Title: The prevalence and acceptability of mesocosm studies submitted for macrophytes in pesticide risk assessment

Funding: UK Research and Innovation, Natural Environment Research Council NE/VO13041/1

Author Names and affiliations:

Isabel Navarro Law, Department of Environment and Geography, University of York, Wentworth Way York, North Yorkshire, YO10 5DD, United Kingdom

Colin D Brown, Department of Environment and Geography, University of York, Wentworth Way, York, North Yorkshire, YO10 5DD, United Kingdom

Jason Snape: Department of Environment and Geography, University of York, Wentworth Way, York, North Yorkshire, YO10 5DD, United Kingdom

Isabelle Durance: School of Biosciences, Cardiff University, Park Place, Cardiff, CF10 3AT, United Kingdom

Melissa Reed: Chemicals Regulation Division, Health and Safety Executive, Mallard House, Kings Pool, 3 Peasholme Green, York YO1 7PX, United Kingdom

Michael Fryer: Chemicals Regulation Division, Health and Safety Executive, Mallard House, Kings Pool, 3 Peasholme Green, York YO1 7PX, United Kingdom

Corresponding author: Isabel Navarro Law, isabel.navarrolaw@york.ac.uk

Data availability: The data underlying this article cannot be shared publicly due to the study plans submitted by the applicant containing confidential data and GDPR concerns. The study reports written by the regulatory authorities will be shared on reasonable request to the corresponding author.

Conflicts of interest: Michael Fryer and Melissa Reed both work at the Chemicals Regulation Division (the UK regulator for pesticides), overseeing the regulation of PPPs.

Acknowledgments: All authors contributed equally.

The prevalence and acceptability of mesocosm studies submitted for macrophytes in pesticide risk assessment

Abstract

Mesocosms can be used in higher tier aquatic risk assessments to assess the impact of Plant Protection Products (PPPs) on macrophytes. However, it is unclear whether these expensive and time consuming higher tier studies influence regulatory outcomes. This review highlights common shortcomings in the experimental design of mesocosm studies, with the aim of maximising the regulatory value of future mesocosm studies. Fourteen mesocosm studies, which have been submitted for the regulatory risk assessments for macrophytes in the EU or GB, were identified and reviewed. Results show that only five of the 14 mesocosm studies were deemed acceptable by the regulatory authorities, suggesting that mesocosm studies are not currently being used to their full potential. Issues with the submitted studies include not following a realistic PPP exposure profile (including incorrect dose timings and dilutions), only using one macrophyte morphology, not leaving enough time for the macrophytes to establish and a lack of replicates which increases variability within treatments. *Glyceria maxima* and *Myriophyllum spicatum* were frequently the most sensitive macrophyte species, whilst dry weight was often the most sensitive and least variable endpoint. Even though mesocosms provide the opportunity for recovery and community responses to be observed, such information has not been used by regulatory authorities. Future regulatory mesocosm studies can build upon the shortcomings highlighted here, providing a greater chance of regulatory impact.

Keywords

Mesocosms; Higher tier risk assessments; Macrophytes; Herbicides

Introduction

Plant protection products (PPPs), including insecticides, herbicides, fungicides and plant growth regulators, are used to manage weeds and pests, primarily in agriculture but also for domestic, recreational and amenity purposes. In order for a new PPP active substance or product (which will from here on be referred to collectively as ‘PPPs’) to be eligible for use, approval is required by the relevant regulatory authority. For both member states of the

European Union (EU) and Great Britain (GB), the approval process requires, amongst other criteria, assessing the risks posed from use of the PPP to various groups of non-target organisms, including aquatic organisms. Determining the permissible exposure level for aquatic organisms to a PPP requires calculating a Regulatory Acceptable Concentration (RAC). An acceptable risk is identified where the ratio of the predicted PPP concentration to the RAC is ≤ 1 (EFSA, 2013).

To characterise the toxicity of the PPP on the aquatic environment, initial laboratory studies are required for individual organisms. For macrophytes, the first tier study typically undertaken is a laboratory study on *Lemna sp.* following OECD guideline 221 (OECD, 2006a), investigating the impact of the PPP on growth rate. If the risk assessment using the first tier study does not show an acceptable risk to macrophytes, then the applicant can perform second tier studies, which commonly involve using more macrophyte species to either create a species sensitivity distribution curve or to calculate a geometric mean of the results for all species (EFSA, 2013). If this still does not show an acceptable risk to the aquatic environment then mesocosm experiments are advised as a higher tier risk assessment option (EFSA, 2013).

A mesocosm is an artificial experimental system, created to physically model natural aquatic ecosystems (Brock et al., 2009; Giddings et al., 2002). Plant protection products are applied to mesocosms to mimic real world exposure scenarios and investigate the impact of exposure on complex populations and communities (Giddings et al., 2002). Due to their complex nature, mesocosm experiments are often labour intensive and costly (OECD, 2006b). Discussions with regulatory scientists and contract research organisations indicate that depending on the size and set up, mesocosms can often cost hundreds of thousands of pounds.

When using a mesocosm study as part of a higher tier risk assessment, the critical toxicity endpoint from the study (usually expressed as a NOEC (No Observed Effect Concentration)) is divided by an assessment factor to determine the RAC. If there are issues with the reliability or sensitivity of a mesocosm study then higher assessment factors can be used to account for this where necessary (EFSA, 2013). Typically an assessment factor > 1 is used, however any issues with a study can increase the assessment factor. If a mesocosm study is not considered relevant or reliable, then the study will not be used in the regulatory risk

assessment, which can impact whether the PPP can be authorised or any conditions of authorisation (e.g. risk mitigation measures required; EFSA, 2013).

There are no prescriptive guidelines on how to undertake mesocosm studies suitable for PPP regulation. However, four guidance documents have been published over the last 25 years that provide recommendations on how best to design mesocosm studies. Giddings et al. (2002) includes 16 key recommendations which should be used when designing and implementing a mesocosm study, based on the findings from the Community Level Aquatic System Studies Interpretation Criteria (CLASSIC) workshop. In 2002, EU member states produced the SANCO guidance document (2002) to provide guidance on best practice for all aquatic ecotoxicology studies. The OECD have produced a guidance document specifically on best practice for mesocosm studies, focusing on all aquatic organisms (OECD, 2006b). Most recently, the European Food Safety Authority (EFSA) produced a guidance document on aquatic tiered risk assessments (2013). As a whole, these four guidance documents align with one another, however there are some contradictions in advice, as would be expected in documents published across an 11 year period. The EFSA (2013) and SANCO (2002) guidance documents are the only two that contain official guidance.

This review assesses common weaknesses of mesocosm studies that have been received by EU member state and GB regulatory authorities up to January 2025 as part of a higher tier risk assessment to assess impacts on macrophytes. It has been claimed that mesocosms are frequently used as a higher tier test approach to investigate the impact of PPPs on macrophytes (Taylor & Blake, 2013). However, in practice, it is unclear how often they are utilised. There is also currently no information on the success rate of applications supported by these costly and time consuming experiments. Additionally, current regulatory approaches examine risk assessments on a product-by-product basis, without any systematic data collection or analysis between products. Previous reviews have investigated mesocosm experiments to highlight shortcomings (Brock et al., 2000; Brown et al., 2009; Reiber et al., 2022), however none have exclusively focused on either macrophytes or studies that had been submitted for applications to regulatory authorities. By reviewing common shortcomings, this review aims to maximise the regulatory value and impact of mesocosm studies and minimise future opportunities for study failure.

Methods

Following discussion with the Chemicals Regulation Division of the Health and Safety Executive (the UK regulator for pesticides), mesocosm studies were identified which had

been submitted in the context of regulatory risk assessments for macrophytes. Fourteen studies were found, which were for pesticide active substances or products. These studies were conducted either by the applicant or an external contract research organisation and do not enter the peer review process before submission. Only one study was not initially designed for regulation, but rather for a PhD thesis that was submitted as supporting evidence for an application. These studies were conducted between 1993 and 2014. The review was carried out in February 2025. Data from all fourteen studies were systematically extracted into a matrix, which included the following preselected key properties related to mesocosm experimental design: mesocosm size, number of replicates, study duration, macrophyte species and morphologies, whether any other taxa were present, the length of time the plants had to establish before dosing, whether volunteer species were removed, the PPP application method, the number of concentrations, doses and dilutions, the exposure profiles considered, the month of dosing, the type, number and timepoint of endpoint measurements, whether a NOEAC, NOEAEC or community composition analysis was determined, statistical analysis, which endpoints were excluded due to high variability, the least and most sensitive species and endpoints, whether the mesocosm influenced the regulatory outcome and the rationale for the studies inclusion or exclusion. The extracted information was then examined using qualitative document analysis principles (Bowen 2009). Key patterns were identified, so that only properties that contributed to a study not influencing the regulatory outcome were retained for the review.

Results and Discussion

Fourteen mesocosm studies were identified for macrophytes in pesticide risk assessment and only five of these successfully contributed to the critical RAC and hence impacted the regulatory outcome. Of the five studies that did influence the regulatory outcome, four of them increased the RAC by a factor between 1.6 and 10.5. Notably, one mesocosm study led to a decrease in the RAC, however this was due to the impact of the PPP on algae. The submission of 14 studies, compared to the large number of lower tier studies submitted, suggests that in most cases PPPs are either not considered sufficiently toxic to warrant a mesocosm study, having already passed lower tier regulatory assessments, or are regarded by the applicant as posing too high a risk for commercial use and investment in the development of that product has ceased. This review explores the relative shortcomings of the fourteen submitted studies in context of whether or not they impacted on the risk assessment.

GLP

The EFSA guidance (2013) states that risk assessment studies should be conducted under the principles of 'Good Laboratory Practice' (GLP), which helps assure the quality of data produced. Two of the fourteen submitted mesocosm studies were not considered reliable because they were not conducted under GLP. Of the two non-GLP studies, one would have likely been accepted if it had been done to GLP standard. The data analysis and presentation was too poor in the other non-GLP study to be used for regulatory purposes. Following GLP procedure will enhance the likelihood that data analysis and presentation is adequate for regulatory processes.

Macrophytes

Table 1 lists the species used throughout the fourteen mesocosm studies reviewed.

Number of Species

The number of species used in the reviewed mesocosm studies range from three to nine, with the majority of studies using only 3 or 4 species to represent edge-of-field water bodies (Table 1). Four out of the five studies that were deemed acceptable by regulatory authorities only included three species. The number of species used did not influence whether a study was accepted by the regulatory authority.

Species Sensitivity

In three of the studies reviewed, the emergent monocotyledon *Glyceria maxima* was the most sensitive, whilst the submerged dicotyledonous *Myriophyllum spicatum* was the most sensitive in three other studies (Table 1). Two of the studies had the submerged monocotyledon *Elodea canadensis* as their most sensitive species, whilst one had *Sparganium erratum* and another had *Lemna sp.*. The most sensitive species is used to calculate the NOEC value, which is used to determine the RAC.

Macrophyte Morphology

All but one of the mesocosm studies used species representing multiple types of macrophyte morphology. The exception, which exclusively used free-floating species, was not considered reliable as being able to show an acceptable risk to the aquatic environment because two of the three species failed to establish. In mesocosm studies, failing to establish refers to the plants dying even in the absence of chemical exposure, likely because they were unable to root into the sediment. A broader range of morphologies could have improved the study's chances of acceptance.

Establishment

In the majority of reviewed studies, macrophytes were planted in pots before being added to the mesocosm. One study attempted to plant macrophytes directly into the mesocosm, though they did not establish so had to be replanted into pots and delayed the study.

As mesocosms are artificial, yet typically located outside to mimic real world scenarios, it can be difficult to ensure that species establish after planting. A study in which only one of the three species established, only added macrophytes to the mesocosm one day before dosing. Similarly, another study, in which the *Elodea canadensis* was planted one day before dosing, was not considered reliable because the *Elodea canadensis,* which was the most sensitive species, did not grow well enough in the controls to calculate a reliable NOEC.

Five of the mesocosm studies used *Lemna sp.* (Table 1). Throughout this review, *Lemna gibba* and *Lemna minor* are grouped into *Lemna sp*., as they share similar traits (De Lange & Pieterse, 1973). In three of the five studies where *Lemna* was used, the individuals did not establish. However, one of the studies that did successfully prove an acceptable risk to macrophytes used *Lemna* for the NOEC value (Table 1). This study added *Lemna* 5 weeks before dosing.

Volunteer Species

Volunteer species are those that appear in the mesocosm but were not planted, so were most likely introduced via the use of natural or semi-natural sediment or water or via wind dispersal. Most studies removed any volunteer species from the mesocosms before planting the chosen macrophyte species. Some studies left the volunteer species in and took endpoint measurements of both planted and volunteer species. However, in one instance the data from the volunteer species was disregarded due to high levels of variation between treatments. Notably, one of the studies which was accepted by the regulatory authority only used volunteer species to assess the impact of the PPP. Even though this method is acceptable, it comes with risks of lacking uniformity between individuals.

Replication

Two replicates per treatment type is usually advised to enhance the statistical power and reduce the likelihood of losing essential controls (Giddings et al., 2002; OECD, 2006b). Using only one replicate per treatment also carries risk; for example, if that mesocosm is damaged, no data remain for that concentration. This proved true for one study, which was considered unreliable by the regulatory authority after leaks compromised two treatments

and a control. The unpredictable nature of outdoor mesocosm experiments inherently increases the risk of such incidents. Another study was unable to demonstrate an acceptable risk to the environment because there were only two replicates per treatment, meaning that the data were highly variable. The regulatory authority concluded that the high variability, attributed to the lack of replicates, meant that the study did not have sufficient statistical power to be capable of detecting any treatment related effects.

Plant Protection Product Exposure

Mesocosm experiments are designed to physically model natural ecosystems, so PPP exposure in mesocosms should accurately reflect real-world exposure scenarios (Giddings et al., 2002). The EU uses FOCUS scenarios to model the fate of PPPs, using four steps (tiers) that start simple with no specific scenario and end more complex and less conservative (FOCUS, 2001).

When ensuring that PPP exposure in mesocosm studies mimics real world scenarios, it is important to take into account both the magnitude and duration of exposure. For instance, exposure via spray drift will likely only occur at the time of dosage, whereas exposure via drainflow can occur repeatedly and less predictably, typically throughout autumn and winter during high rainfall events. Mesocosm studies should ensure that both the magnitude and duration of the expected exposure is covered by the chosen concentrations.

Exposure Timing

It has been suggested that mesocosm dosing should occur in spring to mid-summer (Giddings et al., 2002; OECD, 2006b), even though most exposure via drainflow occurs due to heavy rainfall between autumn and winter. Between spring and mid-summer plants are actively growing and at their most sensitive life stage, providing a worse case scenario. Most of the mesocosm studies submitted were indeed conducted between May and July when macrophytes are in their fastest growth stage, which satisfied the regulatory authority. However, two of the studies dosed their mesocosms in March. The regulatory authorities were concerned about this exposure timing so enforced a higher assessment factor than usual. A higher assessment factor reduces the RAC to address concerns regarding the potential for lower sensitivity of macrophytes in March. This could lead to the assessment failing to demonstrate an acceptable risk or the need for additional risk mitigation measures.

Dilutions

It is important to ensure that the rate of decline of a PPP is the same in the mesocosm as would be expected in the field. This is typically done by replacing the contaminated water with fresh water, hence decreasing the concentration of the PPP. Only three of the studies chose to dilute the mesocosms to mimic PPP lost to flow and rainwater. Two of these studies were deemed to have appropriate exposure, whilst one was not considered reliable because the dilutions decreased the PPP concentration faster than the FOCUS exposure profiles predicted. Diluting mesocosms can be an accurate way to mimic PPP loss via flow, which will result in macrophytes being exposed to PPPs for a similar length of time as they would be in the field.

Multiple Doses

Only one of the reviewed studies applied multiple doses of PPP, to mimic exposure via drainflow. This study applied two doses, seven days apart, and was deemed sufficient by the regulatory authority. In contrast, a study that only applied one dose of PPP was not considered reliable because real-world scenarios would have multiple exposure peaks due to rainfall events, so more doses were required to accurately represent real-world drainflow exposure. This indicates that multiple doses of PPPs are a good way to accurately reflect the predicted exposure.

Dosing Method

All submitted studies used pulse doses to expose PPP to the mesocosms. There are two main methods for pulse dosing a PPP in mesocosm studies. The toxicological approach involves directly applying the PPP to the water and mixing to achieve an even distribution of chemical whilst the simulation approach mimics the realistic entry route of a PPP in the field, typically by spraying the PPP onto the water surface (OECD 2006b). Only one study used the simulation approach and sprayed the PPP directly onto the water surface to mimic spray drift exposure. This study was not considered reliable for unrelated reasons. There appear to be no limitations for a mesocosm study to use the toxicological approach to dose PPPs.

Matching Exposure Profiles

Three of the mesocosm studies were considered unreliable because they did not cover enough or any of the FOCUS exposure profiles. In one instance, the regulatory authority found that the PPP concentration in the mesocosms declined faster than in any of the modelled FOCUS scenarios. However, it is worth noting that the FOCUS model does not include processes such as photolysis, so the exposure seen in the mesocosm may be a

better indicator of real-world scenarios than the FOCUS modelled exposure. This was recognised by the regulatory authority, however they were still unable to authorise the study whilst the exposure profiles did not match. The other studies that were not considered reliable were due to incorrect dilutions or a lack of consideration of all relevant exposure peaks, as described above.

Test Concentrations

The mesocosm studies reviewed used between 3 and 8 concentrations, with the mesocosms that were accepted by the regulatory authority using between 3 and 6 concentrations. Not all of the accepted studies included a dose-response curve (EFSA, 2013), suggesting that providing key data and a NOEC is sufficient, and a dose-response curve is not essential.

Ideally, selected concentrations should range from the lowest having no treatment related impact on macrophytes, to the highest concentration having a clear impact. In one study, the lowest concentration had a treatment-related effect on the macrophytes, meaning that neither a fixed NOEC or RAC could be determined and the study could not be used for the risk assessment.

Endpoints

Endpoints refer to measurements used to assess the health of an organism after exposure to a treatment. Table 2 details the type of endpoints used in the fourteen mesocosm studies evaluated and their contribution to NOEC values.

NOEC Endpoint

The highest concentration at which no significant effect on an endpoint is referred to as the NOEC. Therefore, the most sensitive endpoints and species are used to dictate the NOEC. Of the fourteen studies, three did not see a significant impact of any test concentration on the species' endpoints. In this instance, the NOEC value was based on the highest concentration tested, so effectively used all endpoints.

In six of the studies, dry weight was the most sensitive endpoint (Table 2). The remaining four studies that were able to detect a most sensitive endpoint identified either a plant health score, root length, cover or stem length as the most sensitive endpoint (Table 2). Notably, total wet weight was the most frequently used endpoint (Table 2), however was never the

most sensitive, so not used to calculate the NOEC. Leaf wet weight was used twice, and both times the results were too variable to derive an effect (Table 2).

NOEC vs NOAEC

Even though the NOEC was used to determine the RAC for most studies, some studies discuss using the No Observed Adverse Effect Concentration (NOAEC). The NOAEC only focuses on adverse effects, so if a concentration causes a perceived positive or neutral effect then it would be disregarded. One of the studies originally attempted to use the NOAEC in a PPP application, as the NOEC was a result of an increase in biomass, which the applicant suggested was not an adverse treatment-related effect. However, after discussion, both the applicant and member state reviewing the study agreed to proceed with the NOEC value. Using a NOAEC raises a lot of uncertainties as to what an adverse effect is. An increase in biomass may appear not to be adverse, however could still have knock-on effects on the resource allocation of the individual plant as well as taxa that rely on macrophytes. This could give an ecological benefit to that species over others and undermine community resilience. Similar discussions are occurring in animal ecotoxicology studies, with the regulatory relevance of behavioural endpoints being questioned if they do not translate to population-level effects (Ågerstrand et al. 2020). Focusing on the reviewed studies, the NOAEC has not been used in a mesocosm study for regulatory purposes.

Recovery

The OECD guidance (2006b) highlights that a positive attribute of mesocosm studies, compared to lower tier laboratory studies, is that they can take into account recovery. However, reviewing the fourteen submitted studies has raised questions regarding how often recovery is investigated in mesocosms.

In the past, the No Observed Ecological Adverse Effect Concentration (NOEAEC) has been recommended for use rather than the NOEC, to take into account recovery (SANCO 2002). The NOEAEC describes a concentration that has no ‘long lasting’ effects. If a macrophyte is able to recover after exposure to a certain concentration, then that concentration would not be considered ecologically adverse. The NOEAEC is not commonly used by regulators in PPP risk assessments at present due to concerns that recovery following exposure to one stressor will not be a good indicator of recovery in a situation with multiple stressors, as occurs often in agricultural landscapes (EFSA 2013).

Five of the studies reviewed here determined a NOEAEC concentration. For two of the studies the NOEC was the same concentration as the NOEAEC, meaning that the NOEAEC did not need to be explicitly taken into account. For one study no treatment related effects were seen when using either the NOEC or the NOEAEC. One of the studies that calculated the NOEAEC found that the root length recovered, however considering that the variation in plant health was so high, this study was not considered relevant. Another study that calculated the NOEAEC found that plant length for one of the species recovered over time, but this study was not considered reliable because the exposure profile did not align with the FOCUS model's predicted degradation rate. There is no evidence that the regulatory authority will accept a NOEAEC value over a NOEC value.

Three of the studies that calculated a NOEAEC were those that diluted the mesocosm units. The remaining two studies did not dilute the mesocosms when investigating recovery. Considering NOEAEC values are not used by regulators, there is no consensus as to whether macrophyte recovery should be recorded during or after herbicide exposure.

Community

Principal Response Curves (PRCs; Van den Brink & Braak, 1999) have been recommended for use in mesocosm studies (OECD, 2006b; SANCO, 2002) to investigate the influence of a treatment type on macrophyte community composition. PRCs are often used in regulatory studies on macroinvertebrates or phytoplankton to determine a community NOEC (the concentration where there is no treatment related effect on the community). Yet they were not used in any of the fourteen macrophyte studies reviewed on macrophytes. One potential reason is the small number of species used in macrophyte studies. Some of the submitted macrophyte studies used as little as 3 species, which is too few to undertake community composition analysis. On the other hand, five studies did use ≥ 6 species, so would have been sufficient for PRC analysis. However, even though PRCs reveal interesting community related impacts they are currently unlikely to set the overall RAC. Regulatory authorities instead currently focus only on the most sensitive species, rather than the entire community.

This means that there is currently no drive for an applicant to design a mesocosm study incorporating community analysis.

Conclusion

This review assessed the current use of mesocosm studies in regulation and identified common weaknesses (Figure 1).

One study was not considered suitable because the lowest concentration yielded significant treatment-related impacts, so a NOEC could not be calculated. Two were not considered reliable because the variation within treatment levels was too high, linked to there not being enough replicates and macrophytes not having enough time to fully establish before dosing. One study was not considered reliable because it was not conducted under GLP. Three were not considered reliable because the herbicide exposure did not cover any of the EU FOCUS profiles. One was not considered reliable for a number of reasons including no replicates or statistical analysis and the exposure profile being incorrect. The remaining mesocosm study produced NOEC values that were the same as the lower tier *Lemna* studies, so the mesocosm study did not change the outcome of the risk assessment.

Mesocosms act as a physical model to represent exposure scenarios in natural ecosystems. This review shows that they are currently not being used as frequently or effectively as they could be, with experimental design being the main barrier preventing regulatory acceptance. This places the responsibility on applicants to ensure the quality of their mesocosm studies prior to submission and the need for dialogue with regulatory agencies in the study design. PPPs that do not pass lower tier risk assessments should be tested in well designed mesocosm studies to increase the likelihood of being accepted for the risk assessment. This review highlights common shortcomings in mesocosm studies used for regulatory decision making.

References

Ågerstrand, M., Arnold, K., Balshine, S., Brodin, T., Brooks, B. W., Maack, G., McCallum, E. S., Pyle, G., Saaristo, M., & Ford, A. T. (2020). Emerging investigator series: use of behavioural endpoints in the regulation of chemicals. *Environmental Science. Processes & Impacts*, *22*(1), 49–65. https://doi.org/10.1039/c9em00463g

388 Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative*
389 *Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/qrj0902027

390 Brock, T., Alix, A., Brown, C., Capri, E., Gottesburen, B., Heimbach, F., Lythgo, C., & Schulz,
391 R, Streloke, M. (2009). *Linking aquatic exposure and effects: Risk assessment of*
392 *pesticides*. CRC Press. https://doi.org/10.1201/9781439813492
393 Brock, T. C., Lahr, J., & Van den Brink, P. J. (2000). *Ecological risks of pesticides in*
394 *freshwater ecosystems. Part 1: Herbicides*. Wageningen, Alterra, Green World
395 Research.
396 Brown, K., Tomlinson, J., Duncan, J., Hinchcliffe, A., & Palmquist, K. (2009). *Critical*
397 *comparison of available and potential higher tier testing approaches for the risk*
398 *assessment of plant protection products, considering at least field and semi-field*
399 *experimental designs, extrapolation from dose-response relationship, and increased*
400 *dosages (aquatic and terrestrial)* (Vol. 6). European Food Safety Authority.
401 https://doi.org/10.2903/sp.efsa.2009.en-16
402 De Lange, L., & Pieterse, A. H. (1973). A comparative study of the morphology of Lemna
403 gibba l. and Lemna minor l. *Acta Botanica Neerlandica*, *22*(5), 510–517.
404 EFSA. (2013). *Guidance on tiered risk assessment for plant protection products for aquatic*
405 *organisms in edge-of-field surface waters* (Vol. 11, p. 3290). Wiley.
406 https://doi.org/10.2903/j.efsa.2013.3290
407 FOCUS. (2001). *FOCUS surface water scenarios in the EU evaluation process under*
408 *91/414/EEC. Report of the FOCUS working group on surface water scenarios, EC*
409 *Document Reference SANCO/4802/2001-rev.2*.
410 Giddings, J. M., Brock, T. C. M., Heger, W., & Heimbach, F. (2002). *Community-level*
411 *aquatic system studies-interpretation criteria (CLASSIC)*. Society of Environmental
412 Toxicology and Chemistry.
413 OECD. (2006a). *OECD Guidelines for the Testing of Chemicals, Section 2 Test No. 221:*
414 *Lemna sp. Growth Inhibition Test*. OECD Publishing.

OECD. (2006b). *OECD Series on Testing and Assessment: Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms)*.

Reiber, L., Foit, K., Liess, M., Karaoglan, B., Wogram, J., & Duquesne, S. (2022). Close to reality? Micro-/mesocosm communities do not represent natural macroinvertebrate communities. *Environmental Sciences Europe*, *34*(1). https://doi.org/10.1186/s12302-022-00643-x

SANCO. (2002). *Guidance Document on Aquatic Ecotoxicology Sanco/3268/2001 rev.4*. European Commission Health and Consumer Protection.

Taylor, S., & Blake, N. (2013). *PS2339: Assessing the temporal / seasonal changes in the ecology of untreated mesocosms and natural water bodies to inform the uncertainty associated with aquatic risk assessments of plant protection products: Report for the Chemical Regulations Directorate (UK) by Cambridge Environmental Assessments - ADAS*. Cambridge Environmental Assessments - ADAS.

Van den Brink, P. J., & Braak, C. J. F. T. (1999). Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry / SETAC*, *18*(2), 138–148.

Figure 1: Common shortcomings observed in the fourteen mesocosm studies submitted for regulatory purposes. PPP: Plant Protection Products.

Table 1: Each species (sub-categorised by morphology) used in one of the fourteen mesocosm studies is listed, along with the number of studies it was used in, the number of times the species did not establish in the controls within a study, the number of times the species was the most sensitive (so used to calculate the NOEC value) and the number of times the species was not impacted by PPP exposure. For the latter three values, the percent of overall use is also provided.

441

442 **Table 2:** The endpoints used across the fourteen mesocosm studies, along with the number
443 of studies they were used in, the number of times the endpoint was the most sensitive (so
444 used to calculate the NOEC) and the number of times there was too much variation to be
445 used.

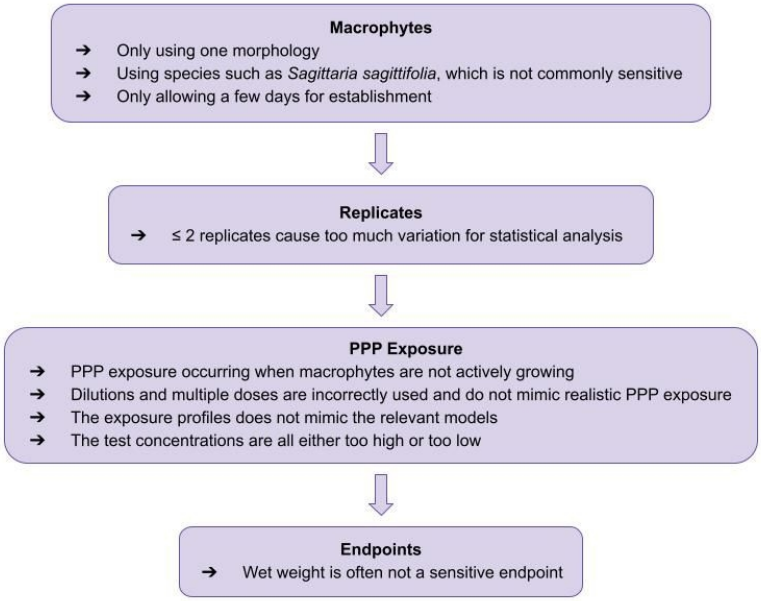

Figure 1: Common shortcomings observed in the fourteen mesocosm studies submitted for regulatory purposes.

81x60mm (300 x 300 DPI)

Table 1: Each species (sub-categorised by morphology) used in one of the fourteen mesocosm studies is listed, along with the number of studies it was used in, the number of times the species did not establish in the controls within a study, the number of times the species was the most sensitive (so used to calculate the NOEC value) and the number of times the species was not impacted by PPP exposure. For the latter three values, the percent of overall use is also provided.

Species	No. of studies it was used in	No. of times it did not establish (% of use)	No. of times it was the most sensitive species (% of use)	No. of times it was not impacted at all (% of use)
Emergent				
Glyceria Maxima	4	0	3 (75%)	1 (25%)
Hippuris vulgaris	2	0	0	0
Eleocharis palustris	1	0	0	0
Veronica beccabunga	1	0	0	1 (100%)
Sagittaria sagittifolia	4	0	0	2 (50%)
Sagittaria japonica	1	0	0	1 (100%)
Myosotis palustris	1	0	0	0
Sparganium erratum	3	0	1 (33%)	2 (66%)
Stratiotes aloides	1	0	0	0
Mentha aquatica	1	0	0	1 (100%)
Zannichellia palustris	1	0	0	1 (100%)
Free floating				
Lemna sp.	5	3 (60%)	1 (20%)	1 (20%)
Callitriche stagnalis	1	0	0	1 (100%)
Spirodela polyrhiza	1	1 (100%)	0	0
Rooted with floating leaves				
Potamogeton natans	5	0	0	3 (60%)
Persicaria amphibia	2	0	0	1 (50%)
Submerged				
Potamogeton crispus	2	0	0	0

Lagarosiphon major	1	0	0	1 (100%)
Ceratophyllum submersum	1	0	0	0
Ceratophyllum demersum	2	1 (50%)	0	1 (50%)
Myriophyllum spicatum	12	1 (8%)	3 (25%)	4 (33%)
Myriophyllum aquaticum	2	0	0	2 (100%)
Elodea canadensis	8	1 (13%)	2 (25%)	3 (39%)
Elodea nuttalli	2	0	0	2 (100%)

Note. NOEC = no observed effect concentration; PPP = plant protection products.

Table 2: The endpoints used across the fourteen mesocosm studies, along with the number of studies they were used in, the number of times the endpoint was the most sensitive (so used to calculate the NOEC) and the number of times there was too much variation to be used.

Endpoint	No. of studies it was used in	No. of times it was the most sensitive endpoint (% of use)	No. of times there was too much variation to be used (% of use)
Non-destructive endpoints, that can be used multiple times throughout a study			
Stem length	5	1 (20%)	0
Stem number	4	0	0
A plant score using health and abundance	1	1 (100%)	0
Cover	8	1 (13%)	3 (38%)
Leaf area	2	0	0
General appearance (using coverage, structure and foliage)	1	0	0
Length of the longest leaf	1	0	0
Number of leaves	4	0	0
Number of flowers	1	0	0
Number of side shoots on main stem	2	0	0
Total shoot length	2	0	0
Destructive endpoints, that can only be used once in a study (typically at the end)			
Total dry weight	7	6 (86%)	1 (14%)
Total wet weight	10	0	3 (30%)
Leaf dry weight	1	0	0
Leaf wet weight	2	0	2 (100%)
Root dry weight	1	0	0
Root wet weight	2	0	1 (50%)

Root length	2	1 (50%)	0

Note. NOEC = no observed effect concentration.