

## Workshop Synthesis

# Intermediate-tier options in the environmental risk assessment of plant protection products for soil invertebrates—Synthesis of a workshop

Gregor Ernst,<sup>1</sup> Mónica J. B. Amorim,<sup>2</sup> Melanie Bottoms,<sup>3</sup> Amy C. Brooks,<sup>4</sup> Mark E. Hodson,<sup>5</sup> Stefan Kimmel,<sup>6</sup> Pia Kotschik,<sup>7</sup> Michael T. Marx,<sup>1</sup> Tiago Natal-da-Luz,<sup>8</sup> Céline Pelosi,<sup>9</sup> Silvia Pieper,<sup>7</sup> Agnes Schimera,<sup>10</sup> Janeck Scott-Fordsmand,<sup>11</sup> Amanda Sharples,<sup>12</sup> José P. Sousa,<sup>8</sup> Cornelis A. M. van Gestel,<sup>13</sup> Bart van Hall,<sup>13</sup> and Matthias Bergtold<sup>14</sup>

<sup>1</sup>Bayer AG, CropScience Division, Monheim, Germany

<sup>2</sup>Department of Biology and CESAM, University of Aveiro, Aveiro, Portugal

<sup>3</sup>Syngenta Ltd., Jealott's Hill International Research Centre, Bracknell, UK

<sup>4</sup>Cambridge Environmental Assessments, Cambridge, UK

<sup>5</sup>Department of Environment and Geography, University of York, York, UK

<sup>6</sup>Corteva Agriscience, München, Germany

<sup>7</sup>German Environment Agency (UBA), Dessau-Roßlau, Germany

<sup>8</sup>Associate Laboratory TERRA, Department of Life Sciences, CFE-Centre for Functional Ecology – Science for the People and the Planet, University of Coimbra, Coimbra, Portugal

<sup>9</sup>INRAE, Avignon Université, UMR EMMAH, Avignon, France

<sup>10</sup>ADAMA Deutschland GmbH, Koeln, Germany

<sup>11</sup>University of Aarhus, Aarhus, Denmark

<sup>12</sup>FMC Corporation, Philadelphia, Pennsylvania, USA

<sup>13</sup>Faculty of Science, Amsterdam Institute for Life and Environment (A-LIFE), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>14</sup>BASF SE, Limburgerhof, Germany

### Abstract

The European environmental risk assessment (ERA) of plant protection products follows a tiered approach. The approach for soil invertebrates currently consists of two steps, starting with a Tier 1 assessment based on reproduction toxicity tests with earthworms, springtails, and predatory mites. In case an unacceptable risk is identified at Tier 1, field studies can be conducted as a higher-tier option. For soil invertebrates, intermediate tiers are not implemented. Hence, there is limited possibility to include additional information for the ERA to address specific concerns when the Tier 1 fails, as an alternative to, for example, a field study. Calibrated intermediate-tier approaches could help to address risks for soil invertebrates with less time and resources but also with sufficient certainty. A multistakeholder workshop was held on 2–4 March 2022 to discuss potential intermediate-tier options, focusing on four possible areas: (1) natural soil testing, (2) single-species tests (other than standard species), (3) assessing recovery in laboratory tests, and (4) the use of assembled soil multispecies test systems. The participants acknowledged a large potential in the intermediate-tier options but concluded that some issues need to be clarified before routine application of these approaches in the ERA is possible, that is, sensitivity, reproducibility, reliability, and standardization of potential new test systems. The definition of suitable assessment factors needed to calibrate the approaches to the protection goals was acknowledged. The aims of the workshop were to foster scientific exchange and a data-driven dialog, to discuss how the different approaches could be used in the risk assessment, and to identify research priorities for future work to address uncertainties and strengthen the tiered approach in the ERA for soil invertebrates. This article outlines the background, proposed methods, technical challenges, difficulties and opportunities in the ERA, and conclusions of the workshop. *Integr Environ Assess Manag* 2023;00: 1–14. © 2023 The Authors. *Integrated Environmental Assessment and Management* published by Wiley Periodicals LLC on behalf of Society of Environmental Toxicology & Chemistry (SETAC).

**KEYWORDS:** Calibration; Risk assessment; Soil organisms; Test development; Tiered approach

This article contains online-only Supporting Information.

Address correspondence to gregor.ernst@bayer.com

Published 11 August 2023 on wileyonlinelibrary.com/journal/ieam.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

### BACKGROUND

Soils represent the basis for agricultural production and food security, and are a habitat for a multitude of soil organisms that contribute to various ecosystem services, for example, nutrient cycling, water regulation, and carbon

sequestration (Creamer et al., 2022; Reid et al., 2005). The protection of soil health is highly important for humankind, as the degradation of soils caused by, for example, erosion, salinization, pollution, loss of organic matter (OM), and soil sealing can have a direct and detrimental impact on agricultural production. An environmental risk assessment (ERA) is conducted prior to the authorization of plant protection products (PPPs) to ensure that no unacceptable effects occur on the environment (European Commission [EC], 2009). In Europe, the current ERA scheme for soil invertebrates assesses the risks for soil organisms in-field and follows mainly a two-step approach. The Tier 1 ERA is conducted using laboratory toxicity tests with the standard soil invertebrate species *Eisenia fetida* or *Eisenia andrei* (earthworm), *Folsomia candida* (springtail), and *Hypoaspis aculeifer* (predatory mite) (EC, 2013a, 2013b). These laboratory tests follow the standard Organization for Economic Co-operation and Development (OECD) test guidelines 222, 232, and 226 (Organization for Economic Co-operation and Development [OECD], 2004, 2008, 2009, respectively), and are conducted in an artificial soil into which the test substances are homogeneously mixed. The derived endpoints describe changes in survival, reproduction, and growth of tested organisms (the latter only for earthworms), exposed to the chemicals. For the ERA, the so-called no observed effect concentrations (NOECs) as well as the effect concentration reducing the measured endpoint by 10% (EC<sub>10</sub>) are compared to predicted environmental concentrations in soil (PECsoil). The PECsoil is derived from the proposed application rate, and considering a soil layer of 0–5 cm thickness, a soil bulk density of 1.5 g/cm<sup>3</sup>, and it includes interception values, corresponding to the intended use and growth stage of the crop (FOCUS, 1997). Risks are considered acceptable if the toxicity exposure ratio (TER) equals or exceeds the assessment factor (AF) of 5 (TER trigger value; EC, 2011). If in this first tier the relevant TER trigger is not passed, a higher-tier ERA is possible. Usually, field studies with earthworm, springtail, and/or soil mite communities are conducted as a refinement to address risks identified at Tier 1. The current ERA scheme considers initial effects in a field study acceptable if recovery can be demonstrated at the latest one year after application. Whether the TER trigger value of 5 is appropriate to achieve the desired protection is the aim of a risk assessment calibration. An ERA calibration can be performed as described in a scheme published by the EFSA Panel of the Plant Protection Products and their Residues (EFSA PPR Panel) (2010, 2017). According to this scheme, a “surrogate reference tier” needs to be defined to link the different risk assessment tiers with the specific protection goals and the general protection goals laid down in the legislation (EC, 2009; European Food Safety Authority [EFSA], 2010; EFSA PPR Panel, 2017). As defined by the EFSA PPR Panel (2017), a surrogate reference tier approximates the real situation in the field (e.g., represented by terrestrial model ecosystems [TMEs] or field studies), while recognizing that a compromise is made between what is desirable and what is practical. In that regard,

Christl et al. (2016) as well as Kotschik et al. (2019) considered field study endpoints, derived from a large set of different environmental situations with different substances in a Tier 1 earthworm calibration case study. In the future, sources of uncertainty in ERA calibration could be addressed with effect and population modeling (Forbes et al., 2021), once validation of population modeling has been carried out.

For the ERA of soil organisms, new developments are expected on both the exposure and effects assessment. A new guidance on exposure assessment was published by the EFSA (2017). Changes compared to the current exposure predictions are driven by amendments in the exposure characterization (lower soil bulk density, the inclusion of wash-off from the crop intercepted fraction, and the consideration of the geomean DT<sub>50</sub> for degradation of the active substances). For a set of 56 randomly chosen active substances, Schimera et al. (2022) concluded that the new exposure assessment according to the EFSA (2017) will lead to higher PECsoils and increased failure at the Tier 1 risk assessment, thereby increasing the need for higher-tier assessments.

In terms of effect assessments, a new OECD guideline for earthworm field testing is expected in the near future, replacing the current International Standard Organisation (ISO) 11268-3 guideline (International Standard Organisation, 2014). It is expected that the new OECD guideline will lead to more effort in conducting the study, driven by amended test designs (e.g., higher number of plot replicates and sampling effort per plot), more efforts on analytical measurements (higher spatial and temporal resolution), and different statistical methods (Römbke et al., 2020).

The upcoming changes in terms of exposure and effect assessments described above are expected to directly impact the ERA of PPPs for soil invertebrates. The availability of more options in the ERA framework in addition to the standard laboratory testing and field assays could be a substantial and valuable contribution to a more workable and efficient tiered ERA approach by serving as a bridge between Tier 1 and higher-tier risk assessment as in other areas of ERA. Introducing intermediate-tier testing and risk assessment options may help to reduce specific sources of uncertainty, such as the test substrate (soil), interspecies sensitivity differences, different life cycle traits, duration of effects, and indirect effects, thereby increasing ecological realism compared to the current simple Tier 1 ERA with standard indicator species tests and strengthening the tiered ERA approach for soil invertebrates.

The EFSA PPR Panel (2017) proposed a tiered risk assessment approach by indicating which type of additional assays can help address uncertainties in the ERA. Two potential intermediate-tier steps are described: (1) testing of more species to address interspecies variability and (2) assembled multispecies tests to cover uncertainty with regard to additional stressors like predation and competition. Currently, it remains unclear how endpoints from additional species can be evaluated in the ERA. Moreover, the limited experience with assembled multispecies tests so far

has hampered a wider application to soil risk assessment (EFSA PPR Panel, 2017).

For an efficient risk assessment scheme, a simple and conservative first assessment step (Tier 1) is needed, which filters out low-risk uses with minimal effort. Highly complex, higher-tier ERA should only be necessary for uses that potentially pose a long-term risk. Intermediate tiers can fill the current gap between Tier 1 and field studies and refine risks with reasonable effort and sufficient certainty if calibrated properly.

## PURPOSE AND OBJECTIVES

A multistakeholder workshop was held on 2–4 March 2022 to discuss potential intermediate-tier options. Within this three-day workshop, discussions were held between participants from academia, industry, regulatory authorities, and contract research organizations (see the list of participants in Supporting information: Table S1) about the potential of four different study types as intermediate-tiered options in the evaluation of risks to soil organisms: (1) natural soil testing, (2) single-species tests (other than standard species), (3) assessing recovery in laboratory studies (use of, for e.g., multigeneration/aged residue studies to assess potential for recovery), and (4) the use of soil multispecies test systems with assembled or natural communities. The workshop was initiated to foster scientific exchange and a data-driven dialog; to review and discuss the relevant studies and technical challenges for each area; to discuss how the different approaches could be used in the ERA; to assess which uncertainties can be addressed in the different tiers; and to identify research priorities for future work, in order to strengthen the tiered approach in the ERA of PPPs.

### *Natural soil testing*

Currently, Tier 1 laboratory toxicity tests following the OECD guidelines are performed in artificial soils consisting of 10% or 5% sphagnum peat, kaolin clay, quartz sand, and calcium carbonate (to adjust the pH value); however, the OECD guidelines for soil invertebrates offer the option to use natural soil. The use of artificial soils in Tier 1 studies has several benefits. For instance, the components are easily obtainable, they are suitable for standard test organisms, and the results obtained from the toxicity tests enable comparisons between different laboratories and over time. Drawbacks are their lack of realism, as evidenced by a different content and type of OM, mineral clay composition, and a different soil texture, when compared with natural soils. As a result, the bioavailability of test chemicals to soil organisms can be different in natural soils compared to artificial soils. To address this issue in the ERA, the toxicity endpoints are divided by a correction factor of 2 for lipophilic substances (if  $\log K_{ow} > 2$ ; SANCO, 2002), independent of the OM or peat content in the artificial soil, in addition to the AF applied in the Tier 1 ERA. Currently, there is a lack of clarity of the influence of different soil properties on the toxicity of PPPs for soil invertebrates. Thus, testing

and understanding the impact of PPPs on soil organisms in different natural soils could help to reduce the uncertainties compared to Tier 1 ERA when extrapolating to field situations.

*Soil parameters influencing the toxicity to soil invertebrates.* Understanding which soil parameters influence the toxicity of PPPs was also considered an important aspect during the workshop. A literature review and subsequent experiments, as presented by Kotschik et al. (2018 at SETAC Europe Annual Conference), identified OM, cation exchange capacity, and pH as soil properties driving toxicity, depending on the chemical properties of the test substances as well as tested organisms. The authors also mention that other soil parameters are rarely measured and thus their influence could not be determined. Hence, a good characterization of natural soils is recommended when used for ecotoxicological testing (OECD, 2008, 2009). Van Hall et al. (2023) conducted a literature review and assessed the correlation between toxicity endpoints and OM content. The correlation was chemical specific; for example, phenmedipham toxicity showed a high correlation with soil OM content, but correlations disappeared when chemicals were grouped together based on their lipophilicity. Additionally, when a correlation was present, it was stronger for mortality than for reproduction endpoints. Van Hall et al. (2023) also noted that the influence of soil OM content on chemical toxicity differed between soft- and hard-bodied invertebrates, as the toxicity was reduced more in soils with high OM contents for soft-bodied than hard-bodied invertebrates. This relationship needs to be better understood prior to incorporating testing of natural soils into risk assessment procedures. In addition, as already pointed out by the EFSA (2017), it throws into question the existing paradigm of adjusting toxicity endpoints from Tier 1 laboratory studies for high-lipophilicity compounds by a default correction factor, as no such consistent correlation is demonstrated in the literature data. Hence, it was stated in the workshop that the grouping based on lipophilicity might not be fully appropriate to address the influence of OM content on PPP toxicity. It was agreed by the participants that testing in natural soils is possible; however, further standardization is needed for routine testing, that is, regarding technical suitability of the different soil types for the performance of the test species and representativity in ERA.

*Technical challenges.* One of the major challenges of using natural soils is the selection of a suitable soil from a testing performance perspective and regarding representativity in the ERA. European soils are diverse, both in terms of their biological and physicochemical parameters. Several important aspects were identified for the use of natural soils for ecotoxicological testing: (1) Selected soils should be representative for the specific situation to be covered in the ERA (regional aspect), (2) appropriate and realistic validity criteria for the performance of the ecotoxicological study

should be available that contribute to the generation of reliable endpoints, (3) soils should be clear of confounding anthropogenic stressors such as contaminants, and (4) different standard soils should be categorized according to their major soil properties to facilitate comparisons.

A European approach led to the definition of so-called EURO-soils (Gawlik et al., 2004; Kuhnt & Muntau, 1994), which are documenting the representativeness of the most frequent and typical soil types for Europe. EURO-soils have been used, for example, in environmental fate research, but they are not an “endless resource.” Therefore, they cannot realistically be used for standard ecotoxicological testing due to the large amounts of soils needed for laboratory toxicity tests. However, it was agreed between participants that natural soils could be categorized into “EURO-soils” categories based on their main properties (i.e., soil type, OM content, pH, C/N ratio, texture). For Germany and Portugal, Römbke and Amorim (2004) selected 13 natural soils, of which 11 could be categorized according to the EURO-soils concept. As such, soils collected regionally (e.g., Germany or Portugal) could still be compared to each other if their characteristics are similar, solving the “endless resource” problem. An interesting example from Brazil was discussed in which the soil type was matched with crop type, showing that more than 70% of the soybean is grown on the same soil type (Ramon et al., 2022). This was seen as a potential approach for the European Union (EU) using relevant data sets (e.g., JRC EFSA spatial data V1.2 and CAPRI or LUCAS topsoil) to define potentially relevant natural soils for each regulatory zone for the major crops. Accordingly, exposure assessments following the new guidance document for calculating the PEC<sub>soil</sub> are based on such data sets (JRC, PERSAM Tool for EFSA, 2017). However, it was considered an important topic for future research whether the ecotoxicological endpoints derived from studies with soils from the same, for example, EURO-soils category but different sites show comparable results.

Although there have been many studies utilizing natural soils, there should also be more comprehensive clarification over whether natural soils are suitable for ecotoxicological tests in the laboratory in terms of meeting the validity criteria for the three standard OECD test species (*E. fetida/andrei*, *F. candida*, and *H. aculeifer*) regarding mortality, number of juveniles, and coefficient of variation (CV) for reproduction in the control. A stable control performance is a prerequisite for routine testing for the risk assessment of PPPs and other chemicals. In the reviewed literature, the validity criteria for *E. fetida/andrei* and *F. candida* were in most cases met in tests using both artificial and natural soils (Amorim et al., 2005; Chelinho et al., 2011; Domene et al., 2011). However, some soils appeared not to be suitable for earthworm testing and it is highly probable that studies that did not meet the validity criteria were not published. In this regard, Chelinho et al. (2011) recommended to avoid extreme textures such as a high content of sand ( $\geq 80\%$ ), strong acidity ( $\text{pH} < 4.2$ ), or low

OM content ( $< 2\%$ ). Not enough data are available for *H. aculeifer* to give any recommendations.

### Single-species tests (other than standard species)

As surrogate species for soil organisms and more specifically for earthworms, springtails, and predatory mites in the ERA of PPPs, the indicator species *E. fetida/andrei*, *F. candida/fimetaria*, and *H. aculeifer* are tested according to OECD test guidelines 222, 232, and 226 (OECD, 2004, 2008, 2009, respectively). Standard surrogate species are not always present in all the natural soils or geographical areas where the PPP is applied. They may not have the same sensitivity toward chemicals compared to other species and may not be representative of all ecological categories within each group. As pointed out by the EFSA (2017), the most relevant criteria to obtain a balanced test battery are the representativeness of the species for the ecosystems to protect and the representativeness of responses resulting from different routes of exposure. The mentioned standard species are easy to rear in the laboratory, and their performance and sensitivity have been ring-tested. Hence, the standard indicator species are well suited for a simple and conservative Tier 1 ERA. If calibrated appropriately, the ERA with the available test battery can be indicative of effects and risks under field conditions.

However, it was agreed among the workshop participants that it would be beneficial to have test methods available to assess effects of a chemical on species other than the standard ones. This can reduce uncertainties regarding potential higher sensitivity of other (often locally more relevant) species and may allow the estimation of more integrated toxicity values, for example, by means of assessing species sensitivity distributions (SSDs).

*Sensitivity differences for earthworms and springtails.* Several studies using nonstandard earthworm and springtail species have been described in the literature. Some examples were reviewed during the workshop to provide a general picture of the current knowledge regarding species sensitivity differences toward PPPs:

**Earthworms.** Regarding acute effects (survival), *E. fetida/andrei* overall seem less sensitive compared to other earthworm species tested (De Silva et al., 2010; Pelosi et al. 2013), confirming the conclusions from the EFSA PPR Panel (2017). It is, however, not to be expected that non-standard species will always be more sensitive than *E. fetida* with regard to chronic endpoints (as shown by Carniel [2019]). Sensitivity differences between different earthworm species regarding the endpoint reproduction are in some cases absent (e.g., Kreuzweiser et al., 2008), sometimes visible, but were shown to be less than a factor of 10 for the reviewed laboratory studies (*Lumbricus terrestris*, *Aporrectodea caliginosa*, *Perionyx excavatus*, *Dendrobaena veneta*; for details, see Carniel [2019], Kreuzweiser et al. [2008], De Silva et al. [2010], Pelosi et al. [2013]).

Springtails. The comparisons of sensitivity of Collembola species (i.e., *F. candida*, *F. fimetaria*, *Sinella curviseta*, *Protaphorura fimata*, *Proisotoma minuta*, *Heteromurus nitidus*) showed that *F. candida* tends to be among the most sensitive species tested, but the relative sensitivity may vary with the mode of action of the test substance (for details, see Bandow, Coors, et al. [2014], Bandow, Karau, et al. [2014], Carniel [2019], Ferreira et al. [2022], De Lima e Silva et al. [2021]). This confirms the conclusions of the EFSA (2017) that *F. candida* or *F. fimetaria* could be representative of other springtails with regard to their toxicological sensitivity. A higher sensitivity of *Yuukianura szeptykii* to fenoxycarb compared to *F. candida* was shown by Lee et al. (2020), but it is not clear if this is compound and/or mode of action specific or indicates a general difference in sensitivity between species. This finding should be discussed in the context of considering the specificities of local species in the ERA of PPPs.

**Standardization.** Several nonstandard species have been considered in international test guidelines (e.g., added to annexes to ISO, OECD, Environment and Climate Change Canada; see Table 1). So far, ring-testing has not been performed with all the newly proposed species. Additional literature data as well as potentially nonpublished data on, for example, control performance should be compiled, and comparative measurements between different laboratories should be conducted that can contribute to an evaluation of the reproducibility and a validation of the test guideline. Further guidance on the evaluation of nonstandard soil invertebrate tests would help to judge on the suitability of their use in ERA.

**Species sensitivity distribution.** Species sensitivity distribution is a statistical evaluation method that integrates toxicity data for species with different sensitivities toward a chemical in one evaluation. With an SSD, the concentration at which a certain proportion of taxa are affected (e.g., hazardous concentration HC<sub>5</sub> for 5%) can be estimated and

used in ERA. It is an approach that is already used within the EU for ERA of other wildlife groups, for example, aquatic organisms and nontarget terrestrial plants and in other risk assessment guidance documents (e.g., EFSA PPR Panel, 2021). The SSD could be used in an intermediate-tier risk assessment approach for PPPs by comparing, for example, the HC<sub>5</sub> value with a PEC value and derive a TER. A critical TER trigger value is not yet available for soil organisms, which has to be calibrated by means of the surrogate reference tier considering agreed protection goals (see the EFSA PPR Panel, 2010, 2017).

Workshop participants discussed how many species should be included in an SSD. A low number of species in an SSD might limit its quality and increase uncertainty in the risk assessment. Six species as a minimum number was proposed as a prerequisite to perform a robust and reliable SSD (besides SSD quality criteria, e.g., goodness of fit). On the other hand, Frampton et al. (2006) suggested that, for soil organisms, SSDs based on five species might be sufficient; this was based, however, on the fact that at the time of their publication, for 96% of PPPs, no toxicity data for more than five test species were available. In any case, the definition of a minimum number of species for an SSD will require a trade-off between being statistically and technically robust and being practical. Available experiences and guidance from other compartments, for example, aquatic organisms (including sediment; Diepens, 2015; EFSA PPR Panel, 2013, 2021; Maltby et al., 2005) or nontarget terrestrial plants (Kwak et al., 2018; Silva et al., 2014), should be used to develop such quality criteria.

The selection of endpoints and species that can be included in an SSD was a major point of discussion. Integrating clearly nonsensitive species in an SSD or by using, for example, the highest tested concentration as proxies for EC<sub>10</sub> or EC<sub>50</sub> in SSD for nonsensitive species, can create a multimodal distribution. Although there are several options to statistically deal with multimodality (Fox et al., 2021), the inclusion of nonsensitive species in the SSD may lead to higher uncertainty, while it increases variability, possibly

**TABLE 1** Additional soil invertebrate test species that are already implemented in test guidelines or are foreseen to be added to annexes of internationally standardized toxicity test guidelines

	OECD	ISO	ECCC
Earthworms, Enchytraeids	OECD 220, <i>Enchytraeus albidus</i> ; <i>E. crypticus</i> , <i>E. buchholzi</i> , <i>E. luxuriosus</i> , <i>E. bulbosus</i> .	ISO 11268-2, <i>Eisenia fetida/andrei</i> ; <i>Dendrodrilus rubidus</i> , <i>Aporrectodea caliginosa</i> , <i>Perionyx excavatus</i> .	STB 1/RM/43—(2): New guideline for <i>E. andrei</i> and <i>D. rubidus</i> .
Springtails	OECD 232, <i>Folsomia candida</i> , <i>Folsomia fimetaria</i> ; <i>Proisotoma minuta</i> , <i>Isotoma viridis</i> , <i>Isotoma anglicana</i> , <i>Orchesella cincta</i> , <i>Sinella curviseta</i> , <i>Paronychiurus kimi</i> , <i>Orthonychiurus folsomi</i> , <i>Mesaphorura macrochaeta</i> .	ISO 11267, <i>F. candida/fimetaria</i> ; <i>Onychiurus yodai</i> , <i>O. folsomi</i> , <i>P. minuta</i> , <i>Protaphorura fimata</i> , <i>Sinella curviseta</i> , <i>Allonychiurus kimi</i> , <i>Yuukianura szeptykii</i> , <i>Heteromurus nitidus</i> , <i>Onychiurus armatus</i> , <i>O. cincta</i> , <i>Hypogastrura assimilis</i> , <i>M. macrochaeta</i> .	EPS 1/RM/47—(4), <i>F. candida</i> , <i>F. fimetaria</i> ; <i>O. folsomi</i> , <i>P. minuta</i> .

Abbreviations: ECCC, Environment and Climate Change Canada; ISO, International Organization for Standardization; OECD, Organization for Economic Co-operation and Development.

leading to lower HC<sub>5</sub> values and an overestimation of the potential risk.

The question of which organism groups should be considered in an SSD was also discussed. There are different possibilities to run an SSD, for example, (a) focus on the most sensitive group that needs to be addressed or (b) include other trophic levels, like plants and soil microorganisms, depending on the purpose of the SSD. Species sensitivity distributions for soil invertebrates were calculated for chlorothalonil by Carniel (2019), separately for Oligochaetes and Collembolans (tested in tropical artificial soil). Other authors calculated SSDs for a wider range of species, including other trophic levels like plants and microorganisms (Kwak et al., 2018; Silva et al., 2014). Using different trophic levels in an SSD could be helpful in case it aims to describe effects on a food web or on the total soil community. Using such a full community SSD in a risk assessment may require a different AF than if focus is set on the most sensitive group. In a scenario where the SSD is used to refine risk, for example, soil microarthropods only, the SSD should be restricted to this group as it aims to reduce the uncertainty related to sensitivity difference of species within the group.

It was mentioned that SSDs should be calculated using comparable endpoints (e.g., EC<sub>10</sub>, reproduction) generated under similar conditions (e.g., same soil type). Including endpoints from tests with different soils introduces an additional source of uncertainty. It was also mentioned that combining endpoints from active substances and, for example, monoformulations, can lead to increased uncertainty and/or wrong interpretation and thus requires careful interpretation. Therefore, the use of both different soils and different test items in one SSD should be avoided (if possible), unless evidence can be presented that no significant impacts on the effects observed are to be expected. Using endpoints that are normalized on the bioavailable fraction (e.g., OM normalized data or endpoints expressed as pore water concentration) might have the potential to overcome the issue of different test substrates. Any uncertainties should be discussed and considered in the conclusions and it should be clarified under which circumstances bridging would be possible.

**Technical challenges.** The rearing of single species other than the standard ones requires optimization of breeding conditions due to their specific ecological and biological properties and requirements (especially earthworms; Lokke & Van Gestel, 1998). Some earthworm species might need deep soil, large amounts of food, accept only low population densities, and/or have long life cycles, which may impede the production of sufficient numbers of juveniles for an ecotoxicological reproduction test within a manageable test period. For Collembola species, it was discussed in the workshop that rearing of different species is feasible with relatively small adaptations.

The validity criteria for tests with nonstandard species might need revision: whereas the criterion on control

mortality might not cause problems, the available criteria for one species may not be appropriate for other species, for example, regarding the minimum number of juveniles and/or the corresponding acceptable CV (e.g., 30%) in the control. For nonstandard species, this CV can be higher due to the higher relative variability related to lower reproductive rates compared to the standard species. However, facilitating routine testing of nonstandard species by implementing less strict validity criteria needs to be carefully balanced with the risk of losing experimental precision, statistical sensitivity and/or robustness, and therefore certainty in the ERA.

For use in the ERA of PPPs, ecotoxicological tests should show a high degree of reproducibility. The tests with the same species in the same soils should lead to similar ecotoxicological results independent of the test facility, region, and season. An optimization of the test conditions with regard to temperature, soil moisture content, variability, control mortality, conditions for growth and reproduction, and food quality and quantity, as well as an unequivocal taxonomic identification of the test species, can be taken into consideration to increase the reproducibility of the tests. Practical experimentation has shown that the number of adults per replicate, the number of replicates per treatment, the test temperature, and the amount of soil per replicate are aspects that often require adaptation in testing earthworms other than the standard species. However, the number of adults per replicate and the need to distinguish males from females for the selection of adults prior to beginning the experiments are issues mainly associated with collembolan testing. Ring-testing procedures are recommended before the implementation of additional species in testing guidelines or ERA guidance documents. This contributes to a higher reproducibility, thus allowing for a more robust test system and a higher acceptability for its use in ERA.

#### **Assessing recovery in laboratory studies**

In general, recovery can be defined in relation to health status (return to health from sickness) or in relation to a previous condition that was lost due to a stressor (return or regain a former or better state of condition). Laboratory tests assessing recovery after PPP exposure may relate to internal (population growth) or external (recolonization) recovery. The different assessments of recovery require different approaches, for example, in the laboratory that can be:

- (1) Laboratory studies with exposure, followed by incubation in clean soil.

Such an approach may assess if affected organisms are capable of recovering from the impact of exposure, for example, by investigating EC<sub>50</sub> increase over time or return to normal growth, reproductive output, or biochemical endpoints in affected organisms (e.g., Feng et al., 2015; Van Gestel et al., 1989). As the PPPs are usually not completely degraded during the test period, the approach of transferring organisms to clean soil (immediate stop of exposure) may not be realistic for

field conditions and hence not suitable to assess the effects and risks of PPPs.

- (2) Laboratory studies with population development monitored over longer periods.

Prolonging the test duration is sometimes helpful in identifying the cause of effects (e.g., due to delayed hatching of cocoons), as was shown for enchytraeids (Kovacevic et al., 2020). This may be indicative of long-term development of a population (by observation of population growth rate) for certain species. However, in laboratory tests, the sensitivity to chemical exposure can be influenced by crowding (density-dependent effects; Noël et al., 2006). Further, density limitation in the laboratory test can have an impact on recovery assessment. In case the population size in the control reaches a plateau due to crowding or food limitation (e.g., as it does under field conditions), the population in the exposed groups could reach the numbers in the control more easily than under density-independent growth. Moreover, natural species might show different population dynamics and observed effects on standard species in the laboratory might not be directly relevant to natural species or organism communities.

- (3) Multigeneration studies combined with aging of residues.

The approach to assess the potential for external recovery in a laboratory test system is an agreed risk assessment approach for nontarget arthropods when evaluating effects in-field following the risk assessment guidance for nontarget arthropods (ESCORT 2; Candolfi et al., 2001). In the proposed multigeneration approach for soil organisms, aged residue studies are performed with standard indicator species in which juveniles from the first test are introduced in a second (or third) test with soil containing aged residues. The individuals are constantly exposed to the potentially continuously degrading and dissipating test substance in soil. The aim is to assess after which time period the individuals are no longer affected by the chemical. Decrease in toxicity was shown for lindane and chlorpyrifos in two-generation tests with *F. candida* (Ernst et al., 2016) and in three-generation tests with *F. candida* exposed to the fast-degrading thiacloprid, but not for the more persistent imidacloprid (Van Gestel et al., 2017).

Recovery (recolonization; external recovery) from off-field areas (which is a valid approach for nontarget arthropods with higher mobility) may be limited for soil organisms, such as Collembola, as not all species are mobile; therefore, internal recovery is assessed with this test (population growth). Two- or three-generation studies can demonstrate intrinsic potential for recovery for the tested species from the treated area itself. As different species show different generation cycles, observed dynamics of surrogate species might not cover realistic effects in agricultural fields.

*Technical challenges.* Multigeneration exposures are feasible for organisms with a relatively short life cycle like many

springtails and enchytraeids (Guimarães et al., 2023), but not for earthworms that have much longer generation times. Proper assessment of the duration of exposure of the different generations as well as selection of the right cohort of juveniles produced for transfer into next-generation exposures are points to be standardized. Another challenge is the possible increase in variation over generations, making it harder to meet validity criteria set for control performance. Larger differences in the age of the parental generation of second (or third) test generations in a multigeneration study can cause larger control variability for reproduction (Ernst et al., 2016).

Different designs for a multigeneration test with *F. candida* were tested by Ernst et al. (2016) to find the best way to reduce uncertainty (optimization of control performance to meet the validity criteria). This was obtained with a design including an intermediate period between exposed generations. This reduced the variability in juvenile numbers as the different age of the transferred individuals no longer causes variability in reproduction (all individuals mature). From a practical perspective, it was concluded that the selection of individuals with the same age or size for different exposure generations can introduce a bias, as larger individuals could be less sensitive to the chemical in the following test run.

#### *Soil multispecies systems*

In soil multispecies systems (SMS), a defined number of preselected species are introduced into a test system in which the impact of a stressor (chemical, climate, etc.) is assessed on total population sizes of the different species under assessment or on the total soil organism community for a defined test period. In SMS, the interactions between different species, that is, competition and predation, lead to a more realistic situation compared to single-species test systems. Interactions between species with different ecological traits in an SMS might lead to higher sensitivity compared to single-species tests. Soil multispecies systems have a nonrandom species composition, being based on field surveys to represent a more realistic scenario, but with the ability to change the experimental setup. Structural (species abundance), species distribution (top, middle, and bottom soil layer), functional (e.g., OM degradation), and explanatory (e.g., biomarkers) endpoints can be investigated. Soil multispecies systems are done under controlled conditions, for instance, a defined amount of (defaunated natural) soil (e.g., 1 kg per replicate); selected species with a defined number per replicate; controlled temperature, light, and water regime; single-dose or dose–response design with flexible replication ( $n \geq 5$ ); various sampling dates; fate measures (timely and spatial resolution) in soil; and defined test duration(s) (e.g., 28, 56, and 84 days; see, e.g., Jensen and Scott-Fordsmand [2012], Mendes et al. [2019], Scott-Fordsmand et al. [2008]). To meet different field relevant scenarios, different soil types and various application methods can be chosen (e.g., surface spray or in-soil mixing) or other relevant parameters can be simulated. Depending

on the experimental design adopted, SMS data can be analyzed using univariate statistics to estimate concentration at which  $x\%$  effect is observed ( $EC_x$ ) or NOEC for single endpoints, for example, abundance of a species, or multivariate statistics (e.g., principal response curve [PRC]) to assess effects on community structure. Depending on the number of species introduced, the calculation of an SSD is also possible.

In an SMS, the use of species from different trophic levels can provide insight into direct as well as indirect effects of the applied stressor (e.g., predator–prey interactions and contaminated food) that are difficult to interpret. However, by attempting to mimic the field situation, an SMS provides a quantitative measure of the interaction (Schnug et al., 2014). The inclusion of these interactions might be a source of variability; however, use of an SMS provides more information than if the same species were tested separately in single-species tests. Including plants in an SMS was seen as an option to increase realism; however, it was realized that the representativity may not be given if just one plant in a small system is included without much space. The inclusion of plants would be best in an adapted version of the SMS, for example, used a larger size of test vessels. Having a standard set of species and gaining a better knowledge about typical species interactions in this setup would facilitate designing the test and the interpretation of the results.

The use of natural communities from the field instead of built communities (i.e., SMS) is also a possibility in laboratory tests with spiked soils (e.g., Chelinho et al., 2014). The use of natural communities increases the ecological relevance of the experiment, but other factors (e.g., seasonality and higher intersample variability) may influence the outcome of the tests, making test results more difficult to interpret. However, results showing a dose–response relationship of carbofuran on different feeding guilds of soil Collembola and mites, as a result of both direct and indirect effects (i.e., competition and shortage of prey), were clearly observed by Chelinho et al. (2014). Also, by using multivariate analysis, this approach allows the derivation of an  $EC_x$ -community endpoint (Renaud et al., 2021).

**Technical challenges.** The reproducibility of SMS tests will be different compared to single-species tests due to interactions between species (competition and predation, facilitation), which could lead to higher variation between tests. A variation in predation efficiency will affect other species, which will be more difficult to standardize. Hence, the precision and reproducibility may be lower (the random variation may be higher) but the ecological relevance will be higher. Therefore, more work is needed to understand the predictive ability of SMS regarding the protection goal in the field.

The selection of test concentrations is considered a challenge due to different sensitivities of single species. However, delayed intoxication (via poisoned prey, for example) may also happen. Hence, predictions from single-species tests might not give a correct prediction of the effects on a community. A limitation of an SMS is that it may

not be possible to see a reasonable response for all species in one system, depending on the range and spacing of the test concentrations. This limits the possibilities to calculate, for example, SSDs. With a proper number of replicates and concentrations (which is more time and labor intensive), the SMS may provide a response for a sufficient number of species to be used in an SSD. The workload, however, might be less than if a comparable number of single-species tests is conducted, and species interactions are taken into account.

The choice of the test duration and sampling times in an SMS is important and may vary between species and the purpose of the test. However, the stability of such test systems can represent a challenge in relation to community development. The test duration can have an impact on possible side effects through, for example, fungal growth, which could be indicative of an unstable test system.

#### *Other lower-tier refinement options*

In addition to the approaches mentioned above, other options could be considered as further refinements in the ERA of PPPs. Laboratory tests could be performed using more realistic exposure conditions for specific intended uses (e.g., application of seed treatments, drip applications, granules, spray application instead of mixing the PPP with the soil). Following more realistic types of applications, spatiotemporal heterogeneous exposure profiles can lead to different effects due to movement and potential avoidance of the organisms. In this regard, ecological modeling in combination with analytical measurements of chemical residues over the test period can help to understand the actual exposure that is experienced by the organism under investigation, once ecological models have been validated. The specific exposure of organisms to contaminants could be refined through simulation of realistic movement of the organisms in the soil profile over time and using toxicokinetic toxicodynamic modeling on the individual level (Forbes et al., 2021; Gergs et al., 2022; Roeben et al., 2020). However, Toschki et al. (2020) showed long-lasting effects of chemicals persisting after surface spray application on soil organisms living also in deeper soil layers. Therefore, refinement of exposure in relation to the occurrence of organisms may be hampered as movements of organisms are difficult to predict. Hence, the reciprocity and magnitude of effects when soil organisms are exposed to spatially and temporally variable concentrations, as well as the vertical and horizontal movement of soil organisms need to be better understood.

## DISCUSSION ON THE TIERED RISK ASSESSMENT APPROACH

The workshop participants considered the introduction of intermediate tiers as a potential improvement of the ERA of PPPs. Intermediate tiers can help to reduce uncertainties in ERA by providing more data on toxicity under more realistic exposure conditions, sensitivity differences between species, duration of effects, species interactions, and indirect effects (Table 2). In general, ERA is done by comparing



**TABLE 2** Overview of options for intermediate-tier testing and ERA for soil organisms exposed to pesticides, showing the specific aspect that can be addressed with the test system, the estimated level of experience (ranging from a low [+] to a high level [+++]), potential assessment endpoints, and the technical and regulatory challenges

Test system	Specific aspect addressed	Level of experience	Potential assessment endpoints	Technical challenges	Regulatory challenges
Realistic application scenario in a standard study	More realistic exposure	+++	No-observed effect concentration, EC <sub>x</sub>	–	Environmental risk assessment calibration
Natural soil tests	More realistic sorption properties and bioavailability	++	No-observed effect concentration, EC <sub>x</sub>	Validity criteria, reproducibility, soil variability	Environmental risk assessment calibration, representativity of natural soils
Nonstandard species tests	Sensitivity differences between different species	++	No-observed effect concentration, EC <sub>x</sub> , Geomean NOEC or EC <sub>x</sub> , HC <sub>5</sub> (SSD based on EC <sub>x</sub> or NOEC)	Validity criteria, reproducibility	Environmental risk assessment calibration, species to consider in an SSD
Multigeneration tests	Internal recovery, multiple applications, transgenerational effects	++	No-observed effect concentration	Validity criteria	Representativity of tested species with regard to recovery traits, ERA calibration
Soil multispecies systems	Species interactions, indirect effects	+	No-observed effect concentration, EC <sub>x</sub> , Geomean NOEC or EC <sub>x</sub> , HC <sub>5</sub> (SSD based on EC <sub>x</sub> or NOEC)	Test stability, reproducibility, interpretation	Representativity, ERA calibration

Abbreviations: EC<sub>x</sub>, concentration at which x % effect is observed; ERA, environmental risk assessment; Geomean NOEC or EC<sub>x</sub>, Geomean from an available set of endpoints from, for example, different species tested; HC<sub>5</sub>, hazardous concentration at which 5% of the species in the SSD show lower endpoints and 95% of the species show higher endpoints; NOAEC, no-observed adverse effect concentration, including population recovery in, for example, a multi-generation test; NOEC, no-observed effect concentration; SSD, species sensitivity distribution.

effect endpoints, for example, NOECs or EC<sub>x</sub> values with exposure values (PECsoil) and using AFs to account for various sources of uncertainty. It was highlighted that the new soil exposure framework (EFSA, 2017) uses soil scenarios (i.e., soil bulk density, OM content, and crop interception) that lead to realistic worst case exposure estimation in the EU, and can lead to substantially increased PECsoil values. Some participants questioned the agronomical relevance and representativity of the underlying soil parameters. However, in this regard, the intermediate-tier options can be seen as a possibility to better link exposure and effect assessment. In general, the uptake routes of PPPs in different soil invertebrates are diverse due to physiological and ecological differences and depend on the bioavailability of PPP, driven by chemical properties of actives as well as soil composition. The question on which metric the risk assessment should be based (total soil concentrations, OM normalized endpoints, or pore water concentrations, as discussed by the EFSA PPR Panel [2017]) is relevant for all ERA tiers and should be taken up in the context of risk assessment calibration by designing an appropriately conservative tiered risk assessment framework. A higher degree of realism (e.g., more precise expression of

exposure metric in ERA) and less uncertainty due to additional data provided (e.g., on sensitivity of different species and communities, species interactions, as well as recovery potential) may justify changing AFs compared to those currently used for Tier 1, if the ERA calibration against a suitable reference tier and to the protection goals allows.

Testing of natural soils instead of artificial soils can add realism and reduce uncertainty in the risk assessment of PPPs. Natural soils are in principle considered suitable for use in ecotoxicological tests and considered in ERA. The workshop revealed two different opinions on how ecotoxicity studies with natural soils should be considered in the ERA: Some participants stated that the current AF (AF = 5) in Tier 1 may not be protective with artificial soil tests, as the tests with natural soils represent a more realistic scenario and can lead to lower endpoints. Therefore, the results obtained with the testing in natural soils should be known to calibrate the Tier 1 assessment step. Other participants argued that for the calibration of the Tier 1 ERA, the surrogate reference tier (e.g., a set of field data) is decisive, and hence, a calibration can be done with any soil type in the Tier 1 study, that is, artificial soil. Therefore, a Tier 1 calibration (if relying on tests performed in artificial soil) could be done

independently from results generated with natural soils. Hence, there was no consensus on this point. Christl et al. (2016) validated the current Tier 1 for earthworms with laboratory data generated with OECD artificial soil against a set of field studies covering a wide range of different soil types, different substances, and climates (surrogate reference tier). The data showed that the AF of 5 is appropriate and that the protection at Tier 1 with standard indicator species tested in artificial soil is achieved for this data set and the selected Specific Protection Goal. Furthermore, some participants pointed out that the protection level at the current Tier 1 ERA is not achieved, by referring to Kotschik et al. (2019; presentation at SETAC EUROPE 29th annual meeting, Helsinki). The discussion on the conservativeness of the Tier 1 risk assessment should also consider the upcoming changes at the exposure side (new guidance document on PECsoil estimation; EFSA, 2017). It was indicated that assuming higher PECsoil compared to past assessments would enhance the protection level at Tier 1 compared to the current assessment scheme, independently from the soil that is used in an ecotoxicological test. The new exposure guidance (EFSA, 2017) includes soil scenarios to cover realistic worst-case PECsoil values for different regulatory zones in the EU. These scenarios are partly characterized by extreme soil conditions, for example, very high OM content leading to low soil bulk density (recalculated via pedotransfer function; Tiktak et al., 2002) and therefore high exposure concentrations. High OM content in ecotoxicological tests with soil organisms can lead to lower bioavailability of PPPs; however, as discussed above, this is not clear cut. In such environmental conditions, toxicity and risk to soil organisms can potentially be much lower (e.g., if such a soil would be used in ecotoxicological tests). It was pointed out at the workshop that the inconsistency between soil scenarios in ecotoxicological tests and those used for exposure estimation can lead to an unrealistic combination of worst-case situations and an overly conservative Tier 1 risk assessment, as from both sides (effect and exposure), worst-case situations are considered (Schimera et al., 2022). As stated above, it needs to be scientifically unraveled which parameters drive bioavailability of different PPPs, followed by the development of an appropriately conservative and consistent tiered ERA scheme.

The question on how to use data from tests with non-standard single species in the ERA was discussed intensely. One option could be to select the lowest endpoint from a set of species and reduce the AF to account for less uncertainty, that is, in the case of limited data availability. The majority of the participants were in favor of using an SSD approach as an intermediate risk assessment tier, for example, by comparing an  $HC_5$  with the PEC.

The EFSA aquatic guidance (EFSA PPR Panel, 2013) was highlighted as providing background on approaches for using SSDs in the ERA. It sets out details on numbers of relevant species needed for SSD derivation depending on the organism group in focus, for example, algae/plants, invertebrates, and fish. Based on the sensitivity of different

groups, it also provides guidance on which group to focus on, for example, for insecticides, the focus could be on Insecta or Crustacea. However, it was noted that aquatic risk is evaluated only for off-field communities as opposed to in-field for soil organisms. Hence, protection goals and AFs might differ between aquatic and soil compartments. The SSD approach is also used in other areas of chemical industry regulations, for example, metals or industrial chemicals. As such, it was mentioned that a precedent is in place for the use of the SSD approach for soil organisms as an intermediate-tier refinement. Experience with these compounds could be reviewed, as, for example, done in the EFSA PPR statement on transition metals (EFSA PPR Panel, 2021).

Assessing recovery in laboratory studies for risk assessment is regarded as a challenge. It was agreed by the workshop participants that, for example, multigeneration tests with soil organisms can provide evidence on the intrinsic recovery potential of the tested indicator species. However, using this approach in risk assessment to assess potential for recovery from PPP use may pose the risk that the recovery is overestimated (compared to the field situation) while using fast-reproducing indicator species. Recovery of one standard species like *F. candida*—which is selected due to the short generation time and not reproducing sexually—in laboratory tests is difficult to correlate with recovery traits of communities in the field. Some species may not recover as fast as *F. candida* does, and it should be assessed whether the AF used in the risk assessment may need to increase for such type of studies. However, the aspect of univoltinism (i.e., only one generation per year) is seen for Collembola mainly in arctic and boreal regions (Hale, 1965; Potapow, 2001) and might not be so common in temperate and warmer regions (Vegter, 1987). Hence, comparisons with field data could provide more clarity on this.

The possibility to apply the test chemical(s) multiple times and with this, to simulate a specific use pattern, was discussed as a potential refinement option. Using natural test soils in multigeneration tests can add another level of realism on the exposure profile over time. It remained unclear whether using natural soils would represent a worst case (in the case of low OM content) or a best case (considering possible stronger microbiological degradation of test items).

The different designs discussed above may be used in a flexible way to address different questions or uncertainties in the risk assessment, for example, potential delayed effects, multiple applications of the same or different PPPs, or assessment of the duration of effects of PPPs. Multigeneration tests were shown to be able to discriminate between compounds with short- and long-term effects (Amorim et al., 2016; Bicho et al., 2017; Van Gestel et al., 2017). It was agreed that the test can provide useful information for the ERA of PPPs. The generation of more data from this ERA approach (i.e., comparison of recovery times in, for example, multigeneration studies and field studies) could help to better understand whether the

potential for recovery from multigeneration tests does in fact correlate with actual field recovery of species with different traits.

Soil multispecies systems were seen as an option by the participants to refine risks seen at lower tiers by including interactions between species. Additional factors, such as competition and predation, the inclusion of potential indirect effects, and exposure via contaminated prey, increase ecological realism of this test system compared to available standard tests. Depending on the test design (number of species and tested concentrations), various endpoints could be derived (e.g., NOEC, EC<sub>x</sub>, output of SSD, or PRC). As this test system is more complex, and includes interactions between different species, standardization was regarded as a huge challenge. However, a step toward standardization of the species composition represents an opportunity to better understand the dynamics of the expected species interactions in tests with different compounds. Although there is some limited experience with this test system in some scientific working groups, it was agreed that further research is necessary with regard to standardization, reproducibility, interpretation, and risk assessment calibration (Table 2). An ERA calibration exercise is necessary to determine if or how the AF can be adjusted to account for the increased realism in an SMS (e.g., predator–prey interactions, competition, indirect effects) compared to single-species tests. It should also be clarified how the results of tests with different species or trophic levels impact the relationship between test outcome and possible field responses.

The participants agreed on how SMS relate to semifield TMEs (Schäffer et al., 2010), which work with intact soil cores. Terrestrial model ecosystems are considered to be closer to a field study, showing higher realism, but with issues of possible larger variability and associated statistical challenges, compared to laboratory-based approaches. Terrestrial model ecosystems assess effects on natural soil organism communities, including more complex interactions. In comparison to artificially constructed communities, the responses of TME systems to PPP application are deemed more realistic, as species interactions are naturally present and pre-established. Compared to the field situation, TMEs represent an isolated system and the population stability of the test system may be reduced due its limited size and ecological context compared to the field situation. Field studies allow more flexibility with regard to soil management practices and number of organism groups to be investigated. Some participants argued that TMEs are similar in complexity, costs, and in particular time effort compared to field studies. Some participants highlighted the higher statistical power of TME studies compared to field tests. However, other participants pointed out that the possible number of subplot replication is limited in a TME and also the capacities for indoor and outdoor TME in the test facilities can be a limiting factor.

Participants agreed that the availability of additional options to integrate knowledge on the impact of PPPs on soil

organisms would be very valuable, and the implementation of intermediate-tiered steps should be further considered in the ERA of PPPs. According to the EFSA PPR Panel (2017), a tiered approach needs to be appropriately protective, internally consistent, and cost-effective. The options discussed above have the potential to reduce uncertainties and add realism to ERA, while in some cases, specific technical and regulatory challenges exist (Table 2).

Further research is necessary in the context of standardization, reproducibility of new test guidelines, and risk assessment calibration of potential intermediate-tier approaches. The generation of case studies (including studies from Tier 1, different intermediate-tier options, and higher-tier studies with the same compound) is necessary to better understand the relationship between the outcome of the different test options and the specific and general protection goals. If evidence is available to support specific approaches, these could potentially be taken up by regulatory authorities.

## AUTHOR CONTRIBUTION

**Gregor Ernst:** Conceptualization; supervision; writing—original draft; writing—review and editing. **Monica J. B. Amorim:** Writing—original draft; writing—review and editing. **Melanie Bottoms:** Conceptualization; writing—original draft; writing—review and editing. **Amy C. Brooks:** Writing—review and editing. **Mark E. Hodson:** Writing—review and editing. **Stefan Kimmel:** Writing—review and editing. **Pia Kotschik:** Writing—review and editing. **Michael T. Marx:** Writing—original draft; writing—review and editing. **Tiago Natal-da-Luz:** Conceptualization; writing—original draft; writing—review and editing. **Céline Pelosi:** Writing—review and editing. **Silvia Pieper:** Writing—review and editing. **Agnes Schimera:** Writing—review and editing. **Janeck Scott-Fordsmann:** Conceptualization; writing—original draft; writing—review and editing. **Amanda Sharples:** Writing—review and editing. **José P. Sousa:** Writing—review and editing. **Cornelis A. M. van Gestel:** Conceptualization; writing—original draft; writing—review and editing. **Bart van Hall:** Conceptualization; writing—original draft; writing—review and editing. **Matthias Bergtold:** Conceptualization; writing—original draft; writing—review and editing.

## ACKNOWLEDGMENT

The authors would like to thank all participants (who are not listed as authors, but acknowledged in the supporting information) for the lively discussion at the workshop.

## CONFLICT OF INTEREST


The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

All data used in this publication are publicly available or can be made available upon request from corresponding author Gregor Ernst (gregor.ernst@bayer.com).

## ORCID

Pia Kotschik  <http://orcid.org/0000-0002-9983-6961>

Cornelis A. M. Gestel  <https://orcid.org/0000-0002-5651-0208>

## SUPPORTING INFORMATION

**Table S1:** List of workshop participants (with affiliations), sorted in the order: industry, academia, regulatory authorities, contract research organizations.

## REFERENCES

- Amorim, M., Pereira, C., Soares, A. M. V. M., & Scott-Fordsmand, J. (2016). Does long term low impact stress cause population extinction? *Environmental Pollution*, 220, 1014–1023.
- Amorim, M. J. B., Römbke, J., & Scheffczyk, A. (2005). Effects of different soil types on the collembolans *Folsomia candida* and *Hypogastrura assimilis* using the herbicide phenmedipham. *Archives of Environmental Contamination and Toxicology*, 49, 343–352. <https://doi.org/10.1007/s00244-004-0220-z>
- Bandow, C., Coors, A., Karau, N., & Römbke, J. (2014). Interactive effects of lambda-cyhalothrin, soil moisture, and temperature on *Folsomia candida* and *Sinella curviseta* (Collembola). *Environmental Toxicology and Chemistry*, 33(3), 654–661.
- Bandow, C., Karau, N., & Römbke, J. (2014). Interactive effects of pyrimethanil, soil moisture and temperature on *Folsomia candida* and *Sinella curviseta* (Collembola). *Applies Soil Ecology*, 81, 22–29.
- Bicho, R., Santos, F. C. F., Scott-Fordsmand, J., & Amorim, M. (2017). Multigenerational effects of copper nanomaterials (CuONMs) are different of those of CuCl<sub>2</sub>: Exposure in the soil invertebrate *Enchytraeus crypticus*. *Scientific Reports*, 7, 8457.
- Candolfi, M. P., Barret, K. L., Campbell, P. J., Forster, R., Grandy, N., Huet, M. C., Lewis, G., Oomen, P. A., Schmuck, R., & Vogt, H. (2001). *Guidance document on regulatory testing and risk assessment procedures for plant protection products with non-target arthropods*. SETAC Office.
- Carniel, L. S. C. (2019). *Novel approaches and future directions for pesticide—Ecological risk assessment to in-soil fauna* [Master's thesis, Universidade do Estado de Santa Catarina].
- Chelinho, S., Domene, X., Anfrés, P., Natal-da-Luz, T., Norte, C., Rufino, C., Lopes, I., Cachada, A., Espindola, E., Ribeiro, R., Duarte, A. C., & Sousa, J. P. (2014). Soil microarthropod community testing: A new approach to increase the ecological relevance of effect data for pesticide risk assessment. *Applied Soil Ecology*, 83, 200–209.
- Chelinho, S., Domene, X., Campana, P., Natal-da-Luz, T., Scheffczyk, A., Römbke, J., Andrés, P., & Sousa, J. P. (2011). Improving ecological risk assessment in the Mediterranean area: Selection of reference soils and evaluating the influence of soil properties on avoidance and reproduction of tow oligochaete springtails. *Environmental Toxicology and Chemistry*, 30(5), 1050–1058.
- Christl, H., Bendall, J., Bergtold, M., Coulson, M., Dinter, A., Garlej, B., Hammel, K., Kabouw, P., Sharples, A., Mérey, G., Vrbka, S., & Ernst, G. (2016). Recalibration of the earthworm Tier 1 risk assessment of plant protection products. *Integrated Environmental Assessment and Management*, 12(4), 643–650.
- Creamer, R. E., Barel, J. M., Bongiorno, G., & Zwetslot, M. J. (2022). The life of soils: Integrating the who and how of multifunctionality. *Soil Biology and Biochemistry*, 166, 108561.
- De Lima e Silva, C., van Haren, C., Mainard, G., de Rooij, W., Ligtelijn, M., van Straalen, N. M., & van Gestel, C. A. M. (2021). Bringing ecology into toxicology: Life-cycle toxicity of two neonicotinoids to four different species of springtails in Lufa 2.2 natural soil. *Chemosphere*, 263, 128245.
- De Silva, P. M. C. S., Pathiratne, A., & van Gestel, C. A. M. (2010). Toxicity of chlorpyrifos, carbofuran, mancozeb and their formulations to the tropical earthworm *Perionyx excavatus*. *Applied Soil Ecology*, 44, 56–60.
- Diepens, N. J. (2015). *Sediment toxicity testing and prospective risk assessment of organic chemicals* [PhD thesis, Wageningen University].
- Domene, X., Chelinho, S., Campana, P., Natal-da-Luz, T., Alcañiz, J. M., Andrés, P., Römbke, J., & Sousa, J. P. (2011). Influence of soil properties on the performance of *Folsomia candida*: Implications for its use in soil ecotoxicology testing. *Environmental Toxicology and Chemistry*, 30(7), 1497–1505. <https://doi.org/10.1002/etc.533>
- EFSA Panel on Plant Protection Products and their Residues (EFSA PPR Panel). (2013). Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. *EFSA Journal*, 11(7), 3290. <https://doi.org/10.2903/j.efsa.2013.3290>
- EFSA Panel on Plant Protection Products and their Residues (EFSA PPR Panel). (2017). Scientific opinion addressing the state of the science on risk assessment of plant protection products for in-soil organisms. *EFSA Journal*, 15(2), 4690.
- EFSA Panel of the Plant Protection Products and their Residues (EFSA PPR Panel). (2021). Statement of the PPR Panel on a framework for conducting the environmental exposure and risk assessment for transition metals when used as active substances in plant protection products (PPP). *EFSA Journal*, 19(3), 6498.
- Ernst, G., Kabouw, P., Barth, M., Marx, M. T., Frommholz, U., Royer, S., & Friedrich, S. (2016). Assessing the potential for intrinsic recovery in a Collembola two-generation study: Possible implementation in a tiered soil risk assessment approach for plant protection products. *Ecotoxicology*, 25, 1–14. <https://doi.org/10.1007/s10646-015-1560-3>
- European Commission (EC). (2009). Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. *Official Journal of the European Union*, L309, 1–50.
- European Commission (EC). (2011). Commission Regulation (EU) No 546/2011 of 10 June 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards uniform principles for evaluation and authorisation of plant protection products. *Official Journal of the European Union*, L155, 127–175.
- European Commission (EC). (2013a). Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. *Official Journal of the European Union*, L93, 1–84.
- European Commission (EC). (2013b). Commission Regulation (EU) No 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. *Official Journal of the European Union*, L93, 85–152.
- European Food Safety Authority (EFSA). (2010). Scientific opinion on the development of specific protection goal options for environmental risk assessment of pesticides, in particular in relation to the revision of the Guidance Documents on Aquatic and Terrestrial Ecotoxicology (SANCO/3268/2001 and SANCO/10329/2002). *EFSA Journal*, 8(10), 1821.
- European Food Safety Authority (EFSA). (2017). EFSA guidance document for predicting environmental concentrations of active substances of plant protection products and transformation products of these active substances in soil. *EFSA Journal*, 15(10), e04982. <https://doi.org/10.2903/j.efsa.2017.4982>
- Feng, L., Zhang, L., Zhang, Y., Zhang, P., & Jiang, H. (2015). Inhibition and recovery of biomarkers of earthworm *Eisenia fetida* after exposure to thiacloprid. *Environmental Science & Pollution Research*, 22, 9475–9482. <https://doi.org/10.1007/s11356-015-4122-6>
- Ferreira, P., Gabriel, A., Sousa, J. P., & Natal-da-Luz, T. (2022). Representativeness of *Folsomia candida* to assess toxicity of a new generation insecticide in different temperature scenarios. *Science of The Total Environment*, 837, 155712.
- FOCUS. (1997). *Soil persistence models and EU registration*. Final report of the work of the Soil Modelling Work group of FOCUS.
- Forbes, V. E., Agatz, A., Ashauer, R., Butt, K. R., Capowicz, Y., Duquesne, S., Ernst, G., Focks, A., Gergs, A., Hodson, M. E., Holmstrup, M., Johnston, A. S. A., Meli, M., Nickisch, D., Pieper, S., Rakel, K. J., Reed, M.,

- Roembke, J., Schäfer, R. B., ... Roeben, V. (2021). Mechanistic effect modeling of earthworms in the context of pesticide risk assessment: Synthesis of the FORESEE Workshop. *Integrated Environmental Assessment and Management*, 17(2), 352–363.
- Fox, D. R., Van Dam, R. A., Fisher, R., Batley, G. E., Tillmanns, A. R., Thorley, J., Schwarz, C. J., Spry, D. J., & McTavish, K. (2021). Recent developments in species sensitivity distribution modeling. *Environmental Toxicology and Chemistry*, 40(2), 293–308.
- Frampton, G. K., Jänsch, S., Scott-Fordsmand, J. J., Römbke, J., & Van den Brink, P. J. (2006). Effects of pesticides on soil invertebrates in laboratory studies: A review and analysis using species sensitivity distributions. *Environmental Toxicology and Chemistry*, 25(9), 2480–2489.
- Gawlik, B. M., Lamberty, A., Pauwels, J., & Muntau, H. (2004). Certified reference materials to support a European thematic strategy for soil protection: The EUROSOLS (IRMM-443) experience. *Analytical and Bioanalytical Chemistry*, 378(5), 1284–1287.
- Gergs, A., Rakel, K., Bussen, D., Capowicz, Y., Ernst, G., & Roeben, V. (2022). Integrating earthworm movement and life history through dynamic energy budgets. *Conservation Physiology*, 10(1), coac042. <https://doi.org/10.1093/conphys/coac042>
- Guimarães, B., Römbke, J., & Amorim, M. J. B. (2023). On the importance of longer-term exposure to stressors—A critical review and proposal for multigenerational testing in standard soil invertebrates. *Science of The Total Environment*, 854, 158680.
- Hale, W. G. (1965). Observations on the breeding biology of Collembola. I, II. *Pedobiologia*, 5(146–152), 161–177.
- International Standard Organisation. (2014). *Soil quality—Effects of pollutants on earthworms* (ISO 11268-3).
- Jensen, J., & Scott-Fordsmand, J. J. (2012). Ecotoxicity of the veterinary pharmaceutical ivermectin tested in a soil multi-species (SMS) system. *Environmental Pollution*, 171, 133–139.
- Kotschik, P., Frische, T., Swarowsky, K., & Pieper, S. (2019). Verification of the current risk assessment for earthworms, SETAC Europe 29th Annual Meeting—Abstract book. Society of Environmental Toxicology and Chemistry. [https://cdn.ymaws.com/www.setac.org/resource/resmgr/abstract\\_books/setac-helsinki-abstract-book.pdf](https://cdn.ymaws.com/www.setac.org/resource/resmgr/abstract_books/setac-helsinki-abstract-book.pdf)
- Kotschik, P., Sousa, J. P., Natal-da-Luz, T., Chelinho, S., Silva, A. L., Capela, N., Römbke, J., & Pieper, S. (2018). How protective is the current risk assessment for soil invertebrates? SETAC Europe 28th Annual Meeting—Abstract book. Society of Environmental Toxicology and Chemistry. [https://cdn.ymaws.com/www.setac.org/resource/resmgr/abstract\\_books/SETAC\\_Europe\\_Rome\\_abstract\\_b.pdf](https://cdn.ymaws.com/www.setac.org/resource/resmgr/abstract_books/SETAC_Europe_Rome_abstract_b.pdf)
- Kovacevic, M., Hackenberger, D. K., & Hackenberger, B. K. (2020). Effects of strobilurin fungicides (azoxystrobin, pyraclostrobin, and trifloxystrobin) on survival, reproduction and hatching success of *Enchytraeus crypticus*. *Science of The Total Environment*, 790, 148143.
- Kreutzweiser, D. P., Good, K. P., Chartrand, D. T., Scarr, T. A., Holms, S. B., & Thompson, D. G. (2008). Effects on litter-dwelling earthworms and microbial decomposition of soil-applied imidacloprid for control of wood-boring insects. *Pest Management Science*, 64, 112–118.
- Kuhnt, G., & Muntau, H. (Eds.). (1994). *EURO-soils: Identification, collection, treatment, characterization*. Joint Research Centre, European Commission, Ispra.
- Kwak, J. I., Moon, J., Kim, D., Cui, R., & An, Y.-J. (2018). Determination of the soil hazardous concentrations of bisphenol A using the species sensitivity approach. *Journal of Hazardous Materials*, 344, 390–397.
- Lee, Y.-S., Son, J., Wee, J., Kim, Y., Hong, J., & Cho, K. (2020). A reconsideration of the safety of fenoxycarb (IGR) in soil environment: The toxicity of fenoxycarb to *Yuukianura szepteyckii* (Collembola). *Journal of Asia-Pacific Entomology*, 23, 214–218.
- Lokke, H., & Van Gestel, C. A. M. (1998). *Handbook of soil invertebrate toxicity tests* (p. 281). John Wiley and Sons.
- Maltby, L., Blake, N., Brock, T. C. M., & Van den Brink, P. J. (2005). Insecticide species sensitivity distributions: Importance of test species selection and relevance to aquatic ecosystems. *Environmental Toxicology*, 24, 379–388.
- Mendes, L. A., Amorim, M. J. B., & Scott-Fordsmand, J. J. (2019). Assessing the toxicity of safer by design CuO surface-modifications using terrestrial multispecies assays. *Science of The Total Environment*, 678, 457–465.
- Noël, H. L., Hopkin, S. P., Hutchinson, T. H., Williams, T. D., & Sibly, R. M. (2006). Population growth rate and carrying capacity for springtails *Folsomia candida* exposed to ivermectin. *Ecological Applications*, 16, 656–665.
- Organization for Economic Co-operation and Development (OECD). (2004). *Earthworm reproduction test (Eisenia fetida/Eisenia andrei)* (No. 222). OECD guidelines for the testing of chemicals.
- Organization for Economic Co-operation and Development (OECD). (2008). *Predatory mite (Hypoaspis [Geolaelaps] aculeifer) reproduction test in soil* (No. 226). OECD guidelines for the testing of chemicals.
- Organization for Economic Co-operation and Development (OECD). (2009). *Collembolan reproduction test in soil* (No. 232). OECD guidelines for the testing of chemicals.
- Pelosi, C., Joimel, S., & Makowski, D. (2013). Searching for a more sensitive earthworm species to be used in pesticide homologation tests—A meta-analysis. *Chemosphere*, 90, 895–900.
- Potapow, M. (2001). *Synopses on Palearctic Collembola. Isotomidae. Staatliches Museum für Naturkunde Götting* (Vol. 3). Winter Druck.
- Ramon, R., Scopel Camiel, L., Tornisiolo, A., Wolff, C., Scorza Junior, R., Torrado, P., Marques Pereira Leal, R., & Jene, B. (2022). The underground of the crops—Representative agricultural soils in Brazil through a GIS based approach, SETAC Europe 32nd Annual Meeting. Society of Environmental Toxicology and Chemistry.
- Reid, W. V., & Carpenter, S. R., Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Island Press.
- Renaud, M., Martins da Silva, P., Natal-da-Luz, T., Siciliano, S. D., & Sousa, J. P. (2021). Community effect concentrations as a new concept to easily incorporate community data in environmental effect assessment of complex metal mixtures. *Journal of Hazardous Materials*, 411, 125088.
- Roeben, V., Oberdoerster, S., Rakel, K., Liesy, D., Capowicz, Y., Ernst, G., Preuss, T., Gergs, A., & Oberdoerster, C. (2020). Towards a spatiotemporally explicit toxicokinetic-toxicodynamic model for earthworm toxicity. *Science of The Total Environment*, 722, 137673.
- Römbke, J., & Amorim, M. (2004). Tackling the heterogeneity of soils in ecotoxicological testing—An EURO-soil base approach. *Journal of Soils and Sediments*, 4, 276–281.
- Römbke, J., Förster, B., Jänsch, S., Kaiser, F., Scheffczyk, A., Roß-Nickoll, M., Daniels, B., Ottermanns, R., & Scholz-Starke, B. (2020). Necessary adaptations for a harmonized field-testing procedure and risk assessment of earthworms (terrestrial) (TEXTE 00/2020). Umweltbundesamt.
- SANCO. (2002, October 17). *Guidance document on terrestrial ecotoxicology under Council Directive 91/414/EEC* (SANCO/10329/2002 rev 2 final). European Commission, Directorate E—Food Safety: Plant health, animal health and welfare, international questions.
- Schäffer, A., van den Brink, P. J., Heimbach, F., Hoy, S. P., de Jong, F. M. W., Römbke, J., Roß-Nickoll, M., & Sousa, J. P. (2010). *Semi-field methods for the environmental risk assessment of pesticides in soil*. CRC Press.
- Schimera, A., Multsch, S., Bottoms, M., Guevara Montemayor, O. I., Hammel, K., Kimmel, S., Loutseti, S., Marx, M. T., Sharples, A., Staab, F., Szegedi, K., Voß, F., & Ernst, G. (2022). New tools for predicting environmental concentrations in soil in regulatory context—Potential implications for the soil organism risk assessment of Plant Protection Products, SETAC Europe 32nd Annual Meeting—Abstract book. Society of Environmental Toxicology and Chemistry. <https://www.setac.org/resource/setac-eu-copenhagen-abstract-book.html>
- Schnug, L., Jensen, J., Scott-Fordsmand, J. J., & Leinaas, H. P. (2014). Toxicity of three biocides to springtails and earthworms in a soil multi-species (SMS) test system. *Soil Biology & Biochemistry*, 74, 115–126. <https://doi.org/10.1016/j.soilbio.2014.03.007>
- Scott-Fordsmand, J. J., Maraldo, K., & van den Brink, P. J. (2008). The toxicity of copper contaminated soil using a gnotobiotic soil multi-species test system (SMS). *Environment International*, 34(4), 524–530. <https://doi.org/10.1016/j.envint.2007.11.008>

- Silva, P. V., Silva, A. R. R., Mendo, S., & Loureiro, S. (2014). Toxicity of tributyltin (TBT) to terrestrial organisms and its species sensitivity distribution. *Science of The Total Environment*, 466–467, 1037–1046.
- Tiktak, A., de Nie, D. S., van der Linde, A. M. A., & Kruijne, R. (2002). Modelling the leaching and drainage of pesticides in the Netherlands: The GeoPEARL model. *Agronomie*, 22, 373–387.
- Toschki, A., Hammers-Wirtz, M., Hommen, U., Klein, M., Poßberg, C., Römbke, J., Roß-Nickoll, M., Schäffer, A., Scheffczyk, A., Schmidt, B., & Scholz Starke, B. (2020). *Evaluation of the risk for soil organisms under real conditions* (p. 302). Umweltbundesamt.
- Van Gestel, C. A. M., De Lima e Silva, C., Lam, T., Koekkoek, J. C., Lamoree, M. H., & Verweij, R. A. (2017). Multigeneration toxicity of imidacloprid and thiacloprid to *Folsomia candida*. *Ecotoxicology*, 26, 320–328. <https://doi.org/10.1007/s10646-017-1765-8>
- Van Gestel, C. A. M., Van Dis, W. A., Van Breemen, E. M., & Sparenburg, P. M. (1989). Development of a standardized reproduction toxicity test with the earthworm species *Eisenia fetida/andrei* using copper, pentachlorophenol, and 2,4-dichloroaniline. *Ecotoxicology and Environmental Safety*, 18, 305–312.
- Van Hall, B., Sweeney, C., Bottoms, M., & Van Gestel, C. A. M. (2023). The influence of soil organic matter content on the toxicity of pesticides to soil invertebrates: A review. *Integrated Environmental Assessment and Management*, 1–16. <https://doi.org/10.1002/ieam.4770>
- Vegter, J. J. (1987). Phenology and seasonal resource partitioning in forest floor *Collembola*. *Oikos*, 48, 175–185.