Environmental Sciences Europe

Open Access

Close to reality? Micro-/mesocosm communities do not represent natural macroinvertebrate communities

Lena Reiber[1,2]*, Kaarina Foit[3], Matthias Liess[1,2]*, Bilgin Karaoglan[3], Joern Wogram[3] and Sabine Duquesne[3]

Abstract

Background: The European environmental risk assessment of plant protection products considers aquatic model ecosystem studies (microcosms/mesocosms, M/M) as suitable higher tier approach to assess treatment-related effects and to derive regulatory acceptable concentrations (RAC). However, it is under debate to what extent these artificial test systems reflect the risks of pesticidal substances with potential harmful effects on natural macroinvertebrate communities, and whether the field communities are adequately protected by the results of the M/M studies. We therefore compared the composition, sensitivity and vulnerability of benthic macroinvertebrates established in control (untreated) groups of 47 selected M/M studies with natural stream communities at 26 reference field sites.

Results: Since 2013 the number of benthic macroinvertebrate taxa present in M/M studies has increased by 39% to a mean of 38 families per study. However, there is only an average of 4 families per study that comply with the recommendations provided by EFSA (EFSA J 11:3290, 2013), i.e.: (i) allowing statistical identification of treatment-related effects of at least 70% according to the *minimum detectable difference* (here criteria are slightly modified) and (ii) belonging to insects or crustaceans (potentially sensitive taxa for pesticidal substances). Applying the criterion of physiological sensitivity according to the SPEAR_{pesticides} concept, the number of families decreases from 4 to 2.3 per study.

Conclusions: Most taxa established in recent M/M studies do not suitably represent natural freshwater communities. First, because their abundances are often not sufficient for statistical detection of treatment-related effects in order to determine an appropriate endpoint and subsequent RAC. Recommendations are given to improve the detectability of such effects and their reliability. Second, the taxa often do not represent especially sensitive or vulnerable taxa in natural communities in terms of their traits. The uncertainties linked to vulnerable taxa in M/M studies are especially high considering their representativity for field assemblages and the comparability of factors determining their recovery time. Thus considering recovery for deriving a RAC (i.e., ERO-RAC) is not recommended. In addition, this paper discusses further concerns regarding M/M studies in a broader regulatory context and recommends the development of alternative assessment tools and a shift towards a new paradigm.

Keywords: Microcosms, Mesocosms, Macroinvertebrates, Vulnerability, Sensitivity, Natural stream communities, Pesticidal substances, Minimum detectable difference, Risk assessment

Background

Pesticides are a major stressor affecting aquatic ecosystems worldwide [e.g., 1–4, 5–9]. They are released into the environment from diffuse agricultural sources, such as surface run-offs after heavy rainfall events [10] and

*Correspondence: lena.reiber@ufz.de; matthias.liess@ufz.de

[1] Department System-Ecotoxicology, UFZ Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
Full list of author information is available at the end of the article

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 2 of 19

spray-drift [11], or in pulses through accidents or illegal discharges [12] or during cleaning of agricultural equipment [13]. Pesticides were shown to reduce biodiversity, impair ecosystem functions and harm especially pesticide-sensitive and vulnerable macroinvertebrates in freshwater streams [5, 14–16]. Improvements in the pesticide authorization process are therefore essential to protect the environment more effectively [4].

In Europe, the process for authorization of active substances and approval of products is based on regulations (EU) No. 545/2011 and No. 546/2011 implementing Regulation (EC) No. 1107/2009 regarding the data requirements and uniform principles for evaluation and authorization of plant protection products. The approach and recommendations for a tiered aquatic risk assessment was proposed by the Panel on Plant Protection Products and their Residues (PPR Panel) in the 'Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge of-field surface waters' [AGD (Aquatic Guidance Document); EFSA [17]. In the tiered approach of the AGD [17], a higher tier is usually proposed as a refinement when the risks based on a lower tier approach (e.g., based on single species tests or species sensitivity distributions) exceed acceptable levels. The highest experimental tier in aquatic risk assessment includes studies in aquatic model ecosystem experiments, referred to as microcosm and mesocosm (M/M) studies. While different definitions have been used for M/M test systems, here the definition of the AGD [17] is followed stating that M/M studies mainly differ with respect to their size with mesocosms usually being larger than 15 m^3 water volume or 15 m length. The tiered approach enables to derive *regulatory acceptable concentrations* (RACs) that can be based on the *ecological threshold option* (ETO) which only allows for no/negligible population effects, or on the *ecological recovery option* (ERO) which allows for some population-level effects if recovery takes place within 8 weeks after treatment [17]. For the ETO-RAC, the physiological (=intrinsic) sensitivity is the essential trait of a taxon and the *no or lowest observed effect concentration* (NOEC/LOEC) is the relevant endpoint. For the ERO-RAC, the vulnerability (characterized by both the physiological sensitivity and by life-cycle traits that condition the recovery potential) is the essential trait of a taxon and the *no observed ecologically adverse effect concentration* (NOEAEC) is the relevant endpoint.

Although used in the risk assessment as a representative for both, lotic and lentic water bodies, M/M studies performed for regulatory purposes are usually carried out in artificial lentic systems. In order to be appropriate for prospective risk assessment of pesticides, the composition of the community established in an M/M should be such that its sensitivity and vulnerability are similar to those of representative field communities from natural ecosystems. Furthermore, only small uncertainty factors (usually from 2 to 3 for ETO-RAC derivation and from 3 to 4 for ERO-RAC derivation) are applied to extrapolate from the effect levels in M/M studies. Therefore, the AGD [17] requires that conditions in the test systems are sufficiently representative of natural ecosystems. According to the AGD [17], a representative assemblage for edge-of-field surface waters contains the important taxonomical groups, trophic groups and ecological traits typical for communities in ponds, ditches and/or streams. At least eight different populations of the sensitive taxonomic groups need to be present in the test system, for which a concentration–response relationship can be derived [17]. For pesticidal substances with potential harmful effects on macroinvertebrate communities such as insecticides and some fungicides, sensitive taxonomic groups are usually insects and/or crustaceans [17]. The establishment of a representative aquatic community in M/M test systems is one major aspect determining the reliability of the assessment when extrapolating results from experiments to natural waterbodies [18]. Different factors were shown to affect the representativeness of M/M studies. Among them the habitat characteristics in the test system are important [2, 18–20], as well as the magnitude of the environmental stress (as it may increase the sensitivity of populations by a factor of more than one order of magnitude), and the method of species' establishment in the test system including natural colonization and artificial "seeding" [20–22]. Beketov et al. [23] raised critical concerns about the typical low proportion of long-living invertebrate taxa in mesocosms. The authors of this study increased the number of these long-living taxa in their experiments by an extended pre-exposure time and started with the establishment of macroinvertebrates approximately one and a half years before contamination. Williams et al. [18] concluded that macroinvertebrate assemblages in microcosms would only partially resemble natural pond assemblages and that the extent to which M/M realistically represent natural systems remains a subject of debate.

Another important point is that the abundance data of M/M studies need to enable statistical evaluation. To support the ecological evaluation of abundance data and to increase transparency and robustness of M/M results, the AGD [17] suggests to report the critical endpoints together with the results on the *minimum detectable difference* (MDD). The MDD defines the difference between the means of a treatment and the control that must exist to detect a statistically significant effect [24]. Hence,

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 3 of 19

it indicates the actual effect that can be detected in the experiment for a given endpoint at a given time.

Taking all these aspects into consideration the aim of this study was to analyze the composition of benthic macroinvertebrate communities in M/M studies conducted for the risk assessment of pesticidal substances and to compare them qualitatively and quantitatively with natural assemblages at reference sites of freshwater streams. For both, M/M study and field site datasets, we evaluated the presence of (i) sensitive taxa defined according to the AGD [17]; (ii) physiologically sensitive taxa defined according to the trait-based indicator SPEAR$_{pesticides}$ [14] and (iii) vulnerable taxa according to SPEAR$_{pesticides}$. SPEAR$_{pesticides}$ (*species at risk of pesticides index*) was developed by Liess and von der Ohe [14] and further adapted by Knillmann et al. [25] and is well-established [e.g., 1, 5, 6, 26–30]. In addition, the suitability of abundance data in M/M studies for demonstrating treatment-related effects were checked by calculating the MDD. Based on these analyses, we provide an overview on taxa with suitable and unsuitable data in M/M studies, with the suitability being defined by the sensitivity criterion according to the AGD [17] and the MDD-values. We also contribute to a grouping of taxa based on traits, as recommended by the AGD [17].

Methods

Benthic macroinvertebrate data from microcosm/mesocosm studies

Our analysis refers to microcosm and mesocosm (M/M) studies conducted for regulatory purposes. The sample available to us for the following data analysis includes macroinvertebrate data from 66 M/M studies submitted to the German Environment Agency for regulatory risk assessment. These M/M studies were conducted from 1986 until 2018 to test the effects of insecticides, insect growth regulators (IGR), fungicides or mixtures of these substance groups, sometimes including additionally herbicides. M/M studies testing only herbicides were excluded from this analysis as they do not require to focus on macroinvertebrates but on algae and/or macrophytes as sensitive taxa [17]. Furthermore, M/M studies with free-living fish were excluded as they are not accepted for pesticide authorization purposes. Available M/M studies were performed indoors or outdoors. Most test systems were lentic ponds. In rare cases, enclosures were placed in artificial ditches. Test systems varied in shape (round or square) and size. Details on all M/M studies included in the analysis can be found in Table 2, Appendix.

For the present analysis, we included only studies with (i) a minimum of 10 different taxa (families or groups) to ensure the establishment of a representative macroinvertebrate community, and (ii) at least two replicates per control/treatment to allow statistical derivation of significant effects (see also section "Calculation of the minimum detectable difference (MDD)"). We further restricted the dataset by including only data (i) from control (untreated) test systems to analyze the general composition of macroinvertebrate communities which may potentially establish in M/M system; (ii) from the first 30 days after pesticide application in the corresponding treated test systems because within this period the community structure of macroinvertebrates should be optimized for effect monitoring; (iii) of sampling methods adapted for benthic macroinvertebrates. After these selection steps, 51 M/M studies remained for analysis of the macroinvertebrate datasets.

For further processing of macroinvertebrate data, we distinguished between "aquatic sampling" (e.g., enhanced surface area substrate sampling, sweep net sampling and sediment sampling) and "emergence sampling" (e.g., floating emergence traps). In order to obtain only one aquatic abundance per taxon and time point, the individual abundances of the different aquatic sampling techniques were summed up. Macroinvertebrate data from "emergence sampling" could only be used for the analysis of M/M studies conducted in 2013 and later, since previous studies rarely contained emergence sampling data. To increase the homogeneity of the dataset, two studies conducted after 2013 were excluded from the analysis because they did not report emergence data. This resulted in a total number of 49 M/M studies, of which seven M/M studies were conducted in 2013 and later which provided both emergence and aquatic data. The latter seven M/M studies are referred to as recent M/M studies in the following.

Macroinvertebrate data were further processed as follows: entries of taxa in the stages clutch/clutches, cocoon, egg/eggs or exuvia (except for entries in the emergence dataset) were excluded. Taxa belonging to other groups than benthic macroinvertebrates such as zooplankton, terrestrial macroinvertebrates or amphibians were excluded. Abundance of species, genii and families was aggregated at family level (or higher in case of lower detection level, such as order level). The level of identification was very different in the initial datasets; on average 47% of all taxa in the M/M studies selected for analysis have been provided on family level or higher. Hence, the aggregation on family level could not be avoided in order to compare taxa (i) of the different M/M studies; (ii) of the aquatic and emergence dataset of one M/M study, and (iii) of the M/M and field site dataset (see section "Benthic macroinvertebrate data from the field

monitoring study"). Per M/M study, a family represents on average a mean of 1.4 taxa and a median of one taxon (at the level of identification provided by the study). Only in two cases, exceptionally high numbers of aggregated taxa were achieved for the family of Chironomidae, with 25 and 28 represented species/genii, respectively. In the following, the taxon level is referred to as family level.

For the emergence dataset, the abundance per taxon and study was cumulated over time to obtain the total number of emerged adults. Abundance (or cumulated abundance in case of emergence data) was furthermore ln-transformed applying the formula $y(x) = \ln(2x + 1)$, where x is the measured abundance [see also 31].

During analysis of the benthic macroinvertebrate community, a special focus was set on insects and crustaceans (i/c), as they are defined as potentially sensitive taxonomic group for pesticidal substances according to the AGD [17].

Calculation of the minimum detectable difference (MDD)

The *minimum detectable difference* (MDD) is the smallest difference between the means of a treatment and the control that must exist to detect a statistically significant effect for a given experimental endpoint at a given time and at a defined degree of certainty [24, 32]. The specification of the MDD is particularly important when no statistically significant effect is observed to distinguish between cases where indeed no effect occurred and cases where an effect occurred but without statistical evidence.

We calculated the MDD, separately for the aquatic and emergence abundance, by:

$$\text{MDD} = tx\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}, \tag{1}$$

where s_1^2 is the variance of abundance in the control, n_1 is the number of control replicates and n_2 the median of treatment replicates. The parameter t can be expressed as $t_{1-\alpha, df}$ and is the quantile of the t-distribution, where df is the degree of freedom. As degree of freedom, we set $n_1 + n_2 - 2$. We applied an α-value of 0.05 for a one-sided t-distribution, because we focused on the detection of adverse effects on the abundance between controls and treatments. This formula is based on the one from Lee and Gurland [33] proposed by the AGD [17] for the calculation of the MDD between the abundance means of a treatment and the control. The formula according to Lee and Gurland [33] contains the variance of abundance and number of replicates of the treatment. As we deliberately chose to assess only the established macroinvertebrate community of the control test systems, we assumed the

same variance in the treatments as in the control test systems (s_1^2) in formula 1. The number of treatment replicates (n_2) was extracted from the M/M studies. In case of different numbers of replicates for different treatment concentrations, we calculated the median number of replicates for the different treatments.

As abundance data were ln-transformed (see section Benthic macroinvertebrate data from microcosm/mesocosm studies), we back-transformed the MDD to the original scale of the abundance data based on Brock et al. [24]. Afterwards, we calculated the MDD as a percentage of the control mean:

$$\text{MDD\%} = \frac{100}{x_1} \times \text{MDD}, \tag{2}$$

where x_1 is the back-transformed abundance mean of the control replicates.

The calculation of MDD% commonly used in the risk assessment of M/M studies, is a simple restructuring of the *T*-test. It therefore only answers the question at which effect threshold a *T*-test would indicate a statistically significant effect. However, this MDD is inappropriate to assess the power of the mesocosm experiment for a given species and on a given day, and thus the robustness of the inferred endpoint (Duquesne et al., 2020). For further explanation on this point, see section "Interpretation of the MDD-values and critical elements".

Categorization of taxa according to their MDD%

We calculated the mean MDD% per taxon and study as an indicator of possible detection of statistically significant effects. For studies performed before 2013, the MDD% was only evaluated for the aquatic data. For studies after 2013, the MDD% was evaluated for both the aquatic and cumulative emerging data separately, and the lowest MDD% was finally selected. An MDD% > 100 could occur if, e.g., a taxon was poorly established in the test systems or absent from single replicates. As these high MDD% values would have strongly biased the mean MDD% per study, they were not considered in the calculation of the mean MDD%. A mean MDD% per taxon and study could only be calculated if at least two samplings with an MDD% ≤ 100 were available. This step reduced the number of M/M studies from 49 to 47 (see section "Benthic macroinvertebrate data from microcosm/mesocosm studies"), as two studies did not contain any taxa for which a mean MDD% could be calculated.

In the AGD [17] it is recommended that MDD% of critical endpoints should ideally be lower than 70%. Accordingly, we classified all taxa per study as (i) families$_{\text{MDD\%low}}$ with a mean MDD% < 70 and (ii) families$_{\text{MDD\%high}}$

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 5 of 19

with a mean MDD% $\geq$ 70. Only taxa of the first group, families$_{MDD\%low}$, were considered as potentially useful for statistical evaluation of a pesticide-induced effect.

In risk assessment, taxa are usually categorized with respect to their suitability for statistical analysis according to Brock et al. [24]. However, this categorization scheme requires a minimum of five samplings which is frequently not fulfilled in our study due to the following constraints: (i) only data within the first 30 days of a study was selected because it is within this period that the effect threshold (NOEC/ LOEC) is usually derived, and (ii) only samplings from time points, when all aquatic sampling methods used in the respective study were applied (to increase homogeneity of the dataset). As we could not implement exactly the categorization of Brock et al. [24], we followed the recommendations of the AGD [17] and used the category families$_{MDD\%low}$ as an indication that the data of the respective taxon might be suitable for the derivation of a treatment-related endpoint. Further information and limitations about the use of MDD according to the AGD [17] and Brock et al. [24] are given in section "Interpretation of the MDD-values and critical elements".

Benthic macroinvertebrate data from the field monitoring study

Macroinvertebrate data from natural freshwater stream sites were obtained from the sampling campaign '*Kleingewässermonitoring*' KGM; [4]. This sampling campaign was conducted in 2018 and 2019 in 12 federal states in Germany. For this analysis, we selected reference sites to obtain a community representative for freshwater stream sites uncontaminated by pesticides, to compare them with the assemblages in M/M control test systems. For this selection, sites were classified into five classes as proposed by the European Water Framework Directive [34] by applying the index SPEAR$_{pesticides}$ [14]. Only sites with the two highest status classes of "good" and "very good" were chosen. For details on the environmental parameters of these sites, see [4].

Samplings in the month of June were selected to obtain a representative community in a time period of potential pesticide exposure, as June is a typical time period for pesticide application, especially for insecticides. In total, datasets of 26 field sites were obtained and the benthic macroinvertebrate data were processed as follows: as for the M/M data (see section "Benthic macroinvertebrate data from microcosm/mesocosm studies" taxa were aggregated at family level (or higher in case of lower detection level, such as order level) to enable comparison of taxa of the M/M and field site dataset. In the following, the taxon level is referred to as family level. For analysis, only taxa present in at least five sampling sites were selected in order to exclude rare taxa that are potentially not representative for reference stream sites.

Analysis of the macroinvertebrate communities in M/M studies and comparison with field assemblages

We assessed the development of all available M/M studies from 1986 until 2018. For this, the studies were grouped into four consecutive (and similarly long) sampling time periods according to their study completion dates, namely before 1998, from 1998 to 2004, from 2005 to 2012 and from 2013 to 2018. The latter time period is expected to represent M/M studies performed in accordance with the AGD [17]. For each time period, the following parameters were calculated: (i) the mean MDD% per study given as the average of all mean MDD% per taxon and study (under exclusion of all MDD% > 100); (ii) the total number of families per study and (iii) the number of families$_{MDD\%low}$ from the taxonomic group insects and crustaceans.

Further data analysis focused on the reduced dataset of the seven selected recent M/M studies (i.e., performed from 2013 until 2018) that also include emergence data. For these recent studies, the number of families$_{MDD\%low}$ and families$_{MDD\%high}$ were calculated and compared for each taxonomic group.

In a next step, the composition of the benthic macroinvertebrate communities that established in M/M studies was compared with natural assemblages at reference stream sites in the field. Thereby, the following issue arises related to the contrasting ecosystem types: M/M studies are mostly carried out in lentic systems, as the studies obtained for this analysis, while edge-of-field surface waters are mostly represented by lotic (flowing) water systems. Due to the generally different biocoenosis of the two ecosystems, a direct comparison of taxa lists would not be useful. Instead, the comparison of the macroinvertebrate taxa was carried out by comparing their sensitivity and vulnerability towards pesticides. For this purpose, only families$_{MDD\%low}$ were considered and the following parameters were calculated for each M/M study and field site separately: (i) number of insects and crustaceans; (ii) number of all other taxonomic groups not belonging to insects and crustaceans; (iii) number of Ephemeroptera, Plecoptera, Trichoptera (EPT); (iv) number of physiological sensitive families defined by the trait-value "physiological sensitivity" (s-value > − 0.36) according to the definition of the index SPEAR$_{pesticides}$, and finally (v) number of families classified as vulnerable according to SPEAR$_{pesticides}$ (where vulnerability is defined by combining the traits physiological sensitivity, generation time, exposure probability and dependence

on refuge areas with the ability to migrate from them). Data were obtained from the SPEAR_pesticides website [35] and are displayed in Fig. 2 and Table 1, Appendix.

Please note that the physiological sensitivity is defined as intrinsic sensitivity.

Further statistical analysis

For all mean comparisons, the respective data were first tested for the homogeneity of variances (Levene's test) and normal distributions of the residuals (Shapiro–Wilk normality test). If assumptions were met, differences between means were tested with an analysis of variance (one-way ANOVA). In cases where the condition of data normality and/or homoscedasticity was not met, differences were calculated using the non-parametric Kruskal–Wallis test. In case of more than two categories for the respective parameter (as for the parameters regarding the development of M/M studies), a non-parametric pairwise test of mean rank sums for multiple comparisons (Dunn's test) was conducted to test for statistically significant differences between the categories. All statistical tests were performed with R Studio (version 1.2.1335).

Results

The evolution of M/M studies over time

The number of benthic macroinvertebrate families in controls of M/M test systems conducted to test pesticidal substances tends to increase after 2013 by 39%, i.e., from a mean of 28 families per study before 2013 (pool of the 3 sampling time periods before 1998, 1998–2004, and 2005–2012) to a mean of 38 families per study after 2013 (Kruskal–Wallis test, $p = 0.2$, see Fig. 1a); please note that 2013 corresponds to the year of the publication of the AGD, [17]. Additionally the number of families with a mean MDD% < 70 (so-called families_MDD%low) belonging to the taxonomic group of insects and crustaceans tends to increase by 58%, i.e., from a mean of 2.5 before 2013 to a mean of 4 respective taxa after 2013 (Kruskal–Wallis test, $p = 0.32$, see Fig. 1b). However, the mean MDD% of all taxa per study remained relatively constant over the four time periods with means of 64% and 70%, before and after 2013, respectively (Kruskal–Wallis test, $p = 0.33$, see Fig. 1c).

Benthic macroinvertebrate families established in recent M/M studies

In the following, we focus on the seven more family-rich M/M studies (see section "Benthic macroinvertebrate data from microcosm/mesocosm studies") performed in 2013 or later. A total of 29 families were monitored in at least one M/M study with a mean MDD% ≤ 100 (Fig. 2). Out of these 29 families, 15 families belong to the taxonomic group insects or crustaceans. Only three families belong to the orders Ephemeroptera and Trichoptera which are known to contain particularly pesticide-sensitive and vulnerable families.

Most of the 29 families were insufficiently established to allow for a statistical analysis of effects. Only one taxon
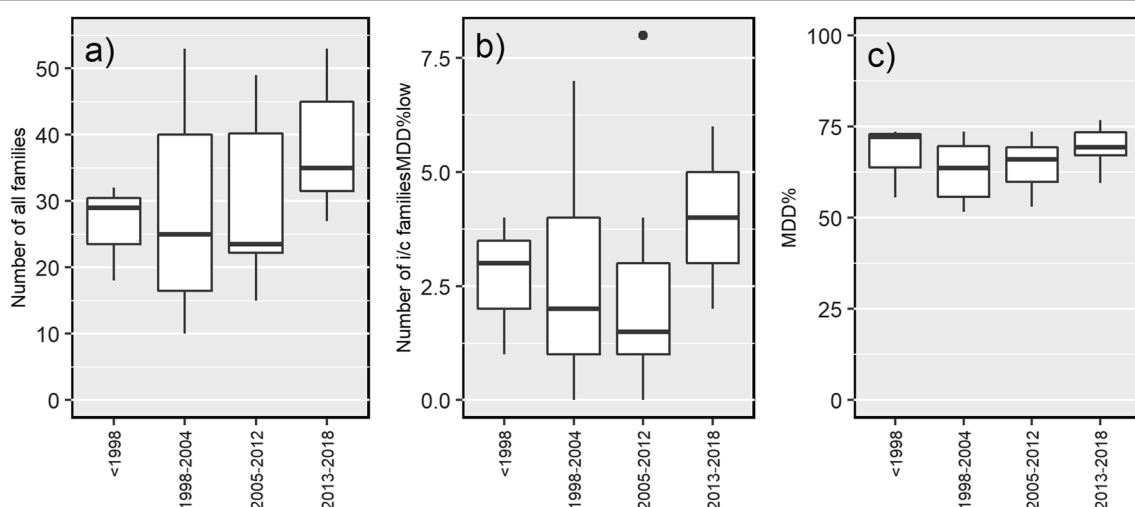

Fig. 1 Evolution of microcosm and mesocosm studies, conducted from 1986 to 2018 to test pesticidal substances, regarding the total number of all families per study (**a**) and the number of families_MDD%low of the taxonomic group insects and crustaceans (i/c) per study (**b**), and the mean MDD% of all taxa per study (**c**). MDD% is the *minimum detectable difference* in % and families_MDD%low are defined as families per study with a mean MDD% < 70. The calculated mean MDD% and the classification as families_MDD%low is based on "aquatic data" only (emergence data were excluded). Please note that the AGD [17] was published in 2013 requiring a minimum of 8 different populations of the taxonomic group insects and crustaceans
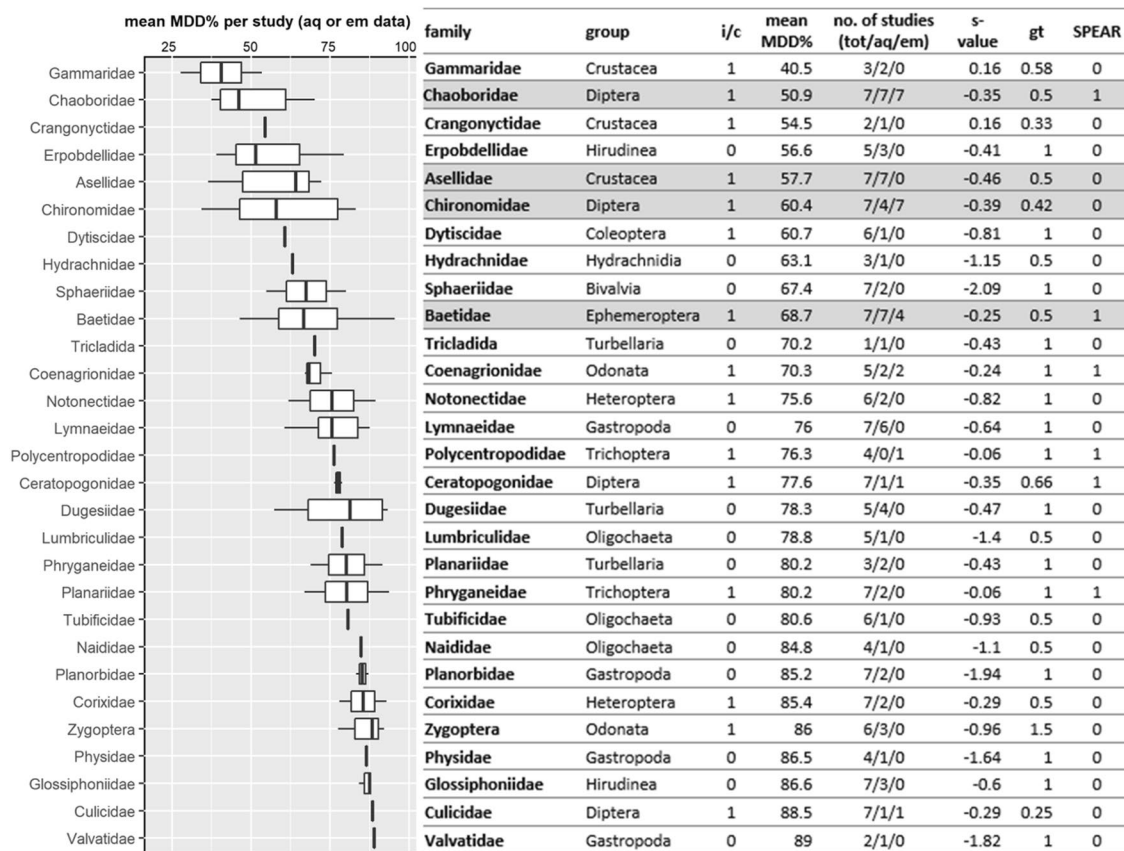
Reiber *et al. Environmental Sciences Europe* (2022) 34:65

Page 7 of 19

family	group	i/c	mean MDD%	no. of studies (tot/aq/em)	s-value	gt	SPEAR
Gammaridae	Crustacea	1	40.5	3/2/0	0.16	0.58	0
Chaoboridae	Diptera	1	50.9	7/7/7	-0.35	0.5	1
Crangonyctidae	Crustacea	1	54.5	2/1/0	0.16	0.33	0
Erpobdellidae	Hirudinea	0	56.6	5/3/0	-0.41	1	0
Asellidae	Crustacea	1	57.7	7/7/0	-0.46	0.5	0
Chironomidae	Diptera	1	60.4	7/4/7	-0.39	0.42	0
Dytiscidae	Coleoptera	1	60.7	6/1/0	-0.81	1	0
Hydrachnidae	Hydrachnidia	0	63.1	3/1/0	-1.15	0.5	0
Sphaeriidae	Bivalvia	0	67.4	7/2/0	-2.09	1	0
Baetidae	Ephemeroptera	1	68.7	7/7/4	-0.25	0.5	1
Tricladida	Turbellaria	0	70.2	1/1/0	-0.43	1	0
Coenagrionidae	Odonata	1	70.3	5/2/2	-0.24	1	1
Notonectidae	Heteroptera	1	75.6	6/2/0	-0.82	1	0
Lymnaeidae	Gastropoda	0	76	7/6/0	-0.64	1	0
Polycentropodidae	Trichoptera	1	76.3	4/0/1	-0.06	1	1
Ceratopogonidae	Diptera	1	77.6	7/1/1	-0.35	0.66	1
Dugesiidae	Turbellaria	0	78.3	5/4/0	-0.47	1	0
Lumbriculidae	Oligochaeta	0	78.8	5/1/0	-1.4	0.5	0
Planariidae	Turbellaria	0	80.2	3/2/0	-0.43	1	0
Phryganeidae	Trichoptera	1	80.2	7/2/0	-0.06	1	1
Tubificidae	Oligochaeta	0	80.6	6/1/0	-0.93	0.5	0
Naididae	Oligochaeta	0	84.8	4/1/0	-1.1	0.5	0
Planorbidae	Gastropoda	0	85.2	7/2/0	-1.94	1	0
Corixidae	Heteroptera	1	85.4	7/2/0	-0.29	0.5	0
Zygoptera	Odonata	1	86	6/3/0	-0.96	1.5	0
Physidae	Gastropoda	0	86.5	4/1/0	-1.64	1	0
Glossiphoniidae	Hirudinea	0	86.6	7/3/0	-0.6	1	0
Culicidae	Diptera	1	88.5	7/1/1	-0.29	0.25	0
Valvatidae	Gastropoda	0	89	2/1/0	-1.82	1	0

Fig. 2 Overview of all taxa with a mean MDD% ≤ 100 based on the seven recent M/M studies (in which the taxon was present with a mean MDD% ≤ 100). To calculate the mean MDD% (column 'mean MDD%'), the lowest mean MDD% from the aquatic and the emergence dataset per taxon and study was selected. The highlighted grey lines show the four families$_{MDD\%low}$ that established in more than three of the seven M/M studies. MDD% is the *minimum detectable difference* in % and families$_{MDD\%low}$ are defined as families per study with a mean MDD% < 70. The column 'no. of studies' lists the number of studies with: taxon was present ('tot') / mean MDD% ≤ 100 for the respective taxon in the aquatic dataset ('aq') / mean MDD% ≤ 100 in the emergence dataset ('em'). The column 'i/c' indicates if a taxon belongs to the taxonomic group insects or crustaceans (1), or to other taxonomic groups (0). Column 'gt' shows the generation time [years]. Column 's-value' shows the physiological sensitivity towards pesticides (value displays the relative sensitivity of a taxon in comparison to that of Daphnia magna, expressed as a logarithmic measure, see Von der Ohe and Liess [36], and column 'SPEAR' the classification as taxon at risk (1) or not at risk (0) towards pesticides according to the index SPEAR$_{pesticides}$ [14]

(Gammaridae) has a mean MDD% < 50; however it was established only in two M/M studies. Only further four families (Chaoboridae, Asellidae, Chironomidae and Baetidae) did establish in more than three of the seven M/M studies and are statistically evaluable with a mean MDD% < 70.

To answer the question whether these four families are sensitive and vulnerable to pesticidal substances, relevant information was compiled and evaluated. All of these four families are insects or crustaceans and hence belong to the classification of potentially insecticide-sensitive according to the AGD [17]. Applying the "physiological sensitivity" according to the indicator SPEAR$_{pesticides}$

(with the physiological sensitivity defined as intrinsic sensitivity), only two of the four families (Baetidae and Chaoboridae) are classified as physiologically sensitive towards pesticidal substances (*s*-value > − 0.36). According to the SPEAR-trait "generation time", all four families have a rather short life cycle with a maximum generation time of 0.5 years and hence a comparatively good potential to recover from a pesticide effect. For the assessment of vulnerability according to SPEAR$_{pesticides}$, all SPEAR-traits are taken into account (see section "Analysis of the macroinvertebrate communities in M/M studies and comparison with field assemblages" for more details); resulting in the fact that only two of the four taxa, namely

Baetidae and Chaoboridae, are classified as vulnerable towards pesticides in the field.

Looking again at all 29 families included in Fig. 2, the highest physiological sensitivity value in M/M studies was identified for the families$_{MDD\%low}$ Gammaridae and Crangonyctidae (*s*-value = 0.16) which, however, were present only in less than half of the recent M/M studies. For the above-mentioned best established four families$_{MDD\%low}$, the highest physiological sensitivity value is −0.25 for Baetidae.

The additional effort of emergence sampling did not lead to a substantial improvement of the entire dataset. Only seven out of 29 families were monitored by emergence traps in at least one M/M study with a mean MDD% ≤ 100. For six of these families, the mean MDD% in general was similar in the aquatic and emergence dataset (paired *t*-test, *p* = 0.56). One family (Polycentropodidae) could only be detected in the emergence dataset. Only for one taxon (Chironomidae), the effort of emergence sampling led to a substantial increase of the number of M/M studies with evaluable MDD% and improved the dataset for statistical analysis of treatment-related effects.

Figure 3 shows the distribution of families$_{MDD\%low}$ and families$_{MDD\%high}$ over 19 taxonomic groups. The highest mean number of families$_{MDD\%low}$ was found for

Diptera (*N* = 1.4) and Crustacea (*N* = 1.3). The highest mean number of families$_{MDD\%high}$ was shown for Diptera (*N* = 5.6), Heteroptera (*N* = 4.3) and Coleoptera (*N* = 4.0). All taxonomic groups except Crustacea contain a smaller proportion of families$_{MDD\%low}$ than families$_{MDD\%high}$. This applies also to the orders containing especially pesticide-sensitive and vulnerable families, such as Ephemeroptera and Trichoptera. Six taxonomic groups (Araneae, Coelenterata, Hymenoptera, Lepidoptera, Meglaoptera and Oligochaeta) did not contain any family$_{MDD\%low}$ in any of the seven M/M studies. Plecoptera are not present in any of the seven M/M studies.

Taxa composition in recent M/M studies compared to natural aquatic ecosystems

The following analysis is focused on families of insects and crustaceans, i.e., on the potentially sensitive taxa for pesticidal substances according to the AGD [17], see Fig. 4). For the M/M studies, all families with a mean MDD% < 70 (families$_{MDD\%low}$) were included. For the reference stream field sites, all families sampled at five of 26 sampling sites were included. Mean comparisons are performed by Kruskal–Wallis test and the statistical results are given in brackets.
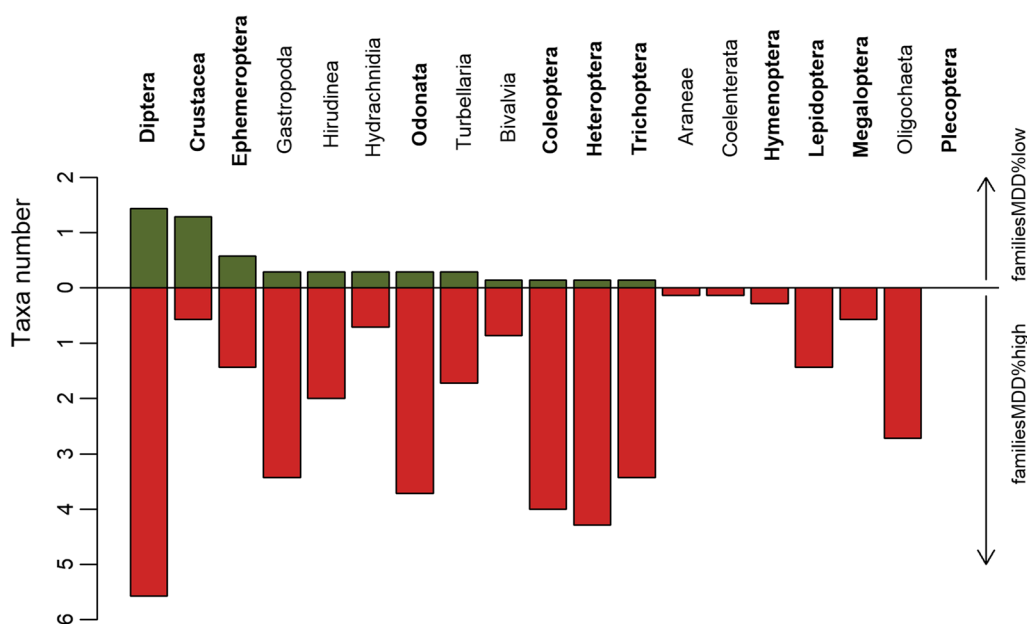

Fig. 3 Comparison of the average number of families$_{MDD\%low}$ (green bars, above *x*-axis) and families$_{MDD\%high}$ (red bars, below *x*-axis) per taxonomic group in the seven recent M/M studies. Families$_{MDD\%low}$ are defined by a mean MDD% < 70. Families$_{MDD\%high}$ are characterized by a mean MDD% ≥ 70, which means that the data of these taxa only enable the statistical detection of treatment-related population effects between 70 and 100%. MDD% is the *minimum detectable difference* in %. Taxonomic groups listed in bold belong to insects and crustaceans
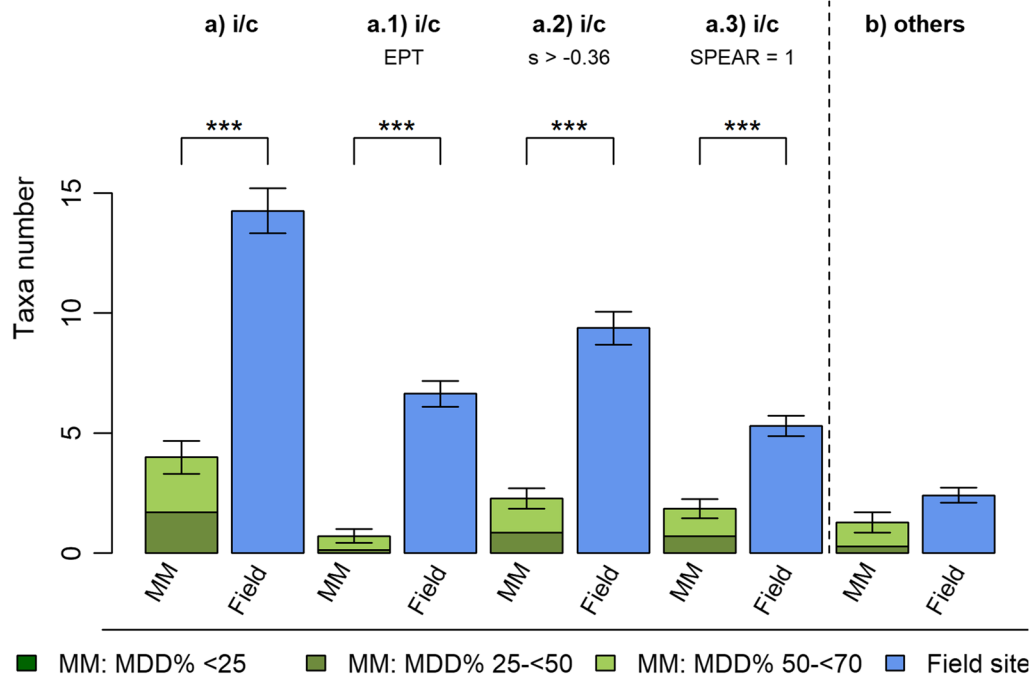
Fig. 4 Average number of families$_{MDD\%low}$ in the seven recent M/M studies, compared to the average number of families with the respective characteristics at reference stream field sites. Families$_{MDD\%low}$ are defined as families per study with a mean MDD% < 70MDD%, whereby MDD% is the *minimum detectable difference* in %. (**a**) families$_{MDD\%low}$ from the taxonomic groups insects and crustaceans (i/c); (**a.1**) i/c families$_{MDD\%low}$ belonging to the EPT taxa (Ephemeroptera, Plecoptera, Trichoptera; (**a.2**) i/c families$_{MDD\%low}$ with an *s*-value (s; physiological sensitivity towards pesticides) > −0.36; (**a.3**) i/c taxa at risk towards pesticides according to the index SPEAR$_{pesticides}$ (SPEAR = 1) [14]; (**b**) families$_{MDD\%low}$ belonging to all taxonomic groups except insects and crustaceans). Error bars indicate the SEM (standard error of the mean)

For all insects and crustaceans, communities at field sites are more family-rich with a 3.6 times higher number of families in the field than families$_{MDD\%low}$ in M/M studies (Kruskal–Wallis test, $p < 0.001$). This increased family-richness at field sites in comparison to well-established families in M/M studies is also reflected in further subsets of the data. For insects and crustaceans belonging to the EPT order (Ephemeroptera, Plecoptera, Trichoptera), the number of families at field sites is 9.3 times higher than the number of families$_{MDD\%low}$ in M/M studies (Kruskal–Wallis test, $p < 0.001$).

According to the SPEAR-classification for "physiological sensitivity" (*s*-value > −0.36), field sites are more family-rich with 9.4 families versus only 2.3 families$_{MDD\%low}$ in M/M studies ($p < 0.001$). According to the SPEAR-classification for vulnerability (SPEAR = 1), field sites are more family-rich with 5.3 families versus only 1.9 families$_{MDD\%low}$ in M/M studies ($p < 0.001$). Families$_{MDD\%low}$ classified as vulnerable according to SPEAR$_{pesticides}$ are Chaoboridae, Baetidae, Coenagrionidae and Phryganeidae, with Baetidae also frequently occurring at reference stream sites in the field (for more details of the comparison see Table 1, Appendix). For all taxonomic groups except insects and crustaceans, a mean

of 1.3 families$_{MDD\%low}$ were present in each of the recent M/M studies, compared to 2.4 families at field sites.

Discussion
Aiming to analyze the composition of the benthic macroinvertebrate communities that established in control (untreated) test systems of M/M studies and to compare them with natural assemblages at reference stream sites in the field, the following aspect needs to be considered: M/M studies are carried out with the aim of demonstrating statistically significant treatment-related effects; the data of potentially sensitive or vulnerable populations must therefore be suitable for this purpose. However, all populations of taxa in the field are defined as an ecological protection good and the risk assessment has to ensure that a potential unacceptable risk of damage to any of them is captured, i.e., not missed because of statistical shortcomings of test systems. For this reason, we first discuss the statistical evaluability of M/M data and then place our results in the context of the *regulatory acceptable concentration* (RAC), separately for (i) the *ecological threshold option* (ETO) and (ii) the *ecological recovery option* (ERO). Finally, we discuss concerns regarding M/M studies in a broader regulatory context.

Taxa number and statistical evaluability of M/M taxon data

The total number of benthic macroinvertebrate families in M/M test systems increased since 2013, which may be induced by the publication of the AGD [17] requiring a minimum of 8 different populations of the sensitive taxonomic group (with "sensitivity" relating to physiological, also called intrinsic sensitivity). Although this increase is not statistically significant due to the high variance between the test systems, the number of families almost doubled which indicates a general development. Recent M/M studies (i.e., conducted since 2013) thus contain an average of 38 families per test system, of which 25 families belong to the generally sensitive group of insects and crustaceans.

However, in order to derive an ETO-RAC or ERO-RAC, sensitive or vulnerable taxa must be successfully established in the M/M test systems to ensure that treatment-related effects are statistically detectable. There are different definitions for this requirement. We followed the general proposal of the AGD [17] that statistical evaluability of data is indicated when the minimum detectable difference (MDD%) of the critical endpoints is below 70%. Specifically, we determined that data of a taxon is suitable for deriving a LOEC/NOEC and consequently an ETO- or ERO-RAC if, within 30 days after a contamination event, the mean MDD% is below 70 (see section "Categorization of taxa according to their MDD%"). Taxa that meet our MDD%-criterion are referred to as families$_{MDD\%low}$. This MDD-based classification of taxa differs from the taxa categorization according to Brock et al. [24] due to specific data constraints (for more details, see section "Categorization of taxa according to their MDD%"). With our requirement of statistical evaluability, the number of suitable families of insects and crustaceans is reduced from 25 to 4 in recent M/M studies. Additionally, there is no improvement in the mean MDD% over the entire period of 32 years. We therefore assume that the recommendations of the AGD [17] increased the number of families in M/M studies in general, but had no relevant positive influence on statistical evaluability. In consequence, the statistical demonstration of treatment-related effects remains one of the greatest challenges of M/M studies for risk assessment.

Presence of sensitive taxa for the derivation of an ETO-RAC

When deriving an ETO-RAC, the level of effect depends on the physiological sensitivity of the respective taxon to the test substance. According to the AGD [17], insects and crustaceans are generally considered as taxa with potentially highest physiological sensitivity to pesticidal substances. Following this definition, there is a mean of 4 families$_{MDD\%low}$ in recent M/M studies compared to a mean of 14.3 families at the reference sites in the field. However, it is well known that insects and crustaceans can differ greatly in their physiological sensitivity to pesticides as shown for example by sensitivity rankings [e.g., 36, 37]. This high variation in physiological sensitivity is also mentioned in the AGD [17]. We therefore applied a stricter definition, using the sensitivity trait of the index SPEAR$_{pesticides}$ [36]. This further reduced the total number of physiologically sensitive taxa to a mean of 2.3 families$_{MDD\%low}$ in recent M/M studies, compared to 9.4 in the field. Moreover, a (due to the logarithmic measure) 4.3 times higher maximum physiological sensitivity according to SPEAR$_{pesticides}$ was identified for taxa at field sites (0.38 for the Plecoptera Leuctridae and Perlodidae [35]) than for Baetidae (−0.25), the most physiologically sensitive taxon of the four best established families$_{MDD\%low}$ in recent M/M studies. Furthermore, several investigations showed that the physiological sensitivity of insects and crustaceans can vary greatly with the activity of the respective test substance [38–41]. It would therefore be advisable for each M/M study to give a justification if (i) especially sensitive taxa towards the mode-of-action of the substance assessed are represented and (ii) their abundances allow for a statistical detection of treatment-related effects. Such a justification could become a key criterion to assess the reliability of an M/M study and for its subsequent use in risk assessment.

In addition it should be noted that the majority of the families most successfully established in recent M/M, i.e., Asellidae, Chironomidae and Baetidae, are also relatively easy to cultivate in the laboratory. Therefore, they are also frequently used in standardized laboratory experiments to derive effect endpoints such as the LC$_{50}$ [see e.g., 42, 43]. Here the question arises if the high effort of M/M studies can be justified when the information related to effect thresholds could also be obtained with far less costly laboratory test systems that usually have a better statistical power, are more targeted towards the toxicant effect and less influenced by complex interactions. One disadvantage of single species laboratory studies is that typically only acute exposure is considered. As there is a growing number of investigations indicating long-term effects of short-term exposure [e.g., 44, 45], long-term effects should hence be also considered in laboratory studies. Furthermore, it could be argued that M/M studies, in contrast to lower tier studies, include factors impacting the physiological sensitivity of organisms in natural aquatic ecosystems, such as environmental stressors [46], competition, food availability or temperature regime [47, 48] or the exposure to multiple

chemicals [e.g., 5, 10, 16, 49]. However, to our knowledge there is no test protocol for M/M studies that standardizes the inclusion of these additional factors in M/M studies or specifies the measurement of their intensity. As additional stressors are not quantified in M/M studies testing pesticides, the question remains if the stress level is representative for natural conditions.

Given the uncertainties discussed, the low number of statistically evaluable taxa (4 and 2.3 families, respectively, see above) that furthermore do not represent the physiologically most sensitive taxa at field sites are valuable arguments against using M/M studies for environmental risk assessment.

Derivation of an ERO-RAC is not recommended

When opting for the derivation of a RAC on the basis of the *ecological recovery optio*n (ERO-RAC), the AGD states that it needs to be critically evaluated whether representatives of potentially vulnerable populations are sufficiently covered in the M/M study [17]. We identified vulnerable taxa using the index SPEAR$_{pesticides}$, which, in addition to the physiological pesticide sensitivity of taxa, also considers their generation time, exposure probability and dependence on refuge areas for the definition of vulnerability [14, 25]. Based on SPEAR$_{pesticides}$, the number of vulnerable taxa is reduced to 1.9 families$_{MDD\%low}$ in M/M studies in comparison to 5.3 families in the field (in addition to the taxa reduction mentioned in section "Presence of sensitive taxa for the derivation of an ETO-RAC"). Among the four families$_{MDD\%low}$ established in more than three of the seven recent M/M studies, two of them namely Baetidae and Chaoboridae are classified as vulnerable according to SPEAR$_{pesticides}$. However, they are characterized by a generation time of 0.5 years which is shorter than that of many vulnerable taxa in the field. Therefore they have a comparatively good potential to recover from a pesticide effect, especially when the populations comprise larvae of different age classes. Another approach assessing the pesticide vulnerability of macroinvertebrates directly under field conditions is the *pesticide associated response* PARe; [3]. PARe-values ≤ -70 indicate a comparatively high pesticide vulnerability in the field. Baetidae is the only familiy$_{MDD\%low}$ in recent M/M studies which contains species with a PARe ≤ -70. Applying the index PARe, the number of vulnerable taxa would be again reduced to 0.6 families$_{MDD\%low}$ in M/M studies, compared to 5.5 vulnerable families in the field. Therefore it is questionable if the representation of vulnerable taxa, as requested by the AGD, is sufficient in recent

M/M studies. The lack of (highly) vulnerable taxa in M/M studies could lead to an underestimation of the vulnerability of macroinvertebrate communities in the field. Thus it is critical to derive endpoints considering recovery, such as a *no observed ecologically adverse effect concentration* (NOEAEC) for the risk assessment.

Moreover, the AGD refers to a period of 8 weeks for recovery [17]. This means that only populations of taxa with very short generation time can recover during that time. However, vulnerable taxa are characterized by a comparatively long generation time of 0.5 years and longer (see SPEAR$_{pesticides}$ [14]). Hence, the ecological recovery option is not applicable for assessing effects on vulnerable taxa. Furthermore, not only the sensitivity of a taxon (see section "Presence of sensitive taxa for the derivation of an ETO-RAC"), but also its vulnerability is altered by the presence of additional stressors, as shown in particular by prolonged population recovery after pesticide effects [e.g., 50, 51]. Therefore, it cannot be ensured that the time for recovery of a taxon in M/M studies is comparable with the time in natural aquatic ecosystems.

Taking the above-mentioned limitations into consideration, we do not recommend the application of an ERO-RAC, but support the use of the ETO-RAC approach for substances with an insecticidal mode-of-action.

Interpretation of the MDD-values and critical elements

The *minimum detectable difference* (MDD) is a measure of the difference needed between the means of a treatment and the control to reveal a specific effect as statistically significant with sufficient probability. Setting an MDD threshold for example at 70% means that only treatment-related population effects (i.e., in case of this study decreased abundance) between 70 and 100% are statistically detectable. However, it is important to consider that a taxon with an average MDD of 70% is not suitable to indicate effects of 50% which correspond to the standard for acute effects of most toxicological endpoints in laboratory systems; it is even less suitable to detect effects of 10% which correspond to the standard for chronic effects. As revealed in our analysis, only four families (namely Chaoboridae, Asellidae, Chironomidae and Baetidae) established in more than three of the seven recent M/M studies and are statistically evaluable with a mean MDD% < 70. In EFSA [17], it is proposed that "the MDD of critical endpoints should ideally exceed class II" which currently corresponds to the threshold of 70%. The poor detectability of effects on the macroinvertebrate taxa—that results from accepting effects as high

as 70%—is impairing the certainty of the protection level achieved; indeed some treatment- related effects may be occurring but cannot be shown due to the inherent variability of the test systems.

In addition, the MDD should be calculated with an appropriate statistical power. In the AGD [17] and Brock et al. [24]—see also Eq. 1 in section "Calculation of the minimum detectable difference (MDD)"—the MDD calculation method does not stipulate the value of the beta-error as raised by Duquesne et al. [32], i.e., the level of probability of type II error/ beta-error is 0.5. However a high degree of certainty should definitely be ensured in order to avoid under-protective regulatory decisions such as authorizing a pesticide that has potentially severe consequences for the environment. Therefore, Duquesne et al. [32] suggested to set the type II error to 0.2, as for a priori analysis performed in lower tier, in order to increase the statistical power (1-β). By doing so, the MDD would result in an 80% probability of detecting the respective effect, versus 50% before. However, this would also result in higher MDD-values and thus have implications in the interpretation of study outcomes. Indeed higher MDD-values are attributed to lower MDD classes. Thus based on the current classification proposed in the AGD [17], a taxon considered as suitable to detect treatment-related effects calculated with an MDD power of 0.5 (i.e., class II, effects below 70%) may not be suitable anymore if the power value is appropriately set at 0.2.

Risk assessment: general considerations and specific recommendations

General considerations about the tiered approach in risk assessment

The current system of plant protection products (PPP) registration considers only the single pesticide. It is based on a tiered approach with the "unless" clause described in the Uniform Principles (Regulation (EU) No. 546/2011) which offers the possibility to perform higher tier risk assessment if an unacceptable risk is identified at a lower tier. The concept of the tiered approach is to act as a filter and perform additional evaluation only if necessary so that it remains a cost-effective procedure. Both types of approaches (lower and higher tier) should result in protective risk assessment decisions addressing the specific protection goals (SPGs). Compared to the lower tier approaches, higher tier approaches are usually more complex, involve more data connected to higher variability and thus require

more considerations and expertise. The tiered approach also has been interpreted as a possibility to deliver infinite data, and in general turns out to be the opposite of cost-effective.

Higher tier approaches such as (semi)-field studies aim at being more realistic than lower tier approaches with a better species representativeness for the field situation, e.g., testing of assemblages in aquatic mesocosms. This should however be put in perspective since—as shown by our analysis of most sensitive and vulnerable taxa in recent M/M studies—the assemblages poorly represent field communities. In addition the higher tier risk assessment is (i) more targeted at specific scenarios which causes problems of extrapolation between scenarios and use patterns, and (ii) less conservative than lower tiers which causes a small margin of safety when concluding on the acceptability of risk (i.e., the risk can be assessed as acceptable but is closer to the threshold of effects). Thus the risk assessment framework should generally be considered in a critical way. Indeed, it is questionable if the difficulties and concerns related to the setup and evaluation of such higher tier risk assessment methods as shown in this paper are counterbalanced by better/safer decisions in terms of avoiding unacceptable effects in the field.

Specific recommendations for better gain of data and knowledge from M/M studies

In the current risk assessment guidance, M/M studies are considered as an acceptable approach to refine a risk for edge-of-field surface water organisms. It is therefore worth to examine which changes in the study design and procedure of recent M/M studies are necessary to ensure a suitable detectability of effects on the exposed populations at risk and facilitate the use of those data for regulatory decisions. Since the guidance of EFSA [17], some issues have been further tackled and possible changes raised [e.g., 24, 52, 53]. Considering these issues and our concerns, the following suggestions can be listed:

1. Measures increasing the number of statistically evaluable sensitive taxa

   – a higher number of replicates per treatment (see [17] and Table 2, Appendix), or a higher number of treatments,
   – an improved sampling technique and time interval between samplings,
   – a sufficient period of establishment (under natural conditions, at least 2 years are needed for a community to re-establish after disturbance [12]),

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 13 of 19

– clustering vulnerable species according to their traits in order to reduce variability and improve detectability of pesticide related effects [54];

2. Measures ensuring that the limits in terms of the realism and the statistical evaluability are suitably detected

– further considerations and development of (regulatory) guidance when it comes to the poor representativity of species from lotic surface waters in lentic M/M studies,
– appropriate calculation of MDD regarding the beta-error and thus use of reliable MDD-values in data interpretation (see section "Interpretation of the MDD-values and critical elements").

However, even with such measures enforced, ensuring a suitable level of taxa representation in M/M studies may not be reached.

An additional possibility would be to optimize the design of M/M studies for deriving a NOEC and effect threshold, i.e., shorter studies focused on direct effects, which would be best suited when testing substances with insecticidal mode-of-action. Indeed this is justified since the suitability of communities established in M/M studies for representing most vulnerable taxa in the field is in most cases questionable, and deriving a NOAEC and using the recovery option is not recommended. Instead the use of the ETO-RAC approach is supported, as concluded in section "Derivation of an ERO-RAC is not recommended".

Another aspect is that according to the SPGs defined for water organisms, the analysis of data and derivation of endpoints related to population-level effects are a necessity. Considering community level effects may be useful but only as supplementary information, as already mentioned in EFSA [17]. This includes, e.g., the principal response curve [PRC; e.g., 55, 56 57], and the trait-based classification of the community following the general concept of the SPEAR$_{pesticide}$ index [14].

If the above suggested elements in terms of study design and procedure as well as an appropriate MDD calculation as described under Interpretation of the MDD-values and critical elements could be implemented, the power of data gained from M/M studies to detect effects would be increased and the interpretation of statistical evaluation would improve. The shortcomings of M/M studies would then be partly overcome and the representativity and outcomes of the higher tier risk assessment for the field communities more reliable.

However, it remains questionable if no unacceptable effects indicated by current higher tier approaches can ensure that no population-relevant unacceptable effects will occur in the field, i.e., if the aim of a more exact and explanatory risk assessment in the current context with complex higher tier approaches should be pursued. Hence, the effort of improving current M/M studies may not be justified. Instead, going beyond the current tiered approaches, e.g., by developing other assessment tools and shifting towards a new paradigm should be explored. For example, the risk assessment of single PPPs could be based on a robust and simplified single tier approach tailored to the mode-of-action of the active substances, making use of all data and knowledge available and designing risk profiles that would facilitate to distinguish and rank the PPPs from better to worse. Such approaches could be implemented in a more holistic context focused on the agricultural landscape (e.g., considering the PPP under assessment in the context of the other PPPs and stressors as well as mitigation and compensatory measures) and associating other approaches such as prospective and retrospective assessment.

Conclusion

It could be demonstrated that recent M/M test systems do not adequately represent sensitive and vulnerable macroinvertebrate taxa at natural freshwater stream sites. Although M/M studies performed in the last decade generally include a higher number of taxa compared to older studies, the data on abundances in most cases are not suitable for the detection of treatment-related effects. Possibilities are suggested to further improve the study design and MDD calculations in order to increase the power of data gained from M/M studies for detecting effects. However, it remains questionable if a risk assessment based on the current higher tiered approach and concluding "no unacceptable effects" can really ensure that no effects occur in the field. Therefore, we recommend the development of other assessment tools or a shift towards a new paradigm.

Appendix

See Tables 1 and 2.

Table 1 Overview on the classification as taxon at risk (column 'SPEAR'= 1) and taxon not at risk (column 'SPEAR'= 0) according to the index SPEAR$_{pesticides}$ [14] of the families present in the seven recent M/M studies (column 'M/M'= x; listed are families$_{MDD\%low}$) and at reference stream field sites (column 'Field'= x)

Family	Group	SPEAR	M/M	Field
Ceratopogonidae	Diptera	1		x
Chaoboridae	Diptera	1	x	
Dixidae	Diptera	1		x
Baetidae	Ephemeroptera	1	x	x
Heptageniidae	Ephemeroptera	1		x
Calopterygidae	Odonata	1		x
Coenagrionidae	Odonata	1	x	
Leuctridae	Plecoptera	1		x
Perlodidae	Plecoptera	1		x
Glossosomatidae	Trichoptera	1		x
Goeridae	Trichoptera	1		x
Limnephilidae	Trichoptera	1		x
Phryganeidae	Trichoptera	1	x	
Polycentropodidae	Trichoptera	1		x
Psychomyiidae	Trichoptera	1		x
Rhyacophilidae	Trichoptera	1		x
Sphaeriidae	Bivalvia	0	x	x
Dytiscidae	Coleoptera	0	x	x
Elmidae	Coleoptera	0		x
Hydrophilidae	Coleoptera	0		x
Scirtidae	Coleoptera	0		x
Asellidae	Crustacea	0	x	x
Crangonyctidae	Crustacea	0	x	
Gammaridae	Crustacea	0	x	x
Chironomidae	Diptera	0	x	x
Limoniidae	Diptera	0		x
Pediciidae	Diptera	0		x
Simuliidae	Diptera	0		x
Tabanidae	Diptera	0		x
Ephemerellidae	Ephemeroptera	0		x
Ephemeridae	Ephemeroptera	0		x
Leptophlebiidae	Ephemeroptera	0		x
Gastropoda	Gastropoda	0	x	
Lymnaeidae	Gastropoda	0	x	
Planorbidae	Gastropoda	0		x
Notonectidae	Heteroptera	0	x	
Erpobdellidae	Hirudinea	0	x	x
Glossiphoniidae	Hirudinea	0		x
Hydrachnidae	Hydrachnidia	0	x	
Hydrachnidia	Hydrachnidia	0	x	
Cordulegastridae	Odonata	0		x
Lumbricidae	Oligochaeta	0		x
Oligochaeta	Oligochaeta	0		x
Tubificidae	Oligochaeta	0		x
Nemouridae	Plecoptera	0		x
Hydropsychidae	Trichoptera	0		x
Leptoceridae	Trichoptera	0		x
Sericostomatidae	Trichoptera	0		x

Table 1 (continued)

Family	Group	SPEAR	M/M	Field
Dugesiidae	Turbellaria	0	x	x
Planariidae	Turbellaria	0	x	

Families$_{MDD\%low}$ are defined as families per study with a mean MDD% < 70MDD%, whereby MDD% is the *minimum detectable difference* in %

Table 2 Overview on details of all M/M studies included in the analysis

Study-ID	Year of conduct	Guidance documents/recommendations followed	Type of aquatic model ecosystem study
Study-01	2002	HARAP; CLASSIC	Outdoor pond microcosm
Study-02	2004	HARAP, CLASSIC	Outdoor pond microcosm
Study-03	2006	HARAP, CLASSIC	Outdoor pond microcosm
Study-04	2002	n.s	Outdoor pond mesocosm enclosures
Study-05	2006	CLASSIC; European Commission [58]	Outdoor pond microcosm enclosures
Study-06	2007	SETAC-Europe [59]; CLASSIC; OECD [60]	Outdoor pond microcosm
Study-07	1986	n.s	Outdoor pond mesocosm enclosures
Study-08	1997	SETAC-Europe [59], SETAC-Resolve [61]; Crossland et al. [62]; World Wildlife Fund [63]	Outdoor pond microcosm enclosures
Study-09	1998	SETAC-Europe [59]; SETAC-Resolve [61]; HARAP; Brock et al. [64]	Outdoor pond microcosm enclosures
Study-10	1998	HARAP	Indoor microcosm (glass aquaria)
Study-11	2007	SETAC-Europe [59]; OECD [60]	Outdoor pond microcosm
Study-12	1999	SETAC-Europe [59]; OECD [65]	Outdoor pond microcosm
Study-13	2001	SETAC-Europe [59]; OECD [65]	Outdoor pond microcosm
Study-14	2004	SETAC-Europe [59]; OECD [65]; CLASSIC	Outdoor pond microcosm
Study-15	2003	SETAC-Europe [59]; OECD [65]; CLASSIC	Outdoor pond microcosm
Study-16	2007	CLASSIC; European Commission [58]; OECD [60]	Outdoor pond microcosm enclosures
Study-17	2008	CLASSIC; European Commission [58]; OECD [60]	Outdoor pond microcosm enclosures
Study-18	2012	HARAP; CLASSIC; OECD [60]; draft of EFSA [17]	Outdoor pond microcosm enclosures
study-19	2012	OECD [60]; EFSA [17]	Outdoor pond microcosm enclosures
Study-20	2014	HARAP; CLASSIC; OECD [60]; EFSA [17]; Brock et al. [24]	Outdoor pond microcosm enclosures
Study-21	2013	HARAP; CLASSIC; OECD [60]; EFSA [17]; Brock et al. [24]	Outdoor pond microcosm enclosures
study-22	2014	HARAP; CLASSIC; OECD [60]; EFSA [17]; Brock et al. [24]	Outdoor pond microcosm enclosures
Study-23	2018	OECD [60]; EFSA [17]	Outdoor pond microcosm enclosures
Study-24	1994	n.s	Outdoor pond microcosm
Study-25	2000	Crossland et al. [62]; HARAP; CLASSIC (Draft version of 2000)	Outdoor pond microcosm
Study-26	2002	SETAC-Europe [59]; HARAP; CLASSIC	Outdoor pond microcosm
Study-27	2001	SETAC-Europe [59]; OECD [65]	Outdoor pond microcosm enclosures
Study-28	2002	SETAC-Europe [59]; OECD [65]	Outdoor pond microcosm enclosures
Study-29	2008	SETAC-Europe [59]; OECD [60]	Outdoor pond microcosm enclosures
Study-30	2007	SETAC-Europe [59]; OECD [60]	Outdoor pond microcosm
Study-31	2000	CLASSIC	Outdoor pond microcosm enclosures
Study-32	2000	CLASSIC	Outdoor pond microcosm enclosures
Study-33	2012	CLASSIC; De Jong et al. [66]; draft of EFSA [17]	Outdoor pond microcosm enclosures
Study-34	2013	HARAP; CLASSIC; European Commission [58]; EFSA [17]	Outdoor pond microcosm enclosures
Study-35	2017	CLASSIC; OECD [60]; EFSA [17]	Outdoor pond microcosm enclosures
Study-36	2001	SETAC-Europe [59]; OECD [65]; HARAP; CLASSIC	Outdoor pond microcosm enclosures
Study-37	2004	SETAC-Europe [59]; OECD [65]; HARAP; CLASSIC	Outdoor pond microcosm enclosures
Study-38	2003	SETAC-Europe [59]; OECD [65]; HARAP; CLASSIC	Outdoor pond microcosm enclosures
Study-39	2003	SETAC-Europe [59]; OECD [65]; HARAP; CLASSIC	Outdoor pond microcosm enclosures
Study-40	2005	SETAC-Europe [59]; OECD [67]; HARAP; CLASSIC	Outdoor pond microcosm enclosures
Study-41	2004	CLASSIC; European Commission [58]; OECD [67]	outdoor pond microcosm enclosures
Study-42	2003	HARAP, CLASSIC	Outdoor pond microcosm

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 16 of 19

Table 2 (continued)

Study-ID	Year of conduct	Guidance documents/recommendations followed	Type of aquatic model ecosystem study
Study-43	2013	SETAC-Europe [59]; CLASSIC; OECD [60]; De Jong et al. [66]; EFSA [17]	Outdoor pond microcosm
Study-44	2009	SETAC-Europe [59]; CLASSIC; OECD [60]	Outdoor pond microcosm
Study-45	2005	CLASSIC; European Commission [58]	Outdoor pond microcosm
Study-46	2000	CLASSIC; European Commission [58]	Outdoor pond microcosm
Study-47	2000	CLASSIC; European Commission [58]	Outdoor pond microcosm

Study-ID	Location of study site	Water volume [m^3]	Establishment period [month], approx	Substance group	Pesticide type	No. repl Ctr	No. repl T
Study-01	UK	1.267	3	Neonicotinoid	I	4	3
Study-02	UK	1.267	2	Macrocyclic lactone insecticides	I	4	3
Study-03	UK	1.267	3	Anilinopyrimidine	F	4	3
Study-04	UK	55	36	Pyrethroid	I	3	3
Study-05	NL	n.s	n.s	Tetronic acid	I	3	2
Study-06	DE	6	n.s	Morpholine	F	3	2
Study-07	UK	30	6–18	Pyrethroid	I	2	2
Study-08	UK	1.23	1.5–2	Strobilurin	F	3	3
Study-09	UK	1.23	3	Strobilurin	F	4	3
Study-10	UK	0.2	1.5	Dithiocarbamate	F	6	4
Study-11	DE	4.9	> 5	Phenylpyridinamine	F	6	4
Study-12	DE	6.24	6	Pyrethroid	I	3	2
Study-13	DE	5.49	7	Strobilurin	F	3	2
Study-14	DE	6	7	Pyrethroid	I	3	2
Study-15	DE	6	6	Diamide insecticides	I	3	2
Study-16	DE	2	> 6	Macrocyclic lactone insecticides	I	4	3
Study-17	DE	2	7	Organophosphate	I	5	2
Study-18	DE	1.6	12	Alkane hydrocarbon	I	4	2
Study-19	DE	4.9	> 15	Unclassified	F	4	3
Study-20	DE	1.76	12	Neonicotinoid	I	6	4
Study-21	DE	1.76	7	Pyrethroid	I	5	2
Study-22	DE	1.8	18	Dithiocarbamate fungicides	F	5	2
Study-23	DE	5	5	Pyrethroid	I	5	2
Study-24	US	10	3	Strobilurin	F	3	2
Study-25	CH	10	1.5	Macrocyclic lactone insecticides	I	3	3
Study-26	DE	6.335	18	Benzoylphenylurea chitin synthesis inhibitors	IGR	4	2
Study-27	DE	3.1–3.8	> 6	Neonicotinoid	I	3	2
Study-28	DE	3.1–3.8	3	Organophosphate	I	3	2
Study-29	DE	4.9	5	Pyrethroid	I	3	2
Study-30	DE	4.9	> 5	Neonicotinoid	I	3	2
Study-31	UK	0.43	n.s	Pyrethroid	I	2	2
Study-32	UK	0.43	n.s	Pyrethroid	I	2	2
Study-33	NL	1	8	Imidazole	F	4	3
Study-34	NL	1.272	9	Botanical insecticides	I	5	3
Study-35	NL	1.526	> 6	Dithiocarbamate fungicides	F	5	3
Study-36	CH	3.2	5	Benzoylphenylurea chitin synthesis inhibitors	IGR	3	2
Study-37	CH	0.95	7	Pyrethroid	I	6	3
Study-38	CH	1	> 5	Pyrethroid	I	6	4
Study-39	CH	1	> 5	Pyrethroid	I	6	4
Study-40	CH	0.95	7	Pyrethroid	I, A	6	3

Table 2 (continued)

Study-ID	Location of study site	Water volume [m³]	Establishment period [month], approx	Substance group	Pesticide type	No. repl Ctr	No. repl T
Study-41	DE	2	>6	Chloronitrile/aromatic fungicides	F	5	2
Study-42	UK	1.267	2	Benzoylphenylurea chitin synthesis inhibitors	IGR	4	3
Study-43	UK	0.972	3–6	Juvenile hormone mimics	IGR	4	4
Study-44	UK	0.8	2–3	Phthalimide fungicides	F	5	3
Study-45	DE	4.9	n.s	Unclassified	F	3	2
Study-46	NL	0.43	n.s	Pyrethroid	I	2	2
Study-47	NL	0.43	n.s	Pyrethroid	I	2	2

In column 'Guidance Documents/recommendations followed', HARAP means higher tier aquatic risk assessment for pesticides [68] and CLASSIC stands for community-level aquatic system studies—interpretation criteria [69]. In column "Type of aquatic model ecosystem study', model ecosystem studies with water volumes greater than 15 m³ are classified as 'mesocosm' in accordance with EFSA [17]. Establishment period ($=$ pre-treatment period) started after the addition of sediment and water in the test units (ponds) and ended with the first test item application. The column 'Pesticide type' lists the type of substance described according to the type of pest or disease they control (A: acaricide; F: fungicide, I: insecticide; IGR: insect growth regulator). The column 'No. repl Ctr' lists the number of control replicates, the column 'No. repl T' includes the number of replicates in each test item treatment group. 'n.s.': not stated

Abbreviations

AGD: Aquatic Guidance Document, namely 'Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge of-field surface waters' by EFSA [17]; ANOVA: Analysis of variance; EFSA: European Food Safety Authority; EPT: Macroinvertebrate orders Ephemeroptera, Plecoptera, Trichoptera; ETO: Ecological threshold option; ETO-RAC: Regulatory acceptable concentration based on the ecological threshold option; ERO: Ecological recovery option; ERO-RAC: Regulatory acceptable concentration based on the ecological recovery option; IGR: Insect growth regulator; KGM: KleinGewässer-Monitoring; LC50: Lethal concentration of a substance that will lead to death of 50% of the dosed population during the observation period; LOEC: Lowest observed effect concentration; M/M: Microcosm and mesocosm; MDD: Minimum detectable difference; NOEAEC: No observed ecologically adverse effect concentration; NOEC: No observed effect concentration; OECD: Organisation for Economic Co-operation and Development; PPR Panel: Panel on Plant Protection Products and their Residues; PPP: Plant protection products; PRC: Principal response curve; RAC: Regulatory acceptable concentration; PARe: Pesticide associated response [see 3]; *s*-value: Physiological sensitivity value according to the index SPEAR$_{pesticides}$ [see 14]; SPEAR$_{pesticides}$: SPEcies At Risk of pesticides index [see 14]; SPG: Specific protection goal.

Availability of data and materials

The Microcosm and Mesocosm study datasets analyzed in the current study are not publicly available. They have been submitted for regulatory risk assessments to the German Environment Agency. The data access is legally restricted.

The field dataset analyzed during this study was obtained from the sampling campaign "Kleingewässermonitoring" and is described in the published study by Liess et al. [4].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department System-Ecotoxicology, UFZ Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany. [2]Institute for Environmental Research (Biology V), RWTH Aachen University, Worringerweg 1, 52056 Aachen, Germany. [3]German Environment Agency (Umweltbundesamt, UBA), Wörlitzer Platz 1, 06844 Dessau-Roßlau, Germany.

References

1. Liess M, Schafer RB, Schriever CA (2008) The footprint of pesticide stress in communities–species traits reveal community effects of toxicants. Sci Total Environ 406(3):484–490. https://doi.org/10.1016/j.scitotenv.2008.05.054
2. Schäfer RB, van den Brink PJ, Liess M (2011) Impacts of pesticides on freshwater ecosystems. Ecol Impacts Toxic Chem. 2011:111–137
3. Reiber L, Knillmann S, Foit K, Liess M (2020) Species occurrence relates to pesticide gradient in streams. Sci Total Environ 735:138807
4. Liess M, Liebmann L, Vormeier P, Weisner O, Altenburger R, Borchardt D et al (2021) Pesticides are the dominant stressors for vulnerable insects in lowland streams. Water Res 201:117262
5. Schäfer RB, Caquet T, Siimes K, Mueller R, Lagadic L, Liess M (2007) Effects of pesticides on community structure and ecosystem functions in agricultural streams of three biogeographical regions in Europe. Sci Total Environ 382(2–3):272–285. https://doi.org/10.1016/j.scitotenv.2007.04.040
6. Hunt L, Bonetto C, Marrochi N, Scalise A, Fanelli S, Liess M et al (2017) Species at Risk (SPEAR) index indicates effects of insecticides on stream

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 18 of 19

invertebrate communities in soy production regions of the Argentine Pampas. Sci Total Environ 580:699–709

7. Chiu M-C, Hunt L, Resh VH (2016) Response of macroinvertebrate communities to temporal dynamics of pesticide mixtures: a case study from the Sacramento River watershed, California. Environ Pollut 219:89–98

8. Malaj E, von der Ohe PC, Grote M, Kuhne R, Mondy CP, Usseglio-Polatera P et al (2014) Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. Proc Natl Acad Sci USA 111(26):9549–9554. https://doi.org/10.1073/pnas.1321082111

9. Schäfer RB, Pettigrove V, Rose G, Allinson G, Wightwick A, von der Ohe PC et al (2011) Effects of pesticides monitored with three sampling methods in 24 sites on macroinvertebrates and microorganisms. Environ Sci Technol 45(4):1665–1672

10. Liess M, Schulz R, Liess MHD, Rother B, Kreuzig R (1999) Determination of insecticide contamination in agricultural headwater streams. Water Res 33(1):239–247. https://doi.org/10.1016/S0043-1354(98)00174-2

11. Maltby L, Hills L (2008) Spray drift of pesticides and stream macroinvertebrates: experimental evidence of impacts and effectiveness of mitigation measures. Environ Pollut 156(3):1112–1120. https://doi.org/10.1016/j.envpol.2008.04.013

12. Reiber L, Knillmann S, Kaske O, Atencio LC, Bittner L, Albrecht JE et al (2021) Long-term effects of a catastrophic insecticide spill on stream invertebrates. Sci Total Environ 768:144456

13. Kreuger J, Nilsson E. (2001) Catchment scale risk-mitigation experiences-key issues for reducing pesticide transport to surface waters. British Crop Protection Council Symposium Proceedings 319–24.

14. Liess M, von der Ohe PC (2005) Analyzing effects of pesticides on invertebrate communities in streams. Environ Toxicol Chem 24(4):954–965

15. Beketov MA, Kefford BJ, Schäfer RB, Liess M (2013) Pesticides reduce regional biodiversity of stream invertebrates. Proc Natl Acad Sci 110(27):11039–11043

16. Münze R, Orlinskiy P, Gunold R, Paschke A, Kaske O, Beketov MA et al (2015) Pesticide impact on aquatic invertebrates identified with Chemcatcher(R) passive samplers and the SPEAR(pesticides) index. Sci Total Environ 537:69–80. https://doi.org/10.1016/j.scitotenv.2015.07.012

17. EFSA (2013) Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA J 11(7):3290

18. Williams P, Whitfield M, Biggs J, Fox G, Nicolet P, Shillabeer N et al (2002) How realistic are outdoor microcosms? A comparison of the biota of microcosms and natural ponds. Environ Toxicol Chem 21(1):143–150

19. Brock TC, Arts GH, Maltby L, Van den Brink PJ (2006) Aquatic risks of pesticides, ecological protection goals, and common aims in European Union legislation. Integr Environ Assess Manag 2(4):e20–e46

20. Ledger M, Harris R, Armitage P, Milner A (2009) Realism of model ecosystems: an evaluation of physicochemistry and macroinvertebrate assemblages in artificial streams. Hydrobiologia 617(1):91–99

21. Landis WG, Matthews RA, Matthews GB (1997) Design and analysis of multispecies toxicity tests for pesticide registration. Ecol Appl 7(4):1111–1116

22. Ledger ME, Harris RM, Milner AM, Armitage PD (2006) Disturbance, biological legacies and community development in stream mesocosms. Oecologia 148(4):682–691

23. Beketov MA, Schäfer RB, Marwitz A, Paschke A, Liess M (2008) Long-term stream invertebrate community alterations induced by the insecticide thiacloprid: effect concentrations and recovery dynamics. Sci Total Environ 405(1–3):96–108

24. Brock T, Hammers-Wirtz M, Hommen U, Preuss T, Ratte H, Roessink I et al (2015) The minimum detectable difference (MDD) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. Environ Sci Pollut Res 22(2):1160–1174

25. Knillmann S, Orlinskiy P, Kaske O, Foit K, Liess M (2018) Indication of pesticide effects and recolonization in streams. Sci Total Environ 630:1619–1627. https://doi.org/10.1016/j.scitotenv.2018.02.056

26. Münze R, Hannemann C, Orlinskiy P, Gunold R, Paschke A, Foit K et al (2017) Pesticides from wastewater treatment plant effluents affect invertebrate communities. Sci Total Environ 599–600:387–399. https://doi.org/10.1016/j.scitotenv.2017.03.008

27. Orlinskiy P, Münze R, Beketov M, Gunold R, Paschke A, Knillmann S et al (2015) Forested headwaters mitigate pesticide effects on

28. Schäfer RB, von der Ohe PC, Rasmussen J, Kefford BJ, Beketov MA, Schulz R et al (2012) Thresholds for the effects of pesticides on invertebrate communities and leaf breakdown in stream ecosystems. Environ Sci Technol 46(9):5134–5142. https://doi.org/10.1021/es2039882

29. von der Ohe PC, Prüß A, Schäfer RB, Liess M, de Deckere E, Brack W (2007) Water quality indices across Europe—a comparison of the good ecological status of five river basins. J Environ Monit 9(9):970–978

30. Rasmussen JJ, McKnight US, Loinaz MC, Thomsen NI, Olsson ME, Bjerg PL et al (2013) A catchment scale evaluation of multiple stressor effects in headwater streams. Sci Total Environ 442:420–431. https://doi.org/10.1016/j.scitotenv.2012.10.076

31. van den Brink PJ, van Donk E, Gylstra R, Crum SJ, Brock TC (1995) Effects of chronic low concentrations of the pesticides chlorpyrifos and atrazine in indoor freshwater microcosms. Chemosphere 31(5):3181–3200

32. Duquesne S, Alalouni U, Gräff T, Frische T, Pieper S, Egerer S et al (2020) Better define beta–optimizing MDD (minimum detectable difference) when interpreting treatment-related effects of pesticides in semi-field and field studies. Environ Sci Pollut Res 27(8):8814–8821

33. Lee AF, Gurland J (1975) Size and power of tests for equality of means of two normal populations with unequal variances. J Am Stat Assoc 70(352):933–941

34. European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities.

35. SPEAR$_{pesticides}$ website Indicate. www.systemecology.eu/indicate. Accessed October 2020.

36. Von der Ohe PC, Liess M (2004) Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. Environ Toxicol Chem 23(1):150–156. https://doi.org/10.1897/02-577

37. Rubach MN, Baird DJ, Van den Brink PJ (2010) A new method for ranking mode-specific sensitivity of freshwater arthropods to insecticides and its relationship to biological traits. Environ Toxicol Chem 29(2):476–487. https://doi.org/10.1002/etc.55

38. Vaal MA, Van Leeuwen CJ, Hoekstra JA, Hermens JL (2000) Variation in sensitivity of aquatic species to toxicants: practical consequences for effect assessment of chemical substances. Environ Manage 25(4):415–423

39. Escher BI, Hermens JL (2002) Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. Environ Sci Technol 36(20):4201–4217

40. Ippolito A, Todeschini R, Vighi M (2012) Sensitivity assessment of freshwater macroinvertebrates to pesticides using biological traits. Ecotoxicology 21(2):336–352. https://doi.org/10.1007/s10646-011-0795-x

41. Rico A, Van den Brink PJ (2015) Evaluating aquatic invertebrate vulnerability to insecticides based on intrinsic sensitivity, biological traits, and toxic mode of action. Environ Toxicol Chem 34(8):1907–1917

42. van Wijngaarden RP, Maltby L, Brock TC (2015) Acute tier-1 and tier-2 effect assessment approaches in the EFSA Aquatic Guidance Document: are they sufficiently protective for insecticides? Pest Manag Sci 71(8):1059–1067

43. Ecotoxicology database of the US Environmental Protection Agency ECOTOX. https://cfpub.epa.gov/ecotox/.

44. Liess M (2002) Population response to toxicants is altered by intraspecific interaction. Environ Toxicol Chem 21(1):138–142

45. Roessink I, Merga LB, Zweers HJ, Van den Brink PJ (2013) The neonicotinoid imidacloprid shows high chronic toxicity to mayfly nymphs. Environ Toxicol Chem 32(5):1096–1100

46. Liess M, Foit K, Knillmann S, Schäfer RB, Liess H-D (2016) Predicting the synergy of multiple stress effects. Sci Rep 6:32965. https://doi.org/10.1038/srep32965

47. Heugens EHW, Hendriks AJ, Dekker T, Straalen NM, Admiraal W (2001) A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. Crit Rev Toxicol 31(3):247–284. https://doi.org/10.1080/20014091111695

48. Stampfli NC, Knillmann S, Liess M, Beketov MA (2011) Environmental context determines community sensitivity of freshwater zooplankton to a pesticide. Aquat Toxicol 104(1–2):116–124. https://doi.org/10.1016/j.aquatox.2011.04.004

Reiber *et al. Environmental Sciences Europe*   (2022) 34:65

Page 19 of 19

49. Kreuger J (1998) Pesticides in stream water within an agricultural catchment in southern Sweden, 1990–1996. Sci Total Environ 216(3):227–251
50. Foit K, Kaske O, Liess M (2012) Competition increases toxicant sensitivity and delays the recovery of two interacting populations. Aquat Toxicol 106–107:25–31. https://doi.org/10.1016/j.aquatox.2011.09.012
51. Liess M, Foit K, Becker A, Hassold E, Dolciotti I, Kattwinkel M et al (2013) Culmination of low-dose pesticide effects. Environ Sci Technol 47(15):8862–8868. https://doi.org/10.1021/es401346d
52. EFSA (2019) Technical report on the outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology. EFSA Supporting Publications 16(7):1673E
53. Beuter L-K, Dören L, Hommen U, Kotthoff M, Schäfers C, Ebke KP (2019) Testing effects of pesticides on macroinvertebrate communities in outdoor stream mesocosms using carbaryl as example test item. Environ Sci Eur 31(1):1–17
54. Liess M, Beketov MA (2012) Rebuttal related to "Traits and stress: Keys to identify community effects of low levels of toxicants in test systems" by Liess and Beketov (2011). Ecotoxicology 21(2):300–303
55. Van Wijngaarden R, Van Den Brink PJ, Oude Voshaar JH, Leeuwangh P (1995) Ordination techniques for analysing response of biological communities to toxic stress in experimental ecosystems. Ecotoxicology 4(1):61–77
56. Van den Brink PJ, Ter Braak CJ (1998) Multivariate analysis of stress in experimental ecosystems by principal response curves and similarity analysis. Aquat Ecol 32(2):163–178
57. Van den Brink PJ, Braak CJT (1999) Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress. Environ Toxicol Chem 18(2):138–148
58. European Commission (2002) Guidance Document on Aquatic Ecotoxicology Under Council Directive 91/414/EEC. SANCO/3268/2001-rev. 4 final, 17 October 2002.
59. SETAC-Europe (1992) Guidance document on testing procedures for pesticides in freshwater mesocosms. A Meeting of Experts on Guidelines for Static Field Mesocosm Tests. Monks Wood Experimental Station, Huntington, U.K., July 1991: Society of Environmental Toxicology & Chemistry-Europe p. 46.
60. OECD (2006) Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms). OECD Publishing, Berlin
61. SETAC-Resolve (1992) Proceedings of a workshop on aquatic microcosms for ecological assessment of pesticides, Wintergreen, Virginia, USA, October 1991. SETAC Foundation for Environmental Education and the RESOLVE Program of the World Wildlife Fund.
62. Crossland N, Heimbach F, Hill I, Boudou A, Leeuwangh P, Matthiessen P, et al. (1992) Summary and recommendations of the European Workshop on Freshwater Field Tests (EWOFFT). Potsdam, Germany
63. World Wildlife Fund (1992) Improving Aquatic Risk Assessment Under FIFRA: Report of the Aquatic Effects Dialogue Group. Washington DC.: Resolve.
64. Brock T, Heger W, Gidddings J, Heimbach F, Maund S, Norman S, et al. (2000) Guidance document on Community Level Aquatic System Studies - Interpretation Criteria. Proceedings of the SETAC-Europe/OECD/EC Workshop held at Schmallenberg, Germany, 30 May–2 June 1999. In preparation for SETAC-Europe Press.
65. OECD (1996) Draft proposal for a guidance document – Freshwater lentic field tests.
66. De Jong F, Brock T, Foekema E, Leeuwangh P (2008) Guidance for summarizing and evaluating aquatic micro-and mesocosm studies. RIVM Rep 601506009:2008
67. OECD (2004) Draft guidance document on simulated freshwater lentic field tests (outdoor micro-and mesocosms).
68. Campbell P, Arnold D, Brock T, Grandy N, Heger W, Heimbach F, et al. (1999) Guidance document on higher-tier aquatic risk assessment for pesticides (HARAP). From the SETAC-Europe. OECD/EC Workshop, Lacanau Océan, France, SETAC-Europe, Brussels, Belgium.
69. Giddings JM, Brock T, Heger W, Heimbach F, Maund S, Norman S, et al. (2002) Community-level aquatic system studies-interpretation criteria.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.