90% drift reduction and 30 m buffer zone should be presented in the core assessment.

It is noted that this is not in agreement with the FOCUS guidance 'Landscape and mitigation', in which it is instructed to calculate FOCUS Step 4 up to 90% drift reduction nozzles in combination with 20 m buffer or 20 vegetated strip.

3.3.9 Extrapolation of studies between different agroclimatic conditions (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

In the case of mesocosms, the majority of the experts at the meeting agreed that the no observable effect concentration (NOEC) and the ecological threshold option (ETO) regulatory acceptable concentration (RAC) can be used in the risk assessment with the assessment factor (AF) recommended by aquatic guidance (EFSA, 2013), and this can be considered as independent of the experimental conditions (e.g. the climatic zone). However, when an ecological recovery option (ERO) RAC is derived, the extrapolation between zones should be considered carefully taking into account the fact that the ability for recovery may vary pending on the agroclimatic conditions. A case-by-case evaluation should be carried out, based on the information available.

3.3.10 Minimal Detectable Difference (MDD) (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

In the EFSA report (EFSA Supporting publication 2019:EN-1673, section 4.8), it is reported that the MDD, presented in the (EFSA, 2013) and the paper by Brock et al. (2015), is considered to be a valid tool to help with the evaluation of the biological results to assess the statistical power – or the absence of power – of a study to detect treatment-related direct effects. It should preferably be reported on non-aggregated data for the relevant taxon and time points. An issue linked to the unclear beta-error associated with the MDD in the available documents (i.e. (EFSA, 2013) and Brock et al., 2015) was raised by Germany. It was concluded that the use of the MDD is supported and that further considerations and clarifications will be addressed in the revision of the aquatic guidance (EFSA, 2013).

Please note that the further clarifications considering the beta- error refer to the paper now published (Duquesne et al., 2020; Environmental Science and Pollution Research https://doi.org/10.1007/s11356-020-07761-0).

3.3.11 PECsw-TWA approach (CIRCABC discussion; CZSC February 2020)

Application of the PECsw-TWA approach for aquatic organisms gives rise to problems that are difficult to overcome, when based on the existing guidance. Therefore:

1. The PECsw-TWA approach should currently not be used in zonal assessments due to lacking guidance and harmonisation at EU-level and other concerns (e.g. most sensitive life-stage tested, extrapolation to other species), to avoid inconsistent evaluations;
2. To gain knowledge and experience with these assessments the information submitted by applicants might be included in the RRs, but together with a general statement that it has not been considered further.

Background information

In a special EFSA ecotox meeting (133): Pesticides Peer Review Meeting on recurring issues in Ecotoxicology (EFSA, 2015), it was decided not to use PECsw-TWA values anymore for EU active substance assessments due to lack of guidance on reciprocity and latency of effects in the EFSA aquatic GD (2013).

Reciprocity refers to Haber's law, which assumes that toxicity depends on the product of concentration and time. There were hopes of the issue being addressed in the planned corrigendum of the aquatic guidance document in 2017, but the update of the GD is postponed due to other priorities. Hence, the question is what to do with this issue when it concerns product assessments on a zonal and national level. The issue was discussed within CTGB and the proposal is that in the absence of proper guidance on reciprocity and latency of effects also on zonal and national level the PECsw-TWA approach cannot be applied, unless the applicant provides information/argumentation with sufficient proof of reciprocity and absence of latency of effects.

Latency of effects can happen in the case of e.g. growth regulators or substances which influence the process of molting or the hormone system. For acute toxicity, in order to determine whether there is latency of effects, tests need to be longer than normal acute testing. With regard to chronic risk assessment; this is based on tests intended to cover the complete life-cycle or the most sensitive stage of a test organism. However, it is difficult to know what is the most sensitive life stage before testing to find out. Also, the effects observed in a test are dependent upon the endpoints which are measured. There might be effects not measured in a test, which could have an impact on the reproduction or survival of a species. Especially in the case of aquatic vertebrates it is difficult to test each stage in the complete life-cycle, and even with "full life cycle tests" it is difficult to ascertain whether the effects seen were as a result of the long-term exposure or exposure during a particular life stage. In any case, due to expense and animal use concerns inherent to full life cycle tests, most available chronic tests are focussed on testing the (assumed)

most sensitive life-stage. Taking into account what is said about this latter issue, it is proposed not to take a PECsw-TWA into account in the risk assessment for aquatic vertebrates, unless sufficient data is provided to show the actual mechanism of toxicological action, also considering the data from the mammalian toxicology package, when full life-cycle testing is available.
In the case of aquatic invertebrates the PECsw-TWA approach should be used with great care, especially when it concerns aquatic insects. For this taxonomic group only a standardized chronic test with chironomids is available. But from other insect species, like the EPT species (which are belonging to the most sensitive species for quite some insecticides), knowledge is lacking, particularly with regard to their most sensitive life stage and how their sensitivity compares to *Chironomus riparius*. Hence, it will not be easy to show sufficient proof of reciprocity and absence of latency of effects. For primary producers this could be somewhat easier and if there is enough evidence, a PECsw-TWA could be applied in the risk assessment.

3.3.12 Use of geometric mean and weight of evidence for acute data (EFSA PPR Meeting on general recurring issues (EFSA, 2019); CZHW Brno (CZ), 2019)

At the EFSA PPR Meeting it was agreed that, in cases where the RACgeomean is greater than the lowest endpoint, the lowest endpoint should be used to calculate the RAClowest. The minimum modified AF for deriving the RAClowest should be 20 for invertebrates and 30 for fish.
The experts suggested that the approach should be further considered with the revision of the EFSA (2013) guidance.
At the CZHW meeting in Brno (2019) the majority of the MSs agreed with the proposed Tier 2A scheme for acute risk assessment (steps 1 and 2):

Step 1-Is lowest EP < $RAC_{geomean}$?

1. Yes: use **RAC_{lowest}** (EFSA, 2013)

   Note: RAC_{lowest} = lowest EP / AF ≥ 20 for invertebrates and ≥ 30 for vertebrates (EFSA, 2019)

1. No: Go to 2

Step 2- $RAC_{geomean}$ and RAC_{lowest}

Compare $RAC_{geomean}$ and RAC_{lowest} (lowest EP /AF 60), report both, use the **lowest RAC**.
If using RAC_{lowest}, add **blocktext***:*
"The RAC_{lowest} (i.e. endpoint of the most sensitive species tested divided by an AF of ≥ 60) is considered as a "safety net" to the $RAC_{geomean}$, especially relevant when the lowest available endpoint of the dataset is in a range close to the $RAC_{geomean}$. In the current situation, the use

of the RAC_{lowest} instead of $RAC_{geomean}$ helps to reduce the shift in the protection level that will be achieved for species situated close to this trigger.

At the EFSA PPR Meeting there was no agreement for using a geometric mean for chronic data. This should be further considered together with the entire approach when the aquatic guidance (EFSA, 2013) is revised.
All MSs at the CZHW meeting in Brno (2019) agreed that the status quo on the use of Tier 2A approach for the chronic risk assessment as in EFSA Technical Report (2019) does not apply to primary producers. The majority of the MSs agreed with the proposed Tier 2A scheme for chronic risk assessment of primary producers (steps 1 and 2):

Step 1- Is lowest EP < $RAC_{geomean}$?

1. Yes: use **RAC_{lowest}** (EFSA, 2013)
   Note: RAC_{lowest} = lowest EP / $AF_{overall}$ ≥ 6
2. No: Go to 2

Step 2- $RAC_{geomean}$ and RAC_{lowest}

Compare $RAC_{geomean}$ and RAC_{lowest} (lowest EP /AF 8), report both, use the **lowest RAC**.
If using RAC_{lowest} , add **blocktext***:*
"The RAC_{lowest} (i.e. endpoint of the most sensitive species tested divided by an $AF_{overall}$ of ≥ 8) is considered as a "safety net" to the $RAC_{geomean}$, especially relevant when the lowest available endpoint of the dataset is in a range close to the $RAC_{geomean}$. *In the current situation, the use of the RAC_{lowest} instead of $RAC_{geomean}$ helps to reduce the shift in the protection level that will be achieved for species situated close to this trigger*. "

3.3.13 General recommendations on mesocosm experiments (EFSA PPR Meeting on general recurring issues (EFSA, 2019))

Representativeness and vulnerability of the communities tested

It was agreed that the absence or low abundance of vulnerable groups, i.e. EPT (Ephemeroptera, Plecoptera and Trichoptera) species, should not necessarily result in the invalidation of the experiment. However, their absence should trigger the need for further considerations, e.g. the selection of a higher AF (i.e. higher than the current AF range since this does not address the potential lack of representativeness) and/or request for further testing to confirm that EPT are not among the most sensitive species. In such assessment, particular consideration should be paid to the mode of action of the active substance. Several recommendations for the experimental design, consideration of indirect effects and definition of Tier 3 experiments were discussed and agreed.

3.3.14 Representativeness of mesocosm studies when the risk assessment at lower tiers is triggered by a non-freshwater species (EFSA PPR Meeting on general recurring issues (EFSA, 2019))

A stepwise procedure was agreed upon:

Step 1: check whether in the mesocosm the taxa closely related to A. bahia are included as the minimum representativeness requirement.

1. If the mesocosm does not meet the minimum representativeness requirement, it cannot be considered to cover the risk for the most sensitive taxonomic group.
2. If the mesocosm covers the minimum representativeness requirement, go to step 2.

Step 2: check that the 'representative surrogate taxa' (those taxonomically similar to the marine species driving the risk assessment at Tier 1) respond to the treatment, showing clear effects.

1. If the 'representative surrogate taxa' respond to the treatment, the mesocosm is considered representative and can be used to address the risk assessment.
2. If the 'representative surrogate taxa' do not respond to the treatment, go to step 3.

Step 3: perform further analysis and additional laboratory experiments might be requested with the 'representative surrogate taxa'. This would allow a better interpretation of the mesocosm by verifying whether the sensitivity of the 'representative surrogate taxa' is similar to that of the marine species untested in the mesocosm.

Background information

The current aquatic guidance (EFSA, 2013) was developed to perform risk assessments for freshwater environments, in accordance with the data requirements specified in EU Regulations 283/2014 and 284/2013. The same AGD, however, does not exclude the opportunity of using data from non-freshwater (marine or brackish) species in the risk assessment scheme. On the contrary, endpoints for these species are regularly used in the evaluations of active substances and PPPs. Data from ecotoxicological tests on non-freshwater species can refer to species at all trophic levels (e.g. Skeletonema costatum for primary producers, Americamysis bahia for aquatic invertebrates and Cyprinodon variegatus for fish). It is not unusual that the lower tier risk assessment is driven by non-freshwater species. When the evaluation at these lower tiers highlights a potentially high risk, an option to refine the assessment is to conduct mesocosm studies on freshwater communities. Non-freshwater species are hardly represented in such mesocosms, and therefore it is

questionable whether these studies are adequate to derive an endpoint able to cover the organisms represented at lower tiers by non-freshwater species.
Usually, the presence of other organisms considered taxonomically similar to the most sensitive non-freshwater species is taken into account to solve the issue. However, the concept of 'taxonomically similar' is open to many interpretations: the term 'taxon' indicates a group of organisms with similar characteristics that can be applied to all the hierarchical levels of biological classification.
The role of phylogeny was discussed at the meeting and some experts disagreed about the use of this approach. It was highlighted that phylogeny is very fluid and hence difficult to be relied upon. The proposal of setting a 'fixed' taxonomic hierarchical limit is problematic, as for some groups it is possible to get a better picture (more sub-group represented) than for others. However, a minimum level to be addressed was proposed on the basis of the comparison between A. bahia and the more closely related taxa that are often tested in mesocosms (Gammarids and Isopods). On this basis the minimum level to be matched should be the superorder. However, a general rule should be to consider which is the closest taxon that can reasonably be tested in a mesocosm, considering its autecology.

3.3.15 Alternative test design in Myriophyllum studies (EFSA PPR Meeting on general recurring issues (EFSA, 2019))

It was agreed that Myriophyllum studies performed to OECD TG 239 (OECD, 20014b) but with an alternative test design (i.e. one shoot per pot per test vessel) should be considered acceptable.

Background information

OECD Test Guidelines (TG) 238 and 239 (OECD, 2014a,b) describe the test designs to perform toxicity tests with the rooted aquatic dicotyledon Myriophyllum spicatum in the absence and presence of sediment, respectively. Both test guidelines require at least five tested concentrations (plus the control) for the determination of the ECX. Test Guideline 238 requires 10 replicates for the control(s) and five replicates for the tested levels, with a single lateral branch for each replicate. Test Guideline 239 requires instead a minimum of six replicates for the control(s) and a minimum of four replicates for the tested levels; each replicate, represented by a test vessel, is composed of three shoots that can be managed in accordance with one of two test designs:
- Test Design A: one shoot per pot and three pots per vessel
- Test Design B: three shoots per pot and one pot per vessel.
Test Guideline 239 reports that 'Alternative test designs of one shoot per pot per test vessel are acceptable provided that replication is adjusted as required to achieve the required validity criteria'.

An alternative test design has been used in toxicity tests for (at least) three active substances: halauxifen-methyl (EFSA, 2014b), florpyrauxifen-benzyl (EFSA, 2018) and oxasulfuron (EFSA, 2017b). In each test, a single shoot was used for each replicate, but the number of replicates was increased to 10 for the control and to five for the tested levels. The studies with this modified test design were considered acceptable in two cases (halauxifen-methyl and florpyrauxifen-benzyl) but were rejected in the third since the number of individuals was considered too low.
The comparison of the two test designs (i.e. the one reported in the OECD test guidelines and the one with single shoots per replicate) (Gonsior and Schwalbach, 2014) gave very consistent results. The use of single shoots in each replicate allows the use of 'real' replicates without interaction among individuals and to increase the statistical power of the test, particularly for the control, owing to a higher number of replicates. Given that the proposed alternative test design is in line with the OECD TG 238, the experts at the meeting agreed to consider it acceptable, as proposed by Italy. This allows the use of 'real' replicates without interaction among individuals and to increase the statistical power of the test, particularly for the control, owing to a higher number of replicates."

3.3.16 How to express the endpoint for sediment-dwelling organisms when tested in the presence of sediment (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

It was agreed that endpoints for sediment-dwelling organisms, when tested in the presence of sediment, should be determined using a mass balance calculation. In this view the submission of mass balance calculations as part of the dataset for the sediment-dwellers is highly recommended, particularly in the case of the substances which are difficult to test (concentrations poorly maintained in the test system).

Background information

During the Pesticide Peer Review Meeting 133 (EFSA, 2015) it was discussed how the endpoints for aquatic Tier 1 studies should be expressed. It was agreed that 'the toxicity endpoint for Tier 1 studies (i.e. mean measured, nominal or initial measured), should not depend on the study design, on the physical chemical or environmental fate parameters, on technical difficulties when testing, or on how the endpoint would be used in the first-tier risk assessment. The choice must depend on the actual exposure throughout the whole exposure period of that particular test. Where a suitable exposure throughout the whole period was not demonstrated, none of the endpoints should be used in first-tier risk assessments.' This discussion did not specifically cover the case of the toxicity tests on sediment-dwellers when tested in the presence of sediment.
The studies more frequently available for addressing the effects on sediment dwellers are performed on Chironomus riparius (OECD 2014a,b).
According to OECD TG 218 (sediment–water chironomid toxicity using spiked sediment), in order to assess the behaviour/partitioning of the tested chemical in the water–sediment system, the

concentrations of the test substance should be measured in the sediment, in the pore water and in the overlying water. These analytical determinations indeed allow for the calculation of mass balance and to express the results based on measured concentrations. According to the same guideline, effect concentrations should be expressed and based on dry weight and preferably based on measured sediment concentrations at the beginning of the test (OECD, 2004a). Further recommendations on how to express the endpoint in the cases where the test item concentrations are not maintained (considering the whole system) or on how the mass balance results should be considered in this context are not included in the test guideline.

Similarly, according to OECD TG 219 (sediment–water chironomid toxicity using spiked water), samples of the overlying water, the pore water and the sediment must be analysed in order to assess the behaviour/partitioning of the tested chemical in the water–sediment system. The test guideline recommends that effect concentrations are expressed as concentrations in the overlying water, preferably calculated based on measured concentrations at the beginning of the test (OECD, 2004b). Further recommendations on how to express the endpoint in the cases where the test item concentrations are not maintained (considering the whole system) are not included in the test guideline. In addition, differently from OECD TG 218, there is no recommendation to calculate a mass balance in order to assess the behaviour of the test item in the system.

In the context of the peer review of the active substance risk assessment, the issue of how the concentrations should be expressed in the case of sediment-dweller toxicity testing was often raised. In particular, there have been instances in which it was questionable to express the endpoints as measured concentrations at the beginning of the test, i.e. in the cases where the concentrations were not maintained in the whole system.

EFSA recommended that the decision on how to express the endpoint for the sediment-dwellers is based on the assessment of the mass balance calculation in order to determine the repartition of the substance in the various compartments. In this view the submission of mass balance calculations as part of the dataset for the sediment-dwellers is highly recommended, particularly in the case of the substances that are difficult to test (concentrations poorly maintained in the test system). In the latter cases, it is also relevant that intermediate measurements in the various compartments are performed (see also Regulation (EU) No 283/2013, Section 8.2.5.3). When a mass balance is available, it is possible to consider the recommendations of the Pesticide Peer Review Meeting 133 (EFSA, 2015). It is additionally recommended that the key endpoints from the sediment-dweller studies are always presented in terms of mg substance/kg dry sediment and mg substance/L water. This would ensure that both exposure via water and sediment are covered for sediment-dwellers.

Where the concentrations in the test system are not maintained, the recommendations of the Pesticide Peer Review Meeting 133 (EFSA, 2015) should be considered, i.e. express the endpoint as the mean measured concentration using mg substance/kg dry sediment and/or mg substance/L water, accordingly, if significant levels are detected in the sediment or in the water or in both. The calculations should be based on geometric mean concentrations. It is proposed to further discuss whether, in such cases, the use of these studies in a Tier 2C approach, similar to the proposal in

the EFSA aquatic guidance document (EFSA, 2013) for the refined exposure studies, would be suitable. This means that it should be demonstrated that the exposure in the study simulates a realistic worst-case exposure relative to the predicted exposure. In this view, a comparison between the exposure in the test system and the expected exposure (FOCUS profiles) should be performed. In order to follow this approach, intermediate analytical measurements should be performed in the course of the study.

It is acknowledged that issues similar to those for the sediment-dwellers could also occur for toxicity tests with the rooted macrophyte Myriophyllum spicatum (OECD TG 239; OECD, 2014b). In those cases it is suggested that the same approach as above is applied. It is noted that OECD TG 239 already highlights that 'if there is evidence that the concentration has declined (i.e. is not maintained within 20 % of the nominal or measured initial concentration in the treated compartment) throughout the test, then analysis of the results should be based on the geometric mean concentration during exposure or models describing the decline of the concentration of the test chemical in the treated compartment'.

Overall, the experts agreed with the proposal to use the mass balance for checking whether the concentrations were adequately maintained. Practical examples of the needed calculations are included in Appendices G and J of the EFSA technical Report, 2019.

3.3.17 Mixture risk assessment calculation tool (CZSC May 2021)

A tool for the mixture risk assessment calculations (called "AGD_AquaMix_v1.15") was developed by a group of Member States from the central and northern zone and was published on the 21st of January 2021 in the CIRCABC Expert exchange forum. It can now be downloaded at the EFSA Knowledge Junction (https://zenodo.org/record/4593676).

The tool is intended to be an extension and implementation of the assessment given in the aquatic guidance document (EFSA Journal 2013;11(7):3290) and to facilitate the associated mixture calculations (also for NTTP). Alongside the tool itself an FAQ was developed as separate file, in which proposals are given for the assessment of complex mixture risk assessment topics (e.g. how to handle metabolites).

This tool will be further developed in the future.

3.3.18 Validity criteria of algae test OECD 201 (CZHW Ede (NL), June 2022; CZSC May/June 2023)

The test protocol OECD 201 is not developed for non-standard species, also for standard species some criteria may not be appropriate in all cases. A proposal was made for harmonized interpretation of validity criteria for non-standard species.

The proposal is acceptable and the CZ MS will use the outlined criteria for non-standard species. The proposal is presented in Appendix 1.

3.3.19 Herbicides with unexpectedly low toxicity to macrophytes: necessity for a different exposure design (overspray) in toxicity tests with emergent and floating macrophytes (CZHW Ede (NL), June 2022; CZSC May/June 2023)

For certain herbicides the current testing methodology does not adequately represent the exposure scenario nor the MoA.

The majority of MS agreed if the toxicity of an herbicide is below the level of acute toxicity classification in emergent macrophytes, the applicant would be asked to either provide justification that direct exposure is not relevant or provide a test including an overspray exposure scenario. The risk assessment for macrophytes would then be performed adding an additional scenario, using the spray drift value for NTTPs, the endpoint from an overspray test (in mg a.s./ha) and the trigger of 10 for aquatic plants. This would be in addition to the "usual" aquatic risk assessment according to FOCUS.

Background information

Several cases were encountered in which, in the standard macrophytes testing (7-14 days exposure), some herbicides showed unexpectedly low toxicity to aquatic macrophytes, thus their toxicity could be underestimated. Some of these herbicides are contact or burn-down herbicides and thus the question raised, if the relevant exposure scenario was tested. Other herbicides might indeed not show effects due to reduced bioavailability, uptake into and translocation in the plant or delayed effects, but the reason is often unknown. Currently there are no OECD tests available for the overspray scenario in macrophytes. Therefore, the proposal of this meeting is as follows: if the ErC50 **value** of a herbicide is **above** the level of acute toxicity classification (i.e., toxicity endpoint > 1 mg/L) in emergent (or floating) macrophytes, the applicant would be asked to either provide justification that direct (spray drift) exposure is not relevant or provide a test including an overspray exposure scenario. The acute toxicity classification is taken as a cut-off value to ensure that chronic toxicity with the overspray scenario is not underestimated.

3.3.20 Aquatics and NTTPs – SSD (CZHW Ede (NL), June 2022; CZSC May/June 2023)

A proposal was presented for the harmonized evaluation and interpretation of SSD data.

The MS agreed to use the approach when evaluating SSDs (aquatic and NTTP) in future dossiers and to bring the paper forward to the EFSA to be considered in the next general issues meeting. The proposal is presented in Appendix 2.

3.3.21 Acute fish testing with PPPs: Limit vs DR tests (CZHW Ede (NL), June 2022; CZSC May/June 2023)

A proposal to minimize vertebrate testing in fish (acute toxicity tests with formulations) was presented. Under certain circumstances only limit tests, or no test, could be accepted rather than a full dose-response test.

The CZ MS agree to follow the proposal when considering whether an acute fish toxicity test is needed with the formulation under consideration for future dossiers. The proposal is presented in Appendix 3.

3.4 Bees

The risk assessment methodology for bees has in EU context been elaborated in the Guidance Document on Terrestrial Ecotoxicology (Sanco/10329/2002 rev 2 final).

3.4.1 Data requirements for honey bees (CZSC, November 2020)

The guidance document on bees is a draft and the MS should not use this as a rule, but could take the draft GD into account in cases regarded helpful. The discussion is ongoing. The zRMS asks for the data requirements and evaluates the data. If the MS use the submitted studies for risk assessment, the risk assessment will be done in the core.

3.5 Other non-target arthropods

The risk assessment methodology for non-target arthropods has in EU context been elaborated in the Guidance Document on Terrestrial Ecotoxicology (Sanco/10329/2002 rev 2 final).

3.5.1 Vegetation Distribution Factor (CZHW Brno, 2019; EFSA PPR Meeting on general recurring issues, (EFSA, 2019); CZSC, April 2022)

The majority of the MSs present at the CZHW meeting in Brno and the majority of the experts present at the EFSA PPR meeting (EFSA, 2019) agreed on the recommendation of using a VDF of 5 for all the tiers of the assessment. It was highlighted in the EFSA technical Report (2019) that this recommendation should be considered as an interim solution until the revision of the current risk assessment scheme. Such an interim solution should be reflected in the (European Commission, 2002) document and its implementation should be further considered.

The CZSC has made an urgent request to the Commission to adjust this issue in the guidance document as soon as possible. As long as this adjustment to the guidance document has not been made, a VDF of 10 should be applied in core assessments.

3.5.2 Use of de Jong et al. (2010) guidance for non-target arthropod field studies (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

The experts at the meeting acknowledged that using the guidance by de Jong et al. (2010) is useful and that some aspects of the guidance should be used for EU-level assessments until further guidance for the evaluation of NTA field studies is available (see Appendix H of the EFSA report of this meeting (EFSA, 2019)).

Background information

Currently there is no agreed guidance at the EU level for the evaluation of NTA field studies. This may lead to differing evaluations at EU level and frequently to discussion points in experts' meetings. Harmonisation of the evaluation of field studies would therefore be beneficial. A possible option would be to use aspects of the de Jong guidance (de Jong et al., 2010) which is also suggested by the EFSA NTA Opinion (EFSA PPR Panel, 2015) as a guideline for summarising and evaluating NTA field studies until further guidance is developed.

It was proposed by EFSA to start using aspects of the de Jong et al. (2010) guidance for EU-level assessments in order to have a more harmonised assessment of higher tier NTA studies. The experts acknowledged that using the guidance from de Jong et al. (2010) has advantages and that some aspects of the guidance should be used for EU-level assessments until further guidance is

available for evaluating NTA field studies. The elements agreed have been in included in a template in Appendix H of the EFSA report of this meeting (EFSA, 2019). It is recommended that this template is followed when reporting the studies in the RARs/DARs.
In using the guidance, the experts agreed that the level of aggregation/detail as proposed in Table 2 of de Jong et al. (2010) is useful for summarising the results. Consequently, it should be included in the study summary presented in the RAR/DAR. All experts agreed that the taxa listed in Table 4 of de Jong et al. (2010) should be used as a reference for the reliability assessment. Footnotes to Table 4 were missing from the guidance and EFSA contacted the author who made them available. For ease of reference, the footnotes to Table 4 are summarised at the end of Appendix H. It was agreed that further information and argumentation should be presented when specific taxa are missing in the field study. The experts also agreed that studies should include a toxic reference item or to apply rates of the test item high enough to cause clear effects. If a suitable toxic reference was not available, unless effects were clearly seen with the test item, the study should be classified as 'unreliable'. The experts agreed that presenting the results in terms of effect classes as suggested by de Jong et al. (2010) are recommended but should not be considered mandatory.

During the CZHW in Dessau 2018, it was discussed that MSs should consider establishing an ad-hoc group to gather information, consider case studies and exchange expertise regarding the evaluation of NTA-field studies. The purpose of the ad-hoc group is to increase the limited experience of MSs in the central zone in the evaluation of NTA-field studies and to address/highlight some gaps in the current De Jong guidance. The idea was welcomed and agreed but not yet implemented. DE is planning to initiate such process on short term.

3.5.3 Use of the minimum detectable difference for interpreting field studies on non-target arthropods (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

The MDD is considered by the experts as a valid tool for evaluating the biological results. Although it could give some information for the assessment of higher tier studies, overall it was considered premature to recommend calculation of the MDD for higher tier studies performed with non-target arthropods, because criteria to help interpret these MDD values are currently lacking (e.g. classes of MDD, minimum number of taxa with an acceptable MDD).

3.5.4 Risk assessment for non-target arthropods when oral exposure is relevant (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

It was agreed that, until guidance is developed and adopted, data for herbivorous species should not be requested. In cases where a concern is raised (e.g. based on the mode of action of the active substance), then this should be highlighted in the risk assessment and acknowledged in the EFSA conclusion.

3.5.5 The use of ER50 in the Tier 1 of the risk assessment of NTA (CZHW Ede (NL), June 2022; CZSC May/June 2023)

The MS agreed to use the ER50 from *T. pyri* and *A. rhopalosiphi* in Tier 1 when these are lower than the LR50. Furthermore, it was noted that sublethal effects should always be assessed and reported in the Tier 1 tests.

It was noted that in the meantime (prior to the decisions of this meeting going into force), MS will still potentially receive tests without sublethal/reproduction endpoints reported. It was agreed that this will be addressed in a qualitative way (zRMS to note this in the study evaluation indicating that reproductive effects are more sensitive) and leave this to MS to address (i.e., by requesting extended laboratory tests where possible/desired, by RMM, etc.).

3.6 Earthworms and other soil macro-organisms

The risk assessment methodology for earthworms and other soil macro-organisms has in EU context been elaborated in the Guidance Document on Terrestrial Ecotoxicology (Sanco/10329/2002 rev 2 final).

3.6.1 Natural soils in the refined risk assessment for in-soil organisms (CZHW Dessau (DE), 2018; CZSC February 2020)

The MS agreed that testing one single natural soil additionally to an artificial soil is not suitable for:

1. skipping the correction factor of 2 for the endpoints of lipophilic substances and
2. overwriting toxicity endpoints based on artificial test soils by the test result of a natural soil

A minimal number of different test soils, which would be appropriate in order to refine the risk for soil organisms, was not defined. MS agreed that it is not acceptable to use natural soils for the risk assessment as long as no further results from ongoing research projects are available.

3.6.2 Use of de Jong et al. (2006) guidance for earthworm field studies (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

Earthworm field tests are carried out according to ISO 11268-3 (2014). In 2006, guidance on how to summarise those studies was published by de Jong et al. (2006).
The guidance gives recommendations on a number of items which should be considered when assessing the reliability of an earthworm field study. EFSA proposed to adopt the approach described in this document for summarising and evaluating the earthworm studies in the RARs/DARs. Overall, the experts at the meeting agreed that the recommendations from de Jong et al. (2006) could be very useful, but some modifications were proposed. The elements agreed have been included in a template that is provided in Appendix I of the EFSA report of this meeting (EFSA, 2019). It is recommended that this template is followed when reporting the studies in the RARs/DARs.

3.6.3 Use of the minimum detectable difference for interpreting field studies on earthworms (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

The MDD is considered by the experts as a valid tool for evaluating the biological results. Although it could give some information for the assessment of higher tier studies, overall it was considered premature to recommend calculation of the MDD for higher tier studies performed with soil organisms, because criteria to help interpret these MDD values are currently lacking (e.g. classes of MDD, minimum number of taxa with an acceptable MDD). However, from the discussions during the CZHW in Brno (2019) the following could be concluded:
“The concept of MDD refers to the magnitude of the effect that needs to exist in the treatment population in relation to the control in order to obtain a statistically significant difference in hypothesis testing. The MDD concept is very beneficial for the interpretation of the suitability of field studies for risk assessment purposes. Even if further criteria need to be developed in order to conclude on the full implementation of the information in the assessment, calculation of the MDD nevertheless indicate which effect range can be statistically detected by the assessed study and which not."

3.6.4 Field studies with soil mesofauna (CZHW Ede (NL), June 2022; CZSC May/June 2023)

The MS agreed to the following:

(1) It is recommended that the study will be started in spring, however, exceptions are possible (e.g. summer) when it is proven that the community is adequately representative in terms of abundance and species diversity (see point 2) and that exposure was sufficient. It is further noted that the study should adequately represent the realistic field effects considering the GAP. In particular, the potential for recovery after application during autumn is potentially lower. For a GAP with applications in autumn, the study may therefore need to be performed in autumn, depending upon e.g. the types of effects observed.

(2) It is recommended that Oribatid and Gamasid mites must be included in the determination. If possible, other Co-horts like *Prostigmata* (Trombidiformes), *Astigmata* and *Uropodina* should be included. Determination should be done at species level, if possible.

(3) It is recommended two soil layers will be included: 0-5 and 0-10 cm. In some cases, it might be necessary to check other soil layers of 0-1 or 0-2.5 cm, e.g., in case of (a) highly adsorptive substances and (b) substances with a high potential for accumulation.

3.6.5 Analytical measurements toxicity studies with soil organisms (CZHW Ede (NL), June 2022; CZSC May/June 2023)

1) The majority of CZ MS agreed to consider the substance “unstable” if the tier 1 (laboratory) DT90 does not cover the exposure phase of the test. This decision will be used at the active substance and product level.

2) The analytical measurements must be performed at least at the start, middle, and end of the study. The intermediate measurements should be designed to capture the degradation of the substance (i.e., substance property dependent). Depending upon substance degradation it could be concluded that an “end” measurement is not relevant/necessary.

3) The TWA or geometric mean measured concentration should be calculated over the duration of the test and used if the concentration falls under 80% of nominal.

4) If analytical measurements are not available when they should be according to the aforementioned criteria, the reliability of the test would be lower and in most cases a new study including analytical measurements would be required.

5) For products containing multiple actives where only “unstable” substance(s) is (are) measured, appendix J of EFSA (2019)[6] shall be followed to calculate the appropriate endpoint.

3.7 Micro-organisms

The risk assessment methodology for soil micro-organisms has in EU context been elaborated in the Guidance Document on Terrestrial Ecotoxicology (Sanco/10329/2002 rev 2 final).

3.7.1 Soil nitrification studies-time intervals for effect calculations (CZHW Ede (NL), June 2022; CZSC May/June 2023)

The MS agreed to always use the intermediate time intervals for expression of the effect endpoints for risk assessment.

3.8 Non-target terrestrial plants

The risk assessment methodology for non-target terrestrial plants has in EU context been elaborated in the Guidance Document on Terrestrial Ecotoxicology (Sanco/10329/2002 rev 2 final).

3.8.1 Endpoint based on phytotoxicity (EFSA PPR Meeting on general recurring issues, (EFSA, 2019))

It was agreed that an endpoint based on phytotoxic effects should be reported in the study summary and in the list of endpoints. Moreover, such an endpoint should also be used in the risk assessment where relevant. Such an interim solution should be reflected in the (European Commission, 2002) document and its implementation should be further considered.

[6] EFSA technical report: Outcome of the pesticides peer review meeting on general recurring issues in ecotoxicology, June 2019

The majority of the MSs agreed that phytotoxicity endpoint should be considered in the risk assessment, in line with EFSA Technical Report (2019), i.e. all effects and endpoints will be reported in the study summary and the lowest endpoint should be used by the zRMS ensuring an harmonized risk assessment at zonal level.

Background information

This issue was proposed and presented by a representative from the central zone. In the test guidelines for seedling emergence, OECD TG 208 (OECD 2006a) and vegetative vigour, OECD TG 227 (OECD 2006b), other variables than biomass, such as visual phytotoxicity, and sometimes shoot length, are evaluated according to these respective guidelines. ERX values for visual observations (also referred to as 'visible detrimental effects' or 'visual injury', such as chlorosis, necrosis, wilting, leaf and stem deformation) could be determined, where a dose–response relationship is available. The experts at the meeting discussed the relevance of using this endpoint in the Tier 2 risk assessment. The experts considered that effects on growth may also cover the phytotoxicity endpoint, which may be subjective being based on visual assessment. However, it was noted that the EFSA PPR Panel (2014) reported that for a significant number of cases this endpoint was reported as being lower than the others. Therefore, considering that the endpoint is part of the test guidelines and that the data requirements do not specify the parameters to define the endpoint for risk assessment, the experts concluded that the ERX based on phytotoxicity should be reported in the study summary and in the list of endpoints. Where the derived endpoint is the lowest of those available, it should be considered for the Tier 2 risk assessment. Such an interim solution should be reflected in the (European Commission, 2002) document and its implementation should be further considered.

3.8.2 Multiple applications in NTTP risk assessment (CZHW Ede (NL), June 2022; CZSC May/June 2023)

The majority of MS agreed to use the same MAF as the Northern Zone. It will be clarified in the Central Zone Evaluation Manual that no refinement based upon DT50 is accepted for vegetative vigor, as this is in line with the NZ policies.

3.8.3 Deviation from test conditions (but not from validity criteria) in NTTP testing (CZHW Ede (NL), June 2022; CZSC May/June 2023)

The MS agreed that the CZMS will carefully evaluate NTTP tests for major deviations from recommended conditions (e.g., temperature, humidity, plant density). Furthermore, if unexpectedly low toxicity is observed for herbicides, a comparison will be made with efficacy screening data to check, e.g., whether appropriate sensitive species have been tested. On a

case-by-case basis it may be necessary to have new tests, or to decline from using tests with major deviations in SSDs.

3.8.4 Aquatics and NTTPs – SSD (CZHW Ede (NL), June 2022; CZSC May/June 2023)

A proposal was presented for the harmonized evaluation and interpretation of SSD data.

The MS agreed to use the approach when evaluating SSDs (aquatic and NTTP) in future dossiers and to bring the paper forward to the EFSA to be considered in the next general issues meeting. The proposal is presented in Appendix 2.

Appendix 1: Test Validity of OECD 201 (algae; species other than recommended): stepwise approach when not met

Background information

This working agreement provides guidance in situations where the performance criteria (validity of the test) are not met in tests performed with algal species other than recommended[7] by OECD 201. As result of a species-specific growth pattern the performance criteria might not be met. A stepwise approach was developped to analyse the response of the unexposed control cultures which can help to decide whether or not the test (with species other than recommended) can be considered valid. OECD 201 states the following concerning test validation:

VALIDITY OF THE TEST

11. For the test to be valid, the following performance criteria should be met:

- The biomass in the control cultures should have increased exponentially by a factor of at least 16 within the 72-hour test period. This corresponds to a specific growth rate of 0.92 day^{-1}. For the most frequently used species the growth rate is usually substantially higher (see Annex 2). This criterion may not be met when species that grow slower than those listed in Annex 2 are used. In this case, the test period should be extended to obtain at least a 16-fold growth in control cultures, while the growth has to be exponential throughout the test period. The test period may be shortened to at least 48 hours to maintain unlimited, exponential growth during the test as long as the minimum multiplication factor of 16 is reached.

- The mean coefficient of variation for section-by-section specific growth rates (days 0-1, 1-2 and 2-3, for 72-hour tests) in the control cultures (See Annex 1 under "coefficient of variation") must not exceed 35%. See paragraph 49 for the calculation of section-by-section specific growth rate. This criterion applies to the mean value of coefficients of variation calculated for replicate control cultures.

- The coefficient of variation of average specific growth rates during the whole test period in replicate control cultures must not exceed 7% in tests with *Pseudokirchneriella subcapitata* and *Desmodesmus subspicatus*. For other less frequently tested species, the value should not exceed 10%.

18. If other species are used, the strain and/or origin should be reported. Confirm that exponential growth of the selected test alga can be maintained throughout the test period under the prevailing conditions.

Please note that *less frequently tested species* (e.g., CV of average specif growth rates) concern the recommended species of diatom and cyanobacteria groups.

Afspraak voor risicobeoordeling

In cases where the validity of the test (OECD 201) is not met the following stepwise approach should be followed:

A1 The biomass in the control cultures should have increased exponentially by a factor of at least 16 within the 72-hour test period. If not, the test is not valid.

[7] recommended species: the green algae: *P. subcapitata* and *D. subspicatus*, the diatom: *N. pelliculosa* and the cyanobacteria: *A. flos-aquae* and *S. leopolienesis*). Please note that: *Pseudokirchneriella subcapitata* is called *Raphidocelis subcapitata* now, and was known as *Selenastrum capricornutum. Desmodesmus subspicatus* was formerly known as *Scenedesmus subspicatus*.

If exponential growth is observed → B1

B1 Look at the mean CV[8] for section-by-section specific growth rates (e.g., assess effects on the pattern of growth).

The growth factor per day = biomass (number of cells) at 24h ÷ 0h, 48h ÷ 24h and 72h ÷ 48h). Please note that the product of these daily growth factors equals the overall growth factor of A1.

In case growth factors are variable this might indicate that the observed pattern might be species-specific.

Consider the growth factors of each time frame (e.g. 0-24, 24-48 and 48-72 hours) in combination with that of the whole period. Is growth steady over the time course or is it decreasing? In that case have a look at the growth factor in the last 24 hours. Is it still reasonable (growth factor > ~2.5).

A daily growth factor of ~2.5 equals a 72-hours growth factor of 16, when considering constant exponential growth, which is the minimum requirement of OECD 201 (see A1). Therefore, a growth factor of ~2.5 in the last 24 hours is only just acceptable. However, this should not be applied too strict. Always look at the whole picture. A flattening growth is often observed in studies with non-standard species, also in the case of standard test species.

A cut-off value up to 50% can be considered acceptable.

The cut-off value of 35% of OECD 201 is drawn up and validated for the recommended species. t is therefore, considered secondary in case of other species. In OECD 238 and 239 the following is stated: ECx values are only reliable and appropriate in tests where coefficients of variation in the control fall below the effect level being estimated (after OECD 238 and 239). Therefore, coefficients of variation should be < 50% for robust estimation of an EC_{50}.

B2 Look at the CV of average specific growth rates (e.g., variation between replicates).

[8] CV = variability of a parameter

If the CV of average specific growth rate is acceptable (e.g., < 10%) the variation between replicates is acceptable and the observed variable growth rate is most likely species-specific.In case the CV is much lower than the cut-off value this could support the validity of the endpoints.

Substantial differences between section-by-section specific growth rate and average specific growth rate indicates a deviation from constant exponential growth and close examination of the growth curve is warranted.

B3 Are the confidence intervals of the E_rC_{50} and E_yC_{50} wide or narrow (e.g., degree of precision). In case it is narrow, this could support the validity of the study endpoints. For this the normalised width of the confidence interval approach of EFSA Supporting Publication 2019:EN-1673 (paragraph 2.1 and table E9 of Appendix E) should be used.

NW = *(EC_x, upp – EC_x, low) / EC_x, med*

NW	**Rating**
< 0.2	Excellent
0.2 – 0.5	Good
< 1	Fair
< 2	Poor
≥ 2	Bad

B4 Weigh the overall quality of the study (is the study well performed, e.g. according to the general requirements of OECD 201, was the duration of the study long enough? if prolonged would it have resulted in a lower EC_{50}?). Relevant issues should be discussed.

Use B1 to B4 in a WoE approach to accept or reject the exceedance of the mean CV for section by section specific growth rate and / or the CV of average specific growth rate.

Appendix 2: Proposal for the 6th Central zone harmonization workshop, June 2022. SSD and its exemplary use for aquatic organisms and non-target terrestrial plants- data selection and statistical procedure -

List of abbreviations

AGD	Aquatic Guidance Document
a.s.	Active substance
CI	Confidence Interval
cZone	Central Zone
d.w.	Dry weight
EC	Effect Concentration
ED	Effective Dose
EP	Endpoint
ER	Effect Rate
HC_5	5th percentile of the Hazard Concentration
HR_5	5th percentile of the Hazard Rate
ini	Initial concentration
LC	Lethal Concentration
$LLHC_5$	Lower limit of the confidence interval of the hazardous concentration for 5 % of the species of an SSD
m.m.	mean measured concentration
MoA	Mode of Action
Nom	Nominal concentration
NOEC	No Observed Effect Concentration
NTTP	Non-Target Terrestrial Plants
OECD	Organisation for Economic Co-operation and Development
RA	Risk Assessment
RAR	Regulatory Acceptable Rate
SANCO	Health and Consumer Protection of the European Commission
SE	Seedling Emergence
SSD	Species Sensitivity Distribution
VV	Vegetative Vigour
zRMS	Zonal Rapporteur Member State

Background

This document aims to give detailed guidance for calculating an SSD in ecological risk assessment. Beside some general aspects on the SSD approach, this document deals with the application of the SSD for aquatic organisms and for NTTP. Therefore, it also points out some specific aspects to consider for each of these groups

Recommendations presented in the current document follow those reported in chapter 8. of the Aquatic Guidance Document (AGD) (EFSA Journal 2013;11(7):3290). When judged necessary, further explanations were added based on concrete experiences gained from the

regulatory practice.
The focus is on the selection of data and the statistical procedure.

The application of the SSD approach for NTTP is described in the Guidance Document on Terrestrial Ecotoxicology (TGD, SANCO/10329/2002 rev 2 final). But this document needs to be urgently revised including the section related to SSD that do not provide much recommendations. Therefore, in this document the recommendations provided for aquatic organisms (EFSA 2013) are analysed in order to assess if they could be applied to NTTP.

To facilitate the reading, specific approaches concerning aquatic organisms and NTTP are presented in separate columns.

Crucial aspects for each section

Data selection:

- For aquatic organisms, follow recommendations of EFSA (2013). Special emphasis regarding insecticides, herbicides and fungicides are given in chapters 8.4.3.1, 8.4.3.2 and 8.4.3.3 of the AGD, respectively.
- For NTTP follow recommendations of SANCO/10329/2002 rev 2 final given in chapter 7.1.
- Be aware of the representativeness of the taxa tested regarding the specific MoA of the a.s.
- Select the same estimates (e.g. EC_{10}; ER_{50}) and preferentially identical variables to calculate an SSD. Note that similar variables as dry weight and fresh weight might be mixed to assess the variable biomass for primary producers (aquatic and NTTP) or for invertebrates.
- EPs should also be expressed with same concentration or rate units.
- Verify that the EPs used are reliable (e.g., calculate the normalised CI around the EP)
- Different test designs – i.e. Tier 1 and tier 2C data (aquatic organisms) and VV and SE data or laboratory and field or semi-field studies (NTTP) cannot be mixed.

Statistical procedure:

- Check detailed procedure regarding censored EP and make sure that the minimum data requirement to conduct an SSD for this organism group is fulfilled.
- Check if the data is unimodal and fits adequately the assumed distribution (e.g. log-normal or log-logistic)
- Check the reliability of the results, with a particular emphasis on the fit and thus choice of the model (log-normal, log-logit, Weibull…)

Special case of primary Producer in aquatic

- If the minimum data requirement is not met because of too many censored E_rC_{50}, instead of going back to lower Tier, we propose the possibility to calculate the SSD with E_yC_{50} values.

Application examples:

- Example on how to report the results as zRMS (approaches 1 and 2)

Selection of Toxicity Data

Effect Side

Selecting toxicity data on the basis of toxic mode of action of the substance

Be aware of the representativeness of the taxa tested regarding the specific MoA of the substance.

Aquatic organisms	NTTP
No deviation to AGD. Follow chapters 8.4.2 and 8.4.3 (p. 92): "*If, for example, the First tier toxicity value for Chironomus is an order of magnitude lower than that of Daphnia and/or Americamysis bahia, it is recommended to construct, in the first instance, a SSD with toxicity data for insects, or to explore which insects and crustaceans (e.g. macro-crustaceans) can be combined in a single SSD on the basis of all relevant information available*." (AGD 2013) As another example for primary producers, in case of auxin herbicides, dicotyledonous species are usually more sensitive. Thus, check that this group is sufficiently represented in the data set and consider constructing an SSD with only dicotyledonous species. In addition, check if rooted macrophytes are sufficiently represented as well.	No deviation to SANCO/10329/2002 rev 2 final (chapter 7.1, Tier2): "*In order to generate data that are useful for probabilistic approaches there should not be a focus exclusively on species assumed to be the most sensitive. If, from the screening data, a specific mode of action is evident, or strong differences in the species sensitivities are identified, this evidence should be used in the selection of the appropriate test species*." E.g., if the First-tier toxicity values are lower for dicotyledonous (which might be the case for auxin herbicides), it might be recommended to construct, in the first instance, an SSD with toxicity data for this group if possible.

Further information regarding the sensitivity of the non-target organisms against the a.s. under evaluation can be found in the respective EU-LoEP(s)/D(R)AR(s) and in addition for NTTP in the efficacy data (c.f., CA B3 or D(R)AR Vol.3 CA/CP -B.3 for zonal and EU applications, respectively). Note that screening data submitted for the evaluation of herbicidal activity of metabolites might also be informative.

Estimates and variables

Terminology:

Endpoint: is the combination of an estimate and a measured variable.
Estimates: is referring to the magnitude of effect described (e.g., ECx, NOEC …)
Variables: is the response variable measured

Aquatic organisms	NTTP
Estimates	
E_rC_{50}: EC_{50} calculated with growth rate EyC_{50}: EC_{50} calculated with yield EbC_{50}: EC_{50} calculated with area under the curve EC_{10}: e.g. reproduction, body weight EC_{50}/LC_{50}	ER_{50}
Variables	
Algae: cell counts (surrogate for biomass and thus most frequently called “biomass”) Macrophytes: frond number, frond area, biomass wet weight, biomass dry weight etc...	Seedling emergence: emergence, mortality, biomass (fresh weight/ dry weight), plant height, visual injury Vegetative vigour: biomass (fresh weight/ dry weight), plant height, mortality, visual injury
Selection of estimates and variables in SSD calculation	
Select identical estimates and preferentially identical measured variables However, for aquatic and terrestrial primary producers, wet weight and dry weight might be pooled to assess the variable biomass (see section 7.1).	
Specific recommendations available for aquatic organisms: **Acute risk assessment:** The AGD sees the possibility to construct an SSD based on NOEC/EC_{10} values. However, no further recommendations are provided regarding the decision making for regulation (*i.e.*, which approach should be then preferred?). In general, LC/EC_{50} values are most robust and reliable and should be used for constructing an SSD. An SSD based on NOEC/EC_{10} values might be suitable in cases when LC/EC_{50} are less reliable (e.g. in case of very steep dose-response curves).	No further specific recommendations available. The SSD is simulated with ER_{50} values as recommended in SANCO (2002)
Chronic risk assessment: Classically, NOEC or EC_{10} values are available for multiple biological variables (e.g., reproduction, body weight, body length...). Select same estimates (e.g. only EC_{10} values) and preferentially identical biological variables as underlying data for an SSD. EC_{10} is the preferred estimate.	

Exposure Side

Test design

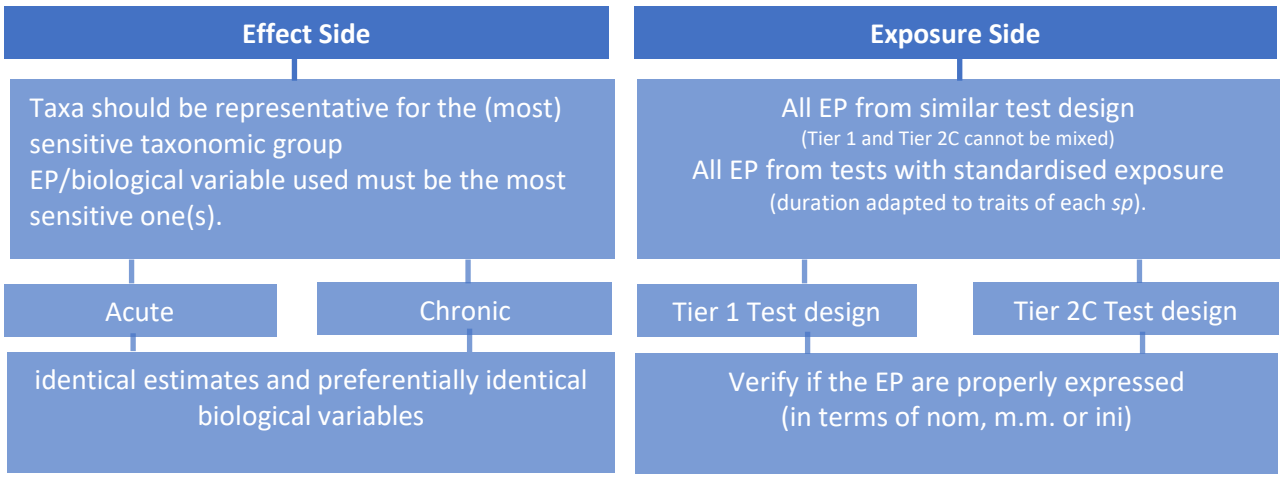
Aquatic organisms	NTTP
Different test designs cannot be mixed	
Note that Tier 1 and Tier 2C data cannot be mixed within an SSD. **SSD based on Tier 1 data:** All endpoints used for the SSD are derived from standard (i.e. OECD) tests; however please note that the duration of the test might differ according to the traits of the tested species (e.g. 48 h for *D. magna* but 96 h for *A. bahia*), as mentioned in the AGD under 8.4.2. Note that for certain insect growth regulators, the standard duration (48–96 hours) of the acute toxicity test may not be sufficient, since latency of effects may occur (refer to AGD 2013, p. 94). **SSD based on Tier 2C data**: In theory, it is possible to calculate an SSD with EPs derived from refined exposure tests (e.g. pulses and/or water-sediments lab tests, i.e. Tier 2 C). In practice, this is problematic since there are a number of critical issues for refined exposure test. In such case, it has to be carefully verified that each single refined exposure test is acceptable for risk assessment.	ER_{50} cannot be mixed within an SSD if they are from - (i) SE and VV tests or - (ii) from tests having different application methods (sprayed *versus* mixed to the soil) or - (iii) from tests having different duration or - (iv) from tests with different settings (e.g. from laboratory and semi- or field conditions)

Expression of endpoints

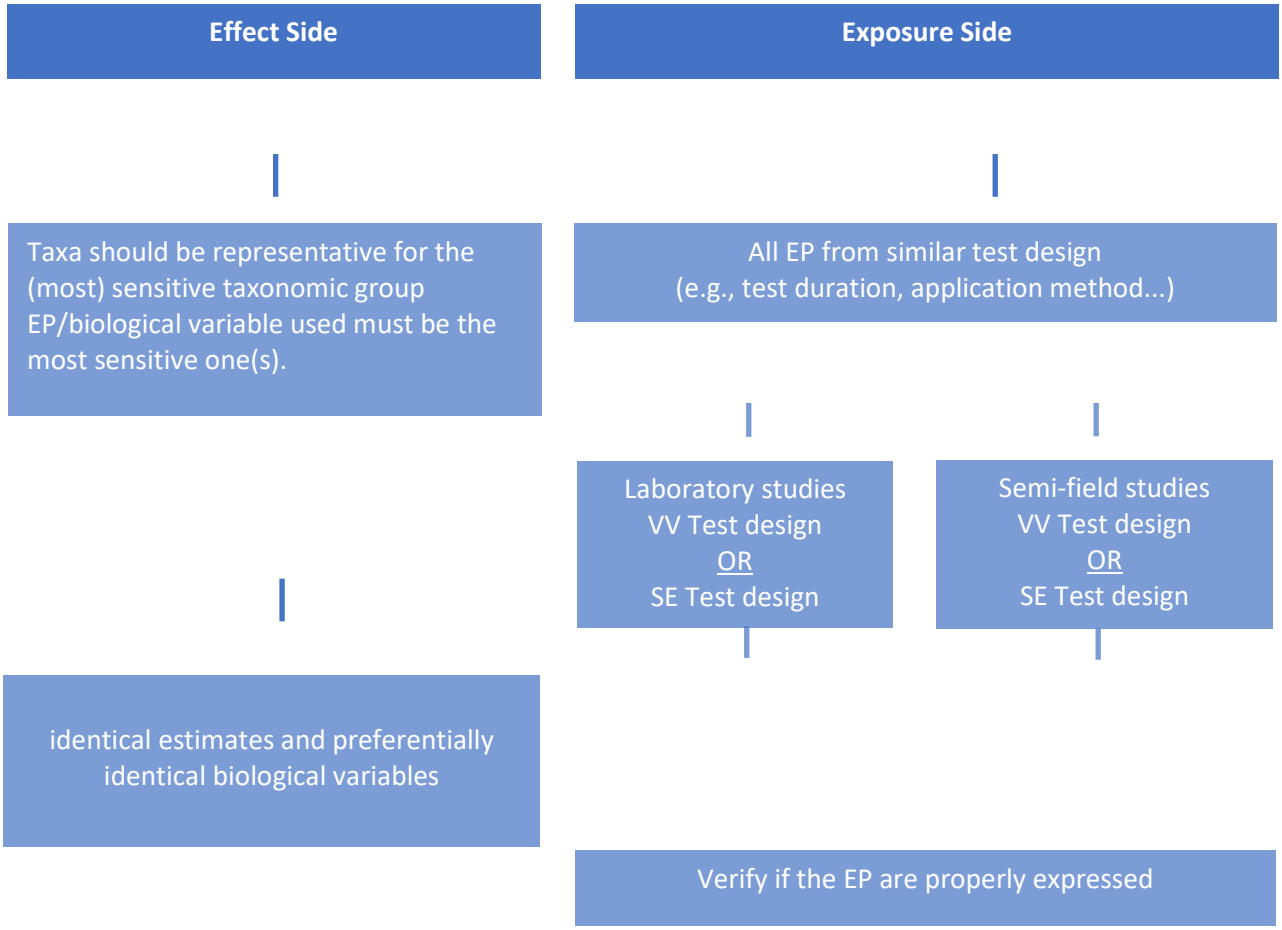
Aquatic organisms	NTTP
For both Tier 1 and Tier 2C tests, carefully verify that the EP is properly expressed in terms of e.g., nom, m.m., or ini. concentrations. . Please refer to section 3.1 in EFSA Supporting publication 2015:EN-924 as well as to Appendix J in EFSA Supporting publication 2019:EN-1673.	All EP should be expressed in the same unit (e.g. in g product / ha).

Summary schemes for data selection

Scheme for data selection for aquatic



Scheme for data selection for NTTP



Statistical procedure

Pooling different types of endpoints

For terminology, please refer to section 5.3.2.

Estimates: Cannot be mixed within an SSD.

Variables: Should in general not be mixed. In case the more sensitive biological variable differs between species (e.g. reproduction for *D. magna versus* body weight for *A. bahia* or plant height *versus* plant biomass for NTTP), different SSD have to be calculated for each variable.
There is an exception for identical variables, such as wet weight and dry weight for aquatic and terrestrial plants (see section 9.1). If available variables differ only slightly, they might be mixed to construct an SSD (e.g. fresh weight and dry weight for primary producers or invertebrates.

For the special case of aquatic primary producers, please refer to section 7.

Censored endpoints

Some endpoints might be expressed as censored values, i.e. less than (<) or greater than (>) values.
Censored EPs are also referred as “unbound values” in AGD.

In principle, censored EP can be dealt as recommended in EFSA (2013),” *i*.e. to include censored EP as “= value” in the SSD data set, only when those EP are out of the range of sensitivity of the species tested. Censored EP within the range of sensitivity of the species tested should be excluded from calculation”. Additionally, EFSA 2013 recommends to conduct an SSD with this potentially restricted data set, only if the minimum number of EP needed for calculation is still required (*i.e.*, n ≥ 8 and n ≥ 5 for fish). See also section 6.3.1 and 2. below for more details. In case this minimum requirement is not fulfilled, the SSD refinement option should be rejected.

We suggest to enlarge these recommendations to NTTP. This means that in case censored ER_{50} are part of the data set, they should be treated as recommended in EFSA (2013), i.e. > or < ER_{50} should be further considered only when they are out of the range of sensitivity of the species tested. The minimum number of EP available for SSD calculation should be n ≥ 6 as reported in SANCO/10329/2002 rev 2 final.

Calculation

Following calculation methods for SSD simulations are possible:

- ETX program: It is the usual approach considering lognormal models and non-censored endpoints.
- R-package fitdistrplus: it is developed by Sandrine Charles from the University of Lyon and implemented in the platform MOSAIC (https://mosaic.univ-lyon1.fr/ssd)[9]. This program has many advantages since:
    - (i) it considers censored values,
    - (ii) it takes confidence interval into account, which is particularly relevant when uncertainties around the EP exist (i.e., large CIs, which often occur in case of NTTP); with this approach, relevant available information regarding the robustness and reliability of the single estimates is included in the SSD, and
    - (iii) it is possible to apply different models (log-normal, log-logistic, Weibull…), whereas in ETX only the log-normal model is used.

UBA developed an Excel Tool connected with R to implement the R-package fitdistrplus. It has been published by UBA on the EFSA Knowledge Junction platform Zenodo on 26 October 2022: https://zenodo.org/record/7249239

[9] Kon Kam King G. Veber P., Charles S., Delignette-Muller M. L. (2014) MOSAIC_SSD: A new web tool for species sensitivity distribution to include censored data by maximum likelihood. Environmental Toxicology and Chemistry 33(9) 2133-2139

Pre-requisite for SSD calculation	
Aquatic organisms	**NTTP**
Sufficient representative toxicity data according to the AGD must be available (see AGD p. 92-93; i.e. n ≥ 5 (only for fish) or n ≥ 8) after that censored EP in the range of species sensitivity have been excluded from data set.	Sufficient representative toxicity data according to SANCO/10329/2002 rev 2 final must be available, the minimum requirement is n ≥ 6 for NTTP. Thus, we suggest a minimum of 6 available ER_{50} after that censored EP in the range of species sensitivity have been excluded from data set.

For calculation, we propose:

- Approach 1: to follow EFSA (2013) that recommends to simulate an SSD only with censored EP that are out of the range of species sensitivity of non-censored EPs (see below)
- Approach 2: additionally, to simulate an SSD with the whole data set (*i.e.*, using all censored and non-censored EP) by using the R-package fitdistrplus (see below). Indeed, in case censored endpoints and/or confidence intervals are available in the SSD data set, approach 2 (R-package fitdistrplus) might be more appropriate more reliable, as the results of the simulations consider more information than only the EP. See also Green, 2016 and 2018[10,11].However, results of the R-package fitdistrplus simulations might be more complex to evaluate.

Decision on which approach (i.e., 1 or 2) as well as which simulation models is the most appropriate (*i.e.*, log-normal, log-logistic, Weibull…) should be done on a case-by-case basis considering the recommendations provided in section 5.4. In case of the inclusion of "bigger than" censored values (e.g., LC_{50} > 10 mg a.s./L), the approach with *fitdistrplus* provides in our view more reliable results as it considers intervals as such (e.g., LC_{50} > 10 mg/L a.s. belongs to the interval 10; +∞; see below)

Approach 1: Data selection according to EFSA (AGD 2013)

Data are selected excluding censored EPs in the range of species sensitivity and the SSD is performed according to AGD (p. 92-93). Censored EPs out of the range of species sensitivity are considered as non-censored EPs in the SSD (e.g. > 42 mg/L is considered as 42 mg/L).

Although no specific program for SSD calculation is recommended in the AGD and in SANCO (2002), the program ETX is commonly used by MS.
However, we also recommend to use the R-package fitdistrplus as it can consider more than only the lognormal model. Moreover, this approach also takes confidence intervals of single estimates into account, which might be particularly relevant for NTTP (see 5.3 above).
Take decision on which model is the most appropriate according to section 5.4.1.

[10] Green (2016) Species Sensitivity Distribution with censored values. SETAC (Nantes) 2016.

[11] Green, Springer & Holbech (2018) Statistical Analysis of Ecotoxicity Studies ISBN: 978-1-119-48881-1| July 2018| 416 Pages|

Approach 2: Including all censored EP

First, data are selected excluding censored EPs in the range of species sensitivity as in approach 1 (see 5.2). Then, Approach 2 is applied only if sufficient toxicity data according to EFSA (2013) and SANCO (2002) are still available.

In approach 2, data used for the SSD include all censored EP (i.e., within and outside the range of species sensitivity of non-censored EPs) and censored EP are considered as such in the SSD (e.g., LC_{50} > 10 mg a.s. is used as interval : 10; +∞). The SSD is modelled with the R-package fitdistrplus (e.g., available in the platform MOSAIC).

The particularity of the R-package fitdistrplus is that the program can treat "interval values". This means that the package can treat Confidence Intervals (CI) as well as Censored Endpoints.
Indeed, censored values belong to an interval. E.g., LC_{50} > 10 mg a.s./L belongs to the interval [10; +∞[; LC_{50} < 10 mg/L belong to the interval]-∞; 10].

(ix) Uncertainty: Perform the SSD analysis with the Confidence Intervals (CI) of EP.
(x) Censored values: Enter all censored endpoints as an interval as described just above.

When reporting the results with R-package fitdistrplus add the following:
"*SSD calculation is conducted with the R-package fitdistrplus, which allows including censored data and consideration of confidence intervals (for details see https://doi.org/10.1002/etc.2644*)

Note that a detailed example of Approach 2 is given in section 8.

Reliability check

Model selection and model fit

If a calculation method is chosen that enables the application of different models (such as the R-package fitdistrplus), it is advised to fit several models (log-normal, log-logistic, Weibull…) and to compare different criteria to select the model (e.g. Akaike Information Criterion (AIC)). The best fitting model should be selected. Also test statistics from the goodness of fit estimations can be considered for model comparison.

The quality of the model, especially the fit of the underlying distribution, should be checked (i) by visual inspection of the output graph and (ii) if possible the qq-plot (*e.g.* does the model reflect the assumed distribution of the EPs?). If available goodness of fit estimations such as the Cramér–von Mises test can be considered to check if the underlying distribution is significantly deviating from the data set. Note that the check of the model fit and selection might result in the rejection of the SSD simulation.

Furthermore, we highly recommended to check the width of the confidence interval around the median HC_5. Indeed, the model underlying an SSD is always linked with uncertainties expressed in an interval – the confidence interval. Thus, the confidence interval provides the uncertainty of the model and is dependent on the model structure, data structure, and fitting method. Given the uncertainty of the model, the median HC5 (or just HC5) is estimated to be correct with a probability of 50%, whereas the lower and upper limit HC5 simulate the HC5

with a probability of 95%. It is important to notice that the confidence interval does not provide the confidence existing around the median HC5 but rather provide confidence in the model fit, given that the underlying assumptions of the model are met.
E.g., we advise to compare the position of the $LLHC_5$ to the median HC_5. In case the $LLHC_5$ is less than 1/3 of the median HC_5, reliability and/or protectiveness of the simulated median HC_5 might be questioned (i.*e*., consider rejecting the SSD or eventually select a higher AF or regulate on another Effective Dose proposed below in 6.4.2 below). This is also addressed in the AGD 2013, since it is suggested under section 2.1.4.2 to consider that for "*The lower limit value of the HC5*. If the lower limit HC5 derived from the curve is less than 1/3 of the median HC5, a higher AF in the proposed range may be warranted."

Note also that:

(i) Violation of goodness of fit might be acceptable if the distribution of the data in the lower tail of the SSD is considered as relatively conservative (see AGD 8.4.1).
(ii) In some cases, a split of dataset and conduction of specific SSD might be required (see section 5.3.1 of this position paper or 8.4.1 and 8.4.3 of the AGD).

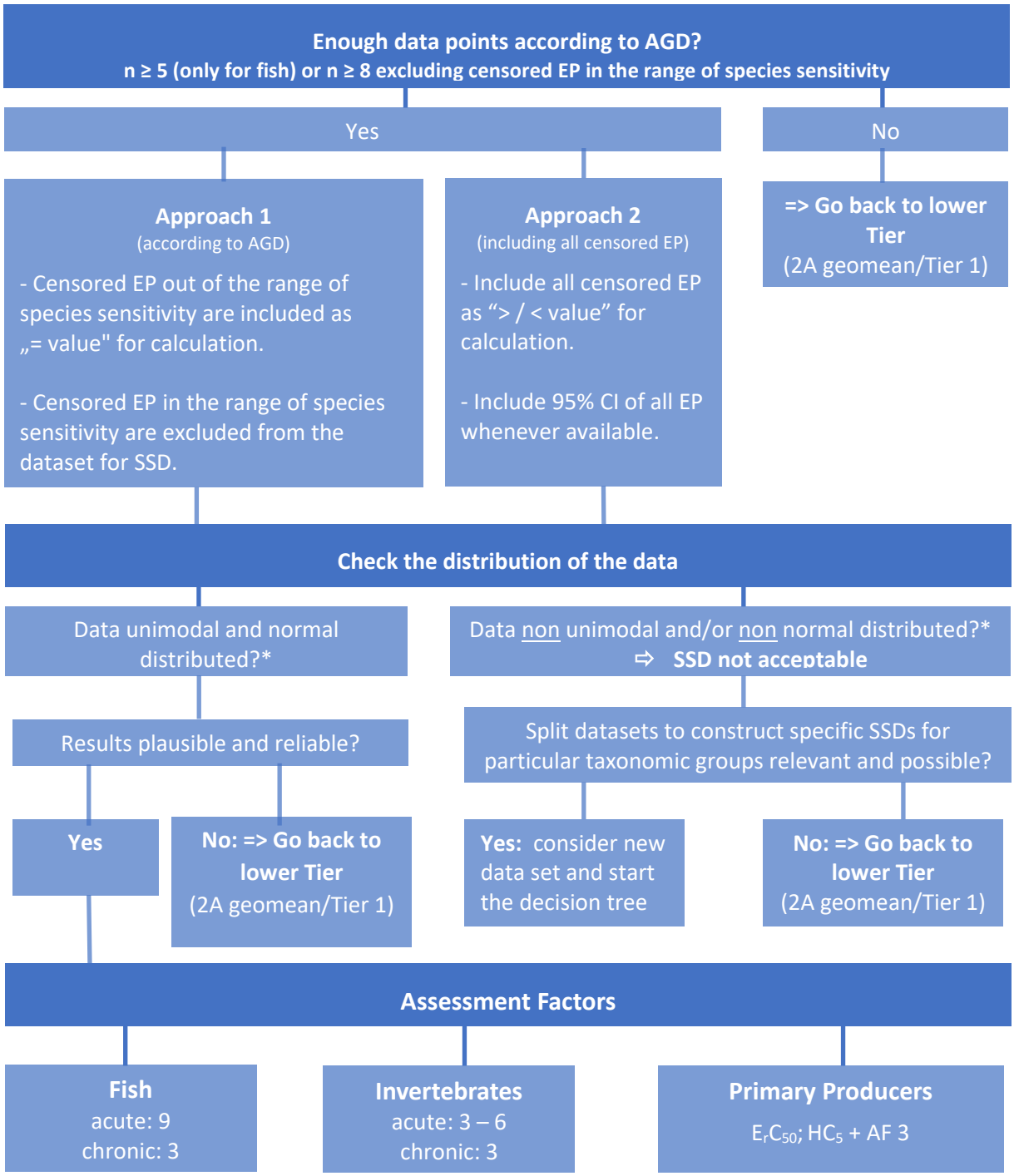
Choice of the AF (aquatic organisms) or relevant Effective Dose (NTTP)

For aquatic organisms, we follow the recommendations provided in EFSA (2013).

For NTTP, SANCO (2002) reports that: "if *the ED50 (Effective dose 50 %) for less than 5 % of the species is below the highest predicted exposure level, the risk for terrestrial plants is assumed to be acceptable*", which corresponds to an AF =1. However, SANCO 2002 does not precise whether the Effective Dose should rely on the median or LLHR5. Thus, we suggest to carefully check which ED (median or $LLHR_5$) is the most appropriate according to some recommendations provided in the check list reported in the table below. Note that these recommendations are adapted from those provided in EFSA (2013).

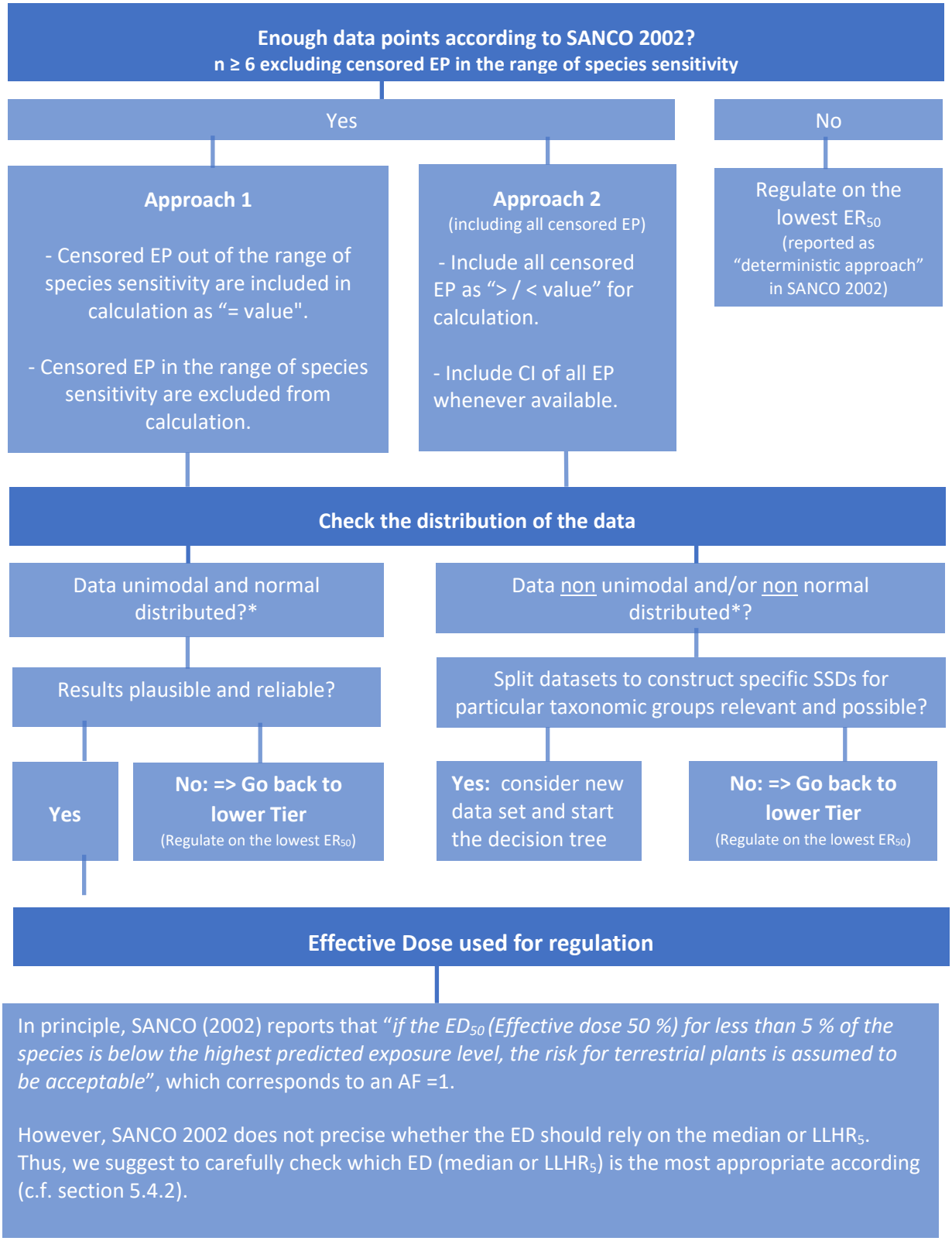
Aquatic organisms	NTTP
Follow recommendations as provided in EFSA (2013) section 2.1.4.2 (p. 20)	We propose to adapt the recommendations provided in EFSA (2013) in section 2.1.4.2, as follow: - If the $LLHR_5$ is less than 1/3 of the median HR_5, then the protectiveness of the median HR_5 should be questioned; the $LLHR_5$ might me better appropriate. - If the median HR_5 is lower than the RAR derived at the lower Tier (i.e., lowest ER_{50}/5), then the relevance of the SSD approach should be questioned. Indeed, in principle following the tiered approach, a RAR higher Tier should be higher than a RAR lower Tier. - Consider the position of the toxicity data in the lower part of the tail of the SSD (around the HR_5). Indeed overall, if they are positioned on the right side of the SSD curve, the derived HR_5 estimate may be considered relatively "conservative" for the most sensitive species. This may indicate that the median HR5 is appropriate. In contrast, if in the lower tail the toxicity data are, overall, positioned on the left side of the SSD curve, this may be a reason to question the protectiveness of the median HR_5. $LLHR_5$ might me better appropriate. - *The steepness of the SSD curve*. In the case of a relatively steep SSD curve (e.g. less than a factor of 100 between lowest and highest ER_{50} value used to construct the SSD curve), the $LLHR_5$ might me better appropriate since exposure concentrations that exceed the RAR may have ecotoxicological consequences for a larger number of taxa. - *Read-across information for compounds with a similar toxic mode of action*. For a PPP with a well-known mode of action, sufficient information on related compounds may be available that allows the evaluation of the predictive value of the median HR_5 and/or lower limit of the HR_5 (e.g. known strong sensitivity of some species but not tested with the PPP under evaluation). This information may be used to decide on the protectiveness of median HR_5 vs $LLHR_5$ or of the whole SSD approach.

Summary schemes of the SSD procedure
Aquatic organisms: scheme for statistical procedure



* Please note that this is a simplification. SSDs should follow the modelled underlying distribution (usually log-logistic or log-normal, which are similar to the normal distribution).

NTTP: scheme for statistical procedure



* Please note that this is a simplification. SSDs should follow the modelled underlying distribution (usually log-logistic or log-normal, which are similar to the normal distribution).

Special case of primary producers in aquatic

Pooling endpoints for algae and macrophytes

Variables for aquatic plants do often differ and the AGD is not specific regarding the pooling of such variables. In case several variables are measured, preferably calculate the SSD for each variable independently and regulate on the lowest HC_5. In case only different variables are measured, a pragmatic approach is used to separate the variables for primary producers in two categories:

i) "weight related" (dry weight, wet weight, biomass)
ii) "growth related" (frond number, shoot length, shoot number…)

SSDs can only be conducted for variables from one category (*i.e.*, i or ii).

The AGD recommends to pool algae and macrophytes in a single SSD for primary producers only under the following conditions:

i) In Tier 1 tests, data (EP) on macrophytes and algae differ less than a factor of 10.
ii) No difference in mode of action leading to a sensitivity difference is described or observed (i.e. algae and macrophytes should be randomly distributed along the SSD curve).

Censored endpoints

The occurrence of censored endpoints is usually more common for the E_rC_{50} estimate than for the E_yC_{50} (or E_bC_{50}) estimates. EFSA (2013) is preferably using the E_rC_{50} estimates but at the same time, EFSA is excluding censored EP from the SSD analysis when they are in the range of sensitivity of uncensored endpoints. Therefore, this might lead in some cases to a restricted data set (n <8) and no possibility to apply the SSD

In case the dataset is too small for an E_rC_{50}-SSD analysis (if for E_rC_{50} EP, n < 8 once censored EP in the range of sensitivity have been excluded), alternatively an E_yC_{50}-SSD might be calculated (if for E_yC50 EP, n ≥ 8, as E_yC50 EP are usually not (or less) frequently censored).

Application examples

Higher tier refinement – SSD aquatic invertebrates

The applicant proposed to refine the short-term risk to aquatic invertebrates by conducting an SSD (Tier 2b). Acute data on aquatic invertebrates (either 48 or 96 hours) are shown in the Table below.

Table: Short-term toxicity data to aquatic invertebrates.

Species	EC_{50} in mg/L	95% confidence intervals
Daphnia magna	0.48	0.34 – 0.69
Asellus aquaticus	3.43	2.75 – 4.26
Gammarus pulex	0.23	0.20–0.25

Neocaridina denticulata	>5	Not available
Procambarus sp.	1.2	0.75–1.93
Chironomus riparius	0.44	0.32–0.59
Anax imperator	1.63	Not available
Cloeon dipterum	0.31	0.26–0.38
Notonecta maculata	2.78	Not available
Paraponyx stratiotata	>4	Not available
Plea minutissima	1.29	0.92–1.80
Ranatra linearis	3.33	2.95–3.76
Sialis lutaria	0.96	Not available

Two approaches are used to model the HC_5:

- The inclusion of censored values outside the range of species sensitivity as non-censored values, using software ETX fitting a log-normal distribution to the toxicity data (i.e., equivalent to Approach 1 in 5.3.1) and
- The inclusion of all censored data and the consideration of confidence intervals, using the R-package *fitdistrplus* (for details see http://ubanet/websites/IV1.3/SG1/FG_Aquatik/FGDokumente/Background%20information/documents-%20publications/Kon%20Kam%20King%20et%20al.%20-%202014%20-%20Environmental%20toxicology%20and%20chemistry%20SETAC.pdf)

The available confidence intervals and censored endpoints shown in the Table are taken into account when fitting the SSD model with the R-package *fitdistrplus* (version 1.0.14).

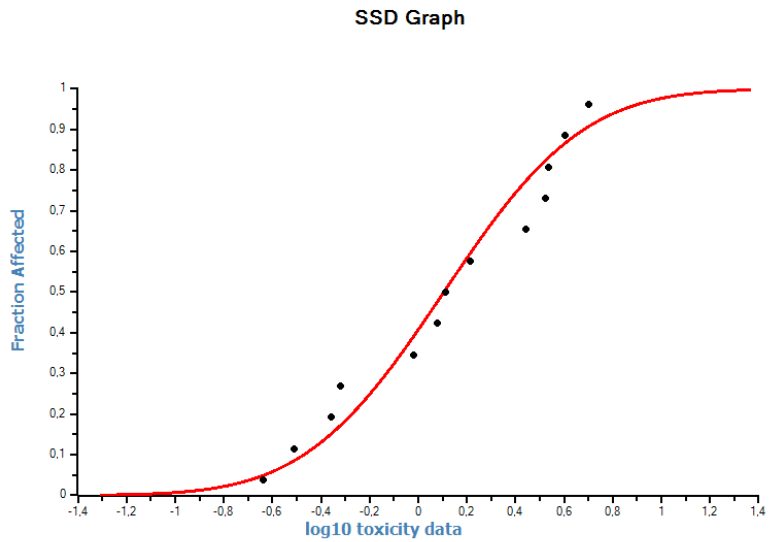
Results ETX:

Test for normality:

Test	**Significance level α = 0.05**
Anderson-Darling	accepted

HC5: 0.223 mg/L (CI: 0.08035 – 0.416)

The fitted model by ETX is shown in the following plot.

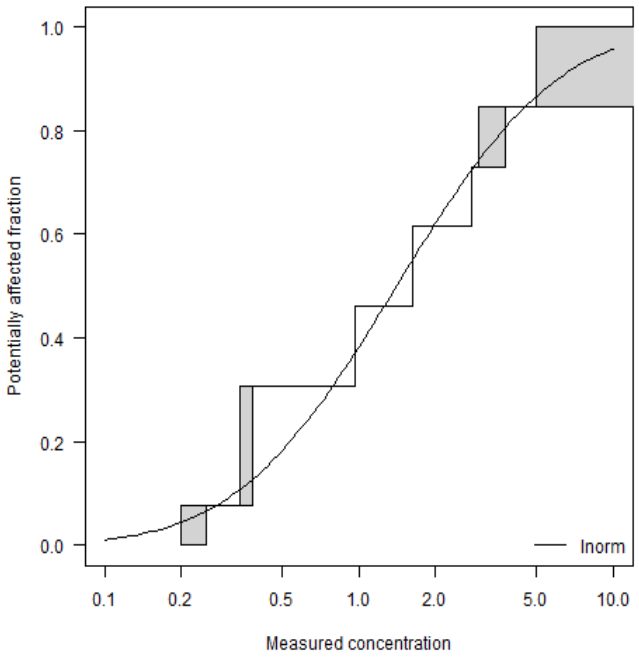


Results R-package *fitdistrplus* (log-logistic model)*:*

Q-Q plot (not displayed here) indicates that a log-normal distribution of the data can be assumed.

HC5: 0.21345 mg/L (CI: 0.11 – 0.56).

The fitted model derived from the R-package *fitdistrplus* including confidence intervals for single endpoints and censored endpoints is shown in the following plot.



Conclusions on the SSD-HC_5:

The derived HC_5 is in general highly dependent on the fitted model and calculation method. To overcome uncertainties, two statistically sound approaches are used and the more reliable approach is selected. The underlying data in the models can be assumed to follow a log-normal distribution. The calculation with *fitdistrplus* allows to take intervals into account, which in this case due to right censored values and available confidence intervals is relevant. Therefore, the calculations with the R-package *fitdistrplus* is more robust and preferred compared to the calculation with ETX. The HC_5 is 0.21 mg/L.

Notes:

- For determination of the precise AF, WoE shown on page 98 and 99 of the AGD should be taken into account.
- Note that in the plot with *fitdistrplus* displays not all data points, as this would result in an unclear graphic illustration. However, all data points are taken into account for fitting the model and calculation of the HC_5.

References:

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013;11(7):3290, 268 pp. doi:10.2903/j.efsa.2013.3290.

EUROPEAN COMMISSION HEALTH & CONSUMER PROTECTION DIRECTORATE-GENERAL Directorate E - Food Safety: plant health, animal health and welfare, international questions E1 - Plant health SANCO/10329/2002 rev 2 final. 17 October 2002.

EFSA (European Food Safety Authority), 2015. Technical report on the outcome of the pesticides peer review meeting on general recurring issues in ecotoxicology. EFSA supporting publication 2015:EN-924. 62 pp.

EFSA (European Food Safety Authority), 2019. Technical report on the outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology. EFSA supporting publication 2019:EN-1673. 117 pp. doi:10.2903/sp.efsa.2019.EN-1673 ISSN: 2397-8325

Appendix 3: Draft proposal for possible use of a limit fish test as alternative to full fish test with formulations

For dossiers without acute fish tox test for the assessed formulation, we follow the requirements/triggers for formulated products as given in Commission Regulation (EU) No 284/2013 and highlighted in the AGD (EFSA 2013), i.e. "*In principle, acute or short-term exposure tests with the formulated products should be carried out on one species from each of the groups of tier 1 aquatic organisms (fish, aquatic invertebrates, algae and/or macrophytes) if the preparation itself may contaminate water. However, where the available information for an a.s. permits the conclusion that one of these groups is clearly more sensitive (factor of 10 difference), only a test using a species of the relevant group needs to be performed*".

Accordingly for acute fish PPP tests, we would consider the following cases (for monoformulation):

1- If the a.s tests show that **fish are at lower risk** (10 fold or more) than other groups of organisms (daphnia/ algae) (i.e. endpoints (EP) deviating of a factor 10 or more): the regulatory praxis is to not follow the data requirement-> no fish PPP test deemed to be necessary.

2- If the a.s. tests indicates that **fish are clearly at risk** (i.e. fish EP is the lowest, deviates of a factor 10 or more): need to perform a PPP fish test -> based on information available, select either a Limit fish test PPP test (may need to be followed by a Dose response-fish test PPP) or directly a Dose response-fish test PPP

3- If the a.s. tests indicates that **fish are potentially at risk** (i.e. fish EP deviating of a factor 10 or less): check if the tests on daphnia and algae indicate a higher tox with the formulation than the a.s. (the tests must be performed with a similar design (e.g. flow-though), and endpoints expressed similarly (e.g. µg a.s./ L)):

- if no higher tox of the formulation is indicated (i.e. deviation of less than a factor 3 between EP of formulation and a.s. tests): it may be assumed that the formulation is also not more acutely toxic to the fish than the a.s. -> no fish PPP test deemed to be necessary
- if a higher tox of the formulation is indicated (i.e. deviation of a factor 3 or more between EP of formulation and a.s. tests): applying the approach proposed ("threshold approach") may be one suitable approach; in such approach, the concentration tested in the fish limit test could be the lowest of the EC50 concentrations available for invertebrate or algae tests performed with the PPP -> Limit fish PPP test requested.
    - If at the concentration tested, the acute toxic effects are lower than 50%, a "> X" LC50 could apply to the fish.
    - If at the concentration tested, the acute toxic effects are higher than 50%, a dose-response test should follow -> Dose response-fish test PPP requested.

Also please note that it is required to conduct chronic studies for formulations where the formulation is more acutely toxic than the a.s. by a factor of 10.

For acute fish test for PPP containing two or more active substances: in principle the same approach as above could apply; but if “the most sensitive taxonomic groups for the individual active substances are not the same, testing on all three/four aquatic groups, that is to say fish, aquatic invertebrates, algae and, where relevant macrophytes, shall be required” (PPP Regulation 284, section 10.2.1).