European Food Safety Authority

EFSA Journal 2014;12(3):3589

SCIENTIFIC OPINION

Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products[1]

EFSA Panel on Plant Protection Products and their Residues (PPR)[2,3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

The Panel has interpreted the Terms of Reference as a stepwise analysis of issues relevant to both the development and the evaluation of models to assess ecological effects of pesticides. The regulatory model should be selected or developed to address the relevant specific protection goal. The basis of good modelling practice must be the knowledge of relevant processes and the availability of data of sufficient quality. The opinion identifies several critical steps in order to set models within risk assessment, namely: problem formulation, considering the specific protection goals for the taxa or functional groups of concern; model domain of applicability, which drives the species and scenarios to model; species (and life stage) selection, considering relevant life history traits and toxicological/toxicokinetics characteristics of the pesticide; selection of the environmental scenario, which is defined by a combination of abiotic, biotic and agronomic parameters to provide a realistic worst-case situation. Model development should follow the modelling cycle, in which every step has to be fully documented: (i) problem definition; (ii) model formulation, i.e. design of a conceptual model; (iii) model formalisation, in which variables and parameters are linked together into mathematical equations or algorithms; (iv) model implementation, in which a computer code is produced and verified; (v) model setup, including sensitivity analysis, uncertainty analysis and comparison with observed data, that delivers the regulatory model; (vi) prior to actual use in risk assessment, the regulatory model should be evaluated for relevance to the specific protection goals; (vii) feedback from risk assessor with possible recommendations for model improvement. Model evaluation by regulatory authorities should consider each step of the modelling cycle: the opinion identifies points of particular attention for the use of mechanistic effect models in pesticide risk assessment. It is recommended that models be documented in a complete and transparent way, that a feedback platform be established involving risk assessors and model developers, and that a set of agreed models be made available.

KEY WORDS

model evaluation, model documentation, effect modelling, environmental scenarios, environmental pesticide risk assessment

---

[1] On request from EFSA, Question No EFSA-Q-2011-00989, adopted on 13 February 2014.

[2] Panel members: Alf Aagaard, Theo Brock, Ettore Capri, Sabine Duquesne, Metka Filipic, Antonio F. Hernandez-Jerez, Karen I. Hirsch-Ernst, Susanne Hougaard Bennekou, Michael Klein, Thomas Kuhl, Ryszard Laskowski, Matthias Liess, Alberto Mantovani, Colin Ockleford, Bernadette Ossendorp, Daniel Pickford, Robert Smith, Paulo Sousa, Ingvar Sundh, Aaldrik Tiktak, Ton Van Der Linden. Correspondence: pesticides.ppr@efsa.europa.eu

[3] Acknowledgement: The Panel wishes to thank the members of the Working Group on good modelling practice: Virginie Ducrot, Sabine Duquesne, Mira Kattwinkel, Matthias Liess, Alberto Mantovani, Melissa Reed, Richard Sibly, Robert Smith, Aaldrik Tiktak, Christopher John Topping for the preparatory work on this scientific opinion and EFSA staff Franz Streissl for the support provided to this scientific opinion.

Suggested citation: EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2014. Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA Journal 2014;12(3):3589, 92 pp. doi:10.2903/j.efsa.2014.3589

Available online: www.efsa.europa.eu/efsajournal

SUMMARY

The European Food Safety Authority (EFSA) asked the Panel on Plant Protection Products and their Residues (PPR) to prepare a scientific opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products.

The Panel has interpreted the Terms of Reference to mean a provision of a stepwise analysis of issues relevant to both the development and the assessment of models to assess ecological effects of pesticides. In accordance with the Terms of Reference, the Panel has pointed out mainly those general issues shared by several types of models and/or application fields, but some specific issues have been considered if highly relevant to a given model type and/or application field.

The PPR Panel emphasises that the availability and quality of data are of the utmost importance at all stages of modelling. If the data are not sufficient, the regulatory model may not be considered acceptable for risk assessment.

Clear and accurate *problem formulation* is the first critical step to implement the use of any model within risk assessment: the risk assessment areas to be addressed by modelling should be identified, considering the review of existing datasets, the risk assessment tier and the specific protection goals to be addressed. The starting point will be the protection goal for the taxa or functional groups of concern: the protection goal has to be defined in terms of which ecological entity, which attribute(s) or characteristic(s) must be protected, the magnitude of effect that can be tolerated for these attribute(s)/characteristic(s), the temporal and spatial scales of effect as well as the degree of certainty that the specified level of effect will not be exceeded. Therefore, the modelling should target specific question(s) defined according to the available data and specific protection goals. Another important aspect of problem formulation is to consider how the output obtained is suited to a given protection goal, which might include modelling of, for example, the proportion of the population affected, the potential for recovery if the protection goal allows for a certain magnitude and duration of effects or the potential impact of negligible effects repeated over several years.

The model *domain of applicability* deserves attention. Modelling offers the opportunity to go beyond the conditions that have been tested in experiments or observed in the field. A sound mechanistic foundation might allow for extrapolations, e.g. from individual to population level, or from one species to another. However, the Panel recognises that care must be taken on the broader conclusions drawn from the modelling results as a model can yield meaningful predictions only for those scales, processes and variables that are taken into account. Thus, the model domain drives the species and scenarios to model. However, models can be developed only if there are either sufficient data available or the potential to generate the required data. Thus, the availability of information could limit the choice of model type, species modelled or scenarios tested.

The Panel notes that selection of relevant species to model is critical, as it is not possible to model all the species present in a given community or ecosystem. It is important to consider relevant life history traits when identifying vulnerable species for modelling. It is recognised that toxicological knowledge is a critical parameter for species selection, as both intended mode of action (toxicodynamic) and toxicokinetic characteristics of the pesticide would determine species sensitivity to the pesticide effect. In particular, sensitivity to chemicals is known to be extremely variable among invertebrate taxa; these differences may concern the species-related biological susceptibility as well as ecological characteristics that are relevant to exposure (e.g. water column-dwelling vs. sediment-dwelling organisms, mobile epigeic vs. low-motility euedaphic species in soil, competitive relationship with other species for plants). Moreover, organisms from the same species may exhibit different susceptibility according to life stages, based on both biological and ecological characteristics (e.g. in insects, terrestrial larval stage vs. aerial adult stage). All such features make up the criteria for species (and life stage) selection in the model and should be clearly described and discussed.

Another critical component is the selection of the *environmental scenario* as the representation of the environmental context in which a model is run. A scenario is defined by a combination of abiotic, biotic and agronomic parameters. It is a conceptual and quantitative description of the environmental system relevant to the risk assessment, including the habitat (at relevant spatial and temporal scales) and the driving environmental variables including external stressors. The chosen scenario(s) should represent a realistic worst-case situation. The approach and decisions that lead to the selection of a certain scenario should be well documented and justified. Scenarios must be defined in relation to the specific protection goal and the level of conservatism defined in the problem definition. In all cases, the PPR Panel recommends that models consider several scenarios, including a control or baseline (without the chemical) and a toxic standard. The PPR Panel endorses the holistic approach to scenario development, which implies that the exposure scenario is developed simultaneously and integrated with the ecological scenario. This would typically be done for higher tiers: the advantage of this holistic procedure is that the environmental scenario applies directly to the endpoints used for assessment in a given area.

It is recognised that different model types are available to assess effects at individual, population or community level. Different models could be also combined together to form linked systems; in this case the individual models need to be considered as separate entities before considering the method by which they are integrated to form the whole model being used for a specific case.

Model development and use of an existing model should follow the modelling cycle or an equivalent scheme (e.g. pattern-oriented modelling). In any case, every step has to be fully documented. Testing and using the model will usually lead to reformulation of parts of the model, which is repeated until the model result is considered to be sufficiently reliable and robust for use in environmental risk assessment. The modelling cycle consists of the following steps:

i. *Problem definition.* The problem formulation sets the scene for the use of the model within the environmental risk assessment. It therefore needs to clearly explain how the modelling fits into the risk assessment and how it can be used to address protection goals.

ii. *Model formulation.* Based on the problem definition, a conceptual model is designed. The conceptual model provides a general and qualitative description of the system to be modelled. It characterises the environmental and biological processes and their interactions and interdependencies.

iii. *Model formalisation.* In this step, model variables and parameters are defined and linked together into mathematical equations or algorithms. The result of this step is called the formal model.

iv. *Model implementation.* In the following step the formal model is transferred into a computer model by implementing the model equations into computer code. The computer code should be verified to check if it correctly represents the conceptual and formal model.

v. *Model setup.* In this step, model parameter values are estimated and the computer model is combined with one or more environmental scenarios. The result is called the regulatory model. Note that the regulatory model includes both the computer model and the environmental scenarios. Model analysis, including sensitivity analysis, uncertainty analysis and comparison with observed data, is an essential part of the procedure to set up the regulatory model.

vi. *Using the model for regulatory purposes.* Before this is actually done, the Panel recommends that the risk assessor evaluates whether the model meets the requirements set in the problem definition. It is particularly important to evaluate whether the regulatory model output is relevant to the specific protection goals. The regulatory model can now be used to assess the risk from application of a specific pesticide.

vii. *Closing the cycle.* Feedback from risk assessor can lead to recommendations for model improvement.

Model evaluation by regulatory authorities should consider the following:

i. quality of the scientific data underlying the model;

ii. plausibility of the model's general behaviour;

iii. documentation substantiating each step of the modelling cycle;

iv. correspondence with available independent observations;

v. if the model is fit for the proposed regulatory purpose.

Therefore, the assessors should evaluate the model in a stepwise way according to the modelling cycle. A checklist, which should assist the evaluation, is provided in Appendix B. The uncertainties in the overall risk assessment should also be assessed (Appendix C).

In general, the following points deserve particular attention for the use of mechanistic effect models in pesticide risk assessment:

i. If the model can be used to extrapolate from one situation to another, the extent of these extrapolations and the resulting effect on the level of uncertainty should be clearly stated.

ii. The consequences of limitations in the datasets on the selection of the species, scenarios and exposure to model need to be clearly set out.

iii. Since sensitivity to chemicals and vulnerability of species are extremely variable, selection of species to be modelled should take account of the available datasets on relevant biological, ecological and toxicological factors.

iv. Published data can be of high value for model development and validation, although the quality of the data needs to be established.

v. When models are used for risk refinement, non-standard test species with increased ecological relevance compared with standard species used in toxicity testing should be considered.

vi. The risk assessment should include an evaluation of the level of conservatism.

vii. Increasing the toxicological database on relevant species would support a robust and reliable modelling. Special attention should be given to toxicity data concerning sub-lethal (e.g. behavioural, reproductive) effects in vulnerable life stages at realistic exposure conditions.

viii. Scenarios for modelling populations or communities should take into account all relevant and quantitatively important biotic and abiotic stressors for the area covered by the scenario, in order to provide a realistic assessment of the resilience of the selected entity.

ix. Scenarios should be constructed appropriate to the distinct regions in which the crop is grown; if distinct schemes of application are considered, then these will be analysed in distinct scenarios.

x. Close co-operation between exposure/fate modellers and eco(toxico)logists is required to implement a holistic approach to scenario development.

xi. A measure of parameter uncertainty (e.g. 95 % confidence or credibility interval) should be provided. If information is available on the usual range of the parameter value (i.e. minimum and maximum values), it should be provided.

The Panel recommends that models should be documented in a transparent way and include all information which is needed for the risk assessor to evaluate the model. All models should be accompanied by a user manual so that the evaluator can use the model, replicate the results and, for example, carry out a reality check.

After evaluation of a model and use of the model results (or rejection of model results), a feedback mechanism for model developers ought to be established, in order to improve the model for regulatory purposes. The following items could facilitate this process:

- Independent scientific reviews and/or articles in peer reviewed scientific journals.
- An online platform where risk assessors and modellers can exchange their experience with the model.
- A steering group consisting of modellers and risk assessors that prioritises model improvements. This group should base their decisions on practical considerations (usually from risk assessors), independent scientific reviews, peer reviewed papers, and of course on experience of the modeller himself.
- A version control system.
- Use of post-registration monitoring data/field studies to improve models.

From a regulatory point of view it would be beneficial to have a set of agreed models that can be used in standard risk assessments. One important step in this process is to compare different models used to address the same risk assessment question. To proceed from research to regulation and decision-making there is a need for:

i. scientifically sound, robust models, which have been thoroughly tested and make valid predictions;

ii. robust, user-friendly and freely available software, which can routinely be used by industry and regulators in all Member States;

iii. a well-defined set of scenarios which represents the full range of relevant species, ecosystems, and eco-regions in the European Union (EU) (DG SANCO, 2012).

Benefits of an agreed set of models would include more efficient use of resources in terms of reduction in cost and labour. Additionally, standardised scenarios would help to harmonise the risk assessment calculations and their interpretation.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

The PPR panel is tasked with the update of the Guidance Document on Terrestrial Ecotoxicology under mandate M-2009-0002. The Guidance Documents which are still in place were developed under Directive 91/414/EEC.

A public consultation on the existing Guidance Documents was held by EFSA in 2008 in order to collect input for the revision of the aquatic and terrestrial Guidance Documents. The following points were most often mentioned in the comments for updating the Guidance Documents:

i. Considerations of the revision of Annexes II and III of Directive 91/414/EEC,

ii. Consideration of the new Regulation (EC) 1107/2009

iii. Harmonisation with other directives and regulations (biocides, REACH)

iv. Clearly defined protection goals

v. Multiple exposure

vi. Inclusion of additional species in the risk assessment (e.g. amphibians, reptiles, bats, molluscs, ferns, mosses, lichens, butterflies, grasshoppers and moths)

vii. More guidance on statistical analysis

viii. Preference of EC_x over NOEC values in the risk assessment

ix. To consider all available information from workshops (EUFRAM, ESCORT, PERAS and other SETAC workshops)

x. Endocrine disruption

xi. Consideration of all routes of exposure

xii. Bee risk assessment

xiii. Non-target arthropods risk assessment

xiv. Soil organism risk assessment

The comments received in the stakeholder consultation will be consulted on again during the revision of the Guidance document.

A survey on the needs and priorities regarding Guidance Documents was conducted among Member States Authorities and a final list was compiled in the Pesticide Steering Committee meeting in November and December 2010.

The following topics were indicated as priorities for the update of the terrestrial Guidance Document:

i. Assessment of impacts on non-target organisms including the on-going behaviour

ii. Impact on biodiversity

iii. Impact on the ecosystem

iv. Effects on bees

v. Effects on amphibians and reptiles

vi. Linking exposure to effects and ecological recovery

vii. The use of field studies in the risk assessment and guidance for interpretation of field studies

viii. Revision of non-target arthropod risk assessment (ESCORT II)

ix. Guidance for risk assessment in greenhouses

x. Definitions of environmental hazard criteria (POP, PBT, vPvB) which will serve as a cut-off criteria according to the new regulation. Guidance on what studies, test conditions and endpoints should be used in determining whether the cut-off values have or have not been met. The Commission will consider the respective competencies of institutions regarding this topic and will check whether it takes the lead in this area.

xi. Definition of hazard criteria in relation to endocrine disruption and guidance on what studies, test conditions and endpoints should be used in determining whether the cut-off values have or have not been met. The Commission has the lead in developing these criteria. It is expected that EFSA will be consulted by the Commission on the final report in October 2011. The outcome of these activities should be incorporated in the Guidance Documents.

Generic questions which arose during the peer-review expert meetings should also be taken into consideration in the update of the guidance document. A compilation of general reports was provided by the pesticides unit. One of the points mentioned was that more detailed guidance is needed for the risk assessment of non-target plants (e.g. sensitivity of test species, use of species sensitivity distributions, exposure estimates).

Regulation (EC) 1107/2009 states that the use of plant protection products should have no unacceptable effects on the environment. The regulation lists in particular effects on non-target species, including their on-going behaviour and impact on biodiversity and the ecosystem.

The assessment of effects on on-going behaviour and biodiversity are not explicitly addressed under the existing Guidance Documents and appropriate risk assessment methodology needs to be developed.

The expertise needed in the different areas of terrestrial ecotoxicology ranges from in-soil biology, non-target arthropods, bees and other pollinating insects, terrestrial non-target plants, amphibians and reptiles and modelling approaches in the risk assessment.

This justifies the need to split the activity in several separate areas due to the complexity of the task and in order to make most efficient use of resources.

A separate question was received from the European Commission to develop a Guidance Document on the Risk Assessment of Plant Protection Products for bees and to deliver an opinion on the science behind the risk assessment Guidance. This question will be dealt with under mandate M-2011-0185.

TERMS OF REFERENCE AS PROVIDED BY EFSA

EFSA tasks the Pesticides Unit and the PPR Panel on the following activities taking into consideration Regulation (EC) 1107/2009, stakeholder comments and the recommendations and priorities identified by Member States: Good modelling practice in effect modelling with the following deliverable:

- Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products to be delivered by the PPR Panel by March 2014.

It is planned to provide in this opinion an overarching concept on how to document and evaluate effect models. It is not intended to focus in this opinion on specific model types or environmental compartments.

ASSESSMENT

1. Introduction

The development of updated guidance documents for pesticide risk assessment needs to take into account the progress in science. The use of modelling in environmental risk assessment is a rapidly developing area. A need was identified for a scientific opinion on how to evaluate models used in a regulatory context. The following text gives some background on the context in which the current opinion was developed.

Every plant protection product has to undergo a thorough risk assessment before its approval for use. The assessment covers human risk assessment as well as assessment of effects on the environment including fate and behaviour of a substance. Regulation (EC) No 1107/2009 states in Article 4.3 (Criteria for approval of plant protection products) that a plant protection product needs to meet the following requirements:

> "… *(e) it shall have no unacceptable effects on the environment, having particular regard to the following considerations where the scientific methods accepted by the Authority to assess such effects are available:*
>
> *(i) its fate and distribution in the environment, particularly contamination of surface waters, including estuarine and coastal waters, groundwater, air and soil taking into account locations distant from its use following long-range environmental transportation;*
>
> *(ii) its impact on non-target species, including on the ongoing behaviour of those species;*
>
> *(iii) its impact on biodiversity and the ecosystem.*"

The risk assessment process is usually conducted in a tiered approach starting with laboratory-based toxicity values and (reasonable) worst-case assumptions in the first tier and progressing to more complex and realistic higher tiers. Modelling can help to extrapolate from effects observed under laboratory conditions to field and landscape effects, and could be useful at different levels of the risk assessment such as:

i. help quantifying specific protection goals and setting of trigger values (e.g. definition of trigger values for bee risk assessment; see EFSA, 2013);

ii. refinement of the risk assessment (effects and/or exposure assessment e.g. body burden modelling, see EFSA, 2009);

iii. interpretation of higher tier studies (e.g. Topping et al., 2013);

iv. complementing and integrating information from higher tier studies;

v. extrapolate to untested situations or where field studies are not feasible.

The advantages of using models include, for example, reduction of animal testing, simulation of combinations of stressors and pesticide applications, and long-term population effects.

The risk assessment aims to assess whether a specific protection goal is met. The protection goal for vertebrates, for example, is prevention of individual death. The specific protection goals are linked to the provision of ecosystem services. Given that most of the services under the selected specific protection goals are performed by populations or groups of populations, development of appropriate population models for use in risk assessment is needed (EFSA PPR Panel, 2010). This need is also

highlighted by the EFSA guidance document on bird and mammals risk assessment (EFSA, 2009) and the recent aquatic risk assessment guidance (EFSA PPR Panel, 2013).

Models are used in many different ways; our focus here is on *mechanistic* models that predict *effects*, in contrast, for example, to statistical models. Mechanistic effect models represent biotic, abiotic and environmental processes and are therefore different from statistical models, although statistical models, such as non-parametric additive modelling of ecological time series, may be used to derive mechanistic models (Moe et al., 2005).

Mechanistic models can be useful at all tiers of current risk assessment schemes. A model type need not be tied to one particular tier in the risk assessment. Indeed, different types of models often have to be combined to extrapolate across levels of biological organisation in a mechanistic way.

An overview on where models and which type of models could typically be used at different tiers is given in Figure 1.

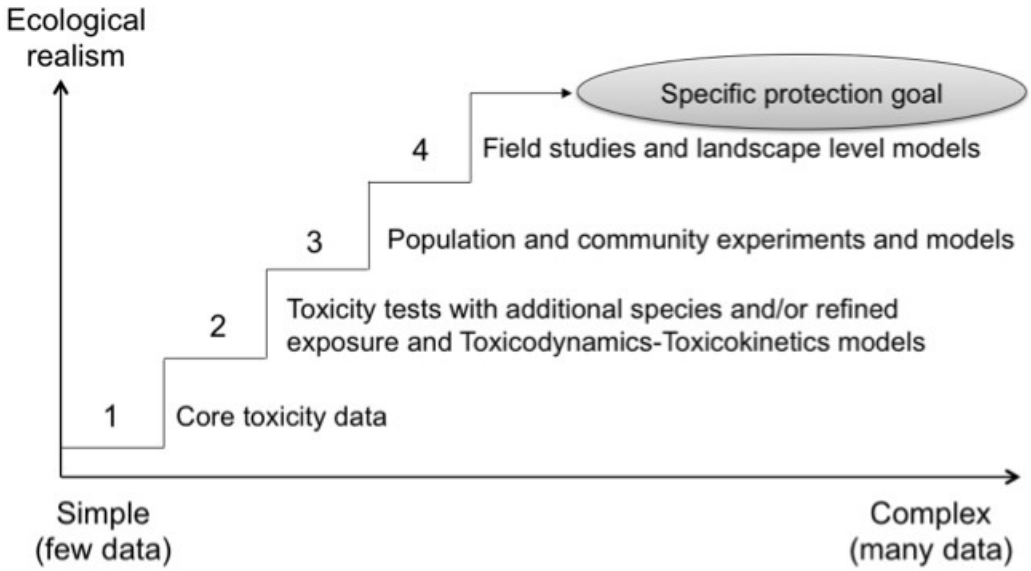


Figure 1: Schematic presentation of the tiered risk assessment approach and where different types of models could typically be used. The tier 1 and tier 2 effects assessments are based on single-species laboratory toxicity tests, but to better address risks of time-variable exposures, the tier 2 assessment may be complemented with toxicokinetic–toxicodynamic (TK-TD) models. Tier 3 (population- and community-level experiments and models) and tier 4 (field studies and landscape level models) may concern a combination of experimental data and modelling to assess population and/or community level responses (e.g. recovery, indirect effects) at relevant spatio-temporal scales. All models included in such a tiered approach need to be properly validated and fulfil required quality criteria

In environmental fate and behaviour assessment there is a well-established set of models and scenarios to estimate the concentrations in the environment while in the effects assessment there are very few models which are accepted and used for effect assessments (e.g. birds and mammals body burden modelling). An agreed set of effect models and scenarios could help to save resources (money and time) and may also contribute to harmonisation of risk assessments.

There is a growing number of available effect models which could potentially be of use in pesticide risk assessment (Galic et al., 2010), and stakeholders expect that effect models will play an important role and become integrated in the risk assessment process in the future (Hunka et al., 2013). Applicants have already used effect models in their dossiers submitted in the context of pesticide registration. However, in a few cases some Member States considered the risk as adequately addressed

with the model while others considered it inadequately addressed (Carpentier, 2013).[4] In the EU peer review for active substance registration, none of the 17 models submitted in dossiers were considered to address the risk sufficiently (Streissl, 2013).[4] This is partly due to a lack of guidance on how to develop, document and evaluate models and how to use the model results in the regulatory risk assessment.

The current document is intended to facilitate the use of effect models in the risk assessment. The document addresses, in the first instance, risk assessors who have to evaluate models submitted in regulatory dossiers. However, the document also includes elements for consideration by risk assessors and modellers in companies and consultancies, e.g. recommendations on how to build a model fit for regulatory purpose and the information which is needed to evaluate the model.

Model development is usually seen as a cycle in which the model is improved until the resulting model is sufficiently robust (the modelling cycle). The structure of this document follows the modelling cycle as described in section 2. The quality of available data to build the model is of great importance. Criteria for data quality are outlined in section 3. Problem formulation and different steps in model development can be found in sections 4, 6, 7 and 8 (see also Appendix A describing pattern-oriented modelling (POM)). Section 9 describes the main elements for model analysis. Guidance on evaluation of models is given in section 10 and a checklist to support the evaluation is included in Appendix B. Finally, section 11 includes recommendations on the documentation of a model.

2. The modelling cycle

2.1. Introduction

This section gives an overview of the steps involved in developing a model for use in environmental risk assessment (ERA). Models may be used by applicants in producing a dossier for a plant protection product (PPP), either to supplement laboratory and field studies or, in some cases, to replace field studies. Replacement of field studies may be acceptable, for example, where the aim is to predict long-term effects of PPP use or where a wider range of input variables or of scenarios than can feasibly be examined experimentally needs to be investigated. A model that is used to extrapolate beyond experimental data must be formulated correctly (i.e. based on knowledge and data) and used appropriately, and the procedures needed to achieve this are addressed in this opinion.

Model development is often seen as a cycle, because testing and use of the model will usually lead to reformulation of part or all of the model, which is repeated until the model results are considered to be sufficiently reliable and robust for use in ERA (e.g. Refsgaard and Henriksen, 2004; NRC, 2007; Schmolke et al., 2010; Perrenet and Zwaneveld, 2012; Augusiak et al., 2013). The modelling cycle consists of the following steps (Figure 2):

i. problem definition;

ii. assembly of hypotheses that constitute the conceptual model of the system in question (model formulation);

iii. translation of the conceptual model via equations into a formal model (model formalisation);

iv. conversion of the formal model into a computer model (model implementation);

v. estimation of values of model parameters and combination of the computer model with one or more environmental scenarios (setup of the regulatory model);

vi. model analysis including sensitivity analysis, uncertainty analysis and evaluation of the model by comparing the model with observed data, preferably under different scenarios;

[4] Presentation at the MODELINK workshop, 22-25 April 2013 Monschau Germany.

vii. use of the regulatory model in environmental risk assessment.

Evaluation is a key element in every step (i.e. steps ii–vi) of the modelling process and is therefore a continuous process (Figure 3). It should be carried out by the model developer to iteratively improve the model until predefined performance criteria are reached. It should also be carried out by risk assessors working in authorities, companies and consultancies to check whether the model is fit for regulatory purpose (see section 10 for guidance on how to practically conduct this evaluation).

The remainder of this section gives a brief summary of the six steps in the modelling cycle together with the required evaluation activities; a more detailed description is found in later sections.

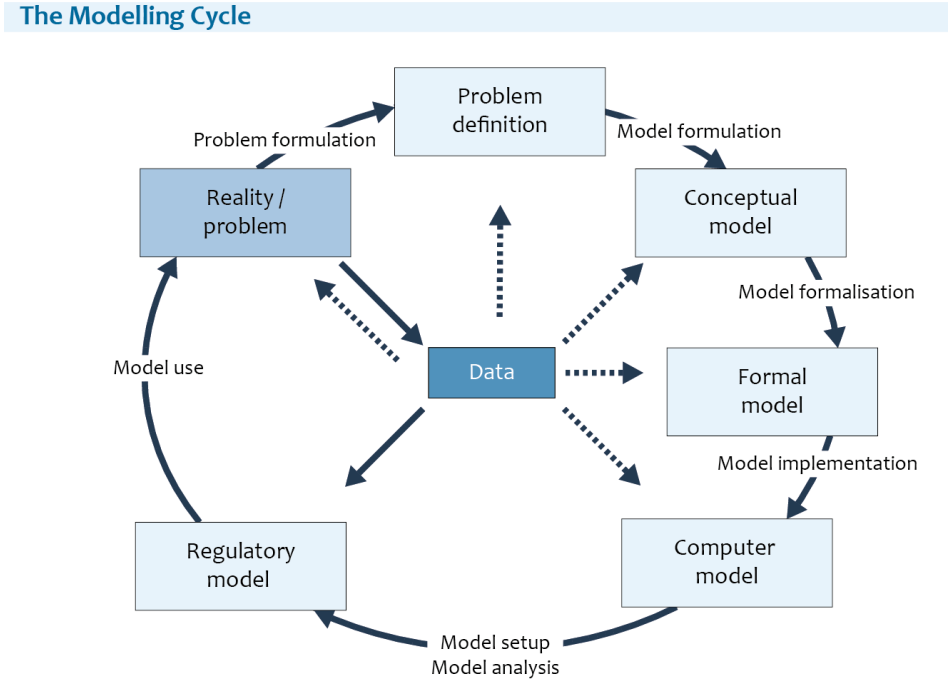


Figure 2: Illustration of the modelling cycle. Each step in developing the model is informed by data. The endpoint or outcome of the process is use of the regulatory model in environmental risk assessment, but there may be several iterations of the cycle before this endpoint is reached. Adapted from PBL (2011). See the glossary for a definition of the regulatory model.

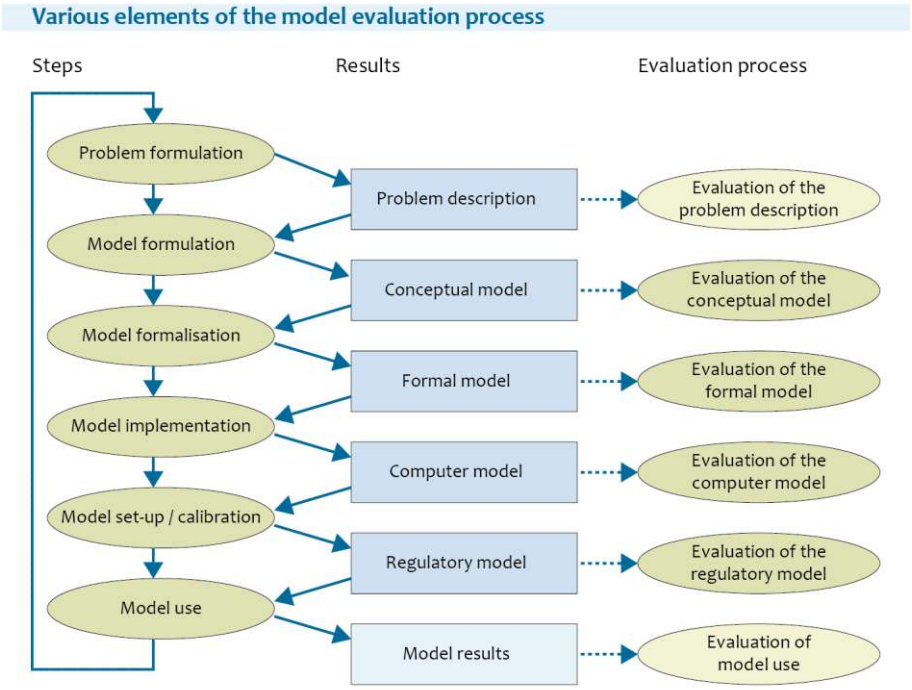


Figure 3: Alternative representation of the modelling cycle with evaluation activities. Model evaluation has to be carried out by both model developers and risk assessors. Adapted from PBL (2011)

2.2. Data analysis

Data are of prime importance at all stages of the modelling cycle (Figure 2). Data and knowledge are particularly crucial in writing the problem definition, in developing the conceptual model, in formalising the model, in estimating the parameters of the model and in evaluating how successful the model is in reproducing general data patterns and in fitting quantitative experimental data. The modeller will therefore check on the quality and availability of data at the start of model development. For further suggestions as to how this should be done, see section 3.

2.3. Problem definition

The modelling process starts with the definition of the problem. The problem formulation sets the scene for the use of the model within the environmental risk assessment. It therefore needs to clearly explain how the modelling fits into the risk assessment and how it can be used to address protection goals. The problem formulation therefore needs to address (see section 4 for further details):

i. the available knowledge and data relevant to the risk assessment question;

ii. the context in which the model will be used (e.g. lower or higher tiers of the risk assessment);

iii. the question(s) that has to be answered with the model;

iv. the outputs required to answer these question(s) including the required performance criteria for the regulatory model;

v. the species to be modelled;

vi. requirements for the environmental scenarios to be used in the environmental risk assessment.

Most of the above points are determined by the specific protection goal (SPG) for the taxa or functional groups of concern (the so-called key drivers). It is recommended that the process outlined in EFSA (2010) be used to define these SPGs, because this process facilitates the communication with risk managers who are responsible for setting the overall level of protection (Figure 4). In this process, the SPGs are defined in six dimensions: biological entity, attribute, magnitude of effect, temporal and spatial scale of the effect, and the degree of certainty that the specified level of effect will not be exceeded (see section 4 for details).

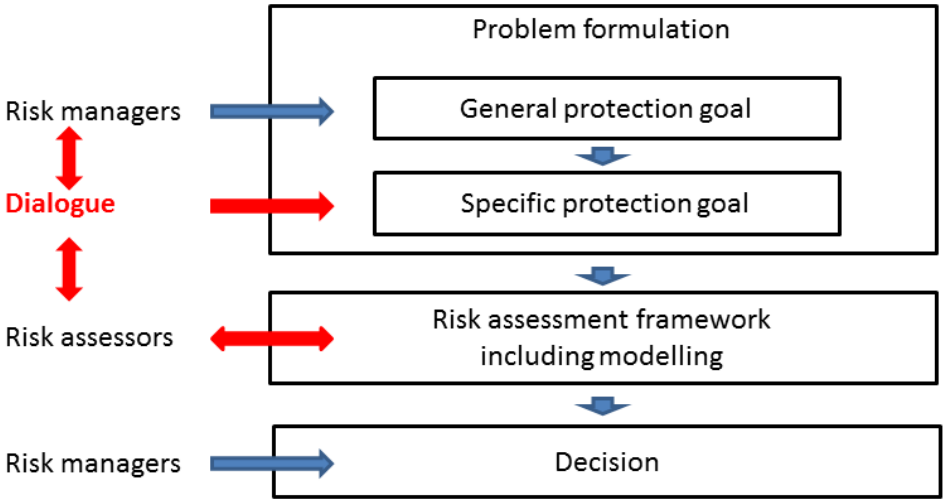


Figure 4: The risk assessment framework and its relationship to problem formulation, protection goals, and risk management in the process of developing specific protection goals. Please notice that this figure highlights the dialogue between risk assessors and risk managers (based on EFSA PPR Panel, 2010)

2.4. Model formulation

Based on the problem definition, the modeller designs a conceptual model ("model formulation"). The conceptual model provides a general and qualitative description of the system to be modelled. It provides insight into the environmental and biological processes and their interactions and interdependencies. Conceptual models are often summarised in diagrams (e.g. Figure 8). Model formulation consists of the following three steps (see section 6 for details):

i. In the first step, the system is characterised. The modeller identifies (i) the main components of the system (e.g. individual animals); (ii) the main processes that govern the functioning and behaviour of the components (e.g. food assimilation, growth, or reproduction); and (iii) the internal and external factors that modulate the functioning of the components (e.g. development stage, season and food availability).

ii. The second step is to decide how to model the harmful effects of a chemical. This is likely to be in the form of a dose–response curve that relates the dose of a chemical taken in by an individual organism to its response.

iii. The third step is to describe how the various components and processes are linked (selection of the model type). A variety of model types exist, which are described in section 6.

2.5. Model formalisation

Once a conceptual model has been formulated, it is necessary to construct the formal model that adequately represents it so that it can be used. Model variables and parameters have to be defined and linked together into model equations and algorithms. This step is called model formalisation, and the result of this step is the formal model, which may be a set of mathematical equations or a specification of the algorithms which will later be translated into a computer program. Formal models consist of the following elements, which all need to be defined and described by the modeller (see section 7.1 for details):

i. model variables, including state variables that describe the state of the system (e.g. population size) and forcing variables describing the scenario under which the model is run (e.g. precipitation, temperature);

ii. model parameters, which are terms in the model that are fixed when conducting a model run (e.g. maximum growth rate at reference temperature);

iii. equation and/or algorithms, which describe the dynamics of the processes included in the model.

2.6. Model implementation

The next step is to convert the formal model into a computer model. This step is referred to as model implementation. The computer model is written in a computer language. Checking the computer model for errors, bugs and inconsistencies is a crucial element of the evaluation process. This process, often referred to as verification, must further identify whether the computer model performs as indicated by the description of the conceptual and formal models. Note that model verification does not check the realism of the model; this is done at later stages of the evaluation process.

2.7. Model setup

2.7.1. Parameter estimation and scenario development

In the following step of the modelling cycle, model parameter values are estimated and the computer model is combined with one or more environmental scenarios. This so-obtained package is referred to as the regulatory model, which consists of the following components: (i) the computer model, (ii) programs for pre- and post-processing (often made available in the form of graphical user interfaces), (iii) model parameters and (iv) the environmental scenarios. By combining the computer model with one or more scenarios, the model will address a certain specific protection goal (see *Section 8.2*). Note that the model can be run only when case-specific data, such as pesticide parameters (relevant to either effects and exposure) and pesticide application regimes, are also input to the model (Figure 5).

An environmental scenario therefore represents the environmental context in which a model is run. An environmental scenario is defined as a representative combination of abiotic, biotic and agronomic parameters to be used in modelling (EFSA PPR Panel, 2010). Scenarios must be defined so that they comply with the specific protection goal. When used for regulatory purposes, they must further conform to the level of conservatism defined in the problem definition. The level of conservatism is determined by the spatio-temporal dimensions of the protection goal, and therefore these must be rigorously assessed in a dialogue between the modeller and the risk assessor (see section 4 for details).

Values of parameters in the environmental scenario or in the computer code have to be estimated before using the model for further simulations. Parameters can be obtained from the literature, from measurements, or estimated by calibration. Calibration is a special method for parameter estimation; adjusting the values of parameters until the model output fits existing data patterns. Further considerations on parameter estimation and calibration are given in section 8.

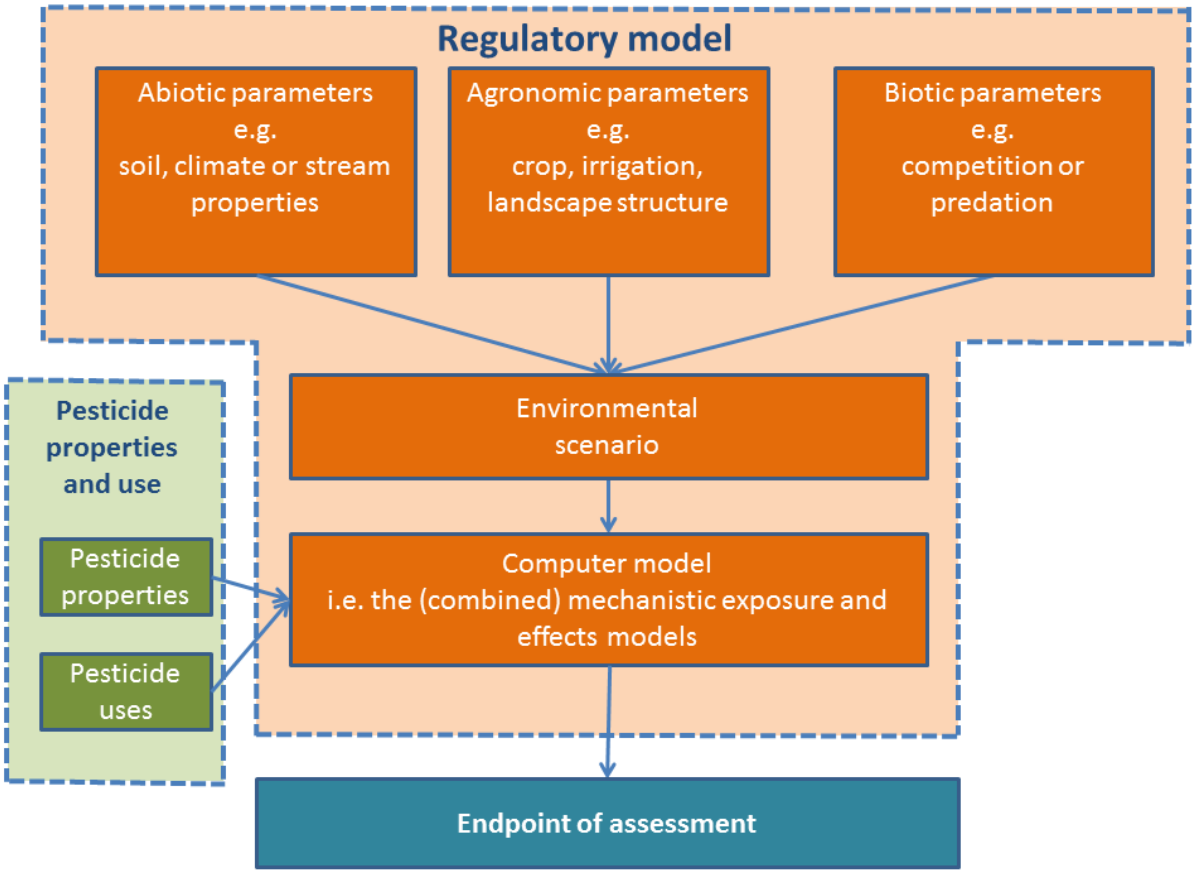


Figure 5: Schematic representation of a regulatory model (orange boxes) which consists of an environmental scenario and the combined exposure/effect model. Combined with information on pesticide properties and uses (green boxes), the regulatory model delivers an output, i.e. the endpoint of the assessment

2.7.2. Model analysis

Model analysis includes sensitivity analysis, uncertainty analysis and evaluation of the model by comparison of the model output with observed data (section 9). Model analysis is a key element of the procedure to set up the model, as it provides insight into the behaviour and performance of the model.

Sensitivity analysis identifies sub-sets of parameters that have relatively large effects on the model outputs. We thereby learn which processes are most important for further consideration. Uncertainty analysis aims to identify how uncertain the model output is (often expressed as confidence intervals). This is important information for both the risk assessor and the risk manager, who may consider this when making decisions, particularly in borderline cases.

The performance of the model is usually evaluated by comparing relevant model outputs with measurements. No model can perfectly describe the system under consideration, but model outputs should provide an adequate match to relevant data patterns. Which match is considered adequate should be predefined in the problem definition. Note that it may require several iterations of the modelling cycle to achieve this predefined performance indicator (see section 8 for details).

2.8. Using the model for regulatory purposes

In the next step of the modelling cycle, the model is used for regulatory purposes. Before this is actually done, the risk assessor should evaluate whether the model conforms to the requirements set in the problem definition (see section 2.3 for details). An important item to evaluate is whether the regulatory model conforms to the SPGs for each group of exposed organisms. The level of conservatism is defined in the SPG and reflected in the scenario but the model should be realistic. The level of conservatism of the regulatory model is primarily determined by the environmental scenarios (section 8.2), and therefore evaluation of these scenarios is crucial (see section 10.5.1 for details).

When using the model, the risk assessor has to input the pesticide properties and the application pattern and/or regime of these pesticides. How this is carried out must be described in a user manual, which should be provided by the modeller (section 10.7). Pesticide properties can be obtained in many different ways (laboratory experiments, field experiments, etc.), and the way in which these properties are derived determines the outcome of a regulatory assessment to a large extent (see, for example, Boesten, 2000; Tiktak, 2000). The methods used should therefore be documented and, whenever possible, harmonised procedures for derivation of these properties should be used.

Risk assessors should finally decide whether or not an authorisation of a PPP could be granted. Regulation (EC) No 546/2011 states that no authorisation shall be granted unless it is "*clearly established that no unacceptable impact occurs*". The term "clearly established" implies a requirement for some degree of certainty. First-tier assessments use standardised scenarios and decision rules which are designed and assumed to provide an appropriate degree of certainty. Assessments based on ecological modelling are generally not standardised, and so the degree of certainty they provide has to be evaluated case by case. Section 12 introduces a tabular approach for qualitative evaluation of uncertainties, which is intended to help in the decision-making.

2.9. Closing the cycle: recommendations for model improvement

Once a model has been used for regulatory purposes, feedback from the risk assessor can lead to recommendations for model improvement. If well organised, this step provides an important feedback mechanism between the modeller and the risk assessor. The following items could facilitate this process:

i. The presence of independent scientific reviews and/or articles in peer-reviewed scientific journals.

ii. The presence of an online platform where risk assessors and modellers can exchange their experience with the model.

iii. The presence of a steering group consisting of modellers and risk assessors that prioritises model improvements. This group should base their decisions on practical considerations (usually from risk assessors), independent scientific reviews and, of course, on the experience of the modeller him/herself.

iv. The presence of a version control system.

3. Data analysis

A major theme running throughout this opinion is the importance of data. Data relevant to the risk assessment question are used at all stages of modelling (Figure 1). They are used in defining the problem to be addressed (*Chapter 4*), at the stages of species and scenario selection (Chapter 5 and section 8.3), and they guide model design and formulation (sections 7.1 and 7.2). The availability and quality of data are of crucial importance during parameter estimation, uncertainty analysis and validation (sections 8 and 9). Thus, the availability and quality of data are of the utmost importance at all stages of modelling. If the data are not sufficient, the regulatory model may not be considered acceptable for risk assessment.

In considering the quality of data, some general questions should be considered:

i. Is sufficient experimental detail given so that the context in which data were obtained is clear, and the experiment could be repeated to check the reproducibility of the results?

ii. Is the uncertainty of data described (e.g. with standard deviations or confidence intervals)? Where possible, these should distinguish between measurement error and biological or environmental variation. Knowledge of the confidence interval allows assessment of the precision of measurement and aids uncertainty analyses.

iii. Is full documentation of data sources given, e.g. citation of references with page numbers?

3.1. General considerations

The section below provides some general considerations when assessing the quality of the data used.

Extrapolation of datasets

Data used for parameter estimation should be relevant to the species and scenario under concern. However, lack of data on the modelled scale (e.g. on the landscape scale) or for the species of concern is a frequent issue in the context of ecological risk assessment. Where data are lacking, it may be permissible to use data from different species or environmental scenarios from the one under consideration. For example, the modeller may look to relationships detailing how life history and physiological traits and rates scale with body mass and temperature. Such relationships are catalogued together with some individual parameter values in Peters (1983), Brown et al. (2004) and Sibly et al. (2012, 2013). When such an approach is used, it should be scientifically justified by the modeller, and the uncertainty introduced by using such data and extrapolation should be characterised.

If the model is calibrated using data representing a particular scenario (e.g. a limited temperature range, a specified food level), transfer to other scenarios outside the range of the calibration scenario may be problematic. In particular, if many parameters represent aggregated processes, applying the model under other scenarios would normally call for recalibration using data from such scenarios.

Inconsistency of datasets

It will not always be the case that datasets are consistent with each other. Where inconsistencies exist, these should be described and the choice of the datasets used in modelling should be justified (e.g. in terms of relevance to the modelled species and/or scenarios; level of detail provided documenting experimental conditions; precision of measurements as assessed by confidence intervals). The datasets should not be chosen with the aim of achieving a given outcome of the risk assessment: this would clearly be poor modelling practice.

Documentation of data

Full documentation of data sources should be given throughout (e.g. citation of references with page numbers). For parameter values taken from the literature, the exact data source must be given, information about the parameter uncertainty should be provided, if available, and potential sources of uncertainty should be mentioned (e.g. differences in the experimental setup or species of interest compared with the scenario/species modelled). In addition, the modeller should provide details on:

i. Raw data used to identify parameter values and their origin (literature, experiments). If literature data were used, citing the paper is not sufficient and detailed information should be provided on the data. If values were directly taken from the literature, references to page numbers should be included. If values are deduced by re-scaling, estimations or extrapolations from the publications, this should be documented in detail.

ii. The conditions under which values reported in the literature/experiment were observed (region, environmental conditions, season, etc.). Differences to the situation to be modelled should be discussed in the uncertainty analysis.

3.2. Considerations on specific data types

Several different data types are needed in modelling. This section gives some specific considerations for each of these data types.

Qualitative data

Examples of qualitative patterns include emergent properties such as density dependence and population cycles. Qualitative data patterns may be used at all stages of model development and evaluation, but in general cannot be used for parameter estimation or other quantitative purposes. Such qualitative patterns may be used for evaluation of model outputs as part of the process that has been termed pattern-oriented modelling (Grimm et al., 2005; Grimm and Railsback, 2011; Railsback and Grimm, 2011; Topping et al., 2012; see Appendix A). However, where quantitative descriptions of data patterns exist, they should generally be used in preference to qualitative descriptions.

Toxicological data

Mechanistic effect models generally require specification of dose–response and/or concentration–response curves detailing how life history traits are affected by the exposure to the chemical. The life history traits may be, for example, daily mortality rates of juveniles and adults, reproduction rates or somatic growth rates, but also traits that indirectly affect life history, including ingestion rate or respiration. Toxicological data may not be available for the study species or for taxonomically related species. In this case, some clear and consistent method of extrapolation between the laboratory species and the study species is needed.

Ecological data

Where available, the modeller will generally use parameter values for the study species in the study areas or areas comparable to the study areas. Measurements may be available for the behaviour, physiology and life histories of individuals and populations maintained in the laboratory or living naturally in the field. Where relevant individual and/or population data exist, they should be described. It is expected that all relevant and consistent data will be used in the modelling cycle; if some data are not used, then the reason should be given.

4. Problem formulation

The problem formulation sets the scene for the use of the model within the risk assessment. It therefore needs to clearly explain how the modelling fits into the risk assessment and how it can be used to address protection goals.

The problem formulation needs to address the context in which the model will be used, to specify the question(s) that should be answered with the model, the outputs required to answer the question(s), the domain of applicability of the model, including the extent of acceptable extrapolations, and the availability of knowledge.

4.1. Regulatory context in which the model will be used

This section should include information about:

i. the reason the model is being used, for example lower tier data for a specific group did not identify an acceptable risk;

ii. which areas of the risk assessment are addressed by the model, for example the risk for earthworm populations at field scale over time.

A basic set of toxicity data will be available for all pesticides (see data requirements for Regulation (EC) No 1107/2009, laid down in Commission Regulation (EU) 283/2013 and in Commission Regulation (EU) 284/2013). In addition, higher tier toxicity or exposure data may also be available. These data are used to perform a risk assessment, which will identify whether or not the risk is acceptable and whether additional information is needed. This should be done for each group of organisms to be protected. If higher tier data are available, the major groups may have been further sub-divided with an acceptable risk being identified for some species but not for others. Before modelling is undertaken, these existing data should be carefully reviewed to identify which issues have not been resolved with the data available. Modelling can be considered once these areas and the implications for higher tier assessment of the species/group to be assessed have been identified. The type of exposure concentration and how this will be incorporated into the model should also be considered at this stage.

4.2. Specification of the question(s) that should be answered with the model

Mechanistic effect models allow for great flexibility in exploring the potential effects of a chemical on an individual, a population or a community. Consequently, it is important that, before modelling is undertaken, the questions that need to be answered are clearly defined. It is also important to target the type of modelling undertaken to the specific question(s) that need to be addressed in the risk assessment.

The starting point will always be the SPG for the taxa or functional groups of concern, so-called key drivers (i.e. microbes, algae, non-target plants, aquatic invertebrates, terrestrial invertebrates, bees and vertebrates). These SPGs should define the ecological entity that is to be protected (individuals, (meta)populations, functional groups or ecosystems), the attribute(s) or characteristic(s) of that entity that must be protected (behaviour, survival/growth, abundance/biomass, processes, biodiversity), the magnitude of effect that can be tolerated for the attributes (biological scale), the temporal scale of effect (e.g. the maximum time over which single or repeated exposure/effect level(s) can be tolerated), the spatial scale of the effect (e.g. the distance from the sites of application where the exposures and critical effect level to be tolerated) and the degree of certainty that the specified level of effect will not be exceeded (e.g. predict with a high certainty the magnitude of effects on population abundance at the edge of field for different exposure scenarios (Figure 6)).

Figure 6: Example of development of a specific protection goal definition. The defined protection goal aims to prevent (positive or negative) effects to the right of any of the circled points

Specific protection goals should be defined in the guidance documents for each group of organisms as they should apply to all pesticides. Therefore, this step does not need to be done each time a risk assessment is conducted, but the relevant protection goal needs to be referred to. Specific protection goals for each area can be used to guide the selection of model type and output.

The combination of the data and risk assessment will guide the choice of specific questions to be answered by the modelling. For example, first-tier bee risk assessment for pesticide A gives a hazard quotient above the trigger value when applied once a year to oilseed rape at 150 g a.s./ha. It is thus concluded that the risk is not acceptable from the tier 1 evaluation and therefore additional information is required for refining the risk for bees, but only lower tier toxicity data are available. The specific protection goal for bees is:

i. Ecological entity: colony.

ii. Attribute: survival and development of colonies, effects on larvae and bee behaviour (i.e. foraging behaviour) and abundance/biomass and reproduction. The viability of colonies is related to the colony strength, defined as colony size.

iii. Effect magnitude and temporal scale: should not exceed a 7% reduction in colony size; Foragers mortality should not be increased compared with controls by a factor of 1.5 for six days or a factor of 2 for three days or a factor of 3 for two days.

iv. The spatial scale is colonies at the edge of field.

v. Degree of certainty: a 90th percentile for exposure is proposed, but it is acknowledged that risk managers might wish to change this.

The question to be addressed by modelling could be: if pesticide A is applied once a year to oilseed rape at 150 g a.s./ha, what will be the effect on overall colony size of honey bees sited at the edge of the field and what will be the forager mortality compared with controls two, three and six days following application?

4.3. Specification of necessary model outputs in relation to protection goals

4.3.1. Using specific protection goals to define type of outputs needed

Another important aspect of problem formulation is to consider what type of output is obtained and how it is suited to the protection goal (to inform the decision). It is important that the protection goal is quantified on the spatial and temporal scale. If the protection goal allows for a certain magnitude (e.g. negligible to small) and duration (e.g. days to weeks) of effects, it might be appropriate to model recovery with an output that identifies both magnitude and duration of effects that can be compared directly with the protection goal.

For example, for vertebrates, the protection goal could specify "no individual mortality" and a toxicokinetic-toxicodynamic TK-TD model might be used. The output could then be the likelihood of an individual dying following the predicted exposure.

For bees, the protection goal is defined in terms of colony size and forager mortality over time; therefore, the output could be colony size and daily forager mortality in the control and treated colonies.

For invertebrates, particularly in-field organisms, some population effects may be acceptable if ecological recovery takes place within a defined time period. However, for off-field populations, according to the new aquatic guidance document (EFSA PPR Panel, 2013), recovery might be a reasonable option only for surface waters adjacent to crops with a limited PPP input. This is because there is more uncertainty whether recovery can be achieved when assessing risk for individual PPPs that are used in intensive crop protection programmes, characterised by simultaneous or repeated use of different PPPs. However, mechanistic effect models may be used even within the range of ecological threshold concentration to explore effects and recovery (e.g. for low but existing effects, for repeated cumulating effects, etc.), and thus to address the specific protection goals.

4.3.2. Using model outputs to assist defining specific protection goals

Mechanistic effect models have so far been discussed in the context of a risk assessment in which they are used to address a specific protection goal. They could also be used to help defining a specific protection goal, as was done in the EFSA guidance on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees).

In the aquatic guidance document (EFSA, 2013) the following specific protection goals are proposed, e.g. for aquatic invertebrates:

i. The ecological threshold option (ETO) allows only negligible effects.

ii. The ecological recovery option (ERO) allows :

   a. small effects for a few months;

   b. medium effects for a few weeks;

   c. large effects for a few days on the abundance and/or biomass of vulnerable populations of invertebrates, as long as their reduction does not result in more persistent indirect effects.

In the Aquatic Guidance Document, and when deriving a regulatory acceptable concentration (RAC) on basis of the Ecologically Recovery Option, the maximum allowed effect period for the most sensitive and vulnerable population in a micro-/mesocosm test is set at eight weeks. No further guidance on the limits/thresholds is given for a realistic edge-of-field scenario (i.e. the magnitudes of negligible, small, medium and large effects are not indicated). Modelling could help to define these thresholds better.

The use of different models under different conditions is required in order to gain experience to translate specific protection goals into model outputs appropriately and so that appropriate criteria can be derived.

The definition of magnitude of effects (i.e. negligible, small, medium and large effects) may be revised as model simulations may help to make predictions on a multi-year basis or a larger spatial scale, e.g. negligible effects repeated over the years may lead to some non-negligible effects and significant local effects may have little impact at the meta-population level.

4.4. Specification of the domain of applicability of the model

Modelling offers the opportunity to go beyond the settings that have been tested in experiments or observed in the field. Hence, it is possible to investigate additional species, different exposure regimes and different spatial and temporal scales. A model can obviously be developed only if sufficient information is available for the setting under consideration. Nevertheless, a sound mechanistic foundation allows for extrapolation and helps to combine different processes that have only been observed separately. For example, the results from tests on different individual life cycle traits together with biological knowledge on the species under consideration can be interconnected to model effects at population level. Likewise, modelling can be used to extrapolate from one species to another (Luttik et al., 2005). For example, if a model was built for a well-investigated species (e.g. *Daphnia magna*), it might be possible to extrapolate to other, related species by using the respective species specific parameter values.

Even if species specific information is scarce, "educated guesses" might be helpful. For instance, the effect on ecologically more vulnerable species can be investigated by assigning smaller values for reproductive capacity or decreased detoxification, depending on the model type used. These values should be within biologically reasonable ranges. The modelled species might not represent real species in every detail but could represent a worst-case evaluation. If such an approach is adopted, difficulties

with model validation are likely to occur (see section 5.2 for discussion about choosing a real or worst-case surrogate species).

However, care must be taken when drawing conclusions from the modelling results. The model can yield meaningful results only for the level for which it was developed and can take into account only take those processes that are included. For example, a population-level model cannot help to identify indirect effects on interacting species; a model without spatial dimension cannot be used to investigate landscape-level effects; a model that was developed from experimental data with food *ad libitum* cannot be used for predictions under food stress, if the relationship to food level is not included somehow. Furthermore, the applicability of the model should be carefully considered if the range of conditions for the external driving variables (e.g. temperature, food level) goes beyond the ones in the experiment for parameter estimation. For example, the growth rate of bacteria might be linearly related to temperature over a range of values but drop at a certain threshold; if this behaviour was not included in the model formulation and temperature goes beyond this threshold, the model predictions would be wrong. If the model is used to extrapolate from one situation to another, the extent of these extrapolations and the resulting effect on the level of uncertainty should be clearly stated.

4.5. Context for the modelling conducted

Once the question to be answered with modelling and type of output needed have been determined, the species, scenarios and exposure all need to be considered

The relevant species to model must be chosen. For example, if the first-tier risk assessment using *Daphnia* to represent all aquatic invertebrates does not show an acceptable risk, assessing the risk to *Daphnia* populations is unlikely to address adequately the risk to all aquatic invertebrates, since, although *Daphnia* is generally sensitive, it has a short life cycle, so populations recover rapidly. The refined risk assessment must cover the appropriate group of aquatic invertebrates. This subject is covered in more detail in section 5.

The relevant scenarios (including spatial and temporal scales) also need to be identified. If a landscape model is being used, the most appropriate type of landscape needs to be identified and explained. The exposure that will represent reasonable worst-case situations needs to be identified and modelled. This is discussed further in section 8.3.

4.6. Assessment of available knowledge and data

The available datasets for model development and evaluation need to be considered during the problem formulation stage. Models can be developed only if sufficient data are available or there is potential to generate the required data. Since different types of models require different types and quantities of data, the availability of information could limit the choice of model type, species modelled or scenarios tested. The consequences of such limitations need to be clearly addressed.

The data used for model development and validation can come from a wide range of sources, not only from standard test guidelines. All data can be useful, although the quality of the data needs to be established. Ideally, toxicity data should come from standardised and higher tier tests following agreed test guidelines used for regulatory risk assessment. For modelled species, it is possible that no toxicity values are available from studies carried out in accordance with good laboratory practice (GLP) or adhering to standard guidelines. These studies will need to be carefully considered by the model developer and the evaluator.

Examples of the information that a model might require include individual growth rates, reproduction rates, energy requirements and ingestion rates. These, in turn, might depend on other information, such as distribution in space and time, distribution in territories, age structure, time series and competition/no competition (density dependence). For example, ingestion rates vary with food density.

5. Selection of species to be modelled

As part of the problem definition, one or more species need to be selected for modelling. This section first provides some general considerations. A practical stepwise approach to species selection is then presented (section 5.3). Finally, some specific recommendations are given for vertebrates and invertebrates.

5.1. General considerations on species choice

The first step in modelling in the context of pesticide risk assessment is to select one or more species for modelling. When considering modelling at the individual (for example TK-TD models) or population level, a specific species has to be chosen. When considering modelling at community or ecosystem level, it will also be necessary to consider which species it is relevant to include in the model as it will not be possible to include all species. Thus, it may be necessary to model more than one species to cover the range of species present in natural systems. This section aims to give guidance about the sort of aspects that should be considered within the context of pesticide risk assessment.

Selecting species for TK-TD models is likely to be driven by knowledge of the pesticide and exposure and available data. Therefore, when TK-TD models are intended for use at lower tiers (e.g. refinement of exposure), standard species can be modelled. When data on several standard test species are available (e.g. from fish acute toxicity tests), the most sensitive species, as determined in toxicity tests, could be used. If TK-TD models are used at higher tiers, non-standard test species with higher ecological relevance than standard species should preferably be used. Selecting species for population, community and ecosystem modelling will require a good knowledge of the species in its environmental context; since the outputs are used to refine the risk identified at a lower tier, this implies increased realism and complexity of the system.

Choosing the most suitable species to model involves considering a wide range of possible factors including:

i. the specific protection goal;

ii. information from toxicity data already collected;

iii. where and when exposure will occur;

iv. knowledge of candidate species:

   a. life history characteristics
   b. behaviour
   c. exposure during contamination (e.g. species present in the field during application)
   d. sensitivity to the chemical of concern (acute tests)
   e. sensitivity of life stages to the chemical of concern (chronic tests)
   f.sensitivity to co-stressors (environmental and ecological stressors);

v. type of effect (mortality, development, behaviour);

vi. acute vs. delayed effects, carryover effects;

vii. availability or reproducibility of data for validation.

If one species is selected to model the risk for a range of species, it is important that it is identified as a vulnerable species.

5.2. Real versus "worst-case" species

In theory, a model could be developed for either a real species or a species that realistically combines traits from a number of relevant vulnerable species to generate a realistic worst case and therefore cover a range of species that occur across Europe. For example, if the modelling needs to cover two species, one species living longer than the other but having fewer offspring, it might not be clear beforehand which of the two populations would be most affected by a specific pesticide. One option is to model both species. An alternative would be to model a vulnerable "made up" species that could have a shorter life and lower number of offspring and would therefore cover both situations. The "made up" species should correspond to a realistic worst case: there is no point in modelling a species that could not survive in nature owing to an unrealistic combination of traits. The risk assessment should include an evaluation of the level of conservatism with this approach (see Appendix C).

The main requirement for models used for regulatory decisions is that there is sufficient confidence in the outcomes of the model to use them for decision-making. This involves an evaluation and validation of the model against real data. This is feasible for a real species for which (field) data are available and published (although difficult, especially considering the variability in the field) but is impossible for a "made up" vulnerable species. Therefore, an initial model should be for a real species which is subject to evaluation and validation.

Once a model has been developed, well tested and validated with data, it might be possible to adapt it to cover other species with similar characteristics in terms of life traits. If the model is to be used for higher tiers, then special care is needed to evaluate if the other species covered by the model are also characterised by a similar environmental context. This could then lead to a suite of well-tested generic models that could be used routinely for lower tier risk assessment as the species modelled (either real or "made up") are vulnerable to a wide range of pesticides. Any limitations with the modelled species (e.g. the species and model are not vulnerable to certain modes of action such as insect growth regulators) would be identified and included in the description of model applicability.

5.3. Practical recommendations on how to choose the species

The choice of species is likely to require collaboration between the modeller, the ecologist and the risk assessor to ensure that the species used helps to fill in gaps in the overall risk assessment. It should be clearly stated how the modelled species fits into the risk assessment and which issues are being addressed by the modelling (see section 4).

In some cases it will be sufficient to model only one species (this is most likely when there are enough toxicity data available to have only a narrow question remaining), but in other cases several species will be needed (e.g. where there are issues to address over a whole group of organisms).

The approach to species selection is likely to vary from group to group depending on the specific protection goals for that group and the guidance available. For example, for birds and mammals, the guidance document (EFSA, 2009) uses generic focal species at the first tier, so before modelling is conducted the type of bird or mammal of concern and some example vulnerable species will already have been identified. With invertebrates, it is likely to be much less defined as the lower tier risk assessments focus on broad groups. If field or semi-field data are available, this could narrow down the species of concern considerably.

It is important to note that the species tested in first tier laboratory tests may often not be the best species to model environmental risk as they are not necessarily ecologically the most vulnerable species. An exception is the use of standard test species modelled in a TK-TD approach in order to refine exposure and/or effects. Such risk refinement strategy is recognized as useful, in the bird and mammals guidance, for example (EFSA, 2009). Furthermore, data from laboratory species might be useful to establish the credibility of the model.

A stepwise approach can be used for species selection (Figure 7). Details on each steps of this approach are provided below.

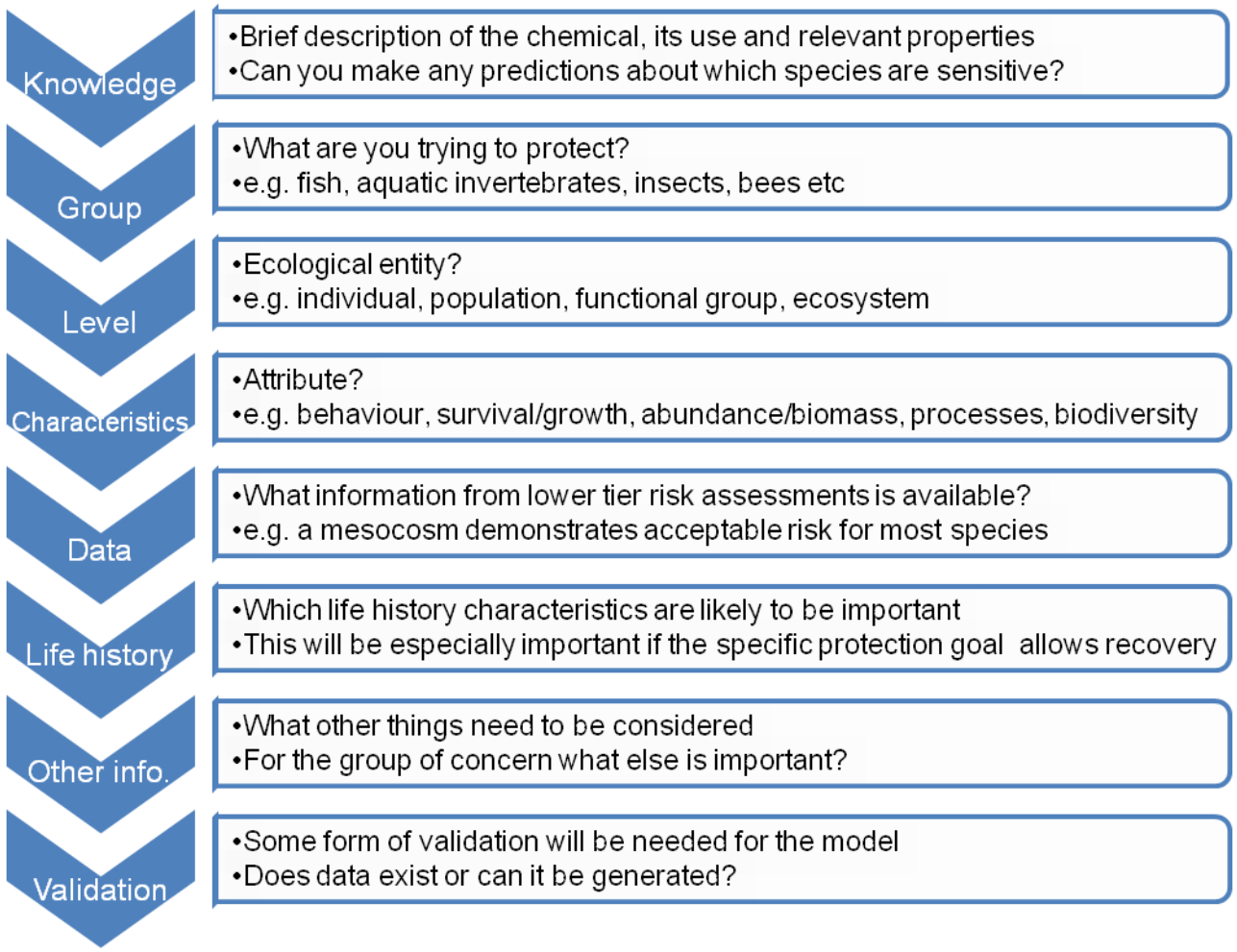


Figure 7: Approach to species-selection.

Step 1: What knowledge is available?

Gathering the information about the chemical and its proposed uses is necessary as background for selecting the species to model. There might be information about the target pests or the mode of action of the pesticide that might help to predict which species might be particularly sensitive. If there are no data about sensitivity of different species, then extrapolation of toxicity data from one species to the other species may be possible, but depending on the number and similarity of the species tested an assessment factor may be needed.

Step 2: Define the group

Broadly define the group of concern (birds, mammals, aquatic invertebrates, fish, etc.). This is necessary to select the relevant specific protection goals.

Step 3: Define the level

This relates to the specific protection goal for the group of concern that needs to be protected. The magnitude and temporal / spatial scale of effects should also be relevant for the species selected. The level could help define the type of model that is best suited to the assessment. For example, if protection goal relates to the individual (and therefore the assessment is done at the individual level),

then TK-TD modelling is likely to be useful. For populations, individual-based models (IBMs) and matrix models are possible options. For functional groups and above, community and ecosystem models will probably be needed. These examples are just for illustration and the type of model should be considered for the specific situation as other options may also be suitable.

Step 4: Which attribute must be modelled?

The specific protection goal will specify certain attributes that need to be protected (for example, for honey bees, the colony size). These attributes need to be relevant for the species selected for modelling (for example, it would not be appropriate to model solitary bees to address the risk to colonies of honey bees as solitary bees do not live in colonies so do not have the attribute "colony size"). For some groups the attribute to be protected might be functional, for example soil processes, in which case it is necessary to decide if it can be addressed by a single representative species or whether a community model would be more appropriate.

Step 5: What information is available from lower tiers?

Information from lower tiers in the risk assessment can be used to narrow the selection of species for modelling. For example, if there is a large laboratory dataset of single-species tests which show that one species is much more sensitive than all the others (with all others showing an acceptable risk), this can be used to select the modelled species. In a similar way, when a semi-field study shows recovery within an acceptable timescale for all species except one, then a model can focus on that species and the differences between the semi-field study and what is likely in real systems.

Step 6: Which life history characteristics are important?

Once all the information available has been considered it is necessary to identify which aspects of life history are likely to be most important. If the protection goal is "no mortality", then the sensitivity and exposure are likely to be the most important aspects (e.g. in an acute bird or mammal risk assessment, small species that eat a high percentage of the exposed food are used). If the protection goal allows for a given level of mortality or temporary effects on other life history traits of individuals, then characteristics such as generation time, number of offspring and dispersal ability are likely to be important too. A trait-based approach could be used in order to select a (group of) species that would cover for the risk of other species belonging to the community but with different biological and ecological traits.

Step 7: What else is important?

This is a stage to think about all the information gathered and check if there are other issues that have not been included. There will be some groups where additional aspects need to be considered. Some examples of additional factors that might need to be considered are:

i. For birds and mammals: does the use of the model introduce a factor that will not apply to all birds/mammals covered by the lower tier assessment? The focal species in the first tier is chosen because it is present in the crop of concern and has the highest exposure of the group of species (for example insectivorous birds) covered. No other ecological traits are considered. Since the model will introduce additional factors (for example number of broods in a year) it is necessary to check that the species selected is still amongst the most vulnerable. If a potentially more vulnerable species passes the first-tier risk assessment then it can be excluded from the species that need to be covered by the refined risk assessment using modelling (since the risk is acceptable assuming worst-case exposure).

ii. For aquatic and terrestrial invertebrates and plants, does the model consider only internal recovery for the modelled population or does it also include external recolonisation? If external recolonisation is considered, what are the uncertainties on the species or environmental settings and how will this be factored into the assessment?

iii. In general, has the conflict between using species for which a large amount of information is available and species likely to be ecologically vulnerable been addressed? This will always be a compromise as species studied in the laboratory are likely to be fast-growing/reproducing and vulnerable species are unlikely to be routinely studied.

Step 8: Which data are available for validation?

The degree to which the model has been validated will be important for deciding how much confidence can be attached to the modelling results. The data used for validation could either be pre-existing datasets or it could be data generated especially to use for validation.

5.4. Additional information for different groups of organisms

There are likely to be different issues to consider for vertebrates, invertebrates and plants. This section considers the vertebrates first, then all the invertebrates (rather than dividing aquatic and terrestrial organisms) and then plants.

5.4.1. Vertebrates

Birds and mammals

When selecting a species to model for birds and mammals, a focal species is chosen. The EFSA Guidance Document on Risk Assessment for birds and mammals (EFSA, 2009) has guidance on selecting a focal species in Appendix M. According to this guidance "*A 'focal species' is a real species that actually occurs in the crop when the pesticide is being used. The aim of using a 'focal species' is to add realism to the risk assessment insofar as the assessment is based on a real species that uses the crop. It is essential that the species actually occurs in the crop at a time when the pesticide is being applied. It is also essential that this species is considered to be representative of all other species from the feeding guild highlighted at the screening level and at Tier 1 that may occur in the crop at that time. As a 'focal species' needs to cover all species present in the crop, it is possible that there may be more than one 'focal species' per crop representing more than one feeding guild*." This is because different aspects affect the vulnerability. For example, without running the risk assessment it is not clear whether a small species with a high feeding rate that spends only half its time in a crop will be more or less at risk than a slightly larger species which spends 90 % of its time in the crop. Therefore, whenever a refinement step is added, it is necessary to check whether the focal species still covers all those it is supposed to represent and add additional species as required. Modelling should be treated in the same way. The approach used to select a species to model should be the same as for selecting a focal species for other refined risk assessments.

Fish

A focal species approach could also be used for fish. Ibrahim et al. (2013) listed European fish species that are susceptible to pesticide exposure in the field and then applied a stepwise filtering approach to select from these. The initial filters were as follows: (i) non-native to Europe, (ii) extinct and (iii) native to Europe but not present in any of the EU Member States. The remaining species were filtered by geographic range and habitat, with species included on the list only if they are widespread in at least one of the three EU zones (northern, central or southern). In addition, only species that live in streams, ditches or ponds were considered as these habitats are considered worst cases because of their high potential to be contaminated with PPPs when adjacent to agricultural land.

These filters resulted in 27 freshwater fish species being identified as potential focal species. This list of species could be used as the basis for identifying focal species. If this analysis is used to aid species selection, the species chosen must still be justified as part of the risk assessment.

5.4.2. Invertebrates

In order to identify appropriate model species, one should consider several biological, ecological and toxicological factors, such as:

i. Which part of the environmental compartment is under concern, e.g. plant surface or soil surface or both?

ii. Will the species be exposed to the chemical, depending on the time of chemical application, the lifespan of the species and its life history? In this respect, voltinism and other reproductive characteristics are very important (Stark et al., 1997, 2004).

iii. Which life cycle stage will be exposed and what is its intrinsic sensitivity? For example, larvae and pupae might be more sensitive than adults; resting stages might be less sensitive than others; aerial and terrestrial stages will not be directly exposed to the active substance in water.

iv. Biological/ecological traits which directly influence species exposure (e.g. respiratory apparatus, feeding regime, avoidance behaviour) and intrinsic sensitivity (size, physiology, e.g. bioaccumulation, biotransformation, elimination capacities).

v. Biological/ecological traits which directly influence individual recovery and population resilience (e.g. dispersion–recolonisation capacities) (Liess and von der Ohe, 2005).

vi. Does the mode of action mean that certain taxa and/or life stages are particularly sensitive (e.g. insect larvae for insect growth regulators)?

Aquatic invertebrates

Sensitivity to chemicals is known to be extremely variable among invertebrate taxa, and related at least partly to the biological and ecological properties of species (Vaal et al., 1997; Escher and Hermens, 2002) and also to the environmental context (Liess and Beketov, 2011).

For higher tier refinements, it is important to consider the population in its environmental context. Therefore, relevant abiotic parameters (e.g. temperature change), environmental parameters (e.g. habitat availability and quality) and biotic processes (e.g. density dependence) should be included when describing the effects on a population. Stress induced by processes could either mitigate or aggravate adverse effects observed on the individual when the level of the population is considered.

Non-target arthropods

There are a large number of non-target arthropod (NTA) species, both in-field and off-field, that might be affected by PPPs. Insecticides, in particular, are likely to be more or less toxic to many, if not most, NTA species. The variety and diversity of NTAs present in and around crops make the choice of species to model difficult. NTAs are important providers of a number of ecosystem services such as pollination, food web support and biological control. Species modelled should relate to the specific protection goals. Data from first tier assessments should be taken into account when selecting species. The final decision should be based on the data which are available, including also information on field effects.

Toxicity data may not be available for all the species modelled. The current data requirements for laboratory toxicity data for terrestrial NTAs are limited, although the core requirements under Escort 2 (Candolfi et al., 2001) are supposed to include more sensitive species.

Soil organisms

Several standard toxicity tests exist. However, the species tested are not necessarily the most relevant species for the field e.g. the Organisation for Economic Cooperation and Development (OECD)-recommended test species *Eisenia fetida* does not naturally occur in the field. Therefore, other species should be considered for modelling. Earthworm species might be allocated to three groups, epigeic (leaf litter- or compost-dwelling worms), endogeics (topsoil- or sub-soil-dwelling worms) and anecic (worms that construct permanent deep burrows that they use to reach up to the surface to obtain plant

material for food). These groups will all have different exposure types. Species should be chosen for modelling on the basis of how common they are in the field, how important they are as food for other organisms (i.e. relevance in terms of potential indirect effects) and their ecological sensitivity.

Folsomia candida is used as test species representing invertebrate soil 'primitive insects' for ecotoxicological testing. *Folsomia candida* is a useful species for laboratory tests because of its toxicological sensitivity, but it would not be representative for the field because of its rapid life cycle (not a vulnerable species). *Hypoaspis aculeifer*, a mite, is a predatory species in the soil environment which is also tested in the laboratory. Tests are also available for other species and groups, e.g. for isopods and nematodes, but are not standard required tests for authorisation of pesticides. In order to apply a trait-based approach, analyses of exposure and life traits (morphological and ecological) of species that are typical for the soil in European agricultural landscape(s) are required. Elements of classification, such as the following, should be considered:

i. mobility and occurrence in different layers of the soils: epigeic species (on the soil surface and in the litter layer, mobile species, e.g. larger collembolans or isopods), hemi-edaphic species (upper soil layers and litter layer, moderate mobility) and euedaphic species (in soil, low mobility);

ii. reproduction cycle, resting stage, behaviour, etc.

5.4.3. Plants

Algae and aquatic plants

For substances such as herbicides or fungicides with a herbicidal mode of action (MOA) or for plant growth regulators, ecotoxicological tests are usually performed on:

i. Algae: green algae (e.g. *Pseudokirchneriella subcapitata*) and another algal species (e.g. diatoms *Navicula pelliculosa* or blue-green algae), and

ii. Macrophytes: the monocotyledonous *Lemna* and another macrophyte species (e.g. the dicotyledonous *Myriophyllum*), when:

   a. *Lemna* is not sensitive, or

   b. efficacy data on terrestrial plant tests indicate that dicotyledonous plants are more sensitive, or

   c. uptake by the roots of submerged macrophytes is identified as a concern although generally the main uptake route is via water owing to the exposure and their morphology (reduced cuticula).

For some types of herbicides (e.g. photosynthesis inhibitors), algae and macrophyte data may have a similar sensitivity. However, herbicides that inhibit amino acid synthesis and herbicides with an auxin simulation mode of action generally seem to be more toxic to aquatic vascular plants than to algae (Giddings et al., 2013). Currently, knowledge of the MOA of several other types of herbicides is limited.

In the selection of a macrophyte species, a range of morphologically and taxonomically different species should be considered, unless the MOA of the herbicide primarily affects a specific group of species (e.g. mosses, monocotyledonous or dicotyledonous vascular plants; floating or rooted).

A trait-based approach here should also be applied to select the relevant (group of) species. Population effects should also be assessed in conditions close to natural systems in terms of competition (e.g. for nutrients, light, etc.), predation and natural stressors (e.g. climatic conditions) in order to obtain

realistic assessment endpoints. Regarding the possibility of recovery, it is important to consider the kind of reproduction and the development time of a species (e.g. while *Lemna* reproduces predominantly vegetatively and has a short development time, *Myriophyllum* has more frequent sexual reproduction and a longer development time). If the potential external recolonisation from transfer from an unexposed water body (e.g. by waterfowls) is addressed, one should consider that this is more likely for algae or species such as *Lemna* than for species such as *Myriophyllum*.

In general, assumptions for algae or the standard aquatic test plant *Lemna* cannot easily be generalised, but instead species-specific characteristics should be considered. Such considerations are common to other groups of organisms; however, in the cases of algae and macrophytes, the ecological roles of the species (e.g. substrate, shelter for macrophytes, e.g. food sources for algae) should be highlighted as many other water organisms depend on them.

Non-target terrestrial plants

Species sensitivity varies considerably with the herbicide tested, and no plant species is consistently the most or least sensitive (Fletcher et al., 1985; Marrs et al., 1989; Boutin et al., 2004; Clark et al., 2004; Strandberg et al., 2012). Species related to the crops for which a specific herbicide is intended for use usually have a higher tolerance to that herbicide (White and Boutin, 2007). Several studies conducted in Canada and Denmark have shown that there is no significant difference between the sensitivity of short- and long-lived species (Boutin et al., 2004; Carpenter and Boutin, 2010; Strandberg et al., 2012). Therefore, it is not possible at this stage to propose one species which is generally more sensitive to toxicants. Considerations for choosing a species need to include, for example, the mode of action (e.g. if a herbicide targets a certain group of plants).

Characteristics of plant species that are important to take into account in modelling are related to lifespan, size of plants, leaf shape/area, pollination strategy, seed production, seed dispersal, seed bank and size of populations.

The ecological success of a plant species in a specific environment may be quantified by its abundance and, depending on the life form of the plant, different measures of abundance may be relevant (e.g. density, biomass, or plant cover). Many natural and semi-natural plant communities are dominated by perennial plant species that form dense vegetation in which it is difficult to distinguish individual plants and, consequently, to determine density. In such cases, plant cover or biomass may be used as measure of plant performance, Further to these general criteria, other relevant traits can be leaf characteristics (leaf area, leaf mass per area, hairiness, etc.), seed production and morphology, height, root/shoot ratio or root morphology and mycorrhizal association and life history attributes, among others (Westoby and Wright, 2006; Dorrough and Scroggie, 2008; Comas and Eissenstat, 2009; Bernhardt-Römermann et al., 2011). Strandberg et al. (2012) studied the importance of selecting the appropriate endpoint relative to time of exposure using four non-target plants, including two annual (*Silene noctiflora*, *Geum molle*) and two taxonomically related perennial species (*S. vulgaris*, *G. robertianum*) with three herbicides with different MOAs (glyphosate, metsulfuron methyl and mecoprop-P) exposed at both vegetative (four- to six-leaf stage) and reproductive stage. Seed production was the most sensitive endpoint regardless of the phenological stage during exposure. Furthermore, biomass was not found to be a useful endpoint for these species and herbicides when plants were exposed at the reproductive stage.

The individual plants in natural and semi-natural habitats will typically compete for resources with both conspecific (intraspecific competition) and heterospecific plants (interspecific competition). Herbicides potentially affect these species interactions. The competitive relationship between plant species can be affected by herbicide use and hence change plant community dynamics. Considerable discrepancies were found between the outcome of the single species dose–response experiments and competition experiments (Strandberg et al., 2007; Damgaard et al., 2008; Strandberg et al., in preparation).

More details on modelling of plant species will be made available in the scientific opinion addressing the state of the science on risk assessment of PPPs for non-target terrestrial plants, which should be published in July 2014.

6. Development of the conceptual model

Based on the problem description, the modeller designs a conceptual model ("model formulation"). The conceptual model provides a general and qualitative description of the system to be modelled. This section first gives an overview of the most important steps of model formulation (section 6.1). Thereafter, a brief overview of existing models will be given.

6.1. Model formulation

When developing the conceptual model, the problem to be addressed has already been defined, and the species to be modelled and the biological entities (individual, population, etc.) have also already been chosen. The conceptual model then provides the conceptual overview of the processes within the model, and is often summarised in a conceptual diagram. Examples of conceptual diagrams are provided in Figure 8.

Models rely on biological and ecological principles and assumptions that describe the system's behaviour and functioning. The conceptual overview reflects the level of understanding and simplification of the system in the model (Schmolke et al., 2010). Note that not all components and processes at play in the system have to be included in the model. Some components or processes can be excluded:

i. if they do not have a major role in explaining the system's functioning/behaviour with regard to the purpose of the model;

ii. if their exclusion does not hamper the internal consistency of the model.

In addition to the conceptual diagram, modellers should provide the following information in relation to model design and formulation:

i. Model assumptions (biological, ecological, toxicological and chemistry) have to be clearly explained and supported by literature references, experimental results or at least a clear reasoning. When several ecological theories or concepts are available to describe a system, or a process in this system, reasons for choosing one particular approach should be provided.

ii. The choice regarding how effects are modelled needs to be well explained and justified.

iii. Origin (experiment, literature) and type of data that support the ecological assumptions have to be indicated, as well as data/information gaps (if any).

iv. When some important component or processes governing the system have been ignored, it should be justified in regard to the purpose of the model and available data. Data gaps and simplifying assumptions that were used to bridge this gap in the model should be clearly identified, as this information will help to assess model uncertainty at a later stage.

The modeller may also comment on the balance between simplicity and complexity given the purpose of the model and the availability of data. In this respect, the modeller should mention if simpler or more complex model alternatives have been explored, and why they were not used.

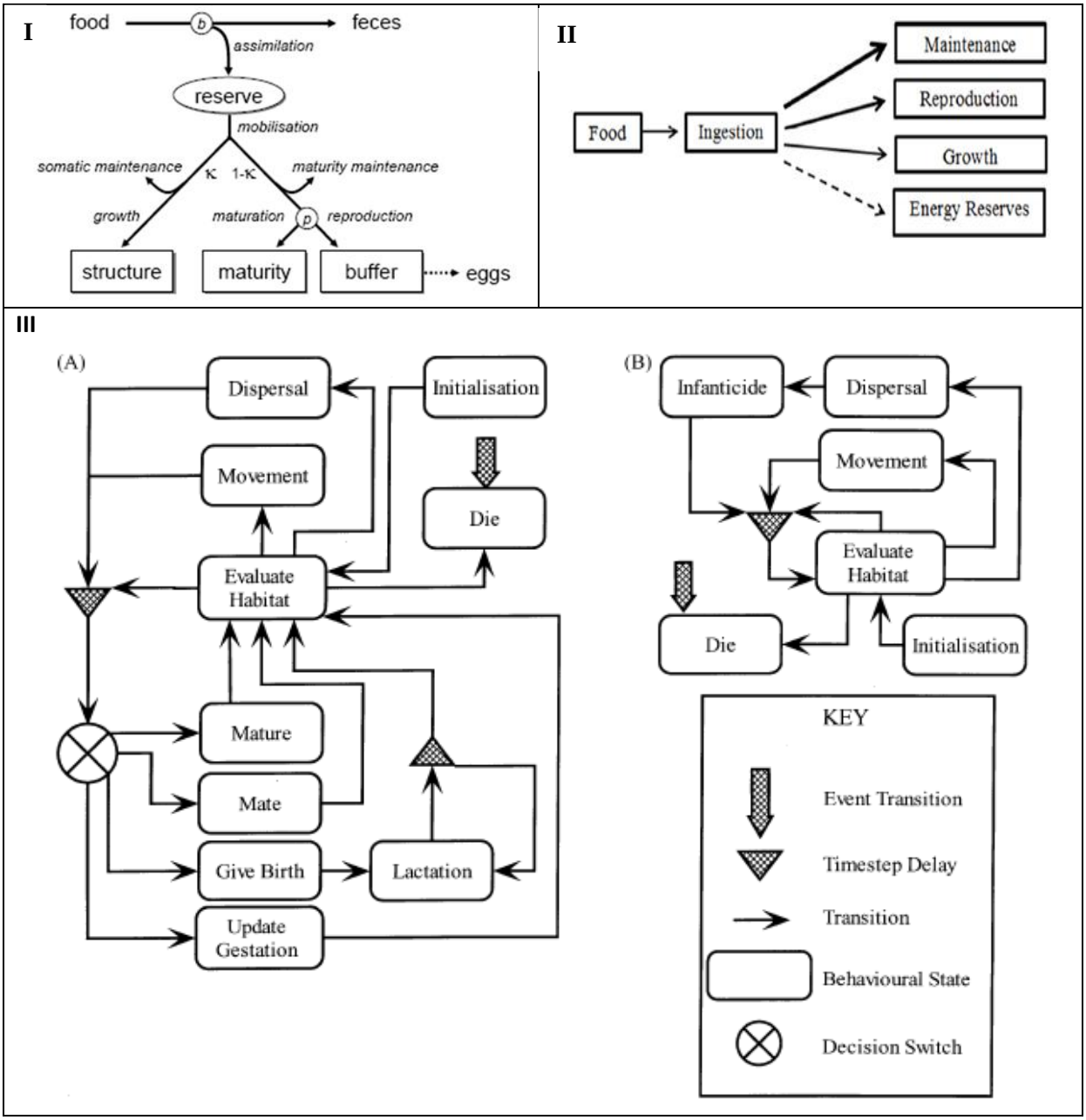


Figure 8: Examples of conceptual diagrams summarising the components, processes and causal links in the model. Panels I and II show two conceptual overviews from different models aimed at assessing energy acquisition and use within individuals (I is from Jager et al. (2004); II is from Johnston et al. (2013)). Panel III shows how individual females (A) and males (B) progress through their lives (from Topping et al., 2003)

6.1.1. System characterisation

The first step in developing the conceptual model is to identify which components and processes to include and which to exclude. This decision should be based on reliable biological, ecological, toxicological and chemical knowledge. Including or excluding components and processes from the model requires making simplifying assumptions on the organisation and functioning of the system. Such assumptions could be based upon ecological principles, theories or concepts, or results from dedicated experiments.

So, the key model items that have to be clearly identified by the modeller are:

i. the main components of the system (e.g. individual animals);

ii. the main processes that govern the functioning of the components (e.g. food assimilation, growth, reproduction);

iii. the internal and external factors that modulate the functioning of the components (e.g. development stage, season, food availability).

When models are used to extrapolate adverse effects across biological levels of organisation, it may be necessary to identify processes that mitigate or aggravate adverse effects from one level to the other, e.g. density dependence may modulate individual-level effects on reproduction or mortality (Liess, 2002). Alternatively, stress due to competition might increase individual sensitivity and thus magnify effects of the pesticide (Forbes et al., 2001; Knillmann et al., 2012a). If relevant to the problem under study, these mitigating/aggravating processes should be included in the model and explicated in the conceptual overview.

6.1.2. How to model the harmful effects of chemicals

The harmful effects of chemicals on individuals generally increase with dose and are modelled in dose–response curves that relate the dose of chemical taken in by an individual organism to its effects on the organism. The relevant effects are those that affect the individual's vital rates, e.g. its rates of growth, reproduction and survival. Delayed effects and carryover effects should be included in the model if considered relevant.

The dose–response relationship should be used if data are available. If this is not possible, approximations based on point estimates of no effect concentration (NOEC), lethal concentration (LC_x) or effective concentration (EC_x) might be used.

It is important to note that the effects may depend critically on how dose–response curves are modelled. For example, acute mortality can be modelled as the stochastic death of a proportion of a population, or as depending on the tolerance of individuals (Jager et al., 2011). Stochastic death would lead to the same proportion of the population dying under repeated exposure, whereas if individuals vary in their tolerance levels then the first exposure pulse of a pesticide may remove the more sensitive individuals, which in turn would result in a lower mortality at repeated exposures.

Likewise, effects can be assumed directly, e.g. a substance reduces individual growth rate directly by a certain percentage, or indirectly acting on growth by inhibition of food uptake (Agatz et al., 2012) with possible consequences not only on growth, but also on other individual performances (e.g. decreased reproduction). In both cases, the reduced fitness of the individuals may lead to some effects at the population level, but these effects may be of different type/intensity/duration depending on the mechanism of effect at the individual level

For individual-based models, movement of animals in the landscape introduces further complications. When modelling the uptake of pesticide by an organism, it is important to ensure that the representation of uptake and effect in the model corresponds to the type of data available, and that the model representation is consistent with the ecotoxicological expectations. For instance, if the median lethal dose (LD_{50}) is based on a laboratory study conducted over several days, the way in which daily mortality is calculated has important consequences. If an animal encounters the LD_{50} each day of the study, with 50 % probability of death each day, then its chances of surviving the whole test is much less than 50 %. An alternative is to use LD_{50}/d, where d is the length of the test period in which the LD_{50} was calculated. This assumes that the probability of death is even for the whole test, however this may not be true. The important point here is how the dose-response curve is modelled as it has important consequences for risk assessment. So in all cases the modeller should account for the choices made in modelling the dose–response curve.

6.1.3. Choice of model type

The third step is to represent the relevant components/processes/links in a model. A variety of model types are available, and these are described in section 6.2. Which is appropriate depends on what data are available and the modelling question. For example, a structured population model which identifies the structure of the population (e.g. through different age classes with specific vulnerability to the chemical) might be preferred to an unstructured population model in which only population size (or biomass) is considered when successive applications of a chemical occur over the lifespan of the species under study. To ensure transparency in choice of model type, this choice should be justified by the modeller. This should be written in such a way that it is comprehensible by all users. Aspects to characterise the model type are:

i. individual/population/community/ecosystem level;

ii. model type within each level (e.g. unstructured/structured population model or IBM at the population level, deterministic/stochastic);

iii. spatial/non-spatial;

iv. temporal scale of the model.

Depending on model formulation, the model can be generic or specific to, for example, a given species, chemical, landscape type. The domain of applicability of the model, including the extent of acceptable extrapolations, should be mentioned by the modeller.

6.2. Overview of different types of effect models

Ecological models are abstractions of real systems which represent biological processes and their consequences within and across levels of biological organisation in a mechanistic way. They are often referred to as "mechanistic effect models"; "mechanistic" because, in contrast to statistical models, they represent biological, physical and chemical processes. The phrase "effect models" distinguishes them from "fate models", which describe the fate of chemicals in the environment. Effect models focus on effects of chemicals on individuals, populations and communities (Grimm and Martin, 2013).

Effects of chemicals on individuals are represented in TK-TD models. In these models, the uptake of a chemical and its distribution between the organs of the body, and its subsequent biotransformation and elimination processes, are collectively referred to as toxicokinetics. Also modelled are the effects of the chemical where it causes harm within the body, with consequences on individual performances and or life cycle trait values, referred to as toxicodynamics.

Effects of chemicals on populations are represented in population models. The types of model, used in population-level risk assessment, have been classified and reviewed (e.g. by Munns et al., 2008). Here we give only an outline summary. Three main types of model are in use; these are: scalar (unstructured), structured (e.g. matrix) and individual based modelling (also known as agent based modelling (ABM)). The types of model differ in how much detail of the population they capture and in how many parameters they use. The simplest are the scalar models. In scalar models all individuals are treated as identical. They do not differ in age or in any other characteristic and so the population can be fully represented by a single scalar variable representing population size. Rate of change of population size with time can be represented by a differential equation, e.g.,

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \quad (1)$$

where N represents population size, r represents population growth rate at low population density and K is the carrying capacity of the modelled environment. Density dependence can be incorporated in scalar models. The model in equation (1) can be integrated to show how population size, N, changes with time, t. It has only two parameters, r and K. These scalar models were developed and analysed

throughout the twentieth century and are presented at length in ecological textbooks (e.g. Begon et al., 2006).

In matrix models, distinction is made between individuals of different ages or sizes, for example. The population is divided into age or size classes—e.g. juveniles and adults—and this gives the population structure, this structure being described by the numbers in each class. All individuals within an age or size class are treated as identical, but there may be variation between the classes in their survival or reproductive rates per unit time. The characteristics of the individuals in each class are entered into the cells of a matrix, and this allows computation of how population structure changes as time progresses. Density dependence can be incorporated into matrix models and easy-to-use software is available to implement matrix modelling (see, for example, http://www.ramas.com/software.htm). Matrix models were developed and analysed in the second half of the twentieth century, and are presented in an authoritative book by Caswell (2001).

In individual based models (IBMs), each individual is modelled separately and individuals are generally located in a spatially explicit, mapped environment. Individuals interact with their environment (e.g. depleting food resources by feeding) and with each other. Individuals vary in features such as their age and size and energy reserves, and each acts according to its modelled needs, e.g. for food, or a mate, or to care for its offspring. The dynamics of populations in specified landscapes are studied using computer simulation. IBMs have been developed and analysed in ecology more recently than other methods. They are presented in detail with recommendations for good practice in Grimm and Railsback (2005) and Railsback and Grimm (2011). A comparison of the use of a matrix model and an IBM in pesticide risk assessment of UK skylark populations is provided by Topping et al. (2005). IBMs provide the most detailed description of the effects of chemicals on populations, and generally use more parameters than other model types. The advantages of IBMs are that they can represent the effects of chemicals applied locally in mapped environments that may change seasonally, they can capture interactions among species (e.g. through changes in food supply or predation) that may result in indirect effects, and they are not limited by mathematical tractability.

Effects of chemicals on communities comprising more than one species can also be modelled. The crucial difference from models of a single population is the inclusion of interspecific interactions such as predation or competition. All the above-described modelling techniques can be combined to model communities. A special type of community model is the food web model, which represents different trophic levels of an ecosystem in a consumer resource system. At the bottom of the food chain are autotroph organisms, which produce organic matter from inorganic substances (e.g. CO_2, nutrients) generally via photosynthesis, and these provide food for heterotroph organisms, which feed on other organisms to ingest organic matter. For example, algae, grazers and predators can be linked to build a (simple) food web. Food web models are often based on differential equations but can also be built using other modelling techniques.

Combining different types of models: Models can be combined to form linked systems. These model systems may include models of different types (e.g. fate, human land and crop management and agent-based models can be combined into a single simulation for risk and impact assessment). In this case the different models need to be considered as separate entities before considering the method by which they are integrated to form the whole model being used for a specific case (e.g. Topping, 2011).

Complementary information about model types can be found in the abundant literature on this subject, as well as in DG SANCO (2013).

6.3. When can models be useful in the risk assessment process?

Mechanistic models can be useful at all tiers of current risk assessment schemes. Figure 9 exemplifies how model can be used at tiers 2–4 in aquatic risk assessment. One model type is not tied to one particular tier in the risk assessment. Indeed, different types of models often have to be combined to extrapolate across levels of biological organisation in a mechanistic way.

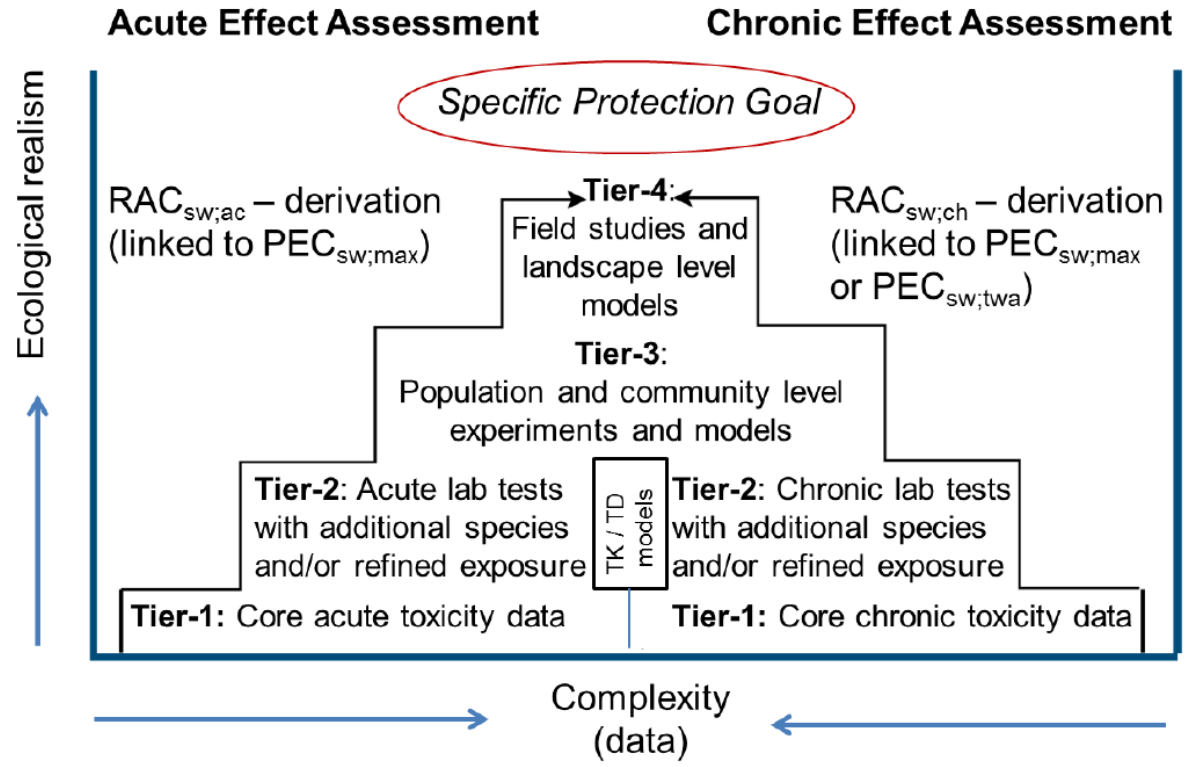


Figure 9: Schematic presentation of the tiered approach within the acute (left part) and chronic (right part) effect assessment for PPPs. For each PPP, both the acute and chronic effects/risks have to be assessed. The tier 1 and tier 2 effects assessments are based on single-species laboratory toxicity tests, but to better address risks of time-variable exposures, the tier 2 assessment may be complemented with TK-TD models. Tier 3 (population- and community-level experiments and models) and tier 4 (field studies and landscape level models) may concern a combination of experimental data and modelling to assess population and/or community level responses (e.g. recovery, indirect effects) at relevant spatio-temporal scales. All models included in such a tiered approach need to be properly validated and fulfil required quality criteria (figure from EFSA PPR Panel, 2013)

The use of TK-TD models is already considered as a relevant risk refinement option in the bird and mammals risk assessment guidance (EFSA, 2009) and is mentioned in the aquatic risk assessment guidance (EFSA PPR Panel, 2013). Furthermore, the EFSA scientific opinion on protection goals (EFSA PPR Panel, 2010) and scientific opinion on new challenges in risk assessment (SANCO, 2012) support the use of models as a risk refinement tool where needed. According to EFSA (2009), body burden models, which are a particular class of TK-TD models, could be helpful to challenge the following areas where the current wildlife risk assessment is not dealing adequately:

i. the relative contribution oral/dermal uptake and inhalation make to systemic exposure;

ii. risk from fast kinetics compounds (i.e. substances that are absorbed and eliminated in hours rather than days);

iii. the role of avoidance in limiting systemic exposure to levels below lethal thresholds;

iv. the risk through bioaccumulation.

Other cases where the use of TK-TD models should be preferred over other existing refinement approaches have also been highlighted (Ashauer and Escher, 2010; Modelink workshop, 2013):

i. when there is strong variation of exposure concentration in time (e.g. exposure via discontinuous feeding patterns; pulses of PPPs);

ii. when behavioural responses may modify exposure (e.g. dehusking behaviour in bird and mammals);

iii. when complex exposure pattern (see item i. and ii. above) might lead to delayed or carryover toxicity;

iv. when compounds have a particularly slow toxicokinetics (i.e. rates of toxicokinetics processes are limiting);

v. when long term risk forecasts are needed but only acute test data is available;

vi. when there are few or no refinement alternatives (e.g. recovery of certain aquatic invertebrates in enclosed systems).

Notice that this review does not preclude the use of TK-TD models for other issues. A forthcoming EFSA Opinion on the state of mechanistic effect modelling approaches for regulatory risk assessment of pesticides for aquatic organisms will provide more insight on when and how to use mechanistic effect models.

Given that most of the services under the selected specific protection goals are performed by populations, or groups of populations, development of appropriate population models for use in risk assessment is needed (EFSA PPR Panel, 2010). This need is endorsed by the recent EFSA guidance document on bird and mammals risk assessment (EFSA, 2009) and aquatic risk assessment guidance (EFSA PPR Panel, 2013). Owing to the complexity of this issue, it is envisaged that each assessment would be on a case-by-case basis.

7. Development of the formal and computer models

When we have formulated a conceptual model, it is necessary to construct the corresponding formal model and computer models, to be used as a basis for effect prediction/simulation. The first step is to develop equation and algorithms. This step is called “model formalisation” and is described in section 7.1. The next step is to implement these equations into computer code (section 7.2). This code needs to be thoroughly evaluated by the modeller (“model verification”) and therefore this process is described in a separate section (section 7.3).

7.1. Model formalisation

Variables vary in time. State variables describe the state of the system (e.g. animal or population size) and the model predicts/simulates the time course of the state variables. Forcing variables (also known as input variables) describes the scenario under which the model is run (e.g. temperature). Parameters are terms in the model that are fixed when conducting a model run or a simulation. In mechanistic models, parameters are the constants in the equations and algorithms that are used to represent the processes in the model. In a mechanistic model, variables and parameters generally have a biological, physical or toxicological meaning and a unit, which should be mentioned when defining the parameter. Note that some parameters or variables may aggregate several processes e.g. “damage” in individual-level effect models derived from the general unified threshold model for survival (Jager et al., 2011) represents the sum of physiological damage occurring due to the toxicant; the different kinds of damage are generally not identified.

Equations and algorithms are derived so that the dynamic of the processes included in the model are correctly represented. Specific care should therefore be given to the choice of mathematical functions

and algorithms describing the processes (e.g. first-order vs. second-order kinetics of a toxicant, linear vs. exponential growth of a population).

7.2. Model implementation

The software/program and the version number used for implementation should be specified (e.g. MatLab, R, etc.). The code/scripts should be thoroughly checked by the modeller and be made available to the evaluator in an annex. The model version used in the regulatory model should be clearly identified. A written model description should be provided by the modeller, which enables users/evaluators who do not have a solid training in mathematics or computer science to understand the model/code structure. An executable version of the computer program implementing the model should be also provided so that evaluators can check whether they can replicate reported results, and can perform their own simulation experiments to better understand the model.

7.3. Model verification

Verification of a model is the process of checking that the computer code correctly represents the conceptual and mathematical model (AIAA, 1998). Verification does not check the realism of the model. Verification of computer code is a large and important professional area. Here, only a summary of key points is offered. An introduction to verification is provided elsewhere (e.g. Grimm et al. (2006) and Railsback and Grimm (2011)).

Verification relies on good understanding of what the model is designed to do and how it works at several hierarchical levels. Most models are made up of a number of interrelated units or modules. The modeller should verify that these units operate correctly in a series of hierarchically linked unit tests. The effort put into each verification step should be documented, explicitly stating how much verification was performed and describing the methods used to achieve it. The unit tests isolate each component of the model in turn and verify its behaviour based on known input–output relationships. When each component has been verified, groups of components should be tested as a unit at a higher hierarchical level. This process is continued until the whole program/model has been tested, at which point further tests become validation exercises. The test plan and unit test results should be documented. For international standards in software testing and documentation approaches see the ISO/IEC 29119 Software Testing Standards (ISO/IEC/IEEE, 2013). These standards are those applied by the software industry in developing software and, in the ideal case, would be followed. Besides verifying the correct implementation of the model as computer code, it may also be necessary to check for the impact of rounding errors and effects of solvers of differential equations. The latter are numerical approximations and therefore correct only to a certain degree. However, most often, these approximations are not a problem.

8. Setup of the regulatory model

In the following step of the modelling cycle, model parameter values are estimated and the computer model is combined with one or more environmental scenarios. The so-obtained regulatory model consists of the following components: (i) the computer model; (ii) programs for pre- and post-processing, often made available in the form of graphical user interfaces; (iii) model parameters; and (iv) the environmental scenarios.

Model parameter values can be obtained in many different ways; section 8.1 gives an overview of these methods with specific attention to model calibration. An environmental scenario describes the environmental context in which the model is run; the development of these scenarios is described in section 8.*2*.

8.1 Model parameter estimation including model calibration

Parameters have constant values while the model is run. The values of parameters have to be estimated before using the model for further simulation/predictions. Depending on parameter values, completely different system dynamics may be computed or simulated by the model. Therefore, selecting values

for a model's parameter should be done with caution. Depending on available data, three cases may occur:

i. The parameter value is taken from literature.

ii. The parameter value is directly measured, or indirectly estimated from measured data (e.g. size was measured in the experiment and weight is estimated using body size scaling relationships), in dedicated experiments which cover certain model processes (e.g. reproduction tests to learn about reproduction for a population-level model).

iii. The parameter value is estimated by calibration, i.e. the value is chosen so that the model fits existing data patterns.

Calibration is a special method for parameter estimation, aimed at adjusting the values of parameters until the model output best fits existing data patterns, which can be of qualitative (e.g. temporal cycles in the population size) or quantitative nature (e.g. population sizes or spatial distribution of a species). This fine-tuning is important if quantitative predictions have to be made, which is the case for models supporting risk assessment. An illustrative example of a calibration could be the tuning of a guitar. If we put on new strings, we do not know which tension produces a correct tone, but we can determine that by calibration, i.e. by gradually changing the tension of the string and simply listening to the produced sound. If that sound matches the reference sound (tuning fork), the tension is correct.

Various statistical methods are available for parameter calibration, from "trial and error" to sophisticated methods such as inverse modelling techniques. The choice of the most appropriate methods depends on the nature of the data available for calibration and the parameter(s) to be estimated. Techniques such as inverse modelling have the advantage that they give not only the parameter values themselves, but also the uncertainty bounds of these parameters. This may be important input for the uncertainty analysis. "Trial and error" could be used in simple cases; however, the disadvantage is that uncertainty bounds are not provided. In both cases it is important to be aware of the potential for local minima (good fits but not the best, which may involve a large change in one parameter to obtain), and fitting the model well, but for the wrong reasons (e.g. good fit resulting from unrealistic parameter values).

Data used for parameter estimation should be relevant to the species and scenario of concern. However, lack of data on the modelled scale (e.g. on the landscape scale) or for the species of concern is a frequent issue in the context of ecological risk assessment. Where data are lacking, it may be permissible to use data from species or environments other than the one under consideration. When such an approach is used, it should be scientifically justified by the modeller, and the uncertainty introduced by using such data should be characterised. If the model is calibrated using data representing a particular environment (e.g. a certain temperature range, certain food level) the transfer to other environments outside the range of the calibration may be problematic.

The modeller should provide a summary table indicating the parameter's symbol, unit, name/biological meaning, estimated value and associated uncertainties. Additionally, the method used for parameter estimation should be given. For parameter values taken from the literature, the exact data source must be given; information about the parameter's uncertainty should be provided, if available; potential sources of uncertainty should be mentioned (e.g. differences in the experimental setup or species of interest compared with the scenario/species modelled).

For parameter values estimated from dedicated experiments or calibration, a measure of parameter uncertainty (e.g. 95 % confidence or credibility interval) should be provided as well as information on the usual range of the parameter value (i.e. minimum and maximum values).

For parameter values estimated from dedicated experiments, potential sources of uncertainty in the data should be mentioned (e.g. the deviations in empirical measurements, potential biases of measurements, uncertainties due to derived parameters).

Additionally the modeller should provide a detailed appendix including:

i. Raw data used to identify parameter values and their origin (literature, experiments). Detailed information has to be provided on the data (see section 3).

ii. Description of the conditions under which the values reported in the literature/experiment were observed (region, environmental conditions, season, etc.). Differences to the situation to be modelled should be discussed in the uncertainty analysis.

iii. The choice of parameter estimation method used (for calibration or estimation from dedicated experiments) should be explained and justified by the modeller.

iv. Visual representation of the match between data patterns used for calibration and model outputs.

v. Statistical indicators for goodness of fit (e.g. correlation between observed and predicted values). A review of suitable statistical indicators, and recommendations for use, have been provided in Jansen and Heuberger (1995).

8.2 Scenario development

A scenario represents the environmental context in which a model is run. It is defined as a representative combination of abiotic, biotic and agronomic parameters for the purpose of modelling. The ecological considerations, or so-called "biotic parameters" in Figure 10, relate to the assembly of populations of different species and the way in which they interact with the species being considered in the effects and exposure models. Abiotic considerations such as climate, soil and stream properties determine the exposure of a pesticide, influence the composition and structure of communities and interact with pesticides to determine their effects on species. Agronomic considerations relate to the management of the agro-ecosystem (the crops and their development over time, tillage and irrigation practices, and the structure of the landscape). Assuming the ecological entity (for example, individual, population, community) is chosen, a scenario is a conceptual and quantitative description of the environmental system relevant to the risk assessment, including the habitat (at relevant spatial and temporal scales) and the driving environmental variables including external stressors. Scenarios must be defined in relation to the SPG and the level of conservatism defined in the problem formulation (section 4.1).

The scenarios selected must be agreed with the risk manager as being sufficiently representative of the specific risk assessment goal, i.e. regarded as appropriately reflecting the characteristics of the agricultural area under focus. In addition, the scenario definition depends on the entity to be modelled. For individuals, populations or communities, a scenario may represent a landscape with its spatial configuration, physical and chemical characteristics.

8.2.1. General considerations on scenario selection

Scenarios are widely used in fate/exposure modelling, so it is worth looking at how fate modellers select scenarios, to see if a similar approach can be adopted in effects modelling. The objective in current fate modelling is to identify the predicted environmental concentration (PEC) following the application of a pesticide according to its intended uses. The exposure concentrations are analysed over the whole area of intended use. From this distribution, the percentiles are derived and a certain percentile (often the 90th percentile) is chosen to reflect realistic worst-case conditions which are used as the scenarios.

It is important to note that the meaning of the term "exposure" differs between fate/exposure modelling and effects modelling. In effects modelling, "exposure" is the concentration that the organism is exposed to. Therefore, exposure results from a combination of the PEC, ecology and behaviour of the organism, and internal toxicokinetics. The result of fate/exposure modelling is the spatial temporal distribution of the PEC, also referred to as the exposure profile (see EFSA PPR Panel, 2013). In ecotoxicological risk assessment, PECs are not currently used as spatially and temporally dynamic variables; therefore, their applicability is limited to some sedentary organisms.

How the results of fate/exposure modelling are currently used may work well for individuals or populations that are stationary (in which case only temporal dynamics matter). In this case, pesticide fate and effects are investigated in separate modelling. However, there could be difficulties in implementing the approach when individuals are mobile. Populations can extend over large areas (e.g. individuals of some bird species can visit different areas in Europe). When individuals are mobile, the spatial aspect of exposure become particularly important (spatial dynamics have to be considered). We recommend the integration of spatial dynamics as a topic for further work. Until this issue has been resolved, we recommend a different approach for populations of mobile organisms. For example, effects cannot rely purely on PECs, but must integrate ecology and behaviour to determine the level of effects and the identity and characteristics of the resource distribution that are important (e.g. food supply, habitat requirements and dynamics).

It would be attractive if the percentile approach to develop scenarios used in fate modelling could be applied in effects modelling. However, abiotic, biotic and agronomic parameters describing the environmental scenario and the behaviour of organisms form complex interactions. Currently, there is no agreed procedure on how to derive a distribution from all these factors and choose a certain percentile from that because of the overwhelming number of potential factors involved. It is clearly a desirable future goal to develop standard environmental scenarios in such a way. However, currently we must rely on carefully argued, realistic, worst-case scenarios.

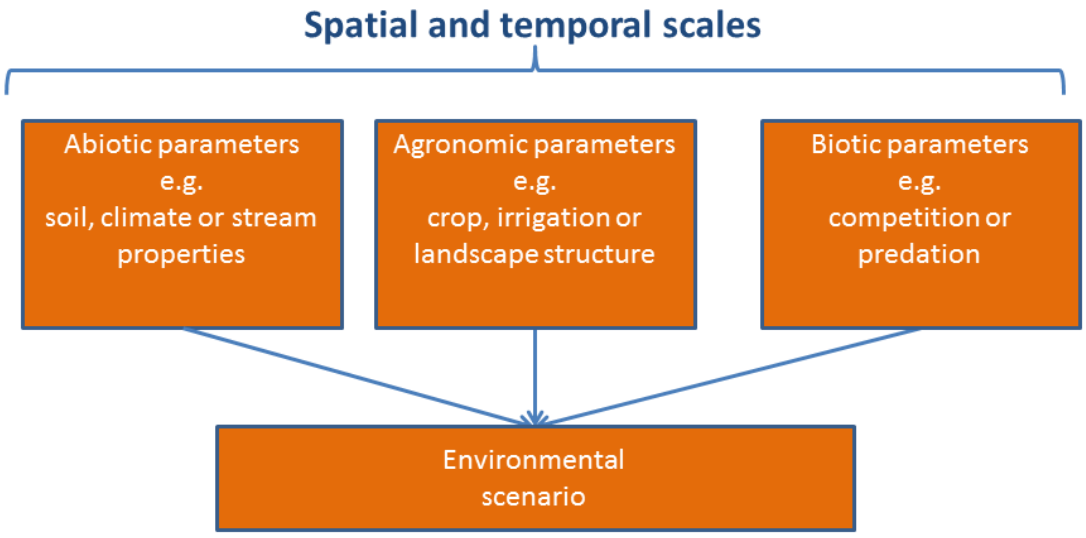


Figure 10: Illustration of the building blocks that together constitute an environmental scenario. Notice that the spatial and temporal scales of the parameters are an important element of the scenario definition

Biotic parameters

The resource distribution and also interaction with other species (predators, competitors) and their spatial and temporal characteristics are important.

To achieve a realistic worst-case scenario, the entity studied (individual, population, community) should be in a realistically resilient state. For example, biological interactions such as interspecific competition and predation have to be considered since interspecific competition may increase time for recovery considerably (Knillmann et al., 2012b). Similarly, predation may also reduce capacity for population recovery (Beketov and Liess, 2006), as will some agricultural practice and environmental stressors. The scenario should therefore describe the relevant aspects of the population and the stressors incorporated in baseline or control conditions (and the models should include relevant age classes, cohorts, etc., and include relevant biotic and abiotic interactions).

Abiotic parameters

Factors affecting population or individual survival or reproduction need to be considered, i.e. resources and stressors (e.g. non-chemical agricultural management). For aquatic biotopes several field investigations have identified the values for abiotic parameters (Liess and von der Ohe, 2005; Liess et al., 2008). For example, long-living K-strategists thrive only if ecosystem disturbance is limited, and may need, for example, continuous water flow.

Key abiotic parameters affecting biological considerations may include:

i. the landscape configuration in terms of the availability, size, shape and distribution of suitable habitats (including source habitats);

ii. connectivity of habitats because this will dramatically affect spatial dynamics and therefore recovery by dispersal potential;

iii. weather parameters, especially where individual growth is strongly influenced by, for example, temperature;

iv. landscape management which may impose mortality or result in resource depletion, e.g. soil cultivation may kill animals in field and may remove food resources;

v. physical structure of the habitat (Bunzel et al., 2014);

vi. permanent water flow for species of flowing waters;

vii. water quality in terms of loads of nutrients and oxygen-consuming substances.

Agronomic parameters

Agronomic considerations relate to the management of the agro-ecosystem. If the new chemical being evaluated is designed to protect a particular crop, then the output of the model may vary with the region in which the crop is grown (e.g. northern vs. southern Europe). Therefore, scenarios should be constructed appropriate to the distinct regions in which the crop is grown. If distinct schemes of application are considered, then these should be analysed in distinct scenarios. Another important component of the environmental considerations is the landscape structure and spatial units chosen (see section 8.3.*2*).

Other exposure considerations

The most relevant exposure routes of the organisms and the environment should be taken into account. Examples of routes of exposure of organisms are oral uptake, inhalation or contact exposure. Examples or routes of exposure of the environment are direct overspray, spray drift, atmospheric

deposition, runoff and drainage. The appropriate type of concentration (referred to in EFSA PPR Panel (2010, 2012, 2013) as the ecologically relevant concentration, or ERC) also needs to be taken into account (e.g. in pore water, in surface water). Parameters describing these processes must be available at the same spatio-temporal resolution as the ecological model. Further considerations of the spatio-temporal dimension of the concentration should be taken into account in the model. When acute effects are modelled, the maximum concentration over time may be most relevant (Liess and von der Ohe, 2005). However, culmination of sequential pulses of pesticides may need to be considered in cases where several pulses of pesticides affect individuals, cohorts or successive generations within a population (Liess et al., 2013). When simulating chronic effects and recovery, the decrease of the exposure concentration resulting from dissipation (the exposure regime) is important as well. In the model, the temporal resolution of the exposure regime should match those of the specific protection goals.

8.2.2. Definition of the spatio-temporal dimensions

As part of the spatial and temporal dimensions of the protection goal, the following elements must be addressed by the modeller when developing realistic worst-case scenarios (EFSA PPR Panel, 2010), so that the dimensions fulfil the specific protection goals:

What spatial scale and resolution to use?

An important property of the spatial unit is the spatial scale of this unit. This scale should reflect the specific protection goal defined in the problem definition for the species under consideration. The currently available exposure scenarios were all developed for the scale of an individual landscape element (e.g. an agricultural field (EFSA PPR Panel, 2012) or an edge-of-field ditch (FOCUS, 2001)). This scale seems to be appropriate also for scenarios to be used in combination with individual effect models such as TK-TD models. If, however, populations or communities are to be modelled, a larger spatial scale (e.g. $2 \bullet 2$ km^2 or $10 \bullet 10$ km^2) seems more appropriate since variability of landscape elements and pesticide application patterns within this landscape are important considerations for processes such as external recovery (Topping and Lagisz, 2012).

When choosing the spatial scale for landscape scale models, boundary effects should be avoided. It is, for example, often necessary to simulate a buffer area around the test location, even if a wrap-around (toroidal) design is used. From an ecological point of view the spatial unit might be determined by the spatial extent of the local population of the species under focus (i.e. ecologically relevant spatial unit). However, depending on the definition of the specific protection goal the extent of the spatial unit might have to be further reduced (e.g. if extinction of local sub-populations is not acceptable even though the whole population is stable).

If recovery by dispersal is included, larger spatial scales may be necessary, including a justification for the habitat configuration and composition. In such cases, impacts off-crop or in non-contaminated areas due to source-sink dynamics should be considered. Conversely, rescue effects (i.e. external recovery) resulting from dispersal from source habitats could also be considered in the scenario design. Regular or unrealistically small landscapes are not suitable for estimation of impacts where spatial dynamics are included due to the potential bias related to specific landscape configuration (Holland et al., 2007). Simple spatial representations (e.g. single field with source habitats) can, under conditions of wider spatial and temporal dynamics, significantly underestimate population level impacts (Dalkvist et al., 2013). The spatial resolution should reflect the mobility of the species under consideration and the spatial variability of the expected exposure pattern.

Spatio-temporal scale for effects considered as acceptable could be agreed with risk managers with respect to protection goals. For example, a reduction in NTA densities of less than 10 % within an area of 10 km^2 at certain times of the year (e.g. when relevant bird species breed) could be considered as a criterion for acceptance. This approach alleviates the issues of having to deal with local or short-term effects which could otherwise become diluted/propagated to a wider landscape.

To what spatial area does the assessment apply?

This could be the whole EU, one of the regulatory zones North–Centre-South or a Member State. The disadvantage of this definition is that it may include large areas of (i) zero pesticide applications, (ii) non-zero application because others pesticides are applied and (iii) zero occurrence of the study organisms. Such a large fraction of zero values would seem to be not meaningful for an adequate risk assessment (EFSA PPR Panel, 2010, 2012). It is proposed to base the definition of the statistical population on the intended area of use; for example, for a pesticide that is applied to potatoes, the population is defined as all fields (or habitats within the range of focal species to those fields) on which potatoes are grown in a zone or a Member State (EFSA PPR Panel, 2012). The spatial distribution of the risk needs to be considered (e.g. aggregation of high risk areas (hot spots) could produce unacceptable long-term effects).

What temporal scale is considered?

The temporal scale of a scenario should be appropriate to the specific protection goals. When populations are simulated, the model may need to run for a number of years since carry-over effects and culmination are important and may take years (several generations) to develop (Topping et al., 2013; Liess et al., 2013). A TK-TD model may need to run only for a single day, e.g. if it used to refine exposure from bolus dose to realistic feeding pattern in bird and mammals risk assessment. In other cases (e.g. fish mortality), simulations should cover at least the lifespan of the species under consideration, to cover potential delayed effects, and several generations if carry-over effects on offspring are expected. However, when considering the temporal scale of the problem, exposure considerations also need to be taken into account. For example, when simulating the fate of persistent pesticides, the exposure part of the model may need to run for several years to simulate the accumulation of the pesticide in the ecosystem of concern, including the transfer along food webs.

The temporal resolution of the model is another important aspect to deal with. Seasonality is important for most organisms and should be included if relevant to the problem investigated. The temporal resolution should further be appropriate to capture the relevant exposure profile. Particularly important is the concentration directly after application and the time course of dissipation after application (EFSA PPR Panel, 2012).

Which value of the percentile is used?

When appropriate spatial and temporal scales have been selected, we can consider which percentile of the exposure concentrations should be used. The percentile chosen determines the conservativeness of the risk assessment. The development of exposure scenarios was based on the 90th percentile exposure concentration (FOCUS, 2009; EFSA PPR Panel, 2010). Whether the 90th percentile can also be used to meet the overall protection goals for effects has to be reviewed by considering the likelihood of exceeding the 90th percentile concentration and the severity of such an occurrence on the effect side and by considering the implication of spatio-temporal aggregations of those units in the upper tail of the distributions. The outcome of this should be communicated to risk managers, whose task it is to decide on the level of protection to be achieved in practice (EFSA PPR Panel, 2010).

What certainty of the model prediction should be reached?

The degree of certainty that the specified level of effect will not be exceeded is part of the definition of the specific protection goal (*Figure 6*) and should therefore be given attention in the derivation of the scenarios (EFSA PPR Panel, 2012). The likelihood that a given percentile is exceeded can be derived from uncertainty analysis. Vanderborght et al. (2011) and van den Berg et al. (2012) performed such an analysis for fate models and showed that parameter uncertainty led to higher concentrations in the higher percentiles (such as the 90th percentile) of spatial concentration distributions, especially for distributions in smaller and more homogeneous regions. Although such an analysis has not been carried out for effect models, it is likely that the higher predicted environmental concentrations would propagate to higher levels of predicted effects. Risk managers might want to deal with this by introducing additional safety, for example by taking the 95th spatial percentile instead of the 90th

spatial percentile (see EFSA PPR Panel (2012) and Tiktak et al. (2013) for procedures on how to deal with parameter uncertainty in the derivation of scenarios).

8.2.3. A systematic approach to scenario development

In this section we provide an example of a structured approach to scenario development based on the development of realistic worst case exposure scenarios (EFSA PPR Panel, 2012; Tiktak et al., 2013). This approach consists of seven steps (see Figure 11):

i. The first step is the problem formulation. As described in section 4.1, this requires a clear definition of the area and time period to be modelled.

ii. The second step is to compile a coherent database covering the area and time period identified in (i). Existing data are transformed to common formats and projections and quality checks are made.

The choice of scenario parameters selected from step (ii) must provide the input variables to support the processes present in the model used to represent the system of study. This will include relevant information on abiotic and biotic components which need to be included. To be able to calculate distribution functions for the entire area of use of a pesticide, simulations need to be done for the whole area. This is not always possible for complex models since the data for these models are not always available. An alternative is then to run a simple model that is compatible with the full model (a so-called metamodel). See Tiktak et al. (2006) for an example of such a model on the exposure side.

During this step, scenarios are selected that meet the target spatial and temporal endpoint. If the same scenario is used for different substances, the modeller must demonstrate that the scenario is sufficiently conservative for all these substances (see EFSA PPR Panel (2012) for procedures). At this point, an evaluation of the necessary spatial scale and the fit between this and the endpoints needs to be undertaken.

i. Once a scenario is selected, appropriate soil, weather, crop, management and community parameters need to be assigned. This now forms the environment scenario in the absence of the pesticide.

ii. Once the environment scenario has been developed, pesticides with specified characteristics can be applied using defined application schedules.

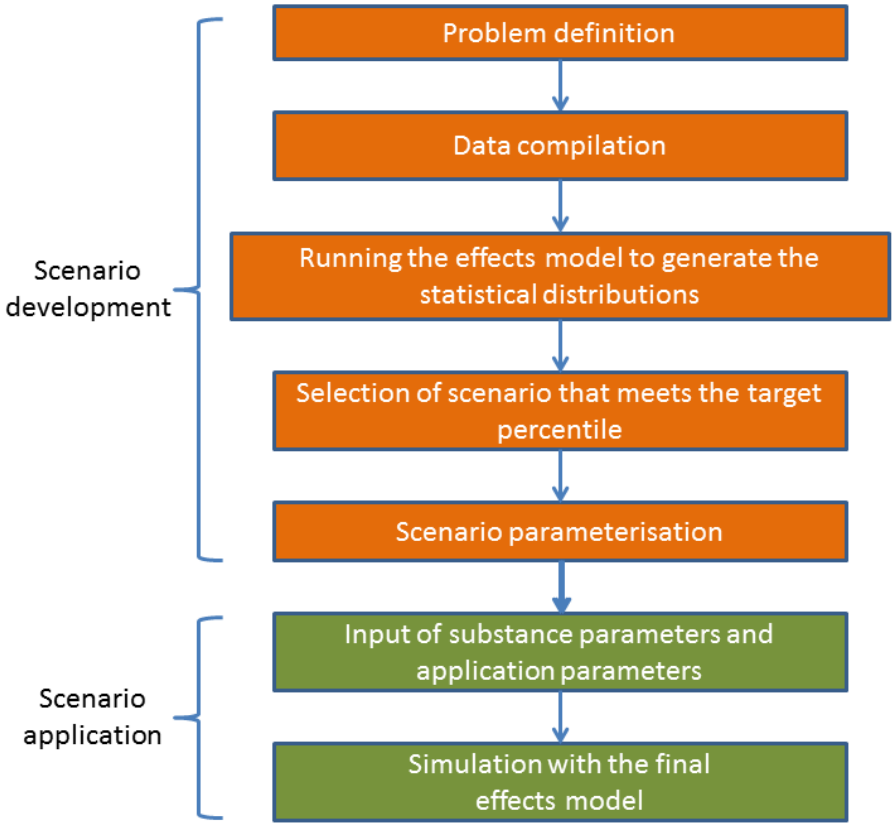


Figure 11: Illustration of the process to develop and apply environmental scenarios based on percentiles of the model results. The process has to be carried out for each relevant effect endpoint (for example, baseline and toxic standards as well as for the product to be tested)

8.2.4. Integration of exposure and effects

Typically in current practice in risk assessment (i.e. not including ecological modelling), the exposure and effect assessments are done separately. The results are then combined to describe the risk. As discussed for mobile organisms above, the disadvantage of this procedure is that it is not known a priori whether a certain percentile of the exposure concentration occurs at the same location as the same percentile of effects. Furthermore, many of the current exposure assessment frameworks (e.g. the FOCUS surface water scenarios (FOCUS, 2001)) are not spatially distributed.

Integrated development of environmental scenarios would improve this situation and would typically be done for higher tiers, for instance simulating the intended use (i.e. good agricultural practice, GAP) at the landscape scale and including fate and exposure, as well as the ecotoxicological model as directly simulated components (Topping et al., 2005). This requires close co-operation between fate modellers and eco(toxico)logists. The advantage of this holistic procedure is that the environmental scenario applies directly to the endpoints used for assessment in a given area.

In those cases where exposure scenarios are already available, the modeller might want to use these scenarios (e.g. edge-of-field surface water scenarios (FOCUS, 2001)) and scenarios for in-field exposure of organisms living in arable soils (EFSA PPR Panel, 2012). These exposure scenarios could be used to provide the exposure part of the environmental scenarios. Note, however, that the currently available exposure scenarios are not developed for the landscape level and are therefore suitable for use only in effect modelling studies in which effects at landscape scale are not considered, e.g. population at edge of field, which is the current specific protection goal for aquatic invertebrates. If

existing exposure scenarios are used, the modeller must demonstrate their suitability for meeting the overall specific protection goal.

In the future applying the percentile approach to exposure and effects modelling could be achieved by evaluating a large number of environmental scenarios representing the range of conditions across the geographical scale relevant for the risk assessment, and ranking them according to their effects. This would allow identification of percentiles of effects distributions.

Where the specific protection goals have not yet been defined as percentiles, it might be necessary to express the output from the model with alternative descriptions, for example the protection goal for reproductive risk to birds and mammals is that there should be no long-term repercussions on abundance and diversity. This does not specify the magnitude or duration acceptable for sub-lethal effects or the degree of certainty. A population model might be able to tell that in 90% of cases there is a 10% or less reduction in number of offspring with no difference in population (all ages) by 18 days after application. In the absence of a protection goal specifying what magnitude, duration and certainty is acceptable, it might also be helpful to express the result with different values for each in turn, for example in 95% of cases there is a 10% or less reduction in number of offspring with no difference in population by 24 days after application or in 90% of cases there is a 5% or less reduction in number of offspring with no difference in population by 35 days after application.

For example, in the bee guidance document (EFSA, 2013), it was acknowledged that by defining a certain percentile exposure assessment goal (e.g. 90th percentile) it is meant that 90 % of all colonies at the edge of a treated field in one regulatory zone should be exposed to less than what is assessed in the risk assessment. For 10 % of the colonies at the edge of a field in the regulatory zone, the exposure could exceed what was assessed in the risk assessment. For these colonies the protection may not be achieved for substances which are highly toxic to bees (e.g. effects could exceed negligible effects). It was proposed to base the exposure estimates at the 90th percentile, as is done for other groups of non-target organisms. However, there was also the suggestion to have a more conservative exposure assessment goal, such as the 95th percentile. The current version of the bee guidance document is based on the 90th percentile, but if risk managers choose a higher percentile then the corresponding exposure values would need to be changed in the final version of the guidance document.

9. Model analysis

Model analysis includes sensitivity analysis (section 9.1), uncertainty analysis (section 9.2) and comparison of the model output with observed data (section 9.3). Model analysis is a key element of the procedure to setup the model, as it provides insight into the behaviour and performance of the model. In the earlier iterations of the modelling cycle, it provides information on how to improve the model. For the regulatory model, it gives information on how much confidence can be placed into the modelling results.

9.1 Sensitivity analysis

A sensitivity analysis assesses how sensitive model outputs are to the values of the model parameters, including input parameters such as those that specify the environment in which the model is run. Sensitivity analyses may also investigate model structure. The results of sensitivity analyses may indicate problems with the model so that the conceptual model has to be reformulated as part of the modelling cycle. The sensitivity analysis helps to increase the understanding of the model behaviour and to separate influential from non-influential parameters. It is closely linked to the uncertainty analysis (see section 9.2). The values of parameters used in a model are always uncertain to a lesser or greater extent, and may vary under differing conditions; hence uncertainty is the main reason that sensitivity analysis is important when assessing a model for regulatory use. For uncertain parameters the sensitivity analysis will provide information on the robustness of the optimal solution chosen (i.e. set of parameter values that provide the best fit to calibration data), and what the consequences would be for decision makers if they used different parameter values.

The simplest sensitivity analysis considers ‘one factor at a time’. One model parameter is varied at a time while keeping all others fixed, and the effect on the model output is recorded. An example is provided in the top two rows of Figure 12. The values used for the parameters may be derived, e.g. from the literature, expert knowledge or parameter calibration. The simplest sensitivity analysis consists in changing one parameter at a time by a certain ratio (e.g. plus and minus 10 %). The quotient of relative change in the parameter value and the relative change in model result for each parameter gives a measure of the relative sensitivity of the model to this parameter, and can be used to rank the parameters as a function of their relative contribution to the model outputs, as shown in the table at the top right in Figure 12. Indicating the rank of the parameters allows rapid identification of those to which model outputs are most sensitive. An elaboration of this approach is to report the changes in model outputs over a range of parameter values, as in the middle row of Figure 12. This may be useful if outputs change non-linearly as the focal parameter changes.

Local sensitivity analysis – 10%

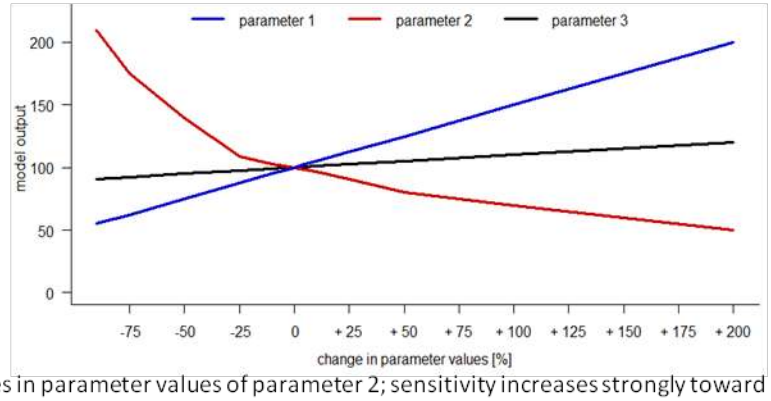
- Set all parameters to reference values, vary one parameter at a time by ± 10%
- Calculate sensitivity as relative change in model output divided by relative change in parameter value , report as table

Para-meter	Description	Reference value	Sensitivity output A	Rank output A	Sensitivity output B	Rank output B
m_b	Baseline mortality	0.2 y^{-1}	2	2	0.5	3
m_t	Acute mortality	0.1 d^{-1}	5	1	1.2	2
b	Average # offspring	20	0.1	3	2	1

- Conclusions:
  - Model output A: very sensitive to changes in the acute mortality (rank 1), relatively insensitive to the number of offspring (rank 3).
  - Model output B: most sensitive to changes in the number of offspring, least sensitive to the acute mortality.
  - Model output A is more sensitive than output B: maximum relative sensitivity is 5 compared to 2.
  - If the final conclusions were mainly based on model output A, values of the acute mortality should be associated with low uncertainty.

Local sensitivity analysis – range of values

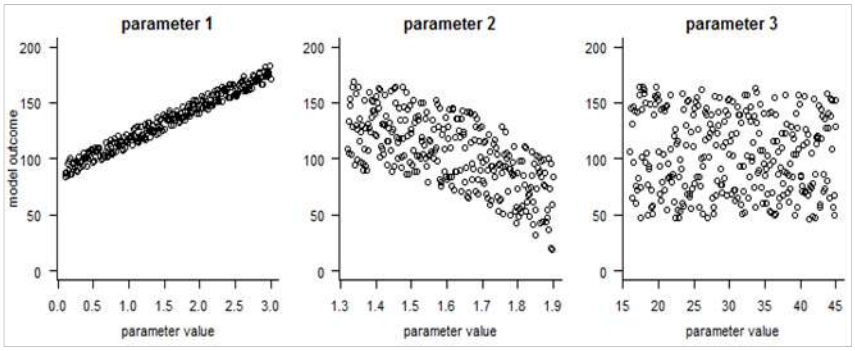
- Reference parameter values, vary one parameter at a time over a reasonable range of values (e.g. ± 2%, 5%, 10%, 20%, 50%, 100%, 200%)
- Report as plot of model output vs. change in parameter value (or absolute values)
- Conclusions:
  - Model output sensitive to parameter changes over the whole range of parameter 1; values of parameter 1 should be relatively certain.



  - Model output little sensitive to small changes in parameter values of parameter 2; sensitivity increases strongly toward larger changes in parameter value. Uncertainty of parameter 2 is important, both regarding the potential range of the parameter values and the likelihood of extreme values.
  - Model output insensitive to parameter 3; hence, if parameter 3 was uncertain, it would not lessen the confidence into the model much.

Regional sensitivity analysis – random sampling

- Probability distribution for parameter **values**, draw random samples for parameter values
- Plot model output vs. parameter value



- Conclusions:
  - Model output sensitive to changes in parameter 1 (left); low scatter, hence parameter 1 has a stronger influence than the other parameters; parameter values of parameter 1 should be certain.
  - Model output sensitive against parameter 2 (in a slightly nonlinear manner indicated by the bend in the point cloud); higher scatter, hence parameter 2 not as dominant as parameter 1 as also the other parameter values influence the outcome; values of parameter 2 should be relatively certain.
  - Parameter 3 with no influence on the model results as there is no trend visible in the point cloud. Hence, its values might be associated with higher uncertainty without compromising the confidence into the model.

Figure 12: Examples of sensitivity analyses. The simplest approach in a local sensitivity analysis varying parameters by, for example, 10 %, and recording the effects on key model outputs. The results can be shown in a table, as in the example in the top row. A local sensitivity analysis may also be conducted over a wider range of parameter values, and then a series of graphs will be needed to portray the results, as in the example in the middle row. More complicated to conduct and present is a global sensitivity analysis, of which an indication is given in the bottom row

The sensitivity analysis should preferably be conducted for all model parameters, but it may not be feasible to include all input variables. Screening can be used initially to identify particularly sensitive parameters and processes in the model before embarking on a full sensitivity analysis. A sensitivity analysis should at least be conducted in the baseline assessment (control) and worst-case assessment (toxic standard, see section 12.2). When possible, an intermediate situation should also be investigated. If several model output variables have to be considered to evaluate risk (e.g. overall abundance, size classes), then a sensitivity analysis should be conducted for all these output variables. Furthermore, it has to be taken into account that the sensitivity depends on the time point chosen for the sensitivity analysis (e.g. spring or autumn, exponential growth phase of near carrying capacity). Therefore, the sensitivity analysis should be conducted at different time points. The sensitivity to initial conditions also needs to be analysed.

It should be noted that there are shortcomings to sensitivity analysis conducted one step at a time because they cannot assess interactions between parameters (Saltelli and Annoni, 2010) and so it may be desirable to supplement one step at a time with a global analysis (example in Figure 12, bottom panel). This is helpful if model outputs depend strongly on the values at which non-varied parameters are fixed. In contrast to local sensitivity analysis, global sensitivity analysis does not depend on the choice of the reference parameter values and takes into account parameter interactions and non-linearities. It investigates the model sensitivity not only at the reference values, but over the whole range of possible parameter values. One approach is to sample from the probability distributions of parameter values, run the model for all samples and construct scatterplots of model results versus parameter values for each parameter. Such scatterplots visualise the dependence of model outputs on each parameter while including the random influence of all other parameters. Plotting model outcome versus two parameters at a time (Figure 12, bottom panel) can help to identify two-way interactions of parameters, and these can also be assessed from a table of pairwise correlations between parameters.

9.2. Uncertainty analysis

The uncertainty analysis of a model aims at describing and evaluating the different factors that make a model result uncertain. Preferably, the uncertainty stemming from different sources should be propagated to the model output. This way, the probability of a specific outcome for a parameter set can be given. Model uncertainty should be assessed not only for the final model, but also during the analysis step in the modelling cycle and to help improve the model.

9.2.1. Sources of model uncertainty

There are different sources of uncertainty, i.e. (i) uncertainty originating from parameter uncertainty; (ii) uncertainty originating from uncertainty of input variables (e.g. temperature series); and (iii) uncertainty arising from model structure. There are other sources of uncertainty as well arising from e.g. stakeholder's view on the problem (Hisschemöller and Hoppe, 1996). This section, however, deals only with parameter uncertainty, input variable uncertainty and uncertainty due to the model structure. Guidance on how to deal with stakeholder's views can be found elsewhere (e.g. MNP, 2008).

The first source of uncertainty is the uncertainty in the model parameter values. Strictly speaking, the values of most parameters are uncertain. This originates from measurement errors, biological variation and extrapolations from one species to another or from one environment to another. Variability or measurement error may be high in experiments reported in the literature, and estimates may be based on laboratory experiments conducted under conditions not relevant for field-scale models (e.g. 20 °C and high food availability).

A second source of uncertainty lies in input variables that drive the model, for instance weather conditions, nutrients or land use. Such input variables are also associated with measurement errors, and there could be gaps in measurements, or the resolution may not be as fine as desired. Similarly, the initial conditions are often unknown or uncertain (e.g. starting population size or spatial distribution).

A third source of uncertainty originates from the model structure. A model is always a simplification of the real world; hence it cannot contain every single process in a system. An adequate model should contain the most important processes and leave out processes that do not affect the model results to a large extent. Even when the most important processes are captured, the mathematical formulation could be inadequate (e.g. assuming a linear equation where an exponential equation would be more appropriate to describe the process). In addition, the spatial and temporal resolution may be a source of structural uncertainty. For example, if the timing of reproduction and contamination is important a yearly time step might be inadequate.

9.2.2. Description of uncertainty

The uncertainty in model parameter values for estimated parameters can be described by confidence intervals or credibility intervals. Uncertainty in input variables can be described if it is possible to formulate a statistical model of the measurement process. For example, it could be assumed that temperature measurements that serve as model input were independent and normally distributed. This statistical model can then be included in the calibration process and its uncertainty can be described in the same way as for the model parameters. To cover a range of (unknown) possible input conditions, scenarios with contrasting settings might be helpful. The range of model outcomes under the different scenarios then gives a measure of the uncertainty due to these external inputs.

The uncertainty in model structure is very difficult to quantify. In principle, different model structures could be compared with measurements and "calibrated" (see, for example, Tiktak, 2000). However, in practice the model structure should be developed iteratively in the modelling cycle. The final model should then be the one with the lowest structural uncertainty.

9.3 Evaluation of models by comparison of model outputs with independent data

It is necessary to consider how well model outputs fit with relevant independent data. Two examples of how this might be done are shown in Figure 13.

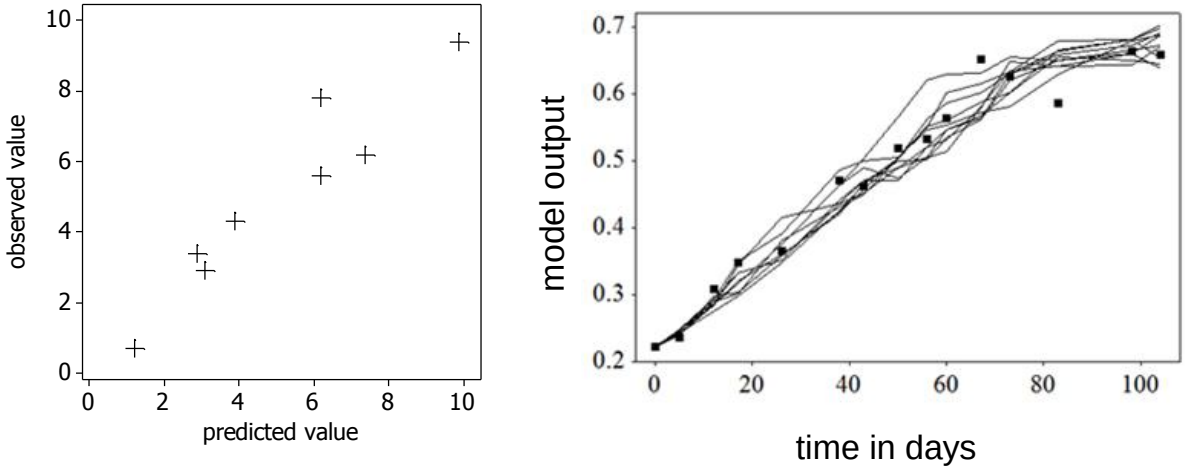


Figure 13: Two examples of how model fits to data may be presented for visual inspection. In the left figure the outputs predicted by the model are plotted against data. In the right figure a model output (e.g. population size or individual body mass) is plotted against time. Here the data are represented by

points and the model outputs by lines. Model outputs may vary between runs, as here, if the model is stochastic

No model can perfectly describe the system under consideration but model outputs should provide an adequate match to relevant data patterns. Perfect matches are not expected because (i) the collection of data is always subject to measurement errors; (ii) some data might not be available, e.g. environmental measurements such as food availability or local density of focal species, which affect model predictions may not be available for observed populations; (iii) models are necessary simplifications of reality, and some aspects of relevant biological processes may not be fully represented. Because perfect matches are not expected, it is necessary to consider carefully how well that the model predictions match the data.

The types of data available for model validation vary. Some relevant data patterns are qualitative but others include quantitative data. Examples of qualitative patterns include emergent properties such as density dependence, population cycles and age structure of populations. In these cases, qualitative comparison with general patterns is required, part of the process that has been termed pattern-oriented modelling (Appendix A). The ways in which qualitative data patterns can be used for the evaluation of models are described in section 10.

In contrast to the qualitative data patterns, some relevant data patterns may be quantitative. Examples include data from laboratory, semi-field and field studies. A combination of visual methods and statistical methods seems to be appropriate for comparing data and model outputs (Janssen and Heuberger, 1995). Many statistical measures to describe the quality of the fit exist describing various aspects of the fit (e.g. some indicators describe the average behaviour of the statistical population while other indicators look at the individual level; see Janssen and Heuberger (1995) for an overview). Which indicator is most appropriate depends on the purpose of the validation and therefore the modeller should carefully select the appropriate indicator. An alternative to classical indicators is the so-called *factor-of-f approach* (Parrish and Smith, 1990). This method acknowledges that not only the model results but also the data with which the model is compared are uncertain (see Tiktak et al. (1999) for an example application). The concept of acknowledging that both the data and the model are uncertain is also built into the pattern-oriented modelling (POM) cycle suggested by Topping et al. (2010) (Appendix A).

The endpoint that is compared with measurements does not need to be the assessment endpoint, but any reasonable model output could be used. Often, summary statistics might be needed to make this comparison. For example, it might not be necessary to match the exact spatial movement pattern of an animal: the mean distance flight per day can be calculated and used as a summary statistics (example taken from Hartig et al., 2011). Another challenge in assessing the fit of predictions occurs when more than one dataset is relevant to model assessment. For example, there may be data on how body sizes change with time and also data on daily individual reproduction. How are fits to two such disparate datasets to be assessed in a single analysis? One possibility might be to use the average of the two R^2 values calculated for each dataset separately. Another possibility is to use integrative models which can account for different data types in a single coherent analysis. For example, it has been suggested that dynamic energy budget models can be used to mechanistically link the effects on various traits over the entire life cycle of an individual (Jager et al., 2004; Ashauer et al., 2011). The available evidence relating to model evaluation should be well documented and made available to the regulator who needs this information to judge whether the model is fit for purpose (section 10.5).

10. Evaluation of the model by risk assessors (authorities, companies, consultancies)

This section aims to provide recommendations to risk assessors on how to practically conduct the evaluation of a model to be used as support for risk assessment. Note that it is not the intention that the risk assessor performs the evaluation himself. Instead, the risk assessor should base his evaluation on documentation provided by the modeller. A summary checklist for model evaluation by the risk assessor is provided in Appendix B.

Model evaluation has to take place in every stage of the modelling cycle and is therefore a continuous process. This is illustrated in Figure 3, which shows the modelling cycle together with the required evaluation activities. The most important questions that the risk assessor should pay attention to are:

i. Is the model based on commonly agreed scientific principles and/or are these principles published in the scientific literature?

ii. Is the general behaviour of the model plausible?

iii. Are all steps of the modelling cycle sufficiently well documented?

iv. Is the correspondence with available independent observations acceptable?

v. Is the model fit for regulatory purpose (in other words, can it be used to provide an answer to the question posed in the problem definition)?

10.1. Evaluation of the supporting data

Data plays a key role in all steps of the modelling cycle (Figure 2). The quality of supporting data needs to be clear and therefore the following items should be evaluated (see section 3 for details):

i. Are the data fit for purpose in view of the problem description? Items that could be evaluated are the spatial extent of the dataset (do the data apply to the region that is intended to be modelled?), the spatial scale of the dataset (data developed for the European scale might not be suitable for developing a field-scale model), the temporal resolution of the dataset (concentrations measured on a monthly basis might not be suitable for detection a peak concentration) and the measurement method (is it based on commonly agreed guidelines?).

ii. Has the quality of data used as support for the modelling been considered and documented by the modeller? For example, have experimental data been collected in a reproducible and transparent manner? Have references been provided when using external datasets? Has the modeller provided a justification for the use of these datasets?

iii. Have all available data (including publically available data) been used? If not, is there a justification why this information has not been used?

10.2. Evaluation of the conceptual model

The modelling cycle starts with the problem description, which should include the requirements for the model to be developed (section 4). More specifically, the risk assessor should evaluate the following items:

i. Are the SPGs sufficiently well described?

ii. Are the spatial and temporal scales described?

iii. Is the modelling approach justified?

iv. Is it clear in which tier the model is intended to be used?

Based on the problem description, the modeller designs a conceptual model ("model formulation"). The conceptual model provides a general and qualitative description of the system to be modelled and insight into the processes and their interactions and interdependencies, which are often summarised in diagrams. Both the diagrams and the text provided by the modeller can be used to evaluate the conceptual model. Whatever format is used, the provided information should be sufficient to ensure understanding of the structure of the system to be modelled. This is particularly true for complex

models that often include a sub-set of different modules (or sub-models). Information provided by the modeller should describe not only the modules themselves, but also the relations between these modules.

The first aspect to evaluate is the a priori capacity of the model to answer the risk assessment question defined in the problem description. This evaluation is related to the particular context and aim of model, to the ecological context, to the amount of available knowledge for model development and to the kind of decision to be based upon the model. Therefore, the following items could be evaluated:

i. Are the specific protection goals sufficiently well addressed by the model? The risk assessor should check that the model entities, attributes, temporal scales, spatial scales and level of certainty are relevant to the risk assessment question under study (EFSA PPR Panel, 2010).

ii. Are the modelling endpoints (e.g. individual mortality, population size, recovery pattern) relevant to the specific protection goal? The evaluator should check that they provide additional useful information to support the risk assessment of the substance under study, given the problem formulation (EFSA, 2013b).

iii. Is the conceptual model logical? The evaluator could perform a reality check of the model concepts towards (i) the biology, ecology and behaviour of the species (e.g. the population cannot recover from local extinction without external migration); and (ii) the properties of the substance under study (e.g. expected toxicokinetics). Are the spatial and temporal scales of the model suitable in regard of the problem formulation? The chosen spatial and temporal scales are crucial for the ecological relevance of the model. Regarding temporal scales, the response time of the various components of the model (e.g. exposed vegetation types, which constitutes the habitat/food source and species to be protected) generally varies greatly. Therefore, the evaluator should check that a suitable timescale has been used when developing (and applying) the models. Regarding spatial scales, the plot or landscape size considered in the model affects population responses and recovery. The stressors are dynamic in space and time (e.g. crops rotate, and the pesticide moves with them. Therefore, the evaluator should check that a suitable spatial scale has been used when developing (and applying) the models, in regard of the scenario under study. When the ecological model is deployed as part of a model chain (e.g. exposure model and effect model), or when it comprises several modules, it must be thoroughly checked that the space and timescales used in the other models or modules are applicable to the ecological model.

The second aspect to evaluate is the relevance of model assumptions about the system composition, organisation and functioning. Therefore, the following items should be evaluated:

i. Are the processes included in the model relevant to the addressed issue? Were scientific references that support model assumptions provided? The evaluator should check that the model includes:

    a. the relevant components and processes at play in the biological system under study;

    b. the internal and external factors that are expected to be the main drivers of the system functioning;

    c. the rationale that supports biological, ecological and toxicological assumptions of the model should be evaluated (and the adequacy and reliability of supporting literature references should be evaluated).

ii. Are the links between different processes to the variables logical? The evaluator should check that variables and processes included in the conceptual model are logically linked (e.g. do the birds indeed built their nests before breeding?).

10.3. Evaluation of the formal model

In the next step of the modelling cycle, the conceptual model is translated into a formal or mathematical model (model formalisation). The evaluator should first check if the formal model is sufficiently well documented. A description of the formal model should at least contain the following items:

i. a description of the variables and parameters including their meaning and unit;

ii. a description of the most important mathematical equations;

iii. justification of the most important model assumptions (e.g. why first-order kinetics).

Based on this documentation, the evaluator could further assess the following items:

i. Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data? Several types of models are generally suitable to address a modelling question: the choice of the modelling approach is tightly linked to the appropriate level of complexity represented in the model.

ii. Have any references or documentation that support model equations or algorithms been provided? The evaluator should check whether literature references supporting the model are available. If desired, a detailed checking of equations and algorithm can be performed by the risk assessor, based upon these references. This refined evaluation should consider the internal consistency of the model: if the units of variable and parameters belonging to the same equation (or between related equations) do not match, there might be consistency issues in the model (e.g. conservation of mass across biological levels of organisation is not respected). The refined evaluation should also consider the use of adequate mathematical functions to describe processes (e.g. non-linear processes should not be modelled using linear equations).

10.4. Evaluation of the computer model

The next step is to convert the mathematical model into a computer model. This step is referred to as model implementation. It is a rather technical process, which requires a high level of specific expertise. The following items could be evaluated:

i. Is there a comprehensive and transparent description of the computer model? This documentation should include (i) a description of the most important algorithms and numerical methods; (ii) a description of the general setup of the model including flow charts; and (iii) description of those model parameters that are part of the computer model (in particular those parameters that are fixed in the model code).

ii. Is the computer code easily readable and is it available? A readable and accessible computer code will increase transparency in the regulatory process. In particular, (i) is there a consistent convention for naming of modules, functions, parameters and variables and (ii) is the code sufficiently well internally commented?

iii. Is it demonstrated that the mathematical model is correctly implemented? Is the model verification described? The evaluator could perform the following tests: (i) run the model and compare the so-obtained model results with results from existing simulations (provided by the modeller) and/or analytical solutions; (ii) check for internal consistency of the model results (e.g. is the mass balance closed?); and (iii) perform reality checks. Such reality checks could include a check for physical impossibilities such as negative concentrations and a global behaviour test. In the last test, the evaluator could for example test what happens to a population if the mortality of individuals would be increased to 100 % or when the birth rate of individuals would be set to zero.

10.5. Evaluation of the regulatory model

In the following step of the modelling cycle, model parameter values are estimated and the computer model is combined with one or more environmental scenarios (this step is referred to as "model set-up"). The so-obtained regulatory model consists of the following components: (i) the computer model; (ii) programs for pre- and post processing (often made available as a graphical user interface; (iii) model parameters; and (iv) the environmental scenarios. Notice that the model can only be run when case specific data, such as pesticide properties and application regimes, is input into the model.

10.5.1. Evaluation of the environmental scenario

As described in *Section 8.3*, environmental scenarios are combinations of abiotic parameters, environmental parameters and biotic parameters to be used in modelling. When used for regulatory purposes, the scenarios should represent realistic worst case conditions as defined in the problem definition. As the level of conservatism of the modelling is to a large extent determined by the scenarios, the scenarios used by the modeller should be evaluated with great care in the context of the problem formulation. Based upon information provided by the modeller, the following items should be evaluated:

i. Is the scenario sufficiently representative for the risk assessment under consideration?

   a. The scenarios should reflect the characteristics of the area under consideration.

   b. The scenarios should be appropriate for the entity to be modelled, e.g. for populations and communities a landscape should be modelled, whereas for individuals a smaller scale may be appropriate.

ii. Has the modeller justified the general biological, abiotic and environmental parameters that constitute the scenario? Refer to section 8.3.1 for details.

iii. Has the modeller ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

iv. Are the spatio-temporal dimensions of the scenario in line with the problem definition? The spatio-temporal dimensions of the scenario determine the overall level of protection of the scenario. The risk assessor should pay attention to the following (see section 8.3.2 for details):

   a. What spatial unit is considered (e.g. field, edge-of-field water body or landscape) and is the spatial scale of this unit in line with the specific protection goal for the species under consideration?

   b. What spatial statistical population of the units is considered? This is the domain of the scenario, for example a country or an EU regulatory zone). Is the statistical population in line with the problem definition?

   c. What temporal statistical population is considered?

   d. Is the conservativeness of the scenario in line with the problem definition? The conservativeness is to a large extent determined by the percentile to which the scenario applies.

   e. What certainty of the model predictions should be reached? The modeller should, preferably by uncertainty analysis, indicate the uncertainty of the predictions and indicate the likelihood that the threshold as indicated in the problem definition is exceeded.

Is the level of conservatism placed into the scenario adequate in regard of the problem formulation? Whereas realism is expected from the model, conservatism level should be placed in the modelling scenario. The appropriate level of conservatism depends on the problem formulation, and should be checked by the evaluator. Indeed, regulatory models can be used at all tiers of the risk assessment scheme, but regulatory models and model adjustment factors used as a support of the first tier should be more conservative than models used as support of higher tiers. Calibration of lower tiers against the reference tier may be necessary (see, for example, EFSA PPR Panel, 2012). In this respect, the evaluator should check if the ranges of input values within the ecological scenario have the adequate level of conservatism (from realistic to worst case).

10.5.2. Evaluation of parameter estimation

Parameter estimation is an important step of model development, since it influences both the capacity of the model to reproduce certain patterns and the reliability of quantitative model outputs, which are critical features regarding the usefulness of the model for supporting risk assessment. In this context, the following items should be evaluated:

i. Has the model parameter estimation been adequately documented? The evaluator should check if the following information has been provided for each model parameter: (i) definition; (ii) units; (iii) parameter estimation method (and corresponding literature reference when the value was derived from the literature); (iv) parameter estimated value and the associated uncertainty; and (iv) common range of parameter variation, if this information is available.

ii. Was the quality of the data supporting parameter estimation (literature or experiment) sufficient? Most of the data should be related to the situation to be simulated in terms of species, region, season, etc. Data obtained in different species or conditions can be used for parameter estimation, but these should be fully justified by the modeller. The lack of data explaining the use of this alternative data should have been clearly identified by the modeller. The choice of alternative data should have been justified and the consequences of using these alternative data should have been considered in the parameter uncertainty estimation.

iii. Were the estimated parameter values realistic? Parameter values should be compared with the range of biological variability, when it is known (which data should have been provided by the modeller). If these data are not available, the evaluator should check that the parameter values are not out of a range that makes biological sense (e.g. a mite weight cannot be 10 tonnes). If the calibrated estimate of a parameter is not realistic, and the model is very sensitive to that parameter, then the model might not be useful as a support for ecological risk assessment.

iv. Has the version number of the database from which data has been taken been documented?

10.6. Evaluation of the model analysis

Model analysis includes sensitivity analysis, uncertainty analysis and evaluation of the model by comparison of the model output with observed data (section 9).

10.6.1. Sensitivity and uncertainty analysis

Sensitivity analysis identifies sub-sets of parameters that have strong effects on the model outputs. We thereby learn which processes are most important for further considerations. The following items could be evaluated by the risk assessor:

i. Has the sensitivity analysis been adequately documented? Methods used and supporting references should be provided by the modeller. The risk assessor could further check whether these methods are appropriate in view of the problem formulation.

ii. The sensitivity analysis may give different results for different parameter settings. It is therefore important to evaluate whether the modeller has documented for which situations the sensitivity analysis is valid.

iii. Have the results of the sensitivity analysis been presented and discussed so that they allow identification of the most sensitive parameters and forcing variables in the model?

Uncertainty analysis aims to identify how uncertain the model output is (often expressed as confidence intervals). This is important information for both the risk assessor and the risk manager, who may consider this when making decisions, particularly in borderline cases. The following questions are relevant:

i. Has the uncertainty analysis been adequately documented? Methods used and supporting references should be provided by the modeller. The risk assessor could further check whether these methods are appropriate in view of the problem formulation.

ii. Have the parameters for which the model outputs are sensitive been included in the uncertainty analysis? This information should be obtained from the sensitivity analysis.

iii. The uncertainty analysis may give different results for different parameter settings. An important question to evaluate is, therefore, whether the modeller has documented for which situations the uncertainty analysis is valid.

iv. Have the results of the uncertainty analysis been presented and discussed so that they allow the identification of the parameters and forcing variables that contribute most to model uncertainty.

v. Have confidence intervals been calculated and has this information been used in further model use?

10.6.2. Evaluation of the model by comparison with measurements

The performance of the model is usually evaluated by comparing relevant model outputs with independent measurements (see section 9.*3* for details) which is also often referred to as model validation. The following items needs to be evaluated:

i. Have the performance criteria for the model been predefined in the problem definition?

ii. Are the model outputs that are compared relevant in view of the problem definition? If, for example, the peak concentration of a pesticide is important, it does not make sense to evaluate the model on the basis of a time-weighted average.

iii. Have the data with which the model is compared been subjected to quality control and is a description of the data available? Are the data available in an appendix?

iv. Is the dataset that has been used for model evaluation relevant in view of the problem definition? This evaluation could also be conducted for endpoints other than the core endpoint of the modelling exercise. The evaluation is more convincing when simulations have been compared with data obtained under different environmental conditions. It is preferred that the data are derived from targeted field studies involving pesticide treatments. If such data are not available, alternative data could be used to evaluate parts of the model. This should, however, be justified by the modeller.

v. Has the performance of the model been reported in an objective and reproducible way? Statistical indicators should be used in addition to visual comparison of data series and model outputs. The choice of these indicators should be justified and discussed in view of the

problem formulation (see Janssen and Heuberger (1995) for more information on statistical indicators).

10.7. Evaluation of model use

When using the model for regulatory purposes, the risk assessor has to input the pesticide properties (relevant to both exposure and effects, thus modelling risk in relation to specific protection goal) and the application pattern and/or regime of these pesticides. How this is carried out must be described in a user manual, which should be provided by the modeller. The user manual should also contain a description of the model outputs. The manual could be made available in print, electronically or as a help file incorporated in the software itself. More specifically, the following items should be described by the modeller (see section 11 for details) in order to reduce model user uncertainty:

i. Description of required skills of intended users.

ii. Description of the installation process.

iii. Description of additional (standard) software (including version numbers) that is needed to run the model (e.g. R, python, etc.);

iv. Description of how the model works (overview screens, how to operate the model, etc.);

v. Description of all input files or screens of a graphical user interface, including a description of individual data fields (meaning, unit, default value, ranges);

vi. Description of how the model inputs can (or if applicable should) be derived (literature, calibration, expert judgement or prescribed procedures such as OECD guidance, EFSA guidance, etc.);

vii. Description of all output files or screens of a graphical user interface, including a description of individual data fields (meaning and unit).

11. Model documentation

All mechanistic effect modelling submitted as part of dossier for pesticide authorisation will need to be documented. Section 11.1 gives an overview of existing protocols. Section 11.2 gives an overview of the model documentation needed. The Panel proposes that, in addition to the model documentation, a summary sheet is prepared in a standard format, which references sections in the model documentation (section 11.3). This allows the evaluator to easily find the information he or she needs. The advantage is that the format of the model documentation does not need to be prescribed as the summary sheet will allow the risk assessor to find the necessary information.

The documentation for the model may be presented as a single document covering the model and its use or as a report on the model use with separate model documentation. In the latter case, the documentation may be formatted to follow a standard documentation protocol if applicable. A single document covering the model and its use may be easier for the risk assessor to use, but separating the model documentation and a report covering its use has the advantage if a model is used multiple times. This is because the documentation can remain the same, with separate reports on each use of the model.

11.1. Standard protocols to document a model

Some standard protocols to model documentation exist. Using standard protocols would have the advantage that the reader knows the structure and hence where to look for information. However, as discussed above, the use of any of these protocols is not prescribed.

For smaller models (e.g. TK-TD, population models), a commonly used format is the "overview, design concepts, detail" (ODD) protocol. It was proposed by Grimm et al. (2006, 2010) and deals with individual based models and population models. The ODdox protocol (www.doxygen.org) combines the ODD approach to provide the overview of the model with model code descriptions generated directly from the computer code itself. This has the advantage that the documentation is automatically updated when the computer code is changed (Topping et al., 2010).

Schmolke et al. (2010) proposed the TRAnsparent and Comprehensive Ecological model documentation framework (TRACE), which extends ODD to document not only a model, but also the process of formulating, implementing, setting up, evaluating and using a model (the "modelling cycle" (section 2; Figure 2). The EPA Guidance on the Development, Evaluation and Application of Environmental Models (EPA/100/K-09/003; EPA, 2009) was developed for environmental decision-making once an environmental issue has been identified; its use is therefore broader than that of environmental risk assessment. It also recommends providing documentation that enables decision-makers and other model users to understand the process by which a model was developed and used.

11.2. Documentation to be provided by the modeller

If the modeller chooses to use a unique format for model documentation, the modeller must ensure that it contains sufficient information for the risk assessor to understand how the modelling was conducted. The modeller must further describe whether the modelling endpoints are in line with the SPGs described in the problem definition. This means that the model documentation should at least contain (see also Table 1):

i. a description of how the model works (the user manual—see section 10.6.2);

ii. a description of the problem definition, including a description of the specific protection goals;

iii. a description of the species to be modelled, including a review of the biology of the species;

iv. a description of available (field) data available for setting up and evaluating the model;

v. a description of the model concepts;

vi. a description of the environmental scenarios;

vii. a description of the model analysis, including evaluation of the model based on comparison of model results with field data;

viii. a description of the regulatory assessment (i.e. a description of the use of a model for a specific case).

As mentioned above, item (viii) could be provided in a separate report. This has the advantage that the model documentation itself does not need to change every time a model is used for regulatory purposes.

11.3. Summary document template

The summary document should be provided by the applicant and should include a brief description of the main elements of the model and reference the main sections of the model documentation (*Table 1*). This allows different model documentation structures to be used, whilst still making it easy for the risk assessor to find the information they require without needing to be familiar with different methods of documenting models. The summary follows the recommendations set out in this opinion, and is based on the modelling cycle.

Table 1: Template of the summary document.

Aspect	Further information
1. Problem definition Context in which the model will be used. Specification of the question(s) that should be answered with the model. Specification of necessary model outputs and protection goals. Domain of applicability of the model Why is the model being used? What protection goal is being addressed? What outputs are required? How was the species chosen? Which other species/groups are being covered by the chosen one(s)? What data will be used to evaluate the model and degree of match to patterns required to be judged adequate.	Reference the sections where this information can be found
2. Supporting data Summary of the key data used in the model for development and evaluation. Assessment of the quality of the data.	
3. Conceptual model Description of the model concepts including a diagram. Identify the main components and processes in the system. How the effects of the chemicals are modelled. How the components and processes are linked.	
4. Formal model Identification of the model variables. Identification of the model parameters. Description of the most important model equations or algorithms.	
5. Computer model Description of the model implementation. Checking the computer model for errors, bugs and inconsistencies in the code. Demonstrate that the computer model performs as indicated by the conceptual and formal models.	
6. Regulatory model—the environmental scenario Description of the environmental scenarios, i.e. the environmental context in which the model is run. Include description and justification of combination of abiotic, biotic and agro-environmental parameters.	
7. Regulatory model—parameter estimation Description of the model parameter estimation. Parameters estimated from the literature—what are the sources and why are these appropriate? Parameters obtained from calibration—how and why this was done?	
8. Regulatory model—Sensitivity and uncertainty analysis Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output. Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain.	
9. Regulatory model—comparison with measurements Description of comparisons of model output with independent data. Demonstration that the model output provides an adequate match to data patterns.	
10. Reality/problem—MODEL use Explanation of how the model conforms to the requirements set in the problem definition. Description how the model works (user manual). Description of the pesticide parameters values used in the model. Description of the specific assessment including a discussion of the most important results.	

Aspect	Further information
11. Reality/problem—conclusion Tie in the results from the modelling with the specific protection goal identified in the problem definition section. Can it be established that it is "clearly established that no unacceptable impact occurs"?	

12. Use of the regulatory model in risk assessment

Before the model is used for regulatory purposes, the risk assessor should evaluate whether it is fit for this. The documentation provided by the modeller, including the summary sheet, is important input for this evaluation. The risk assessor could use the checklist in Appendix B (explained in section 10) to structure his or her evaluation. An important point to determine is whether the model is in line with the specific protection goals (see section 12.1 on the link between model results and specific protection goals).

When using the model, the risk assessor has to input the pesticide properties and the application pattern and/or regime of these pesticides (section 12.*2*). How this is carried out must be described in a user manual, which should be provided by the modeller (section 10.7). Based on the modelling, risk assessors should finally decide whether an authorisation of a PPP should be granted. Section 12.3 introduces a tabular approach for qualitative evaluation of uncertainties, which is intended to help in the decision-making.

12.1 Link to specific protection goals

Models are normally used in higher tiers of the risk assessment. Tiered approaches are the basis of environmental risk assessment schemes that support the registration of PPPs under the PPP Regulation (e.g. Campbell et al., 1999; EC, 2002; Boesten et al., 2007; EFSA PPR Panel, 2010). In this context, a tier is defined as a complete effect or exposure assessment resulting in an appropriate assessment endpoint, for example predicted environmental concentration (PEC) or regulatory acceptable concentration (RAC). The concept of tiered approaches is to start with a simple conservative assessment and carry out additional, more complex work only if this is necessary to achieve a more realistic risk assessment (so it implies a cost-effective procedure for both industry and regulatory agencies). The higher tiers should result in protective risk assessment decisions consistent with the SPGs set by the competent authorities. According to Boesten et al. (2007) and Solomon et al. (2008) the general principles of tiered approaches are:

i. Lower tiers are more conservative than higher tiers.

ii. Higher tiers aim at being more realistic than lower tiers.

iii. Lower tiers usually require less effort than higher tiers.

iv. In each tier all available relevant scientific information is used.

v. All tiers aim to assess the same protection goal.

In the risk assessment conducted under Regulation (EC) No 1107/2009, the basic data requirements for the first tier risk assessment (set out in Regulation (EU) No 283/2013) for approval of active substances are strictly defined, in relation to both exposure and effects assessments.

The 'unless' clauses described in the uniform principles (Regulation (EU) No 546/2011) offer the possibility to perform higher tier risk assessments. In most cases, assessing directly whether the use of a pesticide complies with the SPGs would require refined experimental or modelling methods that would not be practical for routine use in a tier 1 risk assessment procedure. Equally, in general, the standardised studies or models used at a tier 1 level do not measure the specific protection goals

directly. The PPR Panel's solution to this is to identify, for each key driver (taxonomic group or other ecological entity), a surrogate reference tier, based on the most sophisticated experimental or modelling risk assessment method currently available that addresses the specific protection goal (Figure 14). This surrogate reference tier will then be used to calibrate lower tiers using simpler methods that are practical for routine use. Note that from a theoretical point of view the actual ecosystems in the landscape are the final reference.

When mechanistic effect modelling is included in the risk assessment the modelling directly addresses the protection goal (for example the output from a population model could be the likelihood that there is no more than a small effect on abundance in the crop lasting weeks). The use of models for risk assessment will be, therefore, not far from other higher tier approaches such as mesocosms and field studies in which the specific protection goals are also directly addressed.

The templates for regulatory risk assessments (draft assessment report, registration assessment report or registration report) set out the basic structure of the assessment, generally covering the toxicity data, the exposure and then the risk assessment. The risk assessment section will include both the lower tier assessment and higher tier refinements using additional data and modelling. Hence it will cover different approaches, from using simple ratios with large assessment factors at the lower tier, to addressing the protection goals directly (with appropriate assessment factors as required) at the higher tier (including, but likely not limited to, the modelling). It may be appropriate for this to relate to the whole higher tier risk assessment including modelling and this should be addressed in the problem definition described in section 4.

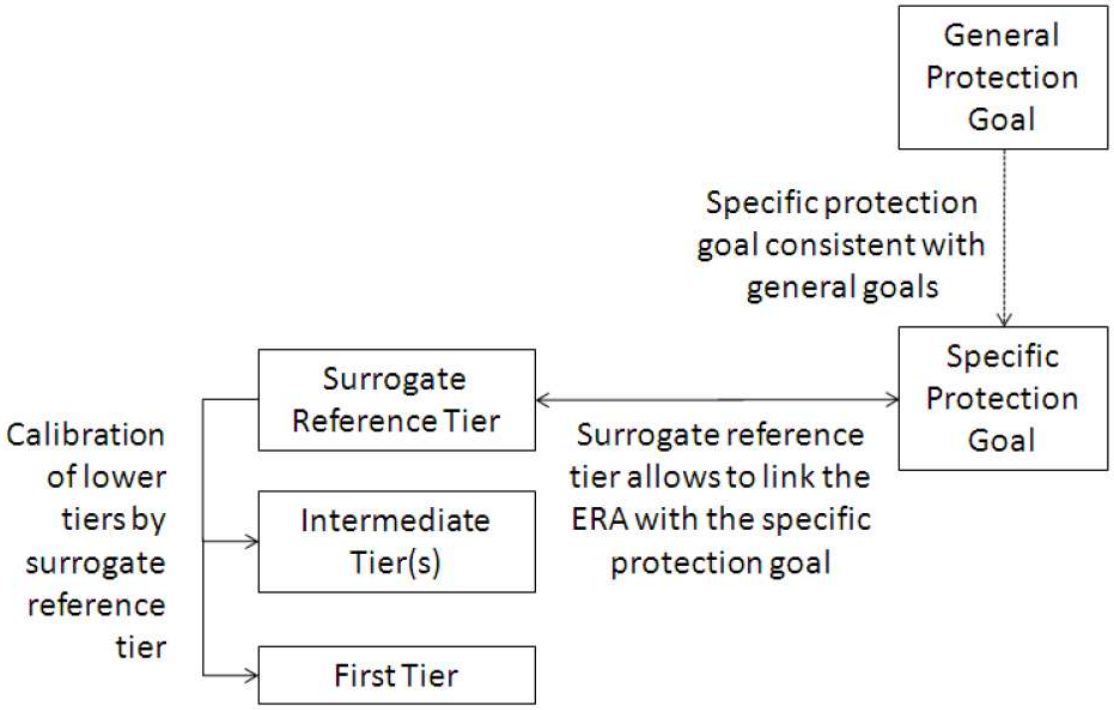


Figure 14: Illustration of the relationship between tiers of the risk assessment process and protection goals in the approach used by the PPR Panel in the guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters (EFSA PPR Panel, 2013)

12.2. Pesticide properties, application regimes and baseline conditions

Pesticide properties can be characterised in many different ways (laboratory experiments, field experiments, etc.), and the way these properties are derived determines the outcome of a regulatory assessment to a large extend (e.g. Boesten, 2000; Tiktak, 2000). The methods used should therefore be

documented and whenever possible harmonised procedures for derivation of these properties should be used.

It is important that the pesticide application regime is carefully considered when carrying out modelling. Ideally the modelling would be done to reflect the proposed GAP, covering the maximum application rate and number of applications and the minimum application interval. Where there are multiple GAPs proposed it might be necessary to restrict the modelling to only certain uses, identified as potentially worst cases. When this is done, it is important to consider that the worst case used at the lower tier, based on simple and highly conservative assumptions, might not be the worst case when mechanistic effect modelling is used for a refined risk assessment.

In addition to modelling the proposed GAP, it might be appropriate to also model multiples of the application rate (for example, 2×, 5× and 10× the proposed rate) in order to demonstrate the margin of safety of the proposed use. This information would be combined with the evaluation of the uncertainty in the modelling to aid decision-making.

In order to demonstrate the level of effect, it will generally be necessary to also run the model without the PPP (as a control). This will demonstrate the level and duration of effects predicted.

It is very important for the risk assessor to know that the model would be able to demonstrate negative effects if they were to occur. This is a similar situation to laboratory and field studies, in which it is necessary to demonstrate that the test system is sensitive to effects from chemicals. This is usually done by including either a toxic standard or a dose of the PPP at a rate that shows clear effects. A similar approach in mechanistic effect modelling will be important for interpreting the results and gaining confidence that the model used is able to demonstrate negative effects should they occur. Therefore, it is recommended that the model is run with a control, several tests of substance at application rates and a toxic standard.

12.3. Qualitative assessment of uncertainties in the risk assessment

This section is intended to help in the decision-making after the model output is used in the risk assessment. Approaches for characterising uncertainty in higher tier risk assessments are covered in more detail in the aquatic guidance document (EFSA PPR Panel, 2013). Here, the main ideas that relate to effect models used in risk assessment are summarised.

Regulation (EC) No 1107/2009 states the following under Annex II criteria for approval of a.s., safeners and synergists under 3.8, Ecotoxicology, point 3.8.1: "…*The assessment must take into account the severity of effects, the uncertainty of the data, and the number of organisms groups which the a.s., safener or synergist is expected to affect adversely by the intended use*". Regulation (EC) No 546/2011 states that no authorisation shall be granted unless it is "clearly established" that no unacceptable impact occurs. The aquatic guidance document states that this implies that uncertainties in the data thus must be considered and that the term "clearly established" implies a requirement for some degree of certainty. First-tier assessments use standardised scenarios and decision rules, which are designed to provide an appropriate degree of certainty. Higher tier assessments are not standardised, and so the degree of certainty they provide has to be evaluated case by case. The need for risk assessments to include characterisation of uncertainty has also been emphasised at senior policy levels in the EU (see also Sterling, 2010).

The risk assessment should be accompanied by an overall assessment of the uncertainty, which will include separate evaluation of uncertainties related to any lines of evidence. Therefore, it should include an evaluation of the uncertainties related to the modelling component if a mechanistic effect model is used as a tool for risk refinement. This section focuses only on the qualitative assessment of uncertainties of mechanistic models as part of the overall risk assessment (which will include the uncertainty in other areas, e.g. laboratory and field data used) rather than a stand-alone assessment for the model.

The guidance documents for each compartment have sections covering the uncertainty analysis, so this opinion will not repeat those details, but will show how the modelling part of the risk assessment can be incorporated into the overall risk assessment tables.

Sources of uncertainty in risk assessments that use models are various and include model uncertainty, parameter uncertainty, uncertainty about species selection and uncertainty about scenario selection.

Quantitative aspects of model uncertainty should be covered in the model evaluation (section 10.6.1) and the output from that analysis should be incorporated into the overall assessment of uncertainty (see example in Appendix C).

Qualitative assessment of uncertainty generally uses words to describe how certain an outcome is, or to describe how different the true outcome might be compared with a predicted outcome. All important uncertainties should be considered and the magnitude and direction of their potential influence on the expected level of impact on risk assessment should be listed. A tabular approach is used in some PPR Panel opinions (EFSA, 2006, 2007a, b, 2008) and we follow that approach here (see Appendix C).

We follow the aquatic guidance document in recommending use of separate columns to describe uncertainties that act in opposite directions. Because we have separate columns for the potential to underestimate or to overestimate the real risk, we propose using symbols simply to show the magnitude of the potential. The proposed scale is: +, low potential for under- or overestimation; ++, medium potential for under- or overestimation; +++, high risk for under- or overestimation. Appendix C tabulates some criteria that might be used to assess uncertainty qualitatively and provides some examples of the sorts of words that might be used to explain a qualitative score.

Thus, the qualitative assessment of uncertainty is meant to give an overview of all possible perceived uncertainties around the risk assessment. This qualitative assessment is important in the context of modelling because the model should be realistic and there is no conservatism built into the model (i.e. no assessment factors are applied in the model). The overall level of uncertainty is often very large, but this should not in any sense mean that the risk assessment has 'failed', i.e. it is allowable to have a few '+++' and several '++' entries in the table. Rather, a high level of uncertainty provides information that can be used in decision-making by risk assessors (Madelin, 2004; Sterling, 2010).

In summary, the output of a mechanistic effects model provides a line of evidence whose uncertainty should be assessed systematically along with other lines of evidence (e.g. laboratory studies, mesocosm tests, field studies). The qualitative assessment of uncertainties should help (a) to assess/evaluate the model so that the risk assessor will better identify the weakness/strengths of the model and, if needed, asks the modeller for further analysis of refined input parameters etc., (b) to explore uncertainties by carrying out more model runs or (c) to refine and develop the model based on new or previously unused data.

12.4. Recommendations for model improvement

Once a model has been used for regulatory purposes, feedback from risk assessor can lead to recommendations for model improvement. If well organised, this step provides an important feedback mechanism between the modeller and the risk assessor. The following items could facilitate this process:

i. The presence of independent scientific reviews and/or articles in peer-reviewed scientific journals.

ii. The presence of an online platform where risk assessors and modellers can exchange their experience with the model.

iii. The presence of a steering group consisting of modellers and risk assessors that prioritises model improvements. This group should base their decisions on practical considerations (usually from risk assessors), independent scientific reviews, and of course on experience of the modeller himself.

iv. The presence of a version control system.

12.5. Version control of the regulatory model

A version control system is needed to avoid confusion caused by the use of different model versions. Elements of good version control are:

i. There is a comprehensive overview of the modules and data files that together constitute one version of a regulatory model.

ii. A system is implemented for archiving model versions including a system for version numbering.

iii. It is clear from the model output which model version has been used to create this output.

iv. There is a description of the authorities and persons who are responsible for releasing new model versions.

v. Differences with earlier model versions are described and justified.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

The model should address a clearly defined question in the context of a regulatory risk assessment and should help to assess whether or not the specific protection goal, as outlined in the relevant guidance documents, is met. The specific protection goal includes considerations about the ecological entities which need to be protected, the magnitude of effects and spatial and temporal scale of effects, as well as exposure considerations and the degree of certainty. If a model does not address the specific risk assessment question including the specific protection goal then it is of very limited use in the context of pesticide authorisation.

Depending on the risk assessment question and protection goal, different types of models may be used. For example, TK-TD models might be used to refine the assessment of individual mortality and matrix or IBMs for population-level effects.

When selecting appropriate species to model, it is important that if one modelled species is selected to cover the risk for a range of species it is identified as a vulnerable species. Vulnerability of a species at the population level can be characterised by the species' exposure to the contaminant, its sensitivity to the toxicant and its population resilience. It should be clearly stated how the modelled species fits into the risk assessment and which issues are being addressed by the modelling. It is important to note that the species tested in first-tier laboratory tests may often not be the best species to model environmental risk as they are not necessarily the most ecologically vulnerable species. An exception is the use of standard test species modelled in a TK-TD approach, in order to refine exposure and/or effects: such risk refinement strategy is recognised as valid, e.g. in the bird and mammals guidance (EFSA, 2009) and in the aquatic guidance (EFSA PPR Panel, 2013). Furthermore, data from laboratory species might be useful to establish the credibility of the model.

If data are not complete for the relevant species, then these gaps need to be clearly listed as well as approaches on how to overcome this lack of information (e.g. toxicity data from closely related species). The uncertainties arising from extrapolation need to be accounted for in the interpretation of

the model outcome. One option is to add uncertainty factors and to repeat the modelling for a range of assumptions (e.g. similar toxicity and 5● or 10● the toxicity). Assessment factors, as proposed in first-tier risk assessments, are intended to cover uncertainties related to the specific risk assessment and may not be appropriate to cover uncertainties related to the model and the data used in the model. Model adjustment factors should be defined in future (as in EFSA PPR Panel (2012), for example).

If there is a large laboratory dataset of single species tests which show that one species is much more sensitive than all the others (with all others showing an acceptable risk), this can be used to select the modelled species. In a similar way, if a semi-field study shows recovery within an acceptable timescale for all species except one, then a model can focus on that species as well as on the differences between the semi-field study and what is likely to occur in real systems.

If the protection goal is "no mortality", then the sensitivity and exposure are likely to be the most important aspects (for example, in an acute bird or mammal risk assessment, small species which eat a high percentage of the exposed food are used). If the protection goal allows mortality or temporary effects on other life history traits of individuals, then characteristics such as generation time, number of offspring and dispersal ability are likely to be very important. It is necessary to identify which aspects of life history are likely to be most important. A trait-based approach should be systematically used in order to select a (group of) species that would cover for the risk of other species belonging to the community, but with different biological and ecological traits. Trait-based approaches should be relevant to describe ecotoxicological vulnerability with respect to the mode of action of the pesticide.

An environmental scenario is a combination of abiotic, biotic and environmental parameters which represent the environmental context in which the model is run. The scenarios must be developed in line with the specific protection goals and the required level of conservatism. Therefore, it is beneficial to define the percentiles used in terms of spatial and temporal scale and endpoints relative to the protection goals (e.g. no increase in background mortality of forager bees by more than a factor of 1.5 over six days in 90 % of colonies at the edge of treated fields). The scenario should take into account a number of factors; for example, the entity studied should be in a sensitive state. The population model should fully describe the state of the population and the stressors incorporated in a baseline or control conditions and include all relevant biotic and abiotic interactions. For example, competition and predation might influence the capacity and time to recover: if considered not relevant, then this should be justified (e.g. if background mortality includes already predation).

All relevant exposure routes must be taken into account in the model. Examples are direct overspray, spray drift, atmospheric deposition, runoff and drainage. The ecotoxicologically relevant exposure concentration also depends on the traits of the modelled organism (feeding pattern, life stage, movement pattern, etc.). It needs to be ensured that the spatio-temporal dimension of the exposure regime matches the specific protection goal. When acute effects are modelled, the maximum concentration over time may be most relevant. When simulating chronic effects and recovery, the decrease in exposure concentration resulting from dissipation (the exposure regime) is also important.

The availability and reliability of data which are used to develop the model are crucial. The quality of the available data should be assessed, e.g. by checking whether the underlying studies gave sufficient experimental details for reproducibility of results, appropriate statistical evaluation and if uncertainties are described with confidence intervals. If inconsistencies exist between datasets, then the choice of dataset used in the modelling needs to be justified (e.g. in terms of relevance to the modelled species and/or scenarios; level of detail provided documenting experimental conditions; precision of measurements as assessed by confidence intervals). The data sources need to be fully documented and descriptions of the conditions under which the values were observed and the differences to the situation to be modelled should be discussed and included in the uncertainty analysis.

A model which is submitted in the context of a regulatory risk assessment will be subject to an evaluation by the risk assessor in an authority and includes among others the following points:

i. documentation (is the model sufficiently well documented in order to allow and evaluation?);

ii. problem formulation (does the model address the risk assessment question and the SPGs?);

iii. data quality (data fit for purpose? e.g. sufficiently robust);

iv. logic of the conceptual model (inclusion of all important biotic and abiotic factors + reality check);

v. evaluation of variables and parameters;

vi. the chosen scenario is in line with protection goals as outlined in the problem formulation;

vii. domain of applicability of scenario (what is covered and what is not covered);

viii. level of conservatism;

ix. evaluation of sensitivity and uncertainty analysis and model validation.

Detailed recommendations are given in section 10 and a checklist was developed to facilitate evaluation of the model.

A model should be documented in a transparent manner in order to facilitate the evaluation. Several proposals for model documentation exist, e.g. TRACE (Schmolke et al., 2010), ODD (Grimm et al., 2006, 2010), ODdox (www.doxygen.org) and EPA Guidance on the Development, Evaluation and Application of Environmental Models (EPA/100/K-09/003 (EPA, 2009). In addition to these, it is recommended to produce a summary document which includes all important steps of the model development and references to the detailed documentation.

The model output needs to be interpreted in the overall context of the risk assessment. In contrast to lower tier testing, a model is intended to be realistic and no conservatism is built into the model itself. A qualitative assessment of uncertainty is therefore important. Such an assessment should cover sources of uncertainty in the model, input parameters, selection of species, environmental scenarios, model validation and sensitivity and uncertainty analysis. An example of criteria for qualitative assessment of the uncertainty is included in section 12.

The opinion is focused on mechanistic effect modelling, nevertheless the principles outlined in the current opinion may also apply to environmental fate modelling.

RECOMMENDATIONS

Models should be documented in a transparent way and include all information which is needed for the risk assessor to evaluate the model. Furthermore, all models should be accompanied by a user manual so that the evaluator can use the model and replicate the results and to do for example a reality check.

A feedback mechanism for model developers ought to be established, in order to improve the model for regulatory purpose. The following items could facilitate this process:

i. Independent scientific reviews and/or articles in peer-reviewed scientific journals.

ii. An online platform where risk assessors and modellers can exchange their experience with the model.

iii. An expert group consisting of modellers and risk assessors that prioritises model improvements. This group should base their decisions on practical considerations (usually

from risk assessors), independent scientific reviews, peer-reviewed papers, and of course on experience of the modeller himself.

iv. A version control system.

v. Use of post-registration monitoring data/field studies to improve models.

From a regulatory point of view it would be beneficial to have a set of agreed models that can be used in standard risk assessments. It is recommended that an expert group which co-ordinates the use of models in risk assessment be established. One of the tasks of such a group could also be the development of objective criteria for evaluating models (see checklist in Appendix B). One important step in this process is to compare different models used to address the same risk assessment question. To proceed from research to regulation and decision-making, there is a need for:

i. scientifically sound, robust models, which have been thoroughly tested and make valid predictions;

ii. robust, user-friendly and freely available software which can routinely be used by industry and regulators in all Member States;

iii. a well-defined set of scenarios which represents the full range of relevant species, ecosystems, and eco-regions in the EU (DG SANCO, 2012).

Benefits of an agreed set of models would include more efficient use of resources in terms of reduction in cost and labour. Additionally, standardised scenarios would help to harmonise the risk assessment calculations and their interpretation.

REFERENCES

AIAA (American Institute of Aeronautics and Astronautics), 1998. Guide for the verification and validation of computational fluid dynamics simulations (G-077-1998e), AIAA Standards, Available online: www.aiaa.org/StandardsDetail.aspx?id=3853, 19 pp., ISBN: 978-1-56347-354-8.

Agatz A, Hammers-Wirtz M, Gabsi F, Ratte H-T, Brown CD, and Preuss TG, 2012. Promoting effects on reproduction increase population vulnerability of *Daphnia magna*. Environmental Toxicology and Chemistry, 31, 1604–1610.

Ashauer R and Escher BI, 2010. Advantages of toxicokinetic and toxicodynamic modelling in aquatic ecotoxicology and risk assessment. Journal of Environmental Monitoring, 12, 2056–2061.

Ashauer R, Agatz A, Albert C, Ducrot V, Galic N, Hendriks J, Jager T, Kretschmann A, O'Connor I, Rubach MN, Nyman A-M, Schmitt W, Stadnicka J, van den Brink PJ and Preuss TG, 2011. Toxicokinetic-toxicodynamic modelling of quantal and graded sublethal endpoints: A brief discussion of concepts. Environmental Toxicology and Chemistry, 30, 2519–2524.

Augusiak J, van den Brink PJ and Grimm V, 2013. Merging validation and evaluation of ecological models to 'evaluation': A review of terminology and a practical approach. Ecological Modelling, in press.

Begon M, Townsend CR and Harper JL, 2006. Ecology: From individuals to ecosystems, 4th edn, Blackwell Publishing, Malden, MA, USA.

Beketov MA and Liess M, 2006. The influence of predation on the chronic response of *Artemia* sp. populations to a toxicant. Journal of Applied Ecology, 43, 1069–1074.

Bernhardt-Römermann M, Gray A, Vanbergen AJ, Berges L, Bohner A, Brooker RW, De Bruyn L, De Cinti B, Dirnbock T, Grandin U, Hester AJ, Kanka R, Klotz S, Loucougaray G, Lundin L, Matteucci G, Meszaros I, Olah V, Preda E, Prevosto B, Pykala J, Schmidt W, Taylor ME, Vadineanu A, Waldmann T and Stadler J, 2011. Functional traits and local environment predict vegetation responses to disturbance: a pan-European multi-site experiment. Journal of Ecology, 99, 777–787.

Boesten JJTI, 2000. Modeller subjectivity in estimating pesticide parameters for leaching models using the same laboratory dataset. Agricultural Water Management, 44, 389–409.

Boesten JJTI, Köpp H, Adriaanse PI, Brock TCM and Forbes VE, 2007. Conceptual model for improving the link between exposure and effects in the aquatic risk assessment of pesticides. Ecotoxicology and Environmental Safety, 66, 291–308.

Boutin C, Elmegaard N and Kjaer C, 2004. Toxicity testing of fifteen non-crop plant species with six herbicides in a greenhouse experiment: implications for risk assessment. Ecotoxicology, 13, 349–369.

Brown JH, Gillooly JF, Allen AP, Savage VM and West GB, 2004. Toward a metabolic theory of ecology. Ecology, 85, 1771–1789.

Bunzel K, Liess M and Kattwinkel M, 2014. Landscape parameters driving aquatic pesticide exposure and effects. Environmental Pollution, 186, 90–97.

Campbell PJ, Arnold DJS, Brock TCM, Grandy NJ, Heger W, Heimbach F, Maund SJ and Streloke M, 1999. Guidance document on higher-tier aquatic risk assessment for pesticides (HARAP), SETAC-Europe, Belgium, Brussels, 179 pp.

Candolfi MP, Barrett KL, Campbell P, Forster R, Grandy N, Huet M-C, Lewis G, Oomen P A, Schmuck R and Vogt H (Eds.), 2001. Guidance document on regulatory testing and risk assessment procedures for plant protection products with non-target arthropods. Report of the SETAC/ESCORT 2 Workshop, Wageningen, The Netherlands, SETAC-Europe, Belgium, Brussels.

Carpenter, D and Boutin, C. 2010. Sublethal effects of the herbicide glufosinate ammonium on crops and wild plants: short-term effects compared to vegetative recovery and plant reproduction. Ecotoxicology 19, 1322–1336.

Caswell H, 2001. Matrix population models, 2nd edn, Sinauer Associates, Sunderland, MA, USA.

Clark J, Ortego LS and Fairbrother A, 2004. Sources of variability in plant toxicity testing. Chemosphere, 57, 1599–1612.

Comas LH and Eissenstat DM, 2009. Patterns in root trait variation among 25 co-existing North American forest species. Phytologist, 182, 919–928.

Dalkvist T, Sibly R and Topping C, 2013. Landscape structure mediates the effects of a stressor on field vole populations. Landscape Ecology, 28, 1961–1974.

Damgaard C, Mathiassen SK, Kudsk P, 2008. Modelling effects of herbicide drift on the competitive interactions between weeds. Environmental Toxicology and Chemistry, 27(6), 1302–1308.

DG SANCO (Health and Consumer Protection Directorate General) SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks) SCHER (Scientific Committee on Health and Environmental Risks) and SCCS (Scientific Committee on Consumer Safety), 2012. Preliminary report on Addressing the New Challenges for Risk Assessment. ISBN 978-92-79-XX, 157 pp. Available online: http://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_037.pdf

Dorrough J and Scroggie MP, 2008. Plant responses to agricultural intensification. Journal of Applied Ecology, 45, 1274–1283.

EC (European Commission), 2002. Guidance Document on aquatic ecotoxicology in the context of the Directive 91/414/EEC (SANCO/3268/2001) rev.4 final, 17.11.2002, 1–62.

EC (European Commission), 2009. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. OJ L 309/1, 24.11.2009, p. 1–50.

EC (European Commission), 2013. Regulation (EU) No 283/2013 of 1 March setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L 93, 3.4.2013, p. 1–84.

EC (European Commission), 2013. Regulation (EU) No 284/2013 of 1 March setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L 93, 3.4.2013, p. 85–152.

EFSA (European Food Safety Authority), 2006. Opinion of the Scientific Panel on Plant Protection Products and their Residues (PPR Panel) on a request from EFSA related to the evaluation of pirimicarb. EFSA Journal 2006, 240, 21pp. doi:10.2903/j.efsa.2005.240

EFSA (European Food Safety Authority), 2007a. Opinion of the Scientific Panel on Plant Protection Products and their Residues (PPR Panel) on a request from the European Commission on acute dietary intake assessment of pesticide residues in fruit and vegetables. EFSA Journal 2007, 538, 88 pp. doi:10.2903/j.efsa.2007.538

EFSA (European Food Safety Authority), 2007b. Opinion of the Scientific Panel on Plant Protection Products and their Residues (PPR Panel) on a request from the European Commission on the risks associated with an increase of the MRL for dieldrin on courgettes. EFSA Journal 2007, 554, 48pp. doi:10.2903/j.efsa.2007.554

EFSA (European Food Safety Authority), 2008. Scientific Opinion of the Panel on Plant Protection Products and their Residues (PPR) on the science behind the Guidance Document on risk assessment for birds and mammals. EFSA Journal 2008, 734, 181 pp. doi:10.2903/j.efsa.2008.734

EFSA (European Food Safety Authority), 2009. Guidance Document on risk assessment for birds and mammals on request from the EFSA. EFSA Journal 2009;7(12):1438, 358 pp. doi:10.2903/j.efsa.2009.1438

EFSA (European Food Safety Authority), 2013. Guidance on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). EFSA Journal 2013;11(7): 3295, 266 pp. doi:10.2903/j.efsa.2013.3295

EFSA PPR Panel (Panel on Plant Protection Products and their Residues), 2010. Scientific Opinion on the development of specific protection goal options for environmental risk assessment of pesticides, in particular in relation to the revision of the Guidance Documents on Aquatic and Terrestrial Ecotoxicology (SANCO/3268/2001 and SANCO/10329/2002). EFSA Journal 2010;8(10):1821, 55 pp. doi:10.2903/j.efsa.2010.1821

EFSA PPR Panel (Panel on Plant Protection Products and their Residues), 2012. Scientific Opinion on the science behind the guidance for scenario selection and scenario parameterisation for predicting environmental concentrations of plant protection products in soil. EFSA Journal 2012;10(2):2562, 76 pp. doi:10.2903/j.efsa.2012.2562

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

EPA (United States Environmental Protection Agency), 2009. Guidance on the development, evaluation, and application of environmental models, EPA/100/K-09/003, 99pp.

Escher BI and Hermens JLM, 2002. Modes of action in ecotoxicology: Their role in body burdens, species sensitivity, QSARs, and mixture effects. Environmental Science and Technology, 36, 4201–4217.

Fletcher JS, Muhitch MJ, Vann DR, McFarlane JC and Benenati FE, 1985. Phytotox database evaluation of surrogate plant species recommended by the U.S. Environmental Protection Agency and the Organization for Economic Cooperation and Development. Environmental Toxicology and Chemistry, 4, 523–532.

Focks A, ter Horst MMS, van den Berg F, Baveco JM, van den Brink PJ, 2013. Integrating chemical fate and population-level effect models forpesticides at landscape scale: New options for risk assessment. Ecological Modelling, in press.

FOCUS (Forum for the Co-ordination of Pesticide Fate Models and their Use), 2001. FOCUS surface water scenarios in EU evaluation process under 91/414/EEC. Report of the FOCUS Working Group on Surface Water Scenarios. EU Document Reference SANCO/4802/2001-rev2. 245 pp.

FOCUS (Forum for the Co-ordination of Pesticide Fate Models and their Use), 2009. Assessing potential for movement of active substances and their metabolites to ground water in the EU. Report of the FOCUS Ground Water Work Group, EC Document Reference Sanco/13144/2010 version 1, 604 pp.

Forbes VE, Sibly RMS and Calow P, 2001. Toxicant impacts on density-limited populations: A critical review of theory, practice, and results. Ecological Applications, 11, 1249–1257.

Galic N, Hommen U and Baveco JM, 2010. Potential application of population models in the European ecological risk assessment of chemicals II: Review of models and their potential to address environmental protection aims. Integrated Environmental Assessment and Management, 6, 338–360.

Giddings M, Arts G and Hommen U, 2013. The relative sensitivity of macrophyte and algal species to herbicides and fungicides: An analysis using species sensitivity distributions. Integrated Environmental Assessment and Management, 9, 308–318.

Grimm V and Martin BT, 2013. Mechanistic effect modeling for ecological risk assessment: Where to go from here? Integrated Environmental Assessment and Management, 9, 58–63.

Grimm V and Railsback SF, 2005. Individual-based modeling and ecology. Princeton University Press, Princeton, NJ, USA, and Oxford, UK.

Grimm V and Railsback SF, 2011. Pattern-oriented modelling: A 'multi-scope' for predictive systems ecology. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 367, 298–310.

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T and DeAngelis DL, 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. Science, 310, 987–991.

Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz SK, Huse G, Huth A, Jepsen JU, Jørgensen C, Mooij WM, Müller B, Pe'er G, Piou C, Railsback SF, Robbins AM, Robbins MM, Rossmanith E, Rüger N, Strand E, Souissi S, Stillman RA, Vabø R, Visser U and DeAngelis DL, 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling, 198, 115–126.

Grimm V, Berger U, DeAngelis DL, Polhill GJ, Giskee J and Railsback SF, 2010. The ODD protocol: A review and first update. Ecological Modelling, 221, 2760–2768.

Hartig F, Calabrese J M, Reineking B, Wiegand T and Huth A, 2011. Statistical inference for stochastic simulation models – theory and application. Ecology Letters 14:816-827.

Hisschemöller M and Hoppe R, 1996. Coping with intractable controversies: the case for problem structuring in policy design and analysis. Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization, 8, 40–60.

Holland EP, Aegerter JN, Dytham C and Smith GC, 2007. Landscape as a model: The importance of geometry. PLoS Computational Biology, 3, 1979–1992.

Ibrahim L, Preuss TG, Ratte HT and Hommen U, 2013. A list of fish species that are potentially exposed to pesticides in edge-of-field water bodies in the European Union—a first step towards identifying vulnerable representatives for risk assessment. Environmental Science and Pollution Research International, 20, 2679–2687.

Hunka A, Meli M, Thit A, Palmqvist A, Thorbek P and Forbes V, 2013. Stakeholders' perspective on ecological modelling in environmental risk assessment of pesticides: Challenges and opportunities. Risk Analysis, 33(1), 68–79.

ISO/IEC/IEEE, 2013. Software and systems engineering–Software testing—Part 1: Concepts and definitions. ISO 29119 series-1, 56 pp.

Jager T, Crommentuijn T, van Gestel CAM and Kooijman SALM, 2004. Simultaneous modeling of multiple endpoints in life-cycle toxicity tests. Environmental Science and Technology, 38, 2894–2900.

Jager T, Albert C, Preuss TG and Ashauer R, 2011. General unified threshold model of survival - a toxicokinetic-toxicodynamic framework for ecotoxicology. Environmental Science and Technology, 45, 2529–2540.

Janssen, PHM and Heuberger PSC, 1995. Calibration of process-oriented models. Ecological Modelling, 83, 55–66.

Johnston ASA, Sibly RM, Hodson ME, Thorbek P and Alvarez T, 2013. Assessing chemical effects on earthworms: An agent based modelling approach. CREAM Open Conference, 10–14 June 2013, Leipzig, Germany.

Knillmann S, Stampfli NC, Beketov MA and Liess M, 2012a. Intraspecific competition increases toxicant effects in outdoor microcosms. Ecotoxicology, 21, 1857–1866.

Knillmann S, Stampfli NC, Noskov YA, Beketov MA and Liess M, 2012b. Interspecific competition delays recovery of *Daphnia* spp. populations from pesticide stress. Ecotoxicology, 21, 1039–1049.

Liess M, 2002. Population response to toxicants is altered by intraspecific interaction. Environmental Toxicology and Chemistry, 21, 138–142.

Liess M and Beketov MA, 2011. Traits and stress: Keys to identify community effects of low levels of toxicants in test systems. Ecotoxicology, 20, 1328–1340.

Liess M and von der Ohe PC, 2005. Analyzing effects of pesticides on invertebrate communities in streams. Environmental Toxicology and Chemistry, 24, 954–965.

Liess M, Schäfer R and Schriever C, 2008. The footprint of pesticide stress in communities - species traits reveal community effects of toxicants. Science of the Total Environment, 406, 484–490.

Liess M, Foit K, Becker A, Hassold E, Dolciotti I, Kattwinkel M, Duquesne S. 2013. Culmination of low-dose pesticide effects. Environmental Science & Technology. 47 (15), 8862–8868.

Luttik R, Mineau P and Roelofs W, 2005. A review of interspecies toxicity extrapolation in birds and mammals and a proposal for long-term toxicity data. Ecotoxicology, 14, 817–832.

Madelin R, 2004. The importance of scientific advice in the Community decision-making process. Opening address to the Inaugural joint meeting of the members of the Non-Food Scientific Committees. Directorate General for Health and Consumer Protection, European Commission, Brussels.

Marrs RH, Williams CT, Frost AJ and Plant RA, 1989. Assessment of the effects of herbicide spray drift on a range of plant species of conservation interest. Environmental Pollution, 59, 71–86.

MNP (Netherlands Environmental Agency), 2008. Stakeholder Participation Guidance for the Netherlands Environmental Assessment Agency: Main Document. MNP publication number 550032007, Netherlands Environmental Assessment Agency, Bilthoven, the Netherlands. Available online http://www.pbl.nl/sites/default/files/cms/publicaties/550032007.pdf

Modelink workshop, 2013. Hommen U, Alix A, Auteri D, Carpentier P, Dohmen P, Ducrot V, Forbes VE, Preuss TG, Reed M, Schmitt W, Thorbek P and Wendt-Rash L. 2013. How to use ecological effect models to link ecotoxicological tests to protection goals. SETAC-EU workshop, Le Croisic (France): 22–5 October 2012 and Monschau (Germany): 22–25 April 2013.

Moe JM, Kristoffersen AB, Smith RH and Stenseth NC, 2005. From patterns to processes and back: Analysing density-dependent responses to an abiotic stressor by statistical and mechanistic modelling. Proceedings of the Royal Society Series B–Biological Sciences, 272, 2133–2142.

Munns WR, Gervais J, Hoffman AA, Hommen U, Nacci DE, Nakamaru M, Sibly RM and Topping C, 2008. Modeling approaches to population-level ecological risk assessment. In: Population level ecological risk assessment. Eds Barnthouse LW, Munns, WR Jr and Sorensen, MT. Taylor & Frances, Boca Raton, FL, USA, 179-210.

NRC (National Research Council), 2007. Models in environmental regulatory decision making. Committee on Models in the Regulatory Decision Process, National Research Council. The National Academies Press, Washington, USA, 286 pp.

Parrish RS and Smith CN, 1990. A method for testing whether model predictions fall within a prescribed factor of true values, with an application to pesticide leaching. Ecological Modelling, 51, 59–72.

PBL Netherlands Environmental Assessment Agency, 2011. The PBL-checklist for assessing model quality. Questionnaire, explanation and vocabulary. PBL Netherlands Environmental Assessment Agency, PBL, Bilthoven, the Netherlands. Available at http://www.pbl.nl/publicaties/het-pbl-normenkader-voor-modellen (in Dutch).

Perrenet J and Zwaneveld B, 2012. The many faces of the modelling cycle. Journal of Mathematical Modelling and Application, 1, 3–21.

Peters RH, 1983. The ecological implications of body size. Cambridge University Press, Cambridge, UK.

Railsback SF and Grimm V, 2011. Agent-based and individual-based modeling: A practical introduction. Princeton University Press, Princeton, NJ, USA.

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal L 309/1 24.11.2009, p. 1–50.

Regulation (EC) No 546/2011 of 10 June 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards uniform principles for evaluation and authorisation of plant protection products.

Refsgaard JC and Henriksen HJ, 2004. Modelling guidelines—terminology and guiding principles. Advances in Water Resources, 27, 71–82.

Saltelli A and Annoni P, 2010. How to avoid a perfunctory sensitivity analysis. Environmental Modelling and Software, 25, 1508–1517.

Schmolke A, Thorbek P, DeAngelis DL and Grimm V, 2010. Ecological models supporting environmental decision-making: A strategy for the future. Ecology and Evolution, 25, 479–486.

Sibly RM, Brown JH and Kodric-Brown A (Eds.), 2012. Metabolic ecology: A scaling approach. Wiley-Blackwell, Oxford, UK.

Sibly RM, Grimm V, Martin BT, Johnston ASA, Kułakowska K, Topping CJ, Calow P, Nabe-Nielsen J, Thorbek P and DeAngelis DL, 2013. Representing the acquisition and use of energy by individuals in agent-based models of animal populations. Methods in Ecology and Evolution, 4, 151–161.

Solomon KR, Brock TCM, De Zwart D, Dyer SD, Posthuma L, Richards SM, Sanderson H, Sibley PK and van den Brink PJ, (Eds.), 2008. Extrapolation practice for ecotoxicological effect characterisation of chemicals. SETAC Press & CRC Press, Boca Raton, FL, USA, 380 pp.

Stark JD, Tanigoshi L, Bounfour M and Antonelli A, 1997. Reproductive potential: Its influence on the susceptibility of a species to pesticides. Ecotoxicology Environmental Safety, 37, 273–279.

Stark JD, Banks JE and Vargas R, 2004. How risky is risk assessment: The role that life history strategies play in susceptibility of species to stress. Proceedings of the National Academy of Sciences of the USA, 101, 732–736.

Sterling A, 2010. Keep it complex. Nature, 468, 1029–1031.

Strandberg B, Mathiassen SK, Damgaard C and Bruus M, 2007. Testing non-target effects of herbicide spray drift. Poster presented at SETAC Europe 17th Annual Meeting, Porto, Portugal.

Strandberg B, Mathiassen SK, Bruus M, Kjaer C, Damgaard C, Andersen HV, Bossi R, Løfstrøm P, Larsen SE, Bak J and Kudsk P, 2012. Effects of herbicides on non-target plants: How do effects in standard plant tests relate to effects in natural habitats? Pesticide Research. Vol. 137, Danish Ministry of the Environment, EPA, 116 pp.

Tiktak A, 2000. Application of pesticide leaching models to the Vredepeel dataset. II Pesticide fate. Agricultural Water Management, 44, 119–134.

Tiktak A, Leijnse A and Vissenberg H, 1999. Uncertainty in regional-scale assessment of cadmium accumulation in the Netherlands. Journal of Environmental Quality, 28, 461–470.

Tiktak A, Boesten JJTI, van der Linden AMA and Vanclooster M, 2006. Mapping ground water vulnerability to pesticide leaching with a process-based metamodel of EuroPEARL. Journal of Environmental Quality, 35, 1213–1226.

Tiktak A, Boesten JJTI, Egsmose M, Gardi C, Klein M and Vanderborght J, 2013. European scenarios for exposure of soil organisms to pesticides. Journal of Environmental Science and Health Part B, 48, 703–716.

Topping CJ, 2011. Evaluation of wildlife management through organic farming. Ecological Engineering, 37, 2009–2017.

Topping CJ and Lagisz M, 2012. Spatial dynamic factors affecting population-level risk assessment for a terrestrial arthropod: An agent-based modeling approach. Human and Ecological Risk Assessment, 18, 168–180.

Topping CJ, Hansen TS, Jensen TS, Jepsen JU, Nikolajsen F and Odderskaer P, 2003. ALMaSS, an agent-based model for animals in temperate European landscapes. Ecological Modelling, 167, 65–82.

Topping CJ, Sibly RM, Akcakaya HR, Smith GC and Crocker DR, 2005. Risk assessment of UK skylark populations using life-history and individual-based landscape models. Ecotoxicology, 14, 925–936.

Topping CJ, Høye TT and Olesen CR, 2010. Opening the black box - Development, testing and documentation of a mechanistically rich agent-based model. Ecological Modelling, 221, 245–255.

Topping CJ, Dalkvist T and Grimm V, 2012. Post-hoc pattern-oriented testing and tuning of an existing large model: Lessons from the field vole. PLoS ONE, 7(9), e45872.

Topping CJ, Kjær LJ, Hommen U, Høye TT, Preuss TG, Sibly RM and van Vliet P, 2013. Recovery based on plot experiments is a poor predictor of landscape-level population impacts of agricultural pesticides. Environmental Toxicology and Chemistry, in press.

Vaal M, van der Wal JT, Hermens J and Hoekstra J, 1997. Pattern analysis of the variation in the sensitivity of aquatic species to toxicants. Chemosphere, 35, 1291–1309.

Van den Berg F, Tiktak A, Heuvelink GBM, Burgers SLGE, Brus DJ, de Vries F, Stolte J and Kroes, JG, 2012. Propagation of uncertainties in soil and pesticide properties to pesticide. Journal of Environmental Quality, 41, 253–261.

Vanderborght J, Tiktak A, Boesten JJTI and Vereecken H, 2011. Effect of pesticide fate parameters and their uncertainty on the selection of 'worst-case' scenarios of pesticide leaching to groundwater. Pest Management Science, 67, 294–306.

Westoby M and Wright J, 2006. Land-plant ecology on the basis of functional traits. Trends in Evolutionary Ecology, 21, 261–268.

White AL and Boutin C, 2007. Herbicidal effects on non-target vegetation: Investigating the limitations of current pesticide registration guidelines. Environmental Toxicology and Chemistry, 26, 2634–2643.

APPENDICES

Appendix A. Pattern-oriented modelling using multi-criteria assessment as an alternative approach to validation

The basic concept of pattern-oriented modelling (POM) corresponds to the overall strategy of science, i.e. to use observed patterns which are characteristic of a certain system for detecting the mechanisms that generate these patterns and, therefore, are likely to be key elements of the internal organisation of the system (Grimm et al., 2005). For complex systems, single patterns are usually not sufficient to narrow down the range of possible generative mechanisms. Therefore, multiple patterns are used, which are observed at different scales and hierarchical levels. For example, cycles in the abundance of small mammals are a striking pattern, but usually do not contain enough information to identify, unambiguously, the mechanism that generates these cycles in reality. Additional patterns are needed, for example changes of cycle characteristics in response to weather, latitude, type of predators, etc., or changes in behaviour in high- and low-density situations.

A common misconception with POM is that it is either a calibration method or a validation method; both are true, but POM is actually a method for developing the model integrating a number of modelling cycle steps. POM refers to the multi-criteria design, selection, development, and calibration of models of complex systems (Grimm and Railsback, 2011), typically IBMs/ABMs. As such it is an integral part of the modelling cycle, the activities being difficult to separate into distinct model development, parameter choice, parameter estimation, and validation sections as might be done for non-ABM models.

POM comprises three interrelated elements, which are briefly explained here (more detailed descriptions are in Grimm and Railsback (2011) and Railsback and Grimm (2011)).

First, for complex systems, multiple patterns should be used for model design, i.e. a model should include not only those factors which are considered essential for the purpose of a model, but also entities and processes that would allow patterns to emerge which are considered characteristic for the structure and functioning of the system. Such patterns can be taken from empirical observations and literature, from discussions with experts, and sometimes from existing theory. These patterns can be complex or simple, striking or relative weak, and possibly contain a lot of information or only a limited amount. Then, criteria are defined for deciding whether the model reproduces each pattern. Simple qualitative criteria should be used first, for example visual inspection of trends or whether or not average outputs are within confidence limits of observed data. The model is then revised (this can and should include changes to the model structure) until the most important patterns observed in reality also emerge in the model.

Second, patterns are used to contrast alternative sub-models representing certain key process. For this, the alternative sub-models, for example of foraging, competition or habitat selection, are implemented one at a time in the full model. Then the alternatives are tested by testing how well the full model reproduces the set of characteristic patterns defined before. Sub-models that cannot reproduce one or more patterns are rejected. This is repeated until the best sub-model has been identified, which might require revising the original set of alternative sub-models or using additional patterns to better be able to distinguish between sub-models.

Third, multiple patterns can also be used for calibration of entire sets of unknown parameters. This works the same way as POM selection: each pattern is used as a criterion for acceptance, but now it is parameter values being accepted or rejected. This approach is similar to "inverse modelling" or "Monte Carlo filtering" techniques used in other disciplines. It includes the following steps: identify parameters that need to be calibrated. These are particularly uncertain parameters, and those to which the model is particularly sensitive; identify calibration criteria. Model outputs should be observed in the same way as their real counterparts. Then, create a large number of parameter sets by varying the

unknown parameters. Sampling techniques such as Latin hypercube sampling, genetic algorithms, Bayesian methods and human guided fitting can be used. The best-fitting parameter set will finally be used to provide the parameter values for use of the model. As part of this process it is unavoidable that there is iterative cycle between sensitivity analysis and calibration. Note also that the POM cycle should include the critical evaluation of the data used for building and testing the model to avoid data inconsistencies and overfitting (Topping et al., 2010, 2012).

Since final sensitivity will often depend on the parameter values chosen, the sensitivity analysis step should follow this POM step (Figure A1).

When the final model is decided upon, the extent to which it matches these multiple patterns, often utilising some measure of goodness of fit, can be used to assess how well the model behaves in the same way as the system it represents (a validation). Note that there is no simple method of generating a statistical measure of fit, nor a confidence limit. Whether a model is considered valid, and to what extent, is a question of critically considering both the relevance of the patterns and the fit to these. It is thus a scientific evaluation process rather than a statistical one.

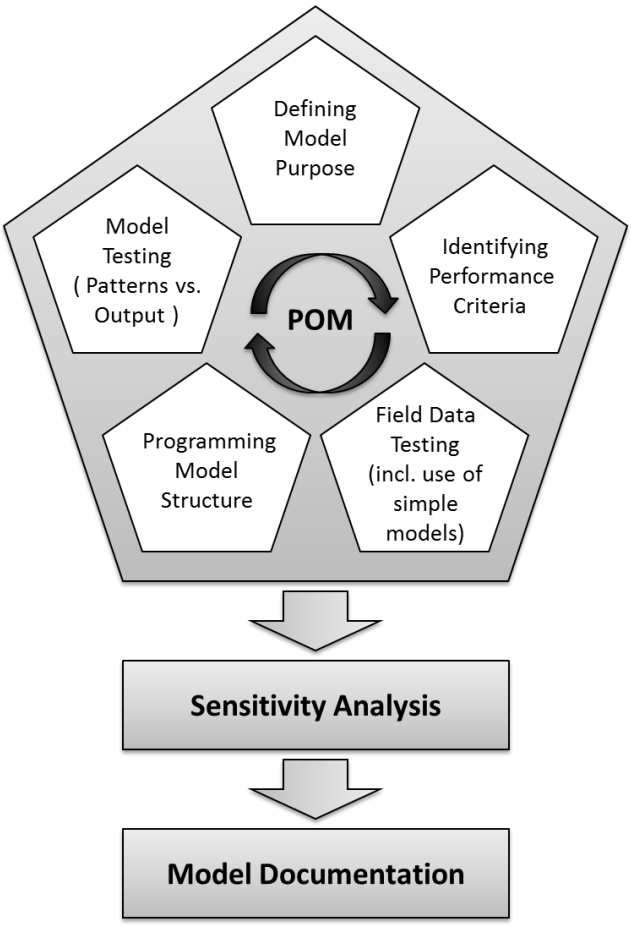


Figure A1: After initialising the cycle by defining a model purpose it is necessary to traverse the complete POM process at least once; thereafter, should the model performance meet the performance criteria, the cycle will be interrupted and sensitivity analysis and documentation performed (from Topping et al., 2010)

Appendix B. Summary checklist for model evaluation by the risk assessor

This checklist aims to help the risk assessor conduct a comprehensive evaluation of the model to be used as support for risk assessment. The checklist follows the evaluation process as described in section 10 (see Figure 13 for an overview of this process). The risk assessors should base their evaluation on documentation provided by the modeller. Each question can be answered with "Yes" or "No". It is recommended that a justification for each answer be provided. It is acknowledged that not all the points have a similar weight for decision-making. It was not possible within the given timeframe to develop objective criteria for the overall assessment. It is recommended that such criteria be developed in future. The checklist can be downloaded on the EFSA webpage.

	ASPECT OF THE MODEL TO BE EVALUATED BY THE RISK ASSESSOR	Yes	No
1. Evaluation of the problem definition The problem definition needs to explain how the modelling fits into the risk assessment and how it can be used to address the specific protection goals (section 4). Please check if due attention is paid to:			
(a)	The available knowledge and data relevant to the risk assessment question		
(b)	The regulatory context in which the model is run		
(c)	The question that has to be answered with the model		
(d)	The outputs required to answer these questions including performance criteria for the regulatory model		
(e)	The species to be modelled (use the checklist in section 5.3)		
(f)	Requirements for the environmental scenarios to be used in the risk assessment		
2. Evaluation of the supporting data The supporting data should be of sufficient quality and be relevant to the risk assessment problem. Please check the following items (see section 10.1 for explanation and section 3 for background information):			
(a)	Are the data fit for purpose in view of the problem definition?		
(b)	Has the quality of the data used been considered and documented?		
(c)	Have all available data been used? If not, is there a justification why this information has not been used?		
3. Evaluation of the conceptual model The conceptual model provides a general and quality description of the system to be modelled. Please check the following items (see section 10.2 for details and section 6 for background information):			
(a)	Are the specific protection goals sufficiently well addressed by the model?		
(b)	Are the modelling endpoints relevant to the specific protection goal?		
(c)	Is the modelling approach justified?		
(d)	Is the conceptual model logical?		
(e)	Are the processes included in the model relevant to the addressed issue?		
(f)	Are the links between different processes to the variables logical?		
(g)	Are the temporal and spatial scales relevant in regard to the problem definition?		
4. Evaluation of the formal model The formal model contains the equations and algorithms to be used in the model. Please check the following items (see section 10.3 for details and section 7.1 for background information):			
(a)	Are the most important model assumptions justified by the modeller?		
(b)	Are the most important mathematical equations described?		
(c)	Is there a description of the variables and parameters including their meaning and unit?		
(d)	Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?		
(e)	Are references supporting the equations been provided?		
5. Evaluation of the computer model The next step is to convert the formal model into a model that can run on a computer (the computer model). Please check the following items (see section 10.4 for details and section 7.2 for background information):			

(a)	Is there a comprehensive and transparent description of the computer model?		
(b)	Is the computer code well readable and is it available?		
(c)	Is it demonstrated that the mathematical model is correctly implemented (model verification)? The following items could be checked: – Has the model been compared with a benchmark? – Has the internal consistency of the model results been checked and reported? – Has a reality check been carried out?		

6. Evaluation of the regulatory model—the environmental scenario

The environmental scenarios determines the environmental context in which the model is run. The scenario determines the conservatism of the scenario and should therefore be thoroughly evaluated (see section 10.5.1 for explanation and section 8.2 for background information). Please check the following items:

(a)	Is the scenario representative for the risk assessment under consideration?		
(b)	Has the modeller justified the general biological, abiotic and environmental parameters that constitute the scenario?		
(c)	Has the modeller ensured that the scenario covers the most relevant exposure pathways for the area under consideration?		
(d)	Is the level of conservatism placed into the scenarios appropriate? The level of conservatism is to a large extent determined by the spatio-temporal dimensions of the protection—see point (iv) in section 10.5.1.		

7. Evaluation of the regulatory model –parameter estimation

Parameter estimation is a crucial step since it determines the behaviour of the regulatory model and hence it applicability for regulatory assessments. Please check the following items (see section 10.5.2 for explanation and section 8 for background information):

(a)	The model parameter estimation has been adequately documented?		
(b)	Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?		
(c)	Were the estimated parameter values realistic?		
(d)	Are the data sources sufficiently documented?		

8. Evaluation of the sensitivity and uncertainty analysis

Sensitivity analysis identifies subsets of parameters that have a strong effect on the model outputs. Uncertainty analysis aims at identifying how uncertain the model output is. Please check the following items (see section 10.6.1 for explanation and section 9 for background information):

(a)	Has the sensitivity analysis been adequately documented?		
(b)	Is the sensitivity analysis applicable to the situations identified in the problem formulation?		
(c)	Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?		
(d)	Has the uncertainty analysis been adequately documented?		
(e)	Is the uncertainty analysis applicable to the situations identified in the problem formulation?		
(f)	Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?		
(g)	Uncertainty is propagated to the model results?		
(h)	Have confidence intervals been estimated and has this information been used in further model use?		

9. Evaluation of the model by comparison with data from independent measurements

The performance of the model is usually evaluated by comparing relevant model outputs with measurements. Please check the following items (see section 10.6.2 for explanation and section 9.3 for details):

(a)	Have the performance criteria for the model been predefined in the problem definition?		
(b)	Are the model outputs that are compared relevant in view of the problem definition?		
(c)	Have the data with which the model is compared been subjected to quality control and is a description of the data available?		
(d)	Is the dataset relevant in view of the problem definition?		
(e)	Is the fit of model output to the data good enough?		
(f)	Has the performance of the model been reported in an objective and reproducible way?		

10. Evaluation of model use

When using a model for regulatory purposes, the risk assessor needs to input the pesticide properties and the application regime of these pesticides. In this stage, it is important that the model is well documented and that it is clear how the model works. Please check the following items (see section 10.6.2 for explanation and section 9.3 for details):

(a)	Is a user manual available? The user manual should contain at least the items described in section 10.*7*.		
(b)	Have all aspects of the modelling cycle been documented? The documentation should contain at least the items described in section 11.2.		
(c)	Has a summary sheet been provided by the modeller? The summary sheet should provide quick access to the comprehensive documentation (see section 11.3).		
(d)	When applicable—is the regulatory assessment described? Please check the following: – Have the pesticide properties been obtained in a justifiable way? – Have commonly agreed standards been used? – Do the application patterns/regimes reflect good agricultural practice? – Is a baseline assessment available? – When applicable, have appropriate assessment factors been used?		
(e)	Have appropriate conclusions been derived from the risk assessment? Section 12.3 and Appendix B are intended to help the risk assessor in decision-making.		

11. Evaluation of the suitability of the model for regulatory purposes

Once a model has been used for regulatory purposes, an evaluation of the suitability of the model for regulatory purposes needs to be carried out. This can lead to recommendations for model improvement. Please check the following items (see sections 12.4 and 12.5):

(a)	Is there a possibility for dialogue between the modeller and the risk assessor? The items described in section 12.4 could facilitate this process.		
(b)	Is a version control system implemented? See section 12.5 for details.		

12. Overall judgement

Based on the results of the checklist, the evaluator could give an overall judgement of the suitability of the model for regulatory purposes. The number of questions answered with “yes” and “no” could be taken in consideration when giving this overall judgement. It is, however, difficult to give a general indication of the number of negative answers that are considered acceptable since not all questions have equal weight.

(a)	Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.		

Appendix C. Qualitative assessment of uncertainty in ecological modelling

This appendix provides some criteria for a qualitative assessment of uncertainty in ecological modelling. The score provides a possibility to judge the potential of underestimating or overestimating the real risk. The proposed scale is: +, low potential for under- or overestimation; ++, medium potential for under- or overestimation; +++, high potential for under- or overestimation.

Not all cells may be relevant for a given assessment and therefore not all cells need to be filled. When necessary, the risk assessor might want to add cells with additional criteria.

		Potential to underestimate the real risk		Potential to overestimate the real risk	
		Score	Explanation	Score	Explanation
Model uncertainty (examples see below)					
Model assumptions					
Any other model uncertainty?					
Parameter uncertainty (examples see below)					
Input variables (toxicity/ effect data)	Variability/ uncertainty of toxicity endpoints	+++	e.g. one or only few species tested	+	e.g. many species tested
	Intra – lab, inter- lab and inter species variations				
	Whole dose response curve- single endpoint derived				
	Type of data, e.g. acute or chronic standard tests- semi-field/ field data				
	Estimation of values (e.g. literature, measured, estimated by calibration)				
	Uptake, elimination rates	+++	e.g. best case species selected	+++	e.g. worst case species selected
Input variables (other than toxicity/ effect data)	Variability of most relevant parameters (tested; expected in the field)				
	Gaps in measurement/ measurement errors				
	Estimation of values (e.g. from literature, directly measured, estimated)				
Any other parameter related uncertainty?					

Model outputs (examples see below)					
Choice of measure of effect (e.g. mean, 90percent)					
Any other uncertainty related to model outputs?					
Selection of modelled species and relevance towards the protection goal (examples see below)					
Life history characteristics (e.g. duration of life cycle)		+++	e.g. multivoltine species	+++	e.g. univoltine species
Life stage sensitivity and size		++	e.g. toxicity data on adults	++	e.g. toxicity data on most sensitive stage
Sensitivity to the chemical (e.g. data from species other than modelled species)					
Presence at time of exposure or when substance accumulate in the environment (e.g. exposed to PEC maximum)					
Potential chronic, delayed, cumulative and carry-over of effects					
Any other species related uncertainties?					
Assessment + environmental scenario (examples see below)					
Biological considerations	Abiotic stressors (context dependency) (e.g. agricultural management, resources, drying, pH, low oxygen, isolation vs. connectivity)				
	Biotic stressors, depending on level of organisation (e.g. intra- or inter-specific competition, predation)				
Exposure considerations	Exposure routes	+++	e.g. only some exposure routes	+	e.g. all relevant exposure routes
	Exposure regime (e.g. pattern, concentration, duration in relation to type of effects)				
	Exposure scale (currently exposure scenarios are not designed for landscape level)				
Spatial scale	spatial scale (e.g. edge-of-field ditch for individual effects; larger scale for effects on populations or communities)				
	Landscape structure (e.g. connectivity, off field size as sources of recolonisation)				

Temporal scale	Temporal scale e.g. seasonality of effects , carry-over effects				
Integration of exposure and effects (e.g. animal behaviour altering exposure)					
Any other uncertainties related to assessment and environmental scenario?					
Multiple PPP exposure (examples see below)					
Different applications of multiple PPP (e.g. combined or successive)					
Output of model validation (examples see below)					
Comparison of model outputs to suitable independent datasets (e.g. baseline data and toxicity data)?		+++	e.g. for invertebrate: only control data , no toxicity data	+	e.g. for invertebrate: complete dataset (control and tox. data)
How well fits the model predictions to observed data patterns?					
To which extent the model output compares to data from all relevant situations with respect to the problem formulation?					
Any other model validation related uncertainties?					
Domain of applicability					
Extrapolations (i.e. list all extrapolations which have been made)		++	e.g. model validated on same crop, climate zone	++	e.g. model validated on different crops, climate zones
Any other type of uncertainties?					
Overall assessment					

GLOSSARY AND ABBREVIATIONS

GLOSSARY

Bayesian inference: This is a way to estimate parameter values from data. This could be in calibration or for estimating particular parameter values from dedicated experiments. The estimation is based on prior knowledge on the parameter (e.g. taken from the literature, expert judgement or previous experiments). In a nutshell, this approach uses the prior knowledge and the comparison between measured data and model results to get a new probability distribution of the parameter value called posterior distribution. Often, the parameter value with the highest posterior probability is used for further modelling. The posterior distribution also gives a measure of the uncertainty of the parameter value; the wider it is, the less certain is the value.

Calibration: The process of adjusting model parameters within physically defensible ranges until the resulting predictions give the best possible fit to the observed data. In some disciplines, calibration is also referred to as "parameter estimation".

Code: Instructions, written in the syntax of a computer language, that provide the computer with a logical process. "Code" can also refer to a computer program or sub-set. The term "code" describes the fact that computer languages use a different vocabulary and syntax from algorithms that may be written in standard language.

Complexity: Complex systems tend to have a large number of variables, multiple parts and mathematical equations of a higher order, and to be more difficult to solve. Used to describe computer models, "complexity" generally refers to the level in difficulty in solving mathematically posed problems as measured by the time, number of steps or arithmetic operations or memory space required (called time complexity, computational complexity and space complexity, respectively).

Conceptual model: A hypothesis regarding the important factors that govern the behaviour of an object or process of interest. This can be an interpretation or working description of the characteristics and dynamics of a physical system

Confidence interval (parameter estimation, frequentist inference): The range of parameter values that would all be accepted under the assumed statistical error distribution and the chosen significance level.

Constant: A fixed value (e.g. the speed of light, the gravitational force) representing known physical, biological, or ecological activities.

Credibility interval (parameter estimation, Bayesian inference): The interval that contains the given parameter value with a certain probability (e.g. 95 %). It can be constructed from the posterior distribution in a Bayesian inference.

Deterministic model: A model that does not includes variability in model parameters. The solution obtained by the model or output is therefore unique for a given set of parameter values and initial conditions.

Ecotoxicologically (environmentally) relevant concentration (ERC): The type of concentration of a pesticide (active ingredient, formulations, and relevant metabolites) that is likely to affect a determinable ecological characteristic of an exposed system.

Environmental scenario: The collection of biotic, abiotic and agronomic parameters describing the environment of the modelled species.

Evaluation (model): The process used to generate information to determine whether a model and its results are of a quality sufficient to serve as the basis for a regulatory decision. Includes assessment model parameter estimation, sensitivity analysis and validation.

Exposure/fate model: Model which can simulate the spatial and temporal distribution of environmental concentrations of a pesticide.

Focal species: A real species exposed when the pesticide is applied and for which the risk assessment is conducted.

Frequentist inference (also known as classical inference): This is a way to estimate parameter values from data. This could be in calibration or for estimating particular parameter values from dedicated experiments. The goal is to find one best parameter value, defined as the one that shows the closest fit between measurements and the outcome of the simulation or statistical model. The uncertainty of this value can be described by the confidence interval (or confidence region). Most often, the error distribution (i.e. the distribution describing the random variation in the measurements) is assumed to be normal and independent for the different measurements. Prior knowledge on the parameter value (e.g. the biologically reasonable range or values from previous experiments) is not taken into account during the estimation.

Implementation: Definition of a given model code in a software (e.g. Excel, Matlab). A model can have several implementations, depending on the code that is used.

Initial conditions: Values of the variables at the beginning of the simulation.

Mechanistic effects model: A model describing mechanistically how a biological entity (individual or population or community) develops over time and how it responds to a toxicant.

Model: A simplification of reality that is constructed to gain insights into select attributes of a physical, biological, economic, or social system. A formal representation of the behaviour of system processes, often in mathematical or statistical terms. The basis can also be physical or conceptual (NRC, 2007).

Parameters: Terms in the model that are fixed when conducting a model run or simulation (but can be changed in the model development step, as a method for conducting sensitivity analysis or to achieve calibration goals).

Parameterisation *sensu stricto*: Parameter definition, the process of defining parameters that are used to represent the (biological) processes in a model. They determine the interactions and controls among model mechanisms.

Parameterisation *sensu lato*: A word that modellers use for "selecting values for a model's parameters". Other equivalent words are parameter estimation or parameter inference.

Parameter uncertainty: Uncertainty (see definition) related to parameter values.

Prior distribution (parameter estimation, Bayesian inference): The formalised knowledge of a parameter value before parameter estimation, e.g. taken from the literature, expert judgement or previous experiments. For some parameters there might be a lot of knowledge available, so that a narrow prior can be constructed (e.g. as a normal distribution with the mean representing the value that was measured most often and the variation representing the spread in the measurements). In other cases nearly nothing might be known (e.g. for the competitive strength of different species) and a vague, flat prior needs to be used that allows for a very wide range of potential values.

Regulatory acceptable concentration (RAC): Effects assessment endpoint expressed in terms of a permissible concentration of an active substance in the environment that is directly used in the risk assessment by comparing it with the appropriate field exposure estimate (e.g. maximal predicted exposure concentration, PEC_{max}). Usually, the RAC is derived by applying an uncertainty factor (also known as an assessment factor) to the relevant measurement endpoint for the most sensitive species, e.g. effective concentration in 50 % of exposed animals (EC_{50}) or no effect concentration (NOEC) from single-species toxicity tests, or effect class concentrations from micro-mesocosm experiments, or by means of an appropriate (statistical) extrapolation procedure (e.g. species sensitivity distribution (SSD)).

Regulatory model: The regulatory model is a package consisting of the following components (i) the mechanistic exposure-effects model, (ii) programs for pre- and post-processing, often made available in the form of graphical user interfaces, (iii) model parameters, and (iv) environmental scenarios. By combining the computer model with scenarios, the model will address a certain goal and can therefore be used regulatory purposes.

Sensitivity: The degree to which the model outputs are affected by changes in selected input parameters.

Sensitivity analysis: The quantification of the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs. By investigating the "relative sensitivity" of model parameters, a user can become knowledgeable about the relative importance of parameters in the model.

Simulation: Development of a solution by incrementing steps through the model domain. Simulations are often used to obtain solutions for models that are too complex to be solved analytically. For most situations, where a differential equation is being approximated, the simulation model will use finite time step (or spatial step) to "simulate" changes in state variables over time (or space).

Stochasticity: Fluctuations in ecological processes due to natural variability and inherent randomness.

Stochastic model: A model that includes variability (see definition) in model parameters. This variability is a function of changing environmental conditions, spatial and temporal aggregation within the model framework, and random variability. The solution obtained by the model or output is therefore a function of model components and random variability.

TRACE—Acronym for "transparent and comprehensive ecological modelling documentation"; TRACE is a framework for documenting models and their analysis, which is based upon the modelling cycle.

Uncertainty: The term used in this document to describe lack of knowledge about models, parameters, constants, data and beliefs. There are many sources of uncertainty, including the science underlying a model, uncertainty in model parameters and input data, observation error and code uncertainty. Additional study and collecting more information allows error that stems from uncertainty to be minimised/reduced (or eliminated). In contrast, variability (see definition) is irreducible but can be better characterised or represented with further study.

Uncertainty analysis: Investigation of the effects of lack of knowledge or potential errors on the model (e.g. the "uncertainty" associated with parameter values). When combined with sensitivity analysis (see definition), uncertainty analysis allows a model user to be more informed about the confidence that can be placed in model results.

Variable: A measured or estimated quantity that describes an object that can be observed in a system and that is subject to change.

Two kinds of variables can be distinguished. The state variables (e.g. population size or body mass) are the dependent variables calculated within a model, which are also often the performance indicators of the models that change over the simulation. The forcing variable corresponds to input data to the model (e.g. toxicity of the substance). This input data may be defined in the exposure/ecological scenario.

Variability: Observed differences attributable to true heterogeneity or diversity. Variability is the result of natural random processes and is usually not reducible by further measurement or study (although it can be better characterised).

Validation: The process of establishing that the model is a sufficiently accurate representation of the real world to be used as the basis for regulatory decisions. It assesses how well the model fits relevant data patterns and if the model provides predicted endpoint/output values with an acceptable error range for risk assessment. This last step is performed through the comparison of model or sub-model outputs with empirical data that were preferably not used for parameter estimation.

Verification (code): Examination of the algorithms and numerical technique in the computer code to ascertain that they truly represent the conceptual model and that there are no inherent numerical problems with obtaining a solution.

ABBREVIATIONS

ABM	agent-based model
EC	effective concentration
EFSA	European Food Safety Authority
EFSA PPR Panel	Panel on Plant Protection Products and their Residues
ERC	ecotoxicologically relevant exposure concentration
ERO:	ecological recovery option
ETO	ecological threshold option
EU	European Union
FOCUS	FOrum for the COordination of Pesticide fate models and its USe.
GAP	good agricultural practice
GLP	good laboratory practice
IBM	individual-based model
LC	lethal concentration
LD_{50}	median lethal dose
MOA	mode of action
NOEC	no effect concentration
NTA	non-target arthropod
ODD	overview, design concepts, detail
OECD	Organisation for Economic Cooperation and Development
PEC	predicted environmental concentration
RAC	regulatory acceptable concentration
SPG	specific protection goal
PPP	plant protection product
TK-TD	toxicokinetics–toxicodynamics
TRACE	TRAnsparent and Comprehensive Ecological model documentation framework