SCIENTIFIC OPINION

Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods[1]

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

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

Following a request from the European Food Safety Authority, the Panel on Plant Protection Products and their Residues developed an opinion on the science to support the development of a risk assessment scheme of plant protection products for non-target arthropods. The current risk assessment scheme is reviewed, taking into consideration recent workshops and progress in science. Proposals are made for specific protection goals which aim to protect important ecosystem services such as food web support, pest control and biodiversity. In order to address recovery and source–sink population dynamics, conducting a landscape-level risk assessment is suggested. A new risk assessment scheme is suggested which integrates modelling approaches. The main exposure routes for non-target arthropods are identified and proposals are made on how to integrate them in the risk assessment. The appropriateness of the currently used vegetation distribution factor was investigated. It is proposed that new tests be included in order to address exposure via oral uptake of residues and uncertainties related to differences in species sensitivity.

KEY WORDS

non-target arthropods, effects, exposure, insects, pesticides, protection goals, risk assessment

[1] On request from EFSA, Question No EFSA-Q-2011-00975, adopted on 19 December 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 (until 30 June 2014), Robert Smith, Paulo Sousa, Ingvar Sundh, Aaldrik Tiktak and Ton Van Der Linden. Correspondence: pesticides.ppr@efsa.europa.eu

[3] Acknowledgement: The Panel wishes to thank the members of the Working Group on Non-target arthropod and In-Soil risk assessment: Peter Craig, Frank de Jong, Michael Klein, Ryszard Laskowski, Ton van der Linden, Barbara Manachini, Silvia Pieper, Robert Smith, Paulo Sousa, Klaus Swarowsky, Aaldrik Tiktak and 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), 2015. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA Journal 2015;13(2):3996, 212 pp. doi:10.2903/j.efsa.2015.3996

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

SUMMARY

The new regulatory framework for plant protection products (PPPs) laid out in Commission Regulation (EC) No 1107/2009 and Commission regulation (EU) No 283/2013 explicitly requires consideration of impacts on non-target species, on their ongoing behaviour and on biodiversity and the ecosystem, including potential indirect effects via alteration of the food web. In view of this new legislative background, the European Food Safety Authority (EFSA) asked the Panel on Plant Protection Products and their Residues (PPR) to develop and update the guidance documents on terrestrial ecotoxicology under mandate M-2009-0002. As the assessment of effects on biodiversity is not explicitly addressed under the existing guidance documents, appropriate risk assessment methodology needs to be developed. As such, expertise was needed in the different areas of terrestrial ecotoxicology, including non-target arthropods (NTAs). This scientific opinion has been written as a precursor to the guidance document on NTAs.

The ESCORT 1 and ESCORT 2 workshops form the basis of current risk assessment methodology, which is focused on beneficial arthropods. In view of the above-mentioned EFSA mandate, the working group of the PPR panel reviewed the current risk assessment, identified NTA key drivers that sustain important ecosystem services in agricultural landscapes and developed specific protection goal options for in-field and off-field areas. A new risk assessment framework was suggested which integrates landscape-level assessments. As key drivers also include species with population ranges larger than the scale of a single field or off-field area, the working group identified the need to additionally study the impact of PPPs on NTAs at landscape level. The working group developed proposals for exposure assessment and testing of effects as well as a method to calibrate the lower tier risk assessments.

The specific protection goals are closely linked to the temporal and spatial boundaries in the context of risk assessment. These boundaries relate to the protection goal (i.e. where is the community of interest?), the life history, behaviour and distribution of the identified key drivers and the route and distance covered by the emission coming from the in-field. It is necessary to make a distinction between the area designated for cultivation (= in-field) and the area surrounding a field (= off-field). The off-field can be either (semi-)natural habitat or simple structures (fence or a bare strip of land). In most cases, the off-field is not to be influenced by the farmer. Another off-field category comprises man-made structures, e.g. an adjacent field, roads, etc. The actual off-field is not known for every field. Therefore, it is proposed that a generic protection goal for the off-field area be defined. Another important spatial element is the buffer strip. It is a cropped or non-cropped zone of a defined width at the edge of a field which is influenced by the farmers action (e.g. spray drift). The buffer strip is normally enforced by authorities and underlies prescribed actions in order to meet the specific protection goal for the off-field. In addition, buffer strips may provide a potential source of NTA species for recovery from impacts in the cropped area. In-crop is the area of the field (= in-field) which is actually cropped and off-crop is any uncropped area.

Biodiversity has to be supported to a certain degree in the in-field areas in order to provide important ecosystem services, such as pollination, food web support, pest control, and to maintain an appropriate level of NTA biodiversity in the landscape. The magnitude of effects on NTAs relates to the most sensitive (ecotoxicologically and/or ecologically) species which drives the ecosystem service to be supported in-field.

It is suggested that in the off-field area only negligible direct effects of pesticides on NTAs are allowed. However, it is known that in-field impacts on non-target species can affect off-field populations. Even if the exposure of individuals in the off-field area is determined to be acceptable, the off-field population can be affected if the treated field acts as a sink. An example of a model illustrating this principle was included in the current opinion. To ensure that effects in-field do not have unacceptable effects on NTA biodiversity, it is suggested that a landscape-level risk assessment be conducted in addition to the local-scale assessment. Such a landscape-level risk assessment could

be done with population models. The local-scale risk assessment is considered sufficient to address impacts on species with a very limited mobility. However, for highly mobile species, the overall population-level impact may be underestimated. Therefore, it is recommended that the risk to mobile species is also addressed at larger spatial scales where treated fields occur.

In order to implement recovery in the risk assessment for NTAs, it is necessary to identify key NTA species that are drivers for ecosystem services to ensure that these species can survive in the agricultural landscape. Direct and indirect effects from multiple stressors (e.g. repeated application of different pesticides) need to be considered in the assessment. The investigation of recovery should focus on species with traits indicating a low recovery potential (e.g. low number of offspring, low dispersal capacity). Modelling and field studies are complementary for assessment of recovery. Field studies can provide information on the magnitude of effects on an in-field community, including indirect effects, while modelling can be used to investigate effects for some species in different landscape scenarios, including source–sink dynamics, effects under different climatic conditions and the impact of standard agricultural management practices. If the provision of a certain level of ecosystem functions (e.g. food web support, pollination, pest control) needs to be maintained, then impacts may be unacceptable even if the NTA community returns to its pre-disturbed state. Furthermore, it would be beneficial to identify and evaluate risk mitigation options at landscape scales to facilitate recovery of NTAs.

The recommendations developed for exposure assessments focus on NTAs living on plants and on soil in the in- and off-field area. Such an exposure assessment potentially does not correctly estimate exposure of highly mobile species. For such species, a landscape-level exposure assessment is developed. For complex landscape models addressing several NTA life stages, existing fate models (for example PEARL, PELMO, PERSAM) can be used to calculate soil predicted environmental concentrations (PECs) at a depth of 1 cm or less as an estimate for contact exposure. In the case of multiple applications in one growing season, accumulation of a substance on crop leaves and on the soil surface may occur. The currently available default values for wash-off and other dissipation routes from leaves are based on limited numbers of experimental values. It is recommended that more independent (experimental) data be gathered to better underpin default values. Furthermore, it should be investigated if scenario conditions, such as temperature, could be taken into account when calculating the dissipation.

Assessing the exposure routes is proposed, i.e. contact to dried and fresh residues on plant surfaces, direct overspray and oral exposure through ingestion of contaminated food. However, it is difficult to relate currently available endpoints from tier 1 assessments to realistic exposure because of limitations of the current tier 1 test design, where animals are exposed to dried residues on glass plates. It is proposed that environmentally relevant concentrations (ERC), for groups of organisms with different life traits, be established so that a direct comparison with calculated PECs is possible. This would enable the use of a so-called criss-cross risk assessment model, where both effect endpoints and exposure endpoints, from various levels of sophistication, can be combined.

Residue per unit dose (RUD) values for insects could potentially provide a good estimate for exposure from overspray and contact to fresh residues on plants and soil surface. Together with endpoints expressed in the same units as the RUD values, a simple and quick screening step assessment could be conducted. It is recommended that it be investigated further whether or not the underlying residue data justify the use of RUDs as a conservative estimate of contact exposure. It also needs to be decided whether the acute contact endpoint from honeybees (LD_{50} µg/bee and recalculated to mg/kg insect) should be used in the assessment or new studies with NTAs should be proposed where the toxicity from contact to fresh residues (including overspray) is investigated.

For the assessment of the effects of exposure via ingestion of contaminated food, RUD values might also be used to assess residue levels in different food matrices. As for the overspray estimation, it should be investigated if available RUDs are a conservative estimate for oral exposure. Currently, it is not be feasible to perform an empirical calibration for the oral exposure routes with results of field experiments.

A vegetation distribution factor (VDF) is used in the current risk assessment to relate the exposure of leaf-dwelling NTAs in the in-field environment to their exposure in an off-field environment, where they are assumed to be exposed less owing to a different vegetation structure than in the field. It may be appropriate to account for lower concentrations due to an uneven distribution and the fact that the total leaf surface is larger than the area on which the vegetation stands, but, in the absence of proper ERCs and exposure concentrations, it is impossible to recommend default values. For example, it should be known whether the average areic mass on a plant or the maximum on its leaves is causing the observed effects. Ideally, in order to apply a VDF as currently used in risk assessment, exposure in the reference tier used for calibration of the risk assessment and the ERCs for the species of relevance should be considered. The possible use of a VDF in the future will depend on the degree to which linking exposure and effects will be possible when calibrating the risk assessment.

In a tiered approach to assess the effects of PPPs on non-target organisms, the lowest tier for assessing effects on a local scale should include a relatively simple, robust set of tests. This set of tests should prevent missing unacceptable effects of intended uses of a PPP on ecosystem services defined to be important in the agricultural landscape. Selection of the species for testing ecotoxicological effects should consider the key drivers for the specific protection goals. Unacceptable effects could occur if the required standard test protocols do not test (i) susceptible taxonomic groups (ii) susceptible life stages or developmental processes targeted by the PPP or (iii) relevant routes of exposure. Based on these selection criteria, the panel recommends carrying out tier 1 toxicity tests on four species (minimum), chosen to represent different lifestyles and taxonomic groups. These recommendations include an oral toxicity study with lepidopteran larvae to represent herbivorous NTAs. Further tests are suggested for specific PPP modes of action or application methods. It is recommended that existing glass-plate protocols should be used to test effects on reproduction as well as mortality. With regard to the assessment of chronic effects, the exposure of NTAs during the reproductive phase should be controlled. As the majority of current test systems with leaf-dwelling NTAs takes only exposure towards dry residues into account, the panel considers that the toxicity endpoints derived from tests with bees (fresh residues) could provide a possible surrogate for the overspray exposure route.

It is suggested that assessment factors be derived on the basis of statistical modelling of the relationships between effects for different species in the various possible lower tier tests and higher tier field studies and the surrogate reference tier. In particular, a Bayesian network model can exploit information from both experimental data and expert judgement and provides a relatively transparent method for deriving assessment factors in order to ensure high probability of acceptable effects for uses which pass the risk assessment. In this context, the panel considers that the species sensitivity distribution (SSD) conceptual model is very useful at the reference tier level but that standard SSD methodology is unlikely to be useful in relation to tier 1 data because variation in sensitivity between NTA species as currently measured in laboratory tests is unlikely to represent variation in sensitivity in the field.

The currently used semi-field and field studies are conducted on small plots. The effects, especially the time to recovery, observed in such small plots can be misleading for mobile species that move in and out of plots during the course of a study. Replicated landscape-scale studies, however, are desirable but usually impractical. A possible compromise is to carry out a field study with a limited number of large plots in combination with a larger number of smaller plots. Modelling is considered a useful tool to extrapolate effects observed in small plots to larger landscapes. It is recommended that such modelling follows the recommendations of the EFSA Scientific Opinion on Good Modelling Practice (2014). The number of species for which potentially useful models are currently available is very limited and it is recommended that this be expanded further in order to cover the majority of vulnerable species identified in the specific protection goals as key drivers for ecosystem services.

The Panel contends that the assessment of effects of PPPs on NTA biodiversity has to focus on landscape level to enable the implementation of effective mitigation policy in the future. The existing data and models clearly show that, even if application rates and/or toxicity of PPPs to NTAs is significantly reduced, maintaining/recovering NTA diversity and ecosystem functions is impossible without preserving (or rebuilding) sufficient habitat diversity.

TABLE OF CONTENTS

Abstract 1
Summary 2
Table of contents 6
Background as provided by EFSA 9
Terms of reference as provided by EFSA 10
Assessment 12
1. **Introduction** 12
1.1. Legislative background 13
1.2. The process adopted to revise the guidance document 13
1.2.1. Reading guidance 13
2. Review of the risk assessment for non-target arthropods exposed to plant protection products 14
2.1. History of non-target arthropod risk assessment 14
2.1.1. Before EU harmonisation (1991) 14
2.1.2. ESCORT 1 (1994) 14
2.2. Current risk assessment for non-target arthropods exposed to plant protection products 15
2.2.1. Background 15
2.2.2. Summary of the present risk assessment scheme 17
2.3. Reviews of the current risk assessment 20
2.3.1. DEFRA report (2004) 20
2.3.2. DEFRA report (2007) 20
2.3.3. ESCORT 3 (2010) 21
2.4. Discussion of relevant points for the revision of the guidance document 22
2.4.1. Testing/study endpoints 23
2.4.2. How sensitive are the standard test species? 23
2.4.3. Exposure routes of non-target arthropods 23
2.4.4. Applicability of species sensitivity distributions to non-target arthropods 23
2.4.5. Vegetation distribution factor 24
2.4.6. Landscape-level effects and source–sink dynamics 24
2.4.7. Sequential and simultaneous use of different pesticides 24
2.4.8. Indirect effects on biodiversity via the food chain 24
2.4.9. Calibration of the tiered risk assessment scheme: lower tier risk assessment 25
3. Definition of spatial aspects considered in the risk assessment 25
3.1. General considerations 25
3.1.1. Spatial aspects in relation to the species to be assessed 25
3.1.2. Spatial aspects in relation to the landscape to be assessed 27
3.1.3. Source–sink dynamics in non-target arthropod populations in agricultural landscapes 27
3.1.4. Consequences for non-target arthropod populations of source–sink dynamics and multi-year exposure 28
3.2. Spatial boundaries at the local scale 28
3.3. Example of use of boundaries in the risk assessment 30
3.4. Spatial boundaries at the landscape scale 30
3.5. Simulations illustrating some spatio-temporal concepts important for non-target arthropod risk assessment 31
3.5.1. Introduction 31
3.5.2. Methods 32
3.5.3. Results 33
3.6. Illustration of the non-target arthropod risk assessment concepts using the Bembidion example 38
3.6.1. Concept 1: the potential of source–sink dynamics to indirectly affect untreated populations 38
3.6.2. Concept 2: the impact of long-term year-on-year application of pesticide 38
3.6.3. Concept 3: the effects of landscape structure on risk assessment 39

3.6.4. Other observations 39
3.7. Conclusions of the simulation exercise 39
4. Specific protection goals in the risk assessment for non-target arthropods exposed to plant protection products 40
4.1. General considerations 40
4.2. Specific protection goals regarding effect assessment 43
4.2.1. Non-target arthropods as drivers in maintaining biodiversity and genetic resources in agricultural landscapes 43
4.2.2. Non-target arthropods as drivers in maintaining cultural services in agricultural landscapes (e.g. aesthetic value) 50
4.2.3. Non-target arthropods as drivers of pest control in agricultural landscapes 51
4.2.4. Non-target arthropods as food web support 62
4.2.5. Non-target arthropods as drivers of plant pollination in agricultural landscapes 70
4.2.6. Summary of proposed specific protection goal options for non-target arthropods. Upper and lower limits of effects on non-target arthropod drivers 74
4.3. Exposure routes of key drivers belonging to the non-target arthropods 77
5. **General framework for risk assessment of non-target arthropods** 77
5.1.1. Assessment at the local scale 79
5.1.2. Assessment at the landscape scale 80
5.1.3. Mitigation of identified risks 80
5.2. The principles of a tiered approach 81
5.3. Tiered approach in the risk assessment for non-target arthropods and definition of (surrogate) reference tiers 82
5.3.1. Surrogate reference tier for effects of plant protection product use on non-target arthropod species at local scale 83
5.3.2. Surrogate reference tier for effects of plant protection products on mobile species at landscape scale 84
5.3.3. Suggestion for a lower tier landscape-scale screening tool for non-target arthropods 85
5.4. Relationship between effect and exposure assessment for non-target arthropods 88
5.5. Calibration of non-target arthropod risk assessment 91
5.5.1. The paradigm for tiered risk assessment based on specific protection goals 91
5.5.2. Adapting the paradigm to risk assessment for non-target arthropods 91
5.5.3. The basis of the calibration of the current risk assessment scheme 91
5.5.4. Fundamentals of calibration for the proposed risk assessment structure 93
5.5.5. Obtaining assessment factors from a statistical model 94
5.5.6. The structure of a Bayesian statistical model for use in non-target arthropod risk assessment 95
5.5.7. Illustrative example 98
5.5.8. Evaluation of additional uncertainties 98
5.6. Recovery 100
5.6.1. Definition of recovery in the context of non-target arthropod risk assessment 100
5.6.2. Protection goals and recovery 100
5.6.3. Recovery and multiple stress 101
5.6.4. Important aspects for assessment of recovery 101
5.6.5. Current methods to assess recovery 102
5.6.6. Conclusions and recommendations for assessing recovery of non-target arthropods 103
5.7. Identification of vulnerable non-target arthropod species to be addressed in the risk assessment 104
6. Exposure assessment for NTAs 106
6.1. Introduction 106
6.2. Environmental compartments to be considered in the exposure assessment of non-target arthropods 107
6.3. Dominant exposure routes for non-target arthropods 107
6.4. Selection of the ecotoxicologically relevant types of concentrations 108
6.5. Specification of the exposure assessment goal 109

6.6. Overall assessment of the exposure based on the different routes 109
6.7. Overview of the assessment of contact exposure 110
6.7.1. Assessment of in-field, on-crop contact exposure 111
6.7.2. Assessment of in-field, on-soil exposure 114
6.7.3. Exposure pathways for off-field exposure 116
6.7.4. Assessment of off-field exposure, on-plant exposure 119
6.7.5. Off-field, on soil 119
6.7.6. Accounting for distribution of the substance over the leaves 119
6.8. Overspray 120
6.9. Oral exposure 121
6.9.1. Herbivorous non-target arthropods 121
6.9.2. Carnivorous non-target arthropods 123
6.9.3. Other non-target arthropods 123
6.10. Landscape scale dynamic exposure modelling 123
6.11. Review of the vegetation distribution factor for use in non-target arthropod risk assessment 124
7. Effects assessment for non-target arthropods 125
7.1. Introduction 125
7.2. Assessment of effects on non-target arthropods at the local scale 126
7.2.1. Introduction 126
7.2.2. Overview on available test systems 126
7.2.3. Laboratory studies for first tier risk assessment 128
7.2.4. (Semi-)field studies for use in higher tier risk assessment 132
7.3. Assessment of effects of plant protection products on non-target arthropods at the landscape scale 136
7.4. Conclusions and recommendations on non-target arthropod testing and effect assessment 137
Conclusions and recommendations 138
References 143
Appendices 161
Appendix A. *Bembidion lampros* model 161
Appendix B. Qualitative assessment of uncertainty in the ecological modelling of *Bembidion lampros* exercise 169
Appendix C. Articles SPG Pest Control 173
Appendix D. Approximation of the ‘hedge’ (off-field) 175
Appendix E. The use of a vegetation distribution factor in the non-target arthropod risk assessment 176
Appendix F. List of taxa that should be evaluated in representative agro-ecosystems in Europe (From de Jong et al., 2010) 180
Appendix G. Summary of (extended) laboratory test systems identified as potentially relevant 182
Appendix H. Semi-field and field test systems 188
Appendix I. Example of how to use Bayesian hierarchical statistical modelling together with elicited expert judgements for calibration of non-target arthropod risk assessment 195
Abbreviations 212

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:

- Considerations of the revision of Annexes II and III of Directive 91/414/EEC,
- Consideration of the new Regulation (EC) 1107/2009
- Harmonisation with other directives and regulations (biocides, REACH)
- Clearly defined protection goals
- Multiple exposure
- Inclusion of additional species in the risk assessment (e.g. amphibians, reptiles, bats, molluscs, ferns, mosses, lichens, butterflies, grasshoppers and moths)
- More guidance on statistical analysis
- Preference of EC_x over NOEC values in the risk assessment
- To consider all available information from workshops (EUFRAM, ESCORT, PERAS and other SETAC workshops)
- Endocrine disruption
- Consideration of all routes of exposure
- Bee risk assessment
- Non-target arthropods risk assessment
- 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:

- Assessment of impacts on non-target organisms including the ongoing behaviour
- Impact on biodiversity
- Impact on the ecosystem
- Effects on bees

- Effects on amphibians and reptiles
- Linking exposure to effects and ecological recovery
- The use of field studies in the risk assessment and guidance for interpretation of field studies
- Revision of non-target arthropod risk assessment (ESCORT II)
- Guidance for risk assessment in greenhouses
- 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.
- 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 ongoing behaviour and impact on biodiversity and the ecosystem.

The assessment of effects on ongoing 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:

Development of Guidance on risk assessment for non-target arthropods, with the following deliverable:

- Opinion addressing the state of the science to be delivered by the PPR Panel by May, 2014.
- Guidance of EFSA to be delivered by July, 2015
- Public consultation on the draft Guidance of EFSA

ASSESSMENT

1. Introduction

Pesticides are deliberately released in the agro-environment to control pest species. As such, plant protection products (PPPs) may affect non-target arthropods (NTAs).

The current opinion is focused on arthropods living on the soil surface or on plants. Other groups of terrestrial invertebrates and in-soil-dwelling arthropods are not covered. It is recognised that the traditional division in the terrestrial risk assessment between NTAs and in-soil organisms does not make sense for all groups of organisms. Some NTAs live part of their life cycle in the soil and part above the soil (e.g. many Coleoptera). Moreover, some above-ground invertebrates are not considered in the NTA assessment even if they are living above ground (e.g. isopods, chilopods, diplopods). The grouping of NTAs and in-soil organisms was performed for better handling along main routes of exposure, as above-ground arthropods are exposed differently to pesticides than to in-soil dwelling organisms.

NTAs occur in-crop and off-crop. The distribution in- and off-field and the spatial and temporal aspects of their life histories need to be considered in the context of the risk assessment. Therefore, a distinction is made between the assessment of effects at the local scale and at the landscape scale. Local scale is defined in this opinion as the treated field and the immediate surroundings and landscape scale is defined as the total (agricultural) landscape in a region where the pesticide is applied. Landscape scale should be large enough to cover source–sink dynamics such as local extinction and recolonisation.

During the development of the current opinion, it became obvious that certain aspects of the environmental risk from intensive agriculture which involves repeated application of a range of different pesticides cannot be solved within the current context of the pesticide risk assessment and authorisation. Such issues refer to some ecosystem services; for example, the maintenance of biodiversity and the provision of suitable habitats where NTA populations can survive and act as sources for recolonisation of in-field habitats. For example, in landscapes with large monocultures and little semi-natural off-field habitats, recolonisation from unaffected off-field populations will take much longer than in structure-rich landscapes with a large proportion of semi-natural off-field habitats. A greater diversity of NTA communities which can be found in such structure-rich landscapes are also more likely to provide a desired ecosystem service after a pesticide impact.

In other risk assessments (e.g. of birds and mammals), the concept of focal species is used. The focal species needs to cover the risk to all species present in the crop. Such an approach is considered very difficult for NTAs because of their great diversity and contrasting traits which determine how severe the pesticide impact will be on a population.

In contrast to the issues above, there are solutions which allow the assessment of simultaneous exposure to different pesticides. Using concentration addition to address mixture toxicity is generally recommended. The approach is explained in detail in other EFSA opinions and therefore is not repeated in the current opinion. See, for example, the EFSA opinions on bee risk assessment (http://www.efsa.europa.eu/en/efsajournal/doc/2668.pdf) and non-target terrestrial plants (http://www.efsa.europa.eu/en/efsajournal/pub/3800.htm), and the guidance document on aquatic organisms (http://www.efsa.europa.eu/en/efsajournal/doc/3290.pdf) for further details.

This opinion does not include proposals for the assessment of endangered NTA species. An EFSA opinion on coverage of endangered species in environmental risk assessment is developed by a working group of the scientific committee of EFSA.

1.1. Legislative background

Active substances used in plant protection products are authorised in the European Union (EU) under Regulation (EC) No 1107/2009[4]. The Regulation requires that ‘substances or products produced or placed on the market do not have any harmful effect on human or animal health or any unacceptable effects on the environment’. With respect to the environment, this includes, in particular, considerations of the impact on non-target species, including on the ongoing behaviour of those species, and the impact on biodiversity and the ecosystem.

New Commission regulations laying down the data requirements for the dossier to be submitted for the approval of active substances contained in PPPs (Commission Regulation (EU) No 283/2013[5] and 284/2013[6]) were published in 2013. These documents provide information on the core data required for the authorisation of PPPs. Furthermore, as a general requirement for substance approval, it is stated in Commission Regulation (EU) No 283/2013 that ‘the potential impact of the active substance on biodiversity and the ecosystem, including potential indirect effects via alteration of the food web, shall be considered’.

Microbial PPPs have specific data requirements for active agents as well as formulated products, but are not specifically addressed in this opinion.

1.2. The process adopted to revise the guidance document

With the publication and entry into force of new Regulation (EC) No 1107/2009 and the revised data requirements, as well as new scientific findings, the PPR Panel was tasked to revise the Guidance Document on Terrestrial Ecotoxicology (EC, 2002). It was decided to split the task and to address individually the risk assessment for separate organism groups, i.e. in-soil organisms, NTAs, amphibians and reptiles, and non-target terrestrial plants. For each of the organism groups, the PPR Panel will first summarise in a scientific opinion the science behind the risk assessment and, in a second step, will develop practical guidance on how to perform the risk assessment. A public consultation on the draft guidance document will give stakeholders and the interested public the opportunity to comment. The feedback received will be taken into account when the guidance document is finalised.

1.2.1. Reading guidance

The history of NTA risk assessment, the current status and reviews of the current risk assessments are summarised in section 2 below (page 14). Major issues for the update of the NTA risk assessment, such as the vegetation distribution factor (VDF), relevant endpoints and landscape-level assessment, are briefly described and reference is made to the relevant sections where these points are discussed in detail. Section 3 below (page 25.) gives definitions of the different spatial elements to be considered in the context of the NTA risk assessment. It is proposed that in-field, off-field and buffer strips should be distinguished between and the whole landscape, in addition to local (field) scale, should be considered in the risk assessment. The different dimensions and options for specific protection goals are outlined in section 4 below (page 40). In this context, the ecosystem services which are considered in more detail in the specific protection goals are pest control, food web support, pollination and biodiversity. for different environmental compartments, exposure routes, in-field and off-field

[4] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing The principles of a tiered risk assessment approach, reference tiers, calibration of lower tiers and proposals for future risk assessment schemes and recovery of affected NTA populations are included in section 5 (page 77). Section 6 (page 106) gives details on exposure assessment 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, pp. 1–50.

[5] Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with the 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.

[6] Commission Regulation (EU) No 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with the 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.

exposures, VDFs and potential uses of residues per unit dose (RUDs) in the NTA risk assessment. The choice of test species and a review of available test methods is included in section 7 below (page 125). The main points are highlighted in the Conclusions and Recommendations sections below (page 137).

2. Review of the risk assessment for non-target arthropods exposed to plant protection products

2.1. History of non-target arthropod risk assessment

2.1.1. Before EU harmonisation (1991)

Before the EU harmonisation of the placing of PPPs on the market (see below), individual Member States were responsible for the registration of pesticides. The focus of the risk assessment was on beneficial arthropods, especially in the context of biological control.

In an international context, the focus was also on beneficial arthropods (e.g. Hassan et al., 1987; IOBC (International Organisation for Biological and Integrated Control of Noxious Animals and Plants), 1988; Beneficial Arthropod Regulatory Testing Group (BART) and European and Mediterranean Plant Protection Organisation (EPPO)). Within these frameworks, test methods were developed and general regulatory requirements were summarised in practical schemes. The focus on beneficial arthropods resulted in the choice of test species of relevance for this particular aim.

2.1.2. ESCORT 1 (1994)

In, 1994, the European Standard Characteristics of beneficial Regulatory Testing (ESCORT) 1 workshop was organised (Barrett et al., 1994). The aim of the workshop was to harmonise the different initiatives of BART, EPPO and the International Organisation for Biological and Integrated Control of Noxious Animals and Plants (IOBC). The workshop was organised in conjunction with the Society of Environmental Toxicology and Chemistry (SETAC) and funded by the European Commission (EC). The outcome of the workshop was a guidance for NTA risk assessment in the framework of EU directive 91/414, which was also to be used for the EPPO scheme for risk assessments of beneficial arthropods. A number of subjects were discussed. These are outlined in the following paragraph.

Test species were selected, on the bases of sensitivity, relevance and amenability. The selection resulted in a recommendation to test four arthropod species: two sensitive standard species, preferably *Aphidius rhopalosiphi* De Stefani-Perez (Hymenoptera: Aphidiidae) and *Typhlodromus pyri* Scheuten (Acari: Phytoseiidae), and two species relevant to the intended use of the product (representing predatory mites, parasitoids, ground-dwelling predators and foliage-dwelling predators) (see Table 1 below).

Table 1: Selection of relevant test species from the ESCORT 1 guidance document. Species should be selected in accordance with the intended use of the product. Species are categorised into two main field environments—orchards and arable crops—in which they are particularly relevant

Crop type	Parasitoids	Predatory mites	Ground dwelling predators	Foliage dwelling predators
Orchard greenhouse, forest and vineyard	*Aphidius rhopalosiphi* *Trichogramma cacoeciae* *Leptomastix dactylopii* *Drino* sp.	*Typhlodromus pyri* *Amblyseius* sp.	*Pardosa* sp. *Poecilus cupreus*	*Orius* sp. *Episyrphus balteatus* *Chrysoperla carnea* *Coccinella septempunctata*
Arable crops	*A. rhopalosiphi* *T. cacoeciae*		*P. cupreus* *Pardosa* sp. *Aleochara bilineata*	*E. balteatus* *C. carnea* *C. septempunctata*

Furthermore, ESCORT 1 gives guidance on how to conduct and interpret laboratory studies, extended laboratory studies, and semi-field and field studies, e.g. testing on glass plates or sand in the case of ground-dwelling species. Validity criteria are also discussed. Test methods are described in Candolfi et al. (2000a), among others.

In the ESCORT 1 report, it is proposed that the risk was based on the classification of effects in low, medium and high risk, based on EPPO (1994). EPPO uses a trigger of 30 % effect as a threshold for further testing (i.e. medium or high risk) for first tier tests on glass plates or sand. The trigger can be refined when more information becomes available. In further testing, a trigger value of 25 % is used. For the purpose of classification, three different situations were distinguished, each with their own acceptability criterion, according to EPPO (1994):

(i) Within crop NTAs non-integrated pest management

Unacceptable[7] if:

- no recovery occurs within a reasonable time (maximum time, e.g. one season);
- it causes an economically important pest resurgence.

(ii) Within-crop natural enemies integrated pest management practised

Unacceptable[7] if:

- measurable effects[82] occur on natural enemies that regulate pest populations which are of economic importance.

(iii) Off-crop NTAs

Unacceptable[1] if:

- ecologically significant effects[2] occur on non-target organisms (only evaluate for products in the high risk category at the maximum use rate).

2.2. Current risk assessment for non-target arthropods exposed to plant protection products

In section 2.2.1, the main documents are summarised in which the current risk assessment is described, i.e. the annexes to the EU directive, the results of the ESCORT 2 workshop and the (draft) Guidance Document for Terrestrial Ecotoxicology (SANCO/10329/2002 rev.2 (final), 17 October 2002). In the following section (2.2.2), the current risk assessment scheme is described in more detail.

2.2.1. Background

Data requirements

The recommendations of ESCORT 1 were adapted in the amending of Annex II (data requirements) of EU Council Directive 91/414/EC in 96/12/EC:

Annex II data requirement active substance

8.3.2. Other arthropods

The test should provide sufficient information to evaluate the toxicity (mortality and sub-lethal effects) of the active substance to selected arthropod species. Effects on non-target terrestrial arthropods (e.g. predators or parasitoids of harmful organisms) must be investigated. The information obtained for these species can also be used to indicate the potential for toxicity to other non-target species inhabiting the

[7] ‘Unacceptable’ effects should not prevent registration, but should be managed through appropriate labelling and management proposals.

[8] Measurable effect based on EPPO low-risk category, i.e. > 30 % 0 % reduction or, when available, species-specific threshold values.

> same environment. […] The test must be performed initially in the laboratory on an artificial substrate (i.e. glass plate or quartz sand, as appropriate) unless adverse effects can be clearly predicted from other studies. In these cases, more realistic substrates may be used. Two sensitive standard species, a parasitoid and predatory mite (e.g. *A. rhopalosiphi* and *T. pyri*), should be tested. In addition to these, two additional species must also be tested, which should be relevant to the intended use of the substance. Where possible and if appropriate, they should represent the other two major functional groups, ground-dwelling predators and foliage-dwelling predators. Where effects are observed with species relevant to the proposed use of the product, further testing may be carried out at the extended laboratory/semi-field level. […]

The text was amended in the present regulation on data requirements. Commission Regulation (EU) No 283/2013 and Commission Regulation (EU) 284/2013 on the data requirements for active substances and PPPs still recommended *A. rhopalosiphi* and *T. pyri* as standard species to be tested in laboratory studies for all substances (except for situations where NTAs are not exposed). The new regulation lists extended laboratory tests, and semi-field and field studies as requirements if the first tier tests did not provide sufficient information to assess the risk. It is also recommended that tests of additional species are carried out in higher tiers if the first tier assessment fails (one additional species should be used if a high in-field risk is indicated or two additional species should be used if a high off-field risk is indicated).

In, 1997, the uniform principles for evaluation and authorisation of PPPs were established as Annex VI of 91/414. In part C, section 2.5.2.5 (EU, 1997), the principles are further specified for arthropods: 'Where there is a possibility of beneficial arthropods other than honeybees being exposed, no authorisation shall be granted if more than 30 % of the test organisms are affected in lethal or sublethal laboratory tests conducted at the maximum proposed application rate, unless it is clearly established through an appropriate risk assessment that under field conditions there is no unacceptable impact on those organisms after the uses of the plant protection product according to the proposed conditions of use'.

The decision criteria did not change in the present Commission Regulation (EU) No 546/2011.

In, 2000, the ESCORT 2 workshop was organised with the aim to review original recommendations, and to specify how the test results should be used in risk assessment. Instead of the four species suggested in ESCORT 1, it was now proposed that two sensitive standard species be tested: the parasitoid *A. rhopalosiphi* and the predatory mite *T. pyri*.

The main changes to the risk assessment in comparison with the past concept according to Directive 91/414 EEC, Annex VI, were:

- an in-field trigger of < 50 instead of < 30 % effects;
- only dose–response tests with *A. rhopalosiphi* and *T. pyr* are required in the first tier (previously, single-dose tests with a parasitoid and a predatory mite (e.g. *A. rhopalosiphi* and *T. pyri*) and two additional species relevant to the intended use of the substance were required);
- a higher tier acceptability criterion is specified, and in-field risk assessment based on aged residue, and semi-field and field studies.

Suggestions for higher tier studies on NTAs comprise mainly field studies in agricultural crops that investigate abundance and diversity of NTAs. For field studies, the ESCORT 2 documents describe the experimental conditions, treatment, application and sampling for this specific type of test. Data analysis and reporting are discussed as well. An additional guidance is given in the document of UK PSD Part 3 Appendix 2, describing specifically methodology for performing cereal studies. The ESCORT 2 document does not give guidance for the evaluation of field studies. Guidelines for laboratory and field studies and the interpretation of the results were drafted as a result of a joint IOBC/BART/EPPO initiative (Candolfi et al., 2000, 2000a).

When unacceptable effects are predicted, risk mitigation options were suggested in the ESCORT 2 document, such as for in-field adaption of application rate, frequency and intervals, timing and unsprayed headlands. For off-field areas buffer zones, wind breaks and drift-reducing application techniques were mentioned as risk mitigation options.

Concerning the methods for assessing the risk to terrestrial NTAs, the NTA guidance currently in practice (‘Guidance Document on Terrestrial Ecotoxicology’) refers to the methods published by the ESCORT 2 workshop. Risk assessment in accordance with the recommendations of the ‘Guidance Document on Terrestrial Ecotoxicology’, as provided by the Commission Services (SANCO/10329/2002 rev.2 (final), October 17, 2002), and in consideration of the recommendations of the guidance document ESCORT 2 is described in detail under section 2.2.2.

In the past, the EU PPP Directive 91/414/EEC of 15 July 1991 was the legal basis for risk assessment with terrestrial NTAs in the EU. The NTA part of the Directive was based on the EPPO/CoE ‘Arthropod Natural Enemies Risk Assessment Scheme’ (EPPO, 1994) and the SETAC/ESCORT ‘Guidance document on regulatory testing procedures for pesticides with NTAs’ (Barrett et al., 1994).

The approach proposed by this directive has in practice (already under EU PPP Directive 91/414/EEC of 15 July, 1991) been replaced by the one proposed in the ‘(draft) Guidance Document on Terrestrial Ecotoxicology’ (EC, 2002). Concerning the methods for assessing the risk to terrestrial NTAs, the ‘Guidance Document on Terrestrial Ecotoxicology’ refers to the methods published by the ESCORT 2 workshop (Candolfi et al., 2001). Until now, this assessment scheme had not been revised and is still in practice under new Regulation (EC) No 1107/2009 concerning the placing of PPPs on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.

2.2.2. Summary of the present risk assessment scheme

The current regulatory NTA risk assessment, performed in accordance with the recommendations of the ‘Guidance Document on Terrestrial Ecotoxicology’, as provided by the Commission Services (SANCO/10329/2002 rev.2 (final), 17 October 2002), and in consideration of the recommendations of the ESCORT 2 guidance document, divides the laboratory, semi-field and field studies into tier 1 tests and higher tier tests.

Testing requirements

Tier 1

Standard test species for assessing the risk of spray applications towards NTAs in TIER 1 are the parasitoid wasp *Aphidius rhopalosiphi* and predatory mite *Typhlodromus pyri*. For these ‘standard indicator species’ tested on glass plates, risk quotient has been empirically calibrated using LR_{50} values for those two species tested on glass plates.

Higher tier

In higher tier risk assessment, it is recommended that further species be tested (e.g. *Orius laevigatus*, *Chrysoperla carnea*, *Coccinella septempunctata* or *Aleochara bilineata*). Higher tier tests can be performed under (extended) laboratory, semi-field or field conditions:

- Laboratory tests. Tests performed on artificial substrate (e.g. glass plates or quartz sand) with species other than the standard test species—*A. rhopalosiphi* and *T. pyri*.
- Extended laboratory tests. Laboratory studies with a refined exposure design in which a natural substrate (i.e. leaf disks or natural soil or a whole plant) is sprayed and the toxicity is

then assessed on freshly dried residues on that substrate. Their design my take into account the dilution of exposure by vegetation ('three-dimensional exposure').

- Aged residue studies. Pesticide spray residues are aged under laboratory or (semi-)field conditions. Subsequently, the time of ageing needed for the residues to cause effects below an acceptable threshold is determined. Aged residue studies have the purpose of demonstrating the potential for recovery in-field.
- Semi-field studies. Single-species studies with exposure under field conditions. For extended laboratory studies and semi-field data, the 50 % effect level is taken as a trigger for acceptability of effects, of both in-field and off-field.
- Field studies. In the present scheme, field studies are aimed at effects under normal agricultural conditions. Both short- and long-term effects (up to one year or even more after application) are studied. The main focus of these studies is (the potential for) recovery. Field size varies, but typically is 25 × 25 m. Movement of specimens is not excluded. The endpoint is population- and community-level effects. Recovery occurs in-field after one year. Acceptability of effects is assessed on case-by case bases.

Endpoints are usually expressed as ER_{50} (mortality, reproduction) in laboratory/extended laboratory studies. For substances with limited toxicity, a limit test is often performed, testing the expected rate (g/ha) occurring on the surface after the last application. Tests consider only a single exposure event.

According to the ESCORT 2 document, for testing for granules and seed dressing, other methods still need to be developed, but, for the time being, it has been recommended that testing should be performed with two appropriate species, e.g. spiders and ground-dwelling beetles.

Risk assessment

In-field

The in-field exposure for NTAs is calculated as the single application rate/by multiplication of the single application rate with a multiple application factor (MAF). The MAF takes into account the number of applications and the dissipation of residues on plant and soil surfaces between applications, nominally considering, for example, a DT_{50} of 2.3 × application interval (days) for plant surfaces (default MAF values are tabled in the ESCORT 2 report).

$$In-field\ exposure = Application\ rate \quad \times MAF$$

For the first-tier assessment, the in-field exposure is divided by 50 % lethal rates (LR_{50}) for the standard species *T. pyri* and *A. rhopalosiphi* as obtained in laboratory tests on inert substrate, to obtain in-field hazard quotients ($HQ_{in\text{-}field}$). However, it is mentioned in the ESCORT 2 document that 'for products where effects on reproduction are expected, assessment of sub-lethal parameters (e.g. oviposition) should also be evaluated' also in tier 1. This specifically applies to substances suspected of having a special mode of action (e.g. insect feeding inhibitors (IGRs)) where tests should include sub-lethal endpoints and may need other modifications.

$$\mathrm{HQ}_{\mathrm{in-field}} = \left(\frac{\mathrm{in-field\ exposure}}{\mathrm{LR}_{50}}\right)$$

In cases where the acceptability criterion HQ ≤ 2 is not accomplished, a higher tier risk assessment is performed by comparing the in-field exposure with the lowest application rate that resulted in effect levels of 50 % (or lower) in extended laboratory tests on natural substrate; for tests carried out in accordance with a dose–response design, this is equivalent to comparing the predicted environmental rate in-field with the LR_{50} or ER_{50}. If necessary, further refinement is possible by taking into account aged residue tests, thereupon using the same approach as for extended laboratory tests.

Off-field

The off-field exposure of NTAs is calculated from the in-field exposure by multiplication with the established BBA drift values (see ESCORT 2 report). To account for interception and dilution in the three-dimensional vegetation in off-crop areas, these values are divided by a VDF. A VDF of 10 is recommended in the ESCORT 2 report when the off-field risk assessment is based on toxicity endpoints obtained in a test design with two-dimensional exposure. However, this value has been determined as not reliable in later documents. For toxicity endpoints obtained in a test design with three-dimensional exposure, the VDF is 1.

For the first-tier assessment, the product of the off-field exposure and a correction factor (CF) of 10 (CF to take account of the uncertainty caused by sensitivity differences in the off-crop NTA community) is divided by LR_{50} for the standard species *T. pyri* and *A. rhopalosiphi* as obtained in laboratory tests on inert substrate, to obtain off-field hazard quotients ($HQ_{off\text{-}field}$):

$$\text{Off}-\text{field exposure} = \text{In}-\text{field exposure} \quad \times \left(\frac{\text{drift factor}}{VDF}\right)$$

$$\text{HQ}_{\text{off}-\text{field}} = \left(\frac{\text{off}-\text{field exposure}}{\text{LR}_{50}}\right) \times \text{CF}$$

In cases where the acceptability criterion HQ ≤ 2 is not accomplished, a higher tier risk assessment is performed by comparing the product of off-field exposure (including CF) with the lowest application rate that resulted in effect levels of 50 % (or lower) in extended laboratory tests on natural substrate; for tests carried out in accordance with a dose–response design, this is equivalent to comparing the product of drift rate and CF with the LR_{50} or ER_{50}. The CF is set to 5, provided that higher tier tests with the species affected in tier 1 and ‘two additional species with different biology’ were submitted; please refer to EC (2002).

Risk mitigation

According to the Guidance Document on Terrestrial Ecotoxicology, ‘in order to reduce effects on non-target arthropods within the cropped area the following use specifications may be modified:

- application frequency and intervals
- timing of application (crop stage)
- unsprayed headlands.

In order to reduce effects in off-field areas there are the following options:

- buffer zones
- wind breaks
- drift-reducing application techniques’.

2.3. Reviews of the current risk assessment

In, 2004, a report of the Department for Environment, Food and Rural Affairs (DEFRA) on biodiversity in a regulatory context was published and, in, 2007, another DEFRA report in which a number of aspects of the current risk assessment are discussed. In, 2010, the ESCORT 3 workshop was organised. All three reports are summarised below.

2.3.1. DEFRA report (2004)

The aim of the project was developing 'a risk assessment scheme for wider biodiversity suitable for use in a regulatory context, with accompanying case studies. A further aim was to review the availability of supporting data, identify knowledge gaps and consider future research needs'. The report reviewed current pesticide risk assessment (including risk assessment for NTAs) with regard to risks to wider biodiversity arising from their use and gave recommendations for further development of current assessment schemes. Specific points discussed were:

- Protection of species and ecosystem services. Species to be protected were classified as (i) biodiversity action plan (BAP) species, (ii) species of concern for conservation and (iii) species of concern for the conservation of other species (in particular, to prevent effects via food chains). Although for (i) and (ii) species are protected directly, for (iii) it was suggested that populations be managed rather than totally protected to prevent indirect effects on groups (i) and (ii). It was recognised that risk assessment needs to consider effects on higher trophic levels via the food chain not addressed explicitly in current risk assessment.
- Test species and endpoints to be used in risk assessment. Invertebrate species recommended by current guidance to be used in tests were reviewed. It was noted that important groups and feeding guilds were not represented in the sensitivity analyses performed for defining current standard indicator species. Hence, it was suggested that further sensitivity tests be carried out on guilds representative of off-field habitats and chick food groups not yet covered in sensitivity analyses, including a lepidopteran larva, a sap-feeding bug, a tipulid and an orthopteran. The use of the chick food index as an endpoint in higher tier tests was proposed.
- Exposure. The use of revised estimates for pesticide drift was suggested. The use of a VDF in risk assessment was not recommended and a maximum value of 3 was proposed if the off-field vegetation is as tall or taller than the crop.
- Risk assessment. It is discussed if different risk assessments should be conducted depending on the mode of action of the pesticide the timing of application and the crop considering specific fauna. For insecticides, it was pointed out that the effects of summer-applied insecticides will be much greater than effects of insecticides applied in autumn, but that a crop-specific risk assessment would not be necessary; in contrast, this is considered to be the case for herbicides.
- Risk mitigation. Options to mitigate/compensate direct and indirect effects are compared in terms of implementation, costs and effects on biodiversity based on available information.

2.3.2. DEFRA report (2007)

In, 2007, the DEFRA report on methods for 'improved pesticide risk assessments for non-target invertebrates' was published. The main focus of the report is to include the variability in both toxicity and exposure assessment. In addition, the uncertainties of combining conservative elements for the overall degree of conservatism are studied. The project comprised five major elements: a detailed examination of factors affecting the exposure of NTAs both in-field and off-field; a review of the applicability of the species sensitivity distribution (SSD) concept to NTAs; a re-evaluation of the sensitivity of the standard test species *T. pyri* and *Aphidius* spp.; an analysis of the relationship between HQs based on the standard species and effects on NTAs in the field; and an analysis of the correlation between sensitivity of the standard species in tier 1 and extended laboratory studies. The main conclusions are:

- the SSD based on the present laboratory studies are not a reliable estimate for variability in the field;
- the standard test species cannot be considered to be sensitive indicator species;
- however, with appropriate extrapolation factors, the standard test species could be used in the risk assessment;
- it is unclear if the present HQ factor is protective for effects in the field, both within and between species;
- using current protection goals, the risk assessment should be re-calibrated;
- exposure estimates should be refined;
- effects in extended laboratory studies should be calibrated against effects in the field.

2.3.3. ESCORT 3 (2010)

As a follow up of the successful ESCORT 1 and 2 workshops, ESCORT 3 was organised as a SETAC workshop, with input from academia, governmental experts and industry. The aim of the workshop was to review methods and risk assessment and to provide input for review of the guidance document. The central question during the ESCORT 3 workshop was related to if the current scheme is sufficiently predictive and protective for NTA communities.

The workshop was organised around four issues: (i) level of protection and testing scheme, (ii) off-crop environment, (iii) recovery and (iv) field studies. For each item, the workshop formulated recommendations for research, regulation and education.

In contrast to the former two ESCORT workshops, the outcome of the workshop was not in the form of guidance, but rather as recommendations for regulation, research and education. In summary, the most important recommendations were:

Recommendations for regulation

- Protection goals. The protection goal for the in-crop was considered to be the maintenance of functions. The protection goal for off-field areas was considered to be the maintenance of NTA biodiversity.
- Special attention for the off-crop areas in the in-field (e.g. unsprayed headlands, windbreaks, etc.). When, for example, a hedgerow is planted by the farmer to protect surrounding areas from exposure by drift, it would be unrealistic to demand buffer strips to protect the hedgerows. If the farmer is compensated for management of unsprayed strips, it might be realistic to demand a higher level of protection. The conclusion is that such landscape elements play a different role in different situations, so that the level of protection should be defined at the national level.

Risk assessment

- Current (ESCORT 1 and 2) guidance is considered appropriate for the risk assessment and to achieve the defined protection goals.
- Include phytophagous species in field studies in general and in the off-field assessment (also with focus on the role in the food web).
- It was not deemed appropriate to use in-crop field studies for the assessment of effects on off-crop communities. Off-crop field studies might be more appropriate; alternatively, appropriate extrapolation factors could be used. Such factors need to be defined.

- Effects might be classified (see for example de Jong et al., 2010), and acceptable effects for different areas may be defined using the effect classes.
- Potential for recovery can be assessed by aged residue studies, or by combining information on degradation with data from effect studies.
- Indirect effects by effects on non-target plants should be assessed in the non-target plant evaluation.
- Dust risk assessment should be incorporated.
- Different exposure routes should be considered:
  - during application: direct exposure during application or by drift, indirect exposure to fresh residues on leaves, flowers or soil;
  - after application: exposure to residues on surfaces and, in the case of systemic products, in flowering parts and pollen.

Recommendations for research

- Identify representative species for all key functions, and link level of effects on current test species with functions.
- Characterise patterns of diversity and abundance in off-field habitats.
- Study potential differences in sensitivity between in- and off-crop communities.
- Quantify relationship between vegetation structure and exposure of NTA fauna for in- and off-crop.
- Assess the impact of NTA biodiversity in the control of the key and secondary pests.
- Assess if current data for pollinator species are sufficient to address the NTA pollinator species.
- Quantify efficacy of drift-reducing measures.
- Exposure to dust from seed treatment, granules, etc. is currently not incorporated. Data are becoming available.
- Research on sources of vapour drift and exposure of NTAs and on off-crop deposition on vegetation.
- Off-crop correction factors should be validated related to greater variability and sensitivity of off-crop species than in-crop species.
- Mode of action and route of uptake should be taken into account while testing all tiers.
- Include sub-lethal effects (especially reproduction) in the risk assessment.

A number of scientific posters and presentations were given during the workshop. In addition, some aspects of the DEFRA 2007 report were presented. During the workshop, however, these aspects were discussed only partly, and the main recommendations were that further research is needed.

2.4. Discussion of relevant points for the revision of the guidance document

Following EFSA mandate, comments by the Member States (see background and terms of references) and the critical reviews (see above) of the present procedure to assess the risk for NTAs exposed to PPPs, several points were identified as being particularly important when revising the Guidance Document on Terrestrial Ecotoxicology.

Critical areas needing particular attention in the revision of the guidance document are related to the species tested and the endpoints measured, questions on the relevant exposure routes for NTAs and the importance of considering effects on NTA response to PPP exposure that derive from the movement of the animals in and out of treated crops in landscapes with different architectures.

The critical points are described in detail below. Reference is made to the sections addressing the approaches that were considered in this scientific opinion.

2.4.1. Testing/study endpoints

The new demands of Regulation (EC) No 1107/2009 need to be implemented in the existing risk assessment by defining appropriate specific protection goals. In order to detect in the risk assessment if the specific protection goals are met, higher tier methodology should pick up relevant effects on ecosystem services and lower tier methods should allow a robust prediction of field effects. Therefore, the current methodology for effect assessment is reviewed and recommendations are given regarding which toxicity inputs should be used in the local scale and landscape scale risk assessment (for details, please refer to section 7).

2.4.2. How sensitive are the standard test species?

Currently, *T. pyri* and *A. rhopalosiphi* are used as standard test species in the risk assessment. This is based on a comparison of the sensitivities of different beneficial species in the laboratory.

One of the conclusions of the ESCORT 3 workshop (2010) was that the current guidance (ESCORT 1 and 2 and SANCO, 2002) is considered appropriate for the risk assessment and to achieve the defined protection goals. In other places in the document, however, it is recommended that the key species for the protection goals be defined, and that the representativeness and or protectiveness of the current test species for the protection goals be studied. Based on this contradiction and supported by Swarowsky et al. (2013) and the DEFRA report (DEFRA, 2007), we do not see that the conclusion of ESCORT 3 is substantiated with data.

It was considered necessary to review this issue with respect to the level of protection of the current risk assessment (for details, please refer to section 5.5).

2.4.3. Exposure routes of non-target arthropods

NTAs living in crop fields and in-field margins might be directly oversprayed with PPPs at field or drift rates, respectively. The direct dripping of NTA with PPP droplets would lead to a far higher immediate uptake in the animals than the uptake via the animals' movement on dried residues. It should be considered, however, that in the tests with dried residues the contact time is increased compared with a single overspray event.

Several publications indicate that oral exposure of NTA feeding directly on sprayed leaves or contaminated animal prey is also a relevant route of PPP uptake. Owing to limitations of current tier 1 tests design, it is difficult to estimate the risk arising from oral and overspray exposure. How future risk assessment needs to be adapted so that different exposure routes can be taken into account is described in section 6 (exposure) and section 7 (effects, particularly in section 7.2.3).

2.4.4. Applicability of species sensitivity distributions to non-target arthropods

The concept of SSDs is applied in other areas of risk assessment (e.g. aquatic risk assessment or terrestrial non-target plant risk assessment) to extrapolate from tier 1 toxicity data of a number of representative species to the overall expected distribution of sensitivities according to a statistical model. For details, please refer to sections 5 and 7.

2.4.5. Vegetation distribution factor

To account for interception and dilution in the three-dimensional vegetation in off-field areas, relevant off-field exposure for NTAs is currently divided by a VDF of 10. Current guidance indicates that 'this figure is considered unreliable, therefore more appropriate data should be used as soon as they become available' (please refer to SANCO/10329/2002 rev 2 final). Therefore, the group discussed this issue with respect to exposure and biological considerations. For details, please refer to section 6.7.6.

2.4.6. Landscape-level effects and source–sink dynamics

An important point regarding the current risk assessment is the lack of consideration of the mobility of species. In the recommendations of the ESCORT 3 workshop report, it is proposed that the mobility of species be taken into account because, in the field studies, species might recover by moving in from nearby untreated fields (e.g. the untreated control), while this might not be the case under conditions of use of pesticides over a large area. For details, please refer to section 3.

2.4.7. Sequential and simultaneous use of different pesticides

Regulation (EC) No 546/2011 requires that Member States base their authorisation decision on the proposed conditions for the use of the PPP. Furthermore, the standard data requirements for PPPs request that 'any information on potentially unacceptable effects of the plant protection product on the environment, on plants and plant products shall be included as well as known and expected cumulative and synergistic effects'. Regulation 1107/2009 requires that the risk assessment methodology should account for the simultaneous use of PPPs (applied in tank mixtures or used in sequence) and that the use of PPPs does not have any long-term repercussions for the abundance and diversity of non-target species (see EFSA PPR Panel, 2010).

As stated in the Guidance Document on Aquatic Ecotoxicology (EFSA PPR Panel, 2013), the question is arising on how protective the risk assessment of a single substance is with regard to the actual use patterns of PPPs in the field. EFSA PPR Panel (2013) states that 'consequences of simultaneous and sequential exposure to different PPPs as well as stress due to other environmental stressors should be further investigated'.

The consequences for NTAs exposed in the field to multiple applications of PPPs in accordance with 'spray schedules' was addressed in this opinion when formulating specific protection goal options (please refer to section 4). Particularly when assessing recovery of NTA biocoenoses from initial effects, sequential uses of PPPs must be taken into account (please refer to section 5.7).

2.4.8. Indirect effects on biodiversity via the food chain

The current legal basis for authorisation of PPPs in Europe (Regulation (EC) No 1107/2009) requires that 'substances or products produced or placed on the market do not have any harmful effect on human or animal health or any unacceptable effects on the environment'. Thereby, as stated in Commission Regulation laying down the data requirement for dossier submission (Commission Regulation (EU) No 283/2013) 'the potential impact of the active substance on biodiversity and the ecosystem, including potential indirect effects via alteration of the food web, shall be considered'.

Arthropods constitute an important part of virtually all ecosystems. Food web effects on higher trophic levels, especially on birds and small mammals, are of particular concern. This issue is not explicitly addressed in the current risk assessment, as it is also recognised by DEFRA (2005) (please refer to section 2.3.1). To provide a basis for updating current risk assessment, section 4.2.2 presents an overview of the contemporary knowledge on links between abundance or biomass of NTAs—the food of many insectivorous and omnivorous species—and birds and small mammals. Moreover, section 4.2.2 defines specific protection goals that allow maintenance of the support of the food web by NTAs.

2.4.9. Calibration of the tiered risk assessment scheme: lower tier risk assessment

In the course of the revision of current guidance documents, EFSA (2010) proposes calibrating risk assessments to make sure that specific protection goals are met at all tiers and gives recommendations on how this could be done. Given the criticisms by DEFRA (2007), section 5.5 reviews current risk assessment regarding this aspect and proposes an approach to calibrate current NTA risk assessment in line with specific protection goals.

3. Definition of spatial aspects considered in the risk assessment

3.1. General considerations

3.1.1. Spatial aspects in relation to the species to be assessed

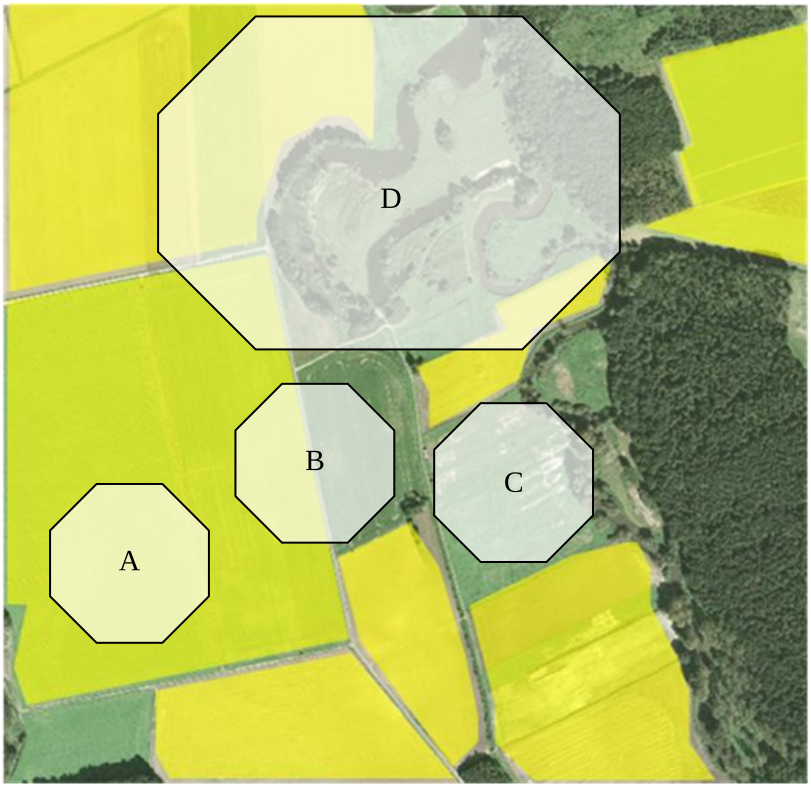
The organism group addressed by the general label NTAs comprises an extremely diverse assemblage of arthropod taxa with very diverse traits affecting their distribution in the landscape. In the frame of this opinion, two main aspects should be considered when the spatial boundaries are defined in which the risk of NTAs being exposed to PPPs will be performed.

Arthropods are almost all mobile or slow-moving species, with different degrees of mobility depending on species and life stage. However, for the purposes of this opinion, we define mobile and non-mobile species pragmatically based on the implications of their mobility for the risk assessment. A non-mobile species is defined as a species or life stage for which movement does not significantly impact the results of the experiment (for supporting data) or risk assessment. Examples of taxa which would typically be classified as non-arthropods are almost all mobile or slow-moving species, with different degrees of mobility depending on species and life stage. A practical definition could be that the individual’s home range during the assessment should be less than the size of the study area (i.e. plot, field or landscape), and that there should be no significant flow of movement through the study area during the study period. Examples of taxa which would typically be classified as non-mobile are scale insects (e.g. Coccidae, Diaspididae) or mealybugs (Pseudococcidae), several mites (Acari) and most insect larvae.

On the one hand, NTA species differ significantly in their mobility and ability to disperse in the landscape. This is of great importance when organisms might be exposed over time to different concentrations, e.g. moving in to (and possibly out of) treated fields. In this respect, only species with low mobility can be suitably addressed by the traditionally strict separation between in- and off-field risk assessment procedures. For further details on so-called source–sink dynamics relevant for NTA, please refer to sections 3.5 and 3.6.

On the other hand - and linked to the traits mentioned above - NTA populations might not be restricted in their boundaries to the size of treated fields. This means that a population of, for example, carabid beetles might cover in its range several landscape elements, including in-field and off-field areas.

Even if several combinations of traits leading to different distribution patterns are possible, not all potential combinations are likely to be found in agricultural landscapes. In Figure 1 below, possible NTA individual ranges and distribution patterns are given.



(©2009 Google – @2009 GeoBasis-DE/BKG (©2009 Vermessungsverwaltungen der Bundesländer und Bundesamt für Kartographie und Geodäsie)

(A) Range in-field; (B) range in- and off-field; (C) range off-field; (D) range bigger than field size. Species with relatively low dispersal and population range might fit to option A, B or C. Species with high dispersal potential might be species seen as having a distribution according to D.

Figure 1: Possible distribution of NTA individual ranges of different species in an agricultural landscape.

Only NTA species with individual range distribution patterns similar to A or C can be addressed with separated in- and off-field risk assessment.

For NTA populations with individual ranges covering various landscape elements (‘B’ or ‘D’ in Figure 1), toxic PPP sprayed in-field will impact the population dimension. If the species is very mobile, then the in-field might act as a sink for individuals that were off-field at the time of PPP application but move in later on.

NTA having individual ranges that fit options B and D in Figure 1 might be considered identical, as B could simply be a small population and thus could be subject to a higher extinction probability just because of its small size. In contrast, option D represents a rather large population in which extinction probability is not driven by its size but, on the other hand, because of the large area occupied, part of the population is always exposed to some pesticides. Given that the range of population D also contains habitats different form agricultural fields, the impact of PPPs on such species might be completely different. Again, these considerations are not necessarily true for small animals.

A NTA species that is able to build large-scale populations over time but has rather long dispersal times will be found in comparable individual densities in-field and off-field, only if it is resistant to all disturbances accompanying agricultural management practices. NTA species with higher sensitivity to mechanical (agricultural management) and/or chemical (e.g. PPP) stressors are not likely to be found in similar densities in- and off-field unless they are highly mobile.

In relation to the aim of this opinion, it was considered helpful to characterise species according to (the set of) traits that determine their distribution and dynamics in agricultural landscapes. This was done in chapter 4, where the NTA species considered to be the drivers of important ecosystem services in agricultural landscapes are grouped according to their traits and the probable resulting population shares in treated areas.

3.1.2. Spatial aspects in relation to the landscape to be assessed

Effects of PPPs on NTA populations cannot be debated without considering the landscape structure. As discussed already in section 2.1.1 and depicted in Figure 1, effects of PPPs depend on the species mobility, species physiology and behaviour, population size, metapopulation structure and sink–source dynamics. However, the sink–source dynamics depend directly on the landscape structure. The two actual landscapes shown in Figure 2 clearly exemplify this problem. The landscape on the left-hand side can support NTA populations by providing diversified refuge areas, consisting of meadows, woods and freshwater bodies. In this landscape, many NTA species would be able to maintain functioning spatially structured populations even with heavy in-crop losses caused by PPPs. In contrast, the right-hand landscape (Figure 2) may not provide enough habitat diversity to maintain the overall population structure. In this case, even relatively minor disturbances due to agricultural practices may bring many populations to extinction. As reported by Topping et al. (2014), experimental work with Staphylinidae, Linyphiidae and Carabidae indicates that the appropriate scale for assessing pesticide effects differs between taxa and is dependant upon the proximity of sources of re-colonisation as well as dispersal ability. Therefore, the authors suggest assessment of a range of landscape structures and management scenarios to ensure that any particularly hazardous combinations are identified. The sink–source dynamics that may take place in different landscapes are discussed in detail in the next section.

(©2009 Google – @2009 GeoBasis-DE/BKG (©2009 Vermessungsverwaltungen der Bundesländer und Bundesamt für Kartographie und Geodäsie)

Figure 2: Differently structured landscapes in Germany; same area depicted (4 hectares). The yellow-marked areas are in-field areas. Region in Rheinland-Pfalz (left) and in Brandenburg (right). Image modified from Brühl et al. (2013)

3.1.3. Source–sink dynamics in non-target arthropod populations in agricultural landscapes

Semi-field or field plots are often used to assess the impact and recovery times of pesticides or genetically modified (GM) crops on NTAs. The spatial dynamic issues surrounding the interpretation of the data generated in these studies were studied from 1988 onwards (Smart et al., 1989; Pullen et al., 1992; Thacker and Jepson, 1993) and the concept of source–sink dynamics (Pulliam, 1988) was introduced. Spatial dynamics whereby populations in areas with a negative population growth rate are maintained by dispersal from source populations is of fundamental importance. The conclusion from this work was that recovery as measured in field plots of, for example, 1 ha provided different relative effects compared with the impacts at a larger scale, and that dispersal capacity was crucial to determine the recovery rates (Thacker and Jepson, 1993). Subsequently, Topping et al. (2014) investigated the implications of these spatial dynamics on landscape-level impacts of pesticides on simulated beetles and spiders if small plots were used to determine the population risk based on recovery. The conclusion was that small plot measurements were an unreliable basis for prediction of population-level risk, potentially dramatically underestimating impacts.

Complex spatial dynamics at the landscape scale were also demonstrated for simulated vole populations (Dalkvist et al., 2013). Here, contrary to expectations, increasing the area treated with an endocrine disruptor by increasing the area of orchards led to lower population impacts and faster recovery. Also surprising was the fact that placing source habitats close to orchards improved recovery and decreased impact owing to rescue effects. These rescue effects can, however, become important depleting dynamics under different circumstances (Topping and Lagisz, 2012). In dispersive spiders, refuges were shown to be able to buffer considerable agricultural mortality impacts (Thorbek and Topping, 2005). The precise effect of landscape structure interacting with source–sink dynamics is therefore context dependent and difficult to generalise without more extensive reference work being available.

3.1.4. Consequences for non-target arthropod populations of source–sink dynamics and multi-year exposure

Source–sink dynamics or 'action at a distance' can have important consequences for NTA populations and the ecosystem services that these populations support. In addition, there can be year-on-year effects of continual exposure and pesticide dynamics that are not apparent immediately but take time to develop.

The persistence of NTA populations at the landscape scale is intricately connected with landscape configuration, specifically the abundance and distribution of exploitable resources (Ricketts, 2001; Fahrig and Nuttle, 2005; Swift and Hannon, 2010; Verboom et al., 2010), but also the dynamics of these (Wahlberg et al., 2002a,b).

Whilst there is focus on these issues for rare species, this is not the case for other more common species which are typically more important for supporting a well-functioning ecosystem and delivering ecosystem services. The exposure to shifting resources and shifting stressors in modern agricultural landscapes may cause declines in species abundance and may also cause non-equilibrium ecological conditions, where species will suffer conditions of extinction debt (Tilman et al., 1994). Extinction debt means that a species can still be present, but only because it takes some time period for the species to become extinct. The ultimate cause of this phenomenon is that the ecological conditions for the species are inappropriate, but, owing to spatial and population processes, the extinction time is long but inexorable.

3.2. Spatial boundaries at the local scale

The considered structures are defined as follows:

In-field: piece of land for cultivation with crops, managed typically by one farmer.

Off-field: area surrounding a field; either (semi-)natural habitats with high ecological value (such as hedgerow or grass strip) or simple structures (fence or a bare strip of land); normally no short-term changes in cultivation, in most cases not to be influenced by the farmer. Another off-field category comprises man-made structures, e.g. an adjacent field, roads, etc.

In-crop: the area actually cropped.

Off-crop: any uncropped area. It includes also uncropped areas in-field, and such areas can be, for example, the minimal required zone for agricultural management, buffer strips or beetle banks.

Buffer strip: in-field; cropped or uncropped zone of a defined width at the edge of a field which is influenced by the farmers action (e.g. spray drift). The buffer strip is normally enforced by authorities and underlies prescribed actions in order to meet the off-field specific protection goal. In addition, buffer strips may provide a recovery potential for the cropped area.

The buffer strip is located in-field and has the same protection goals as the in-field area plus the functions to mitigate exposure of the off-field area (drift and run-off reduction) and may serve as a reservoir for recolonisation of the in-field area if there is no suitable off-field habitat. The off-field protection goal is independent from the actual type of off-field habitat of individual fields.

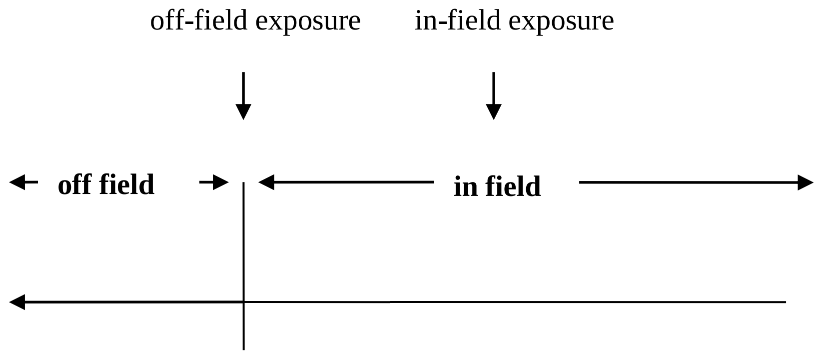
It is necessary to define the temporal and spatial boundaries of the off-field and the way the emission is translated to an exposure in the off-field area. These boundaries relate to the protection goal (where is the community of interest?) in relation to the route and distance covered of the emission coming from the in-field. The choice of such a distance will be the result of both scientific (e.g. is there a critical maximum area that can be at risk, without affecting the population of interest?) and regulatory decisions (is that distance acceptable from a regulatory point of view?).

Predicted environmental concentrations (PECs) could be provided for different distances from the field boundary and choices need to be made depending on the crop, group of non-target organisms and their specific protection goal. This PEC calculation allows definition of buffer strips and the risk assessment in the off-field area at the same time:

Step 1

The exposure and risk is assessed for the in-field and off-field areas. If the specific protection goals in-field and off-field are met, no further risk assessment or risk mitigation is needed.

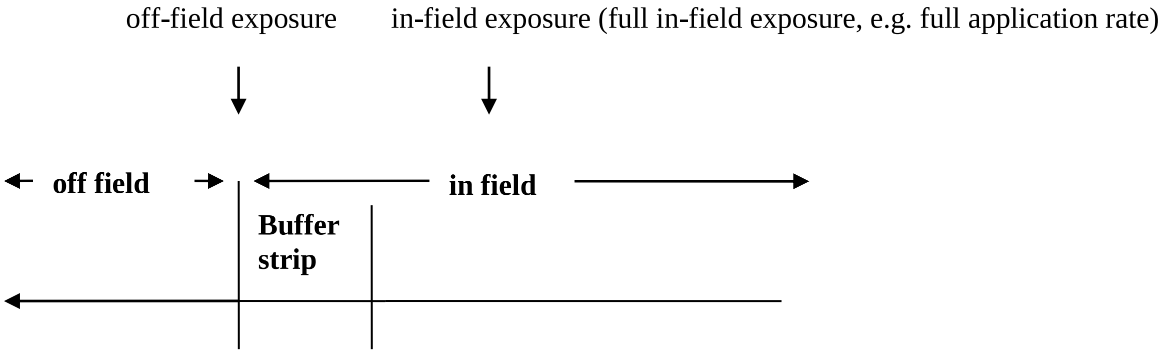
Exposure estimate



Step 2

A buffer strip is necessary if the off-field protection goal is not met in the previous risk assessment. The buffer strip is in-field. The maximum tolerable exposure to meet the off-field protection goal is calculated. The width of the buffer strip is calculated on the basis of this maximum tolerable exposure estimate, the required reduction factor and the reduction potential of the buffer strip. For example, for spray drift, the final width of the buffer strip depends on the combination of the height of the vegetation in the buffer strip and drift reduction techniques. If, for example, a wind break is in the in-field area, the drift to the off-field is significantly reduced compared with a buffer strip without vegetation. A table on reduction of spray drift from the combination of spray drift nozzles and width of the buffer zone can be found in Huijsmans and van de Zande (2011). In order to avoid overspray of the off-field area there is always a certain minimum distance needed from the treated crop to the off-field area.

Exposure estimate

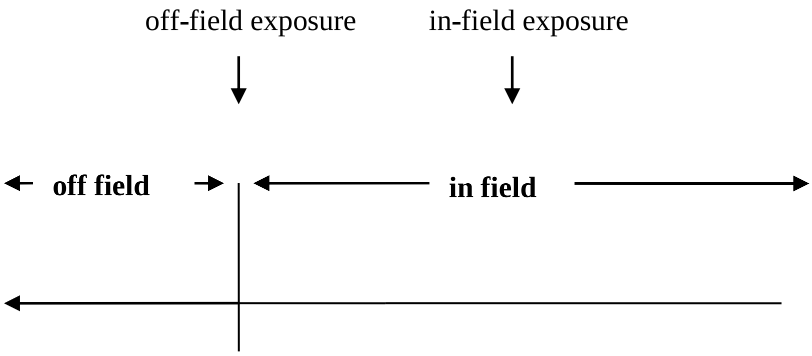


3.3. Example of use of boundaries in the risk assessment

The **initial assessment** should start with calculating the acceptable concentrations in the off-field area.

From this, it can be back-calculated at which distance from the last row of treated crop the off-field protection goal is met.

Exposure estimate



The risk assessment does not assume a pre-defined distance to the off-field. The exposure assessment starts at the field edge and calculates at which distance the off-field protection goal is met. Lets assume that the regulatory acceptable concentration (RAC) is equal to the amount of active substance at a distance of 5 m. The full risk mitigation equivalent to a 5-m buffer zone needs to be achieved in the in-field area. Standard options for reducing the width of the in-field buffer strip could be provided in the risk assessment, e.g. vegetation in the buffer strip of a certain height or wind breaks or drift reduction nozzles. The risk manager decides if the risk is manageable under the national conditions to achieve the required reduction of exposure in the off-field area (e.g. considering agricultural practice or national policy on implementation of buffer zones).

With this approach it is not needed to assume that the off-field protection goal needs to be met 1 m and 3 m away from the last row of the treated crop as was practice before. Thus, a risk management decision can be taken based on knowledge on how much distance is kept in their farms to the edge of the field (it may be different in Member States and crops) and based on national policies for implementation of buffer strips, e.g. obligatory vegetated buffer strips of a certain width.

3.4. Spatial boundaries at the landscape scale

For populations of NTA species displaying high mobility and dispersal rates, the assessment of local effect will not properly address effects of PPP application in-field on the off-field population share and eventually on the population persistence as a whole. Therefore, it is essential that the risk assessment considers the consequences of spatio-temporal variation in stressor dynamics and the interaction with NTA spatial dynamics. The risk assessment should therefore include an assessment of the impact of in-field effects in off-field sub-populations for mobile species with population ranges bigger than field size. If effects at this level are large enough, the risk assessment should also consider risk mitigation effects of landscape elements such as buffer strips/flowering strips for mobile species with population ranges bigger than field size. In cases where NTA abundance is too low to tolerate any PPP impacts, mitigation measures could also be evaluated as a method of raising NTA abundance to a level where the PPP could be applied. These evaluations also need to be considered at this landscape scale.

Modelling could offer a possibility to assess such effects. A simulation approach was used to demonstrate some of the spatial and temporal concepts considered important for a scientifically defendable risk assessment for NTAs at scales at which pesticides are used (i.e. landscapes) (see section 3.5, 3.6 and Appendix A for further details). There were three concepts specifically addressed by this work:

1. The potential of source–sink dynamics to indirectly affect untreated populations.
2. The impact of long-term year-on year application of the pesticide.
3. The effect of landscape structure.

The simulations demonstrate that:

1. There is an off-field effect from in-field mortality, although the exposure in the off-field was set to zero in the model (no spray drift to off-field). Annual effects of up to 70 % or mean effects of 26 % reduction in off-field population size were predicted after 10 years.
2. The assessment based on a single spray would underestimate the long-term effects. This was demonstrated by the fact that, at high toxicities, the population decline was still ongoing after 20 years of pesticide use. Even standard toxicity scenarios required three years for populations to stabilise.
3. Landscape structure also clearly influenced the results as shown by differences between landscapes and between field boundaries and unsprayed margin scenarios.

3.5. Simulations illustrating some spatio-temporal concepts important for non-target arthropod risk assessment

3.5.1. Introduction

Using a carabid beetle example, we employed a simulation approach to demonstrate some of the spatial and temporal concepts considered important for a scientifically defendable risk assessment for NTAs at scales at which pesticides are used (i.e. landscapes). There were three concepts specifically addressed by this work:

1. The potential of source–sink dynamics to indirectly affect untreated populations.
2. The impact of long-term year-on-year application of the pesticide.
3. The effect of landscape structure.

The following spatial elements are used:

- In-crop: actually cropped area, the in-crop area is equal to the in-field area in the simulations without field boundaries and field margins.
- Field boundary: a permanent grass strip surrounding the crop (i.e. an in-field buffer strip managed as a permanent grass strip).
- Field margin: non-sprayed crop (i.e. an in-field buffer strip managed as unsprayed crop).

- In-field: consists of cropped area and, depending on the scenario, it can also include field boundaries and field margins.
- Off-field: everything which is outside of the in-field area (not a cropped area and not a field boundary or a field margin).

3.5.2. Methods

More detailed methods are presented in Appendix A and an overview is given here. The simulations were run using the *Bembidion lampros* model (Bilde and Topping, 2004) of the ALMaSS system (Topping et al., 2003), a model system originally designed to provide answers to policy-level questions related to changing land-use or management and the resultant impacts on animal wildlife. *Bembidion lampros* is considered to be a useful natural enemy of pests in agricultural fields (e.g. Edwards et al., 1979; Ekbom et al., 1992; Humphreys and Mowat, 1994).

3.5.2.1. Scenario set-up

A pesticide causing an 80 % field mortality rate to adults only over seven days, if present in the environment above a toxic threshold, is assumed. A DT_{50} of 10 days and an application rate of twice the LR_{80} to fields was also assumed. As the aim was to demonstrate the concepts listed above and not to undertake a real risk assessment for product, some simplifications of scenarios were made. For example, to isolate the effect of source–sink dynamics, drift to off-field areas was set to zero to prevent interaction between drift-induced direct impacts on off-field areas and source–sink impacts. For all simulations, application of the pesticide was carried out twice during the activity time of the adult beetles: the first applied on 31 May and the second applied 20 days later.

Two landscapes (Præstø and Herning: Pr and He, respectively) with differing field sizes and compositions of elements were used in two forms. In the first, all grassy field boundaries were removed from the landscape (FB0) and, in the second, based on FB0 100 %, fields had grassy boundaries added (FB100). In all cases, it was assumed that all fields were under continuous winter wheat cropping, and if the pesticide was applied it was applied to all fields.

The width of the field boundary was altered as a scenario variable for the FB100 scenarios. Three widths were used: 1 m, 5 m and 10 m (FB100X1, FB100X5 and FB100X10, respectively). The resulting area cover for field boundaries in the two landscapes was markedly different, from 1.0 to 1.75 % cover as a proportion of field area. For FB100X1 scenarios, three variants with different widths of unsprayed crop margins were also used of 2 m, 5 m and 10 m (USM2, USM5 and USM10).

To evaluate long-term usage of the pesticide, the toxic threshold was assumed to be lowered by factors of 2, 5 and 10 for a subset of scenarios. These settings were used with scenarios with zero field boundaries only (FB0, see below) (TX2, TX5 and TX10, respectively).

All landscapes described above were simulated with beetles for 30 years with both baseline and product runs. Baseline run conditions were identical to the product run except no insecticide was applied to the winter wheat fields. Data were extracted from the simulations only after the first 10 years of simulation to allow the populations to equilibrate with the landscape (burn-in period).

Weather conditions were selected to represent the decade, 1990–1999 from central Denmark. Each simulation was run for a total of 30 simulation years, looping the, 1990–1999 weather data three times.

3.5.2.2. Simulation data extraction

Two main sets of data were extracted from all the simulation runs. These were the abundance–occupancy ratio index (AOR index, for full explanation see Appendix A), information which describes changes in abundance and distribution of a population (Hoye et al., 2012), and data on numbers of adult female beetles extant throughout the simulation.

3.5.3. Results

Total beetle populations were comprised of approximately two-thirds in-crop and one-third off-field beetles in both landscapes. Beetle populations in the Herning landscape were 19–27 % higher than the corresponding Præstø landscape scenario.

Standard errors, representing uncertainty owing to within-simulation stochastic variability only, for the in-crop and off-field percentage reductions in Table 2 were computed from 10 replicate runs of the model for each scenario and in all cases were less than 0.17 % giving a margin of error less than 0.40 % for any single scenario and a margin of error less than 0.55 % for comparing any two scenarios. Simulation replicates were therefore very similar and no more than 10 replicates were needed.

The annual variation in beetle numbers measured as the mean abundance over each 12-month period showed considerable variation. This variation was related to changes in weather, with repeating 10-year cycles being clearly visible (Figure 3). There was, however, an interaction with landscape such that between-year differences were consistent within a landscape but not between Præstø and Herning.

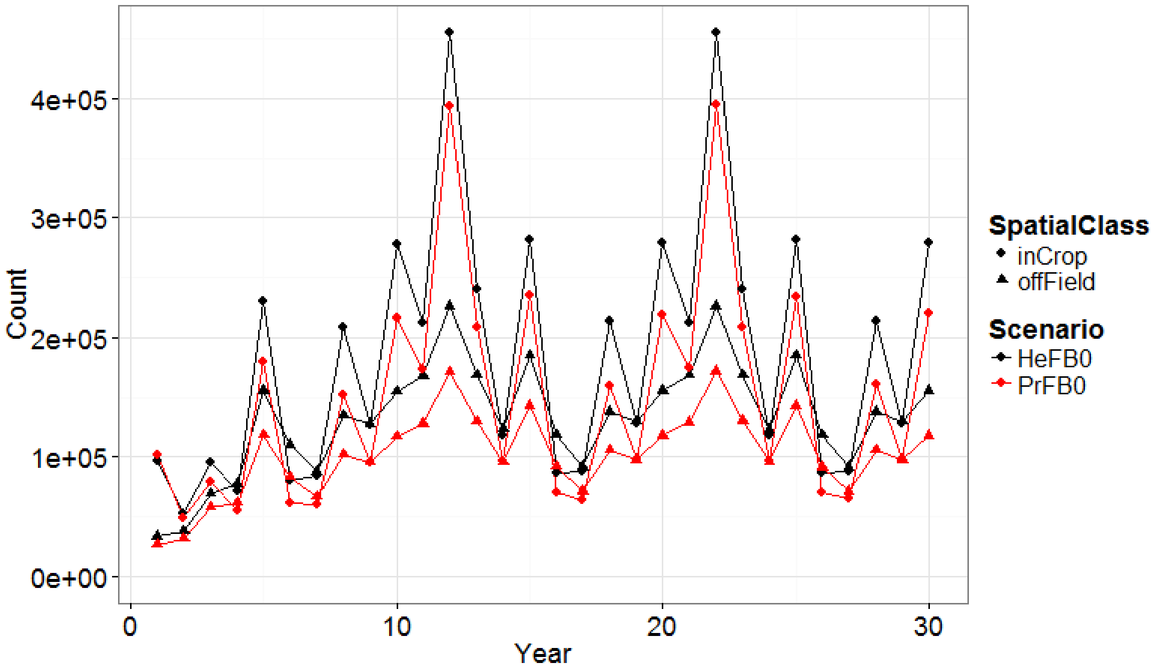


Figure 3: Annual mean 12-month beetle counts for Herning (He) and Præstø (Pr) landscapes with zero field boundaries (FB0) showing in-crop and off-field populations when no test pesticide is applied

To simplify graphical comparisons and to remove the direct effect of weather variation, impacts of the weather scenarios where a test pesticide was applied are shown as population size relative to the appropriate baseline. Hence, a value of 100 % indicates no impact. The impact of the pesticide was to reduce populations by approximately 30 % in-crop in the first year after application and approximately 15 % in off-field for the FB0 scenario in both landscapes (Figure 4). In FB0 scenarios, off-field habitats were hedge-banks, pasture and other grassy areas (e.g. roadside verges). Subsequent in-crop impacts varied between 10 % and 60 % depending upon the weather year. Off-field impacts varied from 10 to 45 %. In-crop impacts were higher in Præstø, but off-field impacts were consistently higher in the Herning landscape (Figure 4).

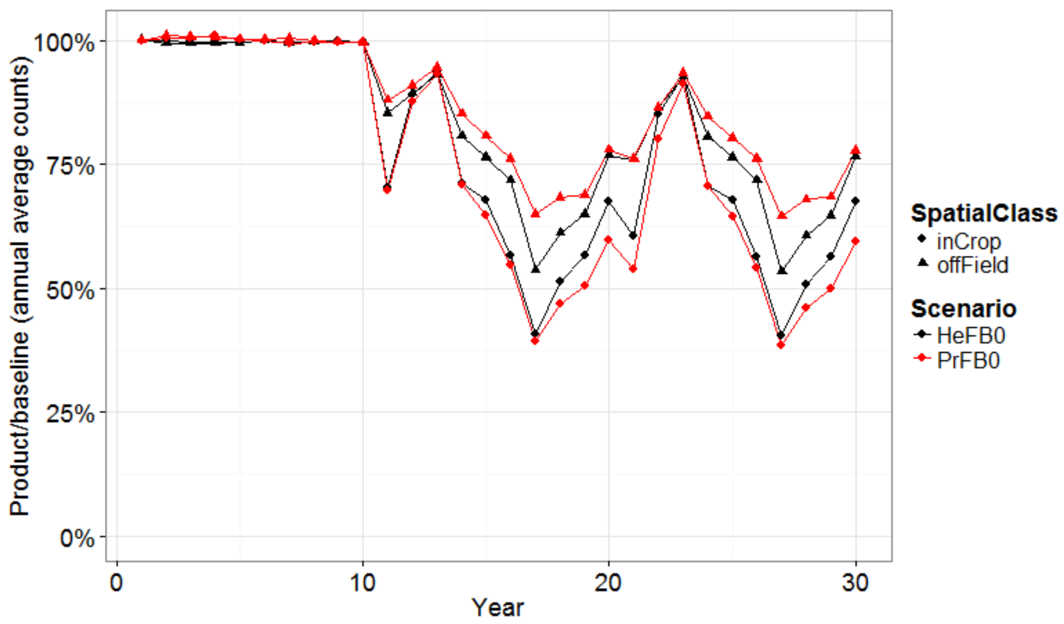


Figure 4: Annual mean 12-month baseline/pesticide applied scenario counts for Herning (He) and Præstø (Pr) landscapes with zero field boundaries (FB0) showing in-crop and off-field

The stabilisation patterns shown for Herning and Præstø were qualitatively similar for the majority of the scenarios considered (varying toxicity threshold being the exception). Results are therefore summarised as mean population impact over the final 10 years of simulation (Table 2). Impact ratios between in-crop and off-field impacts clearly show a difference between the two landscapes, with impacts in off-field populations in Herning always being higher than in Præstø. Conversely, impacts in-crop are consistently higher in Præstø.

Increasing the width of field boundaries decreased both in-crop and off-field impacts. Impacts of adding 1-m boundaries to the landscape reduced in-crop impacts by 25 and 27 % in crop and 44 and 54 % off-field for Herning and Præstø landscapes, respectively. Increasing the size of field boundaries to 5 m decreased impacts by 34 % and 36 % in-crop and 51 % and 59 % off-field, whilst increasing boundary width to 10 m decreased impacts further by 45 % and 46 % in-crop and 59 % and 65 % off-field for Herning and Præstø, respectively.

The effect of adding unsprayed field margins was similar to increasing field boundaries, but was at a lower magnitude. Adding a 10-m unsprayed margin to a 1-m field boundary decreased pesticide impacts by 14 % and 12 % in-crop and 15 % and 14 % off-field for Herning and Præstø, respectively.

Decreasing the pesticide toxicity threshold altered the impact of the pesticide application considerably. Doubling the sensitivity of the beetles increased population impacts by 38 and 40 % in-crop and by 41 and 45 % off-field (Herning and Præstø, respectively). A five-fold increase in sensitivity led to an increased impact of 133 and 172 % in-crop and of 153 and 159 % off-field (Herning and Præstø, respectively) and a 10-fold increase in sensitivity led to an increased impact of 176 and 204 % in-crop and of 224 and 204 % off-field (Herning and Præstø, respectively).

Table 2: Scenario definitions and impacts in proportion reduction of mean annual population of all scenarios relative to their respective baseline when product is applied (In-crop% and Off-field%)

Scenario	Site	Field boundary %	Field boundary width (m)	Unsprayed margin width (m)	Toxicity threshold	In-crop %	Off-field %	Impact ratios
HeFB0	He	0	0	0	25.0	35	26	0.74
HeFB100X1	He	100	1	0	25.0	26	20	0.74
HeFB100X5	He	100	5	0	25.0	23	17	0.75
HeFB100X10	He	100	10	0	25.0	19	15	0.75
HeFB100X1_USM2	He	100	1	2	25.0	25	19	0.74
HeFB100X1_USM5	He	100	1	5	25.0	24	18	0.74
HeFB100X1_USM10	He	100	1	10	25.0	22	16	0.74
PrFB0	Pr	0	1	0	25.0	39	22	0.57
PrFB100X1	Pr	100	1	0	25.0	29	18	0.63
PrFB100X5	Pr	100	5	0	25.0	25	16	0.64
PrFB100X10	Pr	100	10	0	25.0	21	14	0.64
PrFB100X1_USM2	Pr	100	1	2	25.0	28	17	0.62
PrFB100X1_USM5	Pr	100	1	5	25.0	26	16	0.62
PrFB100X1_USM10	Pr	100	1	10	25.0	24	15	0.61
PrFB0_TX2	Pr	0	0	0	12.5	55	32	0.59
PrFB0_TX5	Pr	0	0	0	5.0	88	58	0.66
PrFB0_TX10	Pr	0	0	0	2.5	97	68	0.70
HeFB0_TX2	He	0	0	0	12.5	49	37	0.76
HeFB0_TX5	He	0	0	0	5.0	82	66	0.80
HeFB0_TX10	He	0	0	0	2.5	97	84	0.87

Impact ratios are the ratio between in-crop and off-field impacts.
He, Herning; Pr, Præstø; FB, grassy field boundary; USM, unsprayed field margins; TX, pesticide toxicity threshold decrease factor.

AOR measurements also showed minimal variability. The standard error for each abundance estimate was less than 0.02 and for each occupancy estimate less than 0.05 %; hence, even small differences are the result of scenario factors and not noise in the dataset. Relative impacts can be visualised using standard AOR index plots (Figure 5). In both landscapes, the major impacts are much larger in zero field boundary landscapes, and are reduced maximally by having a 10-m field boundary. A reduction in impacts occurred with increasing field boundary or unsprayed margin width in both landscapes following a similar pattern. There were consistent differences between the two landscapes in the responses to pesticide, higher impacts on abundance in Herning and higher impacts on occupancy in Præstø.

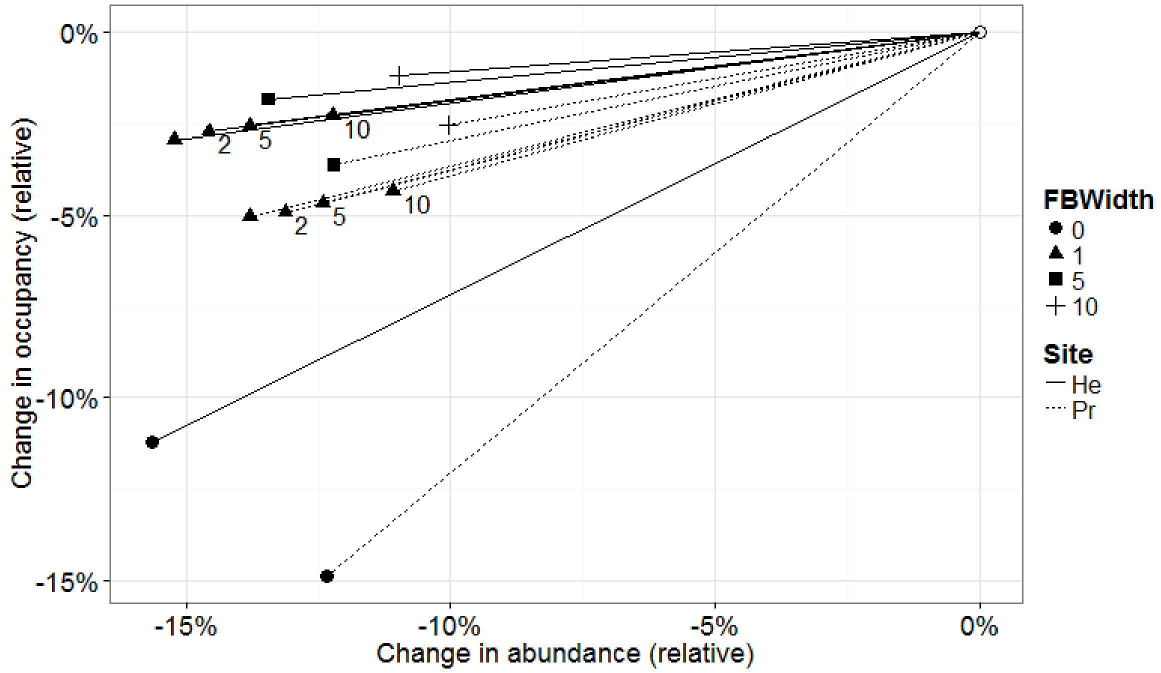


Figure 5: Changes in occupancy and abundance when pesticide is applied to standard toxicity scenarios. Numbers refer to the width of unsprayed margin if present

Increasing toxicity had major impacts on both abundance and occupancy with similar patterns in both landscapes (Figure 6). Impacts on abundance were, however, generally greater in the Herning landscape, whereas impacts on occupancy were broadly similar with a tendency for higher impacts in Præstø.

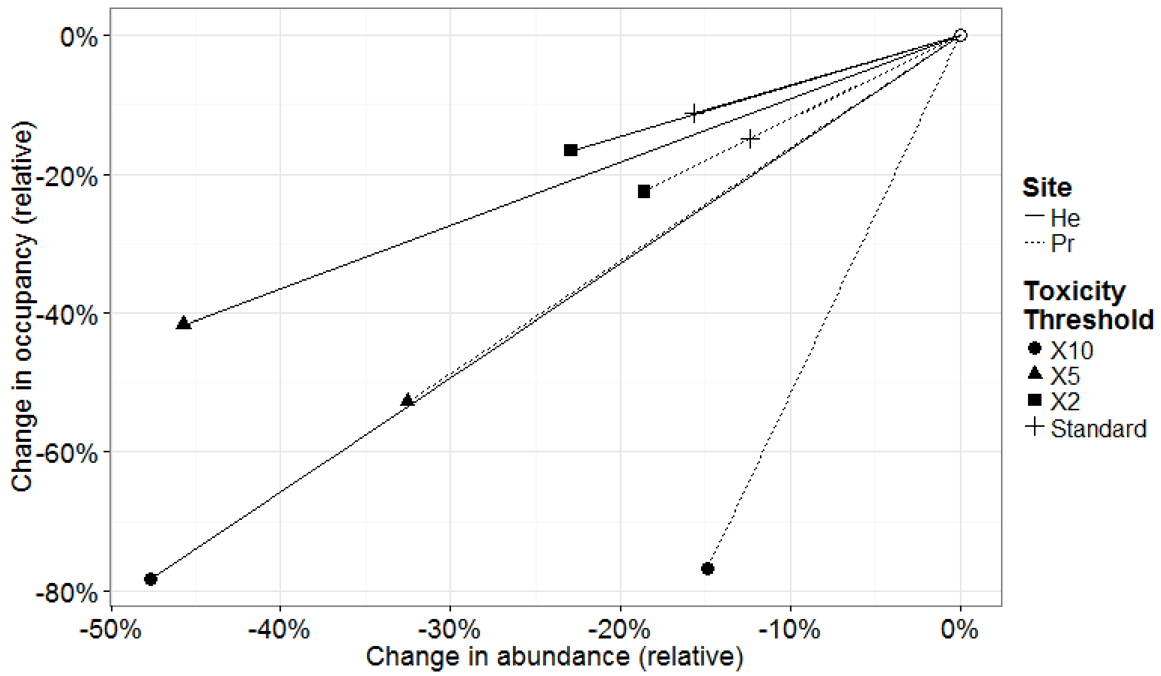


Figure 6: AOR plots for zero field boundary scenarios with differing pesticide toxicity (beetle sensitivity). The toxicity value of 25 is the standard value used in all other scenarios

However, relative impacts hide major baseline differences between scenarios. Figure 7 shows changes in occupancy and abundance for the standard toxicity scenarios when pesticide is applied. The baseline population conditions vary considerably between scenarios. Adding field boundaries increased population abundance by approximately 50 %, whereas subsequent increased width of these field boundaries increased occupancy with a maximum range of 67 to 74 % in Herning.

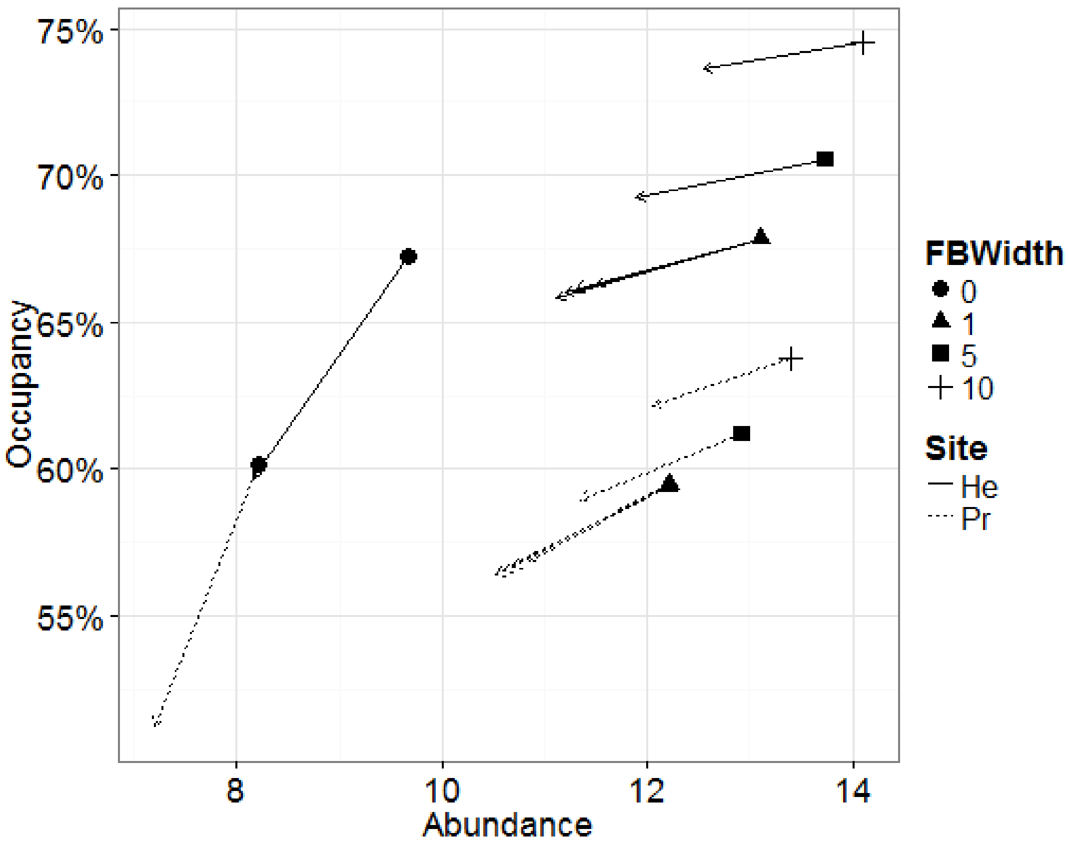


Figure 7: Changes in mean annual occupancy and abundance when pesticide is applied for all standard toxicity scenarios. A 1-m field boundary (FBWidth) includes both unsprayed margin and without unsprayed margin scenarios. Arrows indicate the changes when pesticide is applied

Temporal effects

The ratio of impacts relative to baseline between similar weather years (10 years apart) was not constant but approached it in all standard toxicity scenarios after three years. However, increasing the toxicity of the pesticide increased the time to population stabilisation from typically three years in the standard scenarios to greater than 10 years for very high toxicity scenarios (Figure 8). Speed of relative population stabilisation was similar between the Præstø and Herning landscapes. In all cases, in-crop stabilisation was slower than off-field stabilisation for the same scenario. Slopes of very high toxicity scenarios (value X10) were steeper than the next highest toxicity (value X5).

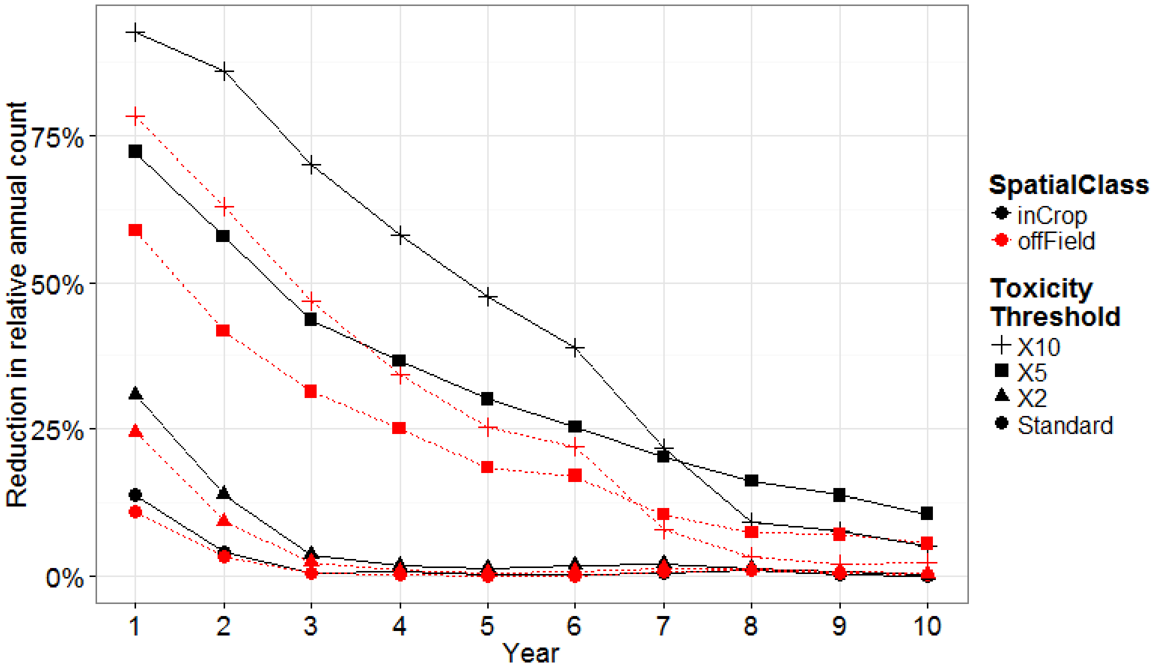


Figure 8: The reduction in relative annual count for years 1–10 after pesticide treatment started compared with years 11–20 for in-crop and off-field areas in Herning. A value of zero indicates that the population level has stabilised relative to the first 10 years after application

3.6. Illustration of the non-target arthropod risk assessment concepts using the Bembidion example

The model used was of an abundant spring breeding carabid beetle; hence, the observations and conclusions drawn are based on the use of this species as an example. While the actual numbers and precise mechanisms used to arrive at these predictions will not match other species with differing life cycles, the fact that the spatial dynamics of this species can result in the observed results should be of general concern. There is no reason to assume that these are concepts that only apply to this species, and some concepts such as source–sink dynamics can be expected to have even greater impacts in species actively attracted into crops (e.g. pollinators).

3.6.1. Concept 1: the potential of source–sink dynamics to indirectly affect untreated populations

All scenarios clearly indicate the impact of source–sink dynamics in these assessments. Impacts in off-field populations can be high, especially where there is little off-field area. Both unsprayed field margins (= unsprayed crop margins) and field boundaries reduced impacts to off-field areas. The effect of unsprayed field margins was positive, but less so than the effect of adding a field boundary (although, in all cases, it was assumed that there was also a 1-m field boundary). Provision of wider margins also reduced the impacts in-crop as a result of healthier off-field populations acting as sources.

It is important to remember, however, that these scenarios assumed zero spray drift to off-field areas; hence, impacts to off-field areas in real situations would be expected to be larger if drift occurs.

3.6.2. Concept 2: the impact of long-term year-on-year application of pesticide

Impacts on the population were not immediate. This is important because field experiments used to evaluate impact of pesticides on non-target organisms normally consider up to only a one-year time frame after first application. Year-on-year application will therefore give a greater overall population impact than would be measured from a single application. It is also important to note that the impact varied with the weather year, e.g. from 30 to 70 %. Experimental systems in place for only a single season will miss this variation.

In this case, toxicity of the pesticide was clearly a critical factor in determining impacts, with a 10-fold increase in beetle sensitivity leading to long-term population declines of over 90 %, or expressed as change in occupancy and abundance a decrease of 80 % in abundance and 50 % occupancy. While this is not surprising in itself, it is not always the case. In voles, similar modelling has demonstrated that ecological factors are at least as important as toxicity in determining impacts at population levels (Dalkvist et al., 2009).

3.6.3. Concept 3: the effects of landscape structure on risk assessment

Overall, the results indicate the importance of taking landscape structure into account in a risk assessment. Landscape structure resulted in differences in effect at large scales. Impacts in Herning were generally on abundance, whereas impacts on Præstø occupancy were higher. In baseline scenarios, abundance was similar in both landscapes, but occupancy was much higher in Herning. As a result, in-crop impacts in Præstø were higher than Herning, but off-field impacts were higher in Herning. This suggests that the larger off-field population in Herning was buffering the in-crop population more efficiently than in Præstø; however, this larger off-field population also suffered the largest proportional impacts. Hence, depending upon the definition of the protection goals, this could result in the populations exhibiting the best post-pesticide application health also being designated as those most at risk. Previous studies have also demonstrated how the geometry of artificial landscape representation can bias simulation outcomes (Holland et al., 2007).

3.6.4. Other observations

Unsprayed margins reduced both in-crop and off-field impacts, but less so than additional or wider field boundaries. This effect was, however, measured without taking into account the reduction in spray drift that unsprayed margins would also provide in the real world. This model result is therefore because the unsprayed margin only acted to remove a proportion of the field population from the effects of the pesticide, whereas the wider field boundary does the same but also provides over-wintering habitat. This particular result is thus a result of the specific life history strategy of the beetle in question. However, in general, it can be expected that habitat provision will have a greater influence than a reduction in mortality for a small proportion of the population over a short period (i.e. when the pesticide is present).

In this case, it is clear that, even with pesticide applied, the condition of the population in landscapes with a minimum of 5-m field boundaries is at least as good as the zero field boundary landscapes without pesticide (compare filled circle positions with the arrows from the square in Figure 7). This indicates the potential to use results such as these to carry out an analysis of potential mitigation strategies. If real landscape conditions were taken into account, addition or widening of field boundaries could be considered as a way to mitigate the impact of a pesticide, and the state of the population with pesticide and mitigation strategy compared with a baseline condition to evaluate overall impact using the model framework. A case could also be made for a 1-m field boundary with pesticides being as good as no field boundaries without pesticides in Herning. Here, although there was a decrease in occupancy, there was still a large increase in abundance which could be considered to be of greater importance. This in effect means that the range of the beetle was reduced but, where it was still present, the densities were higher.

3.7. Conclusions of the simulation exercise

Simulations demonstrate the three concepts stated as the aim of the modelling exercise:

1. We demonstrate an off-field effect from in-crop mortality. Annual effects of up to 70 % or mean effects of 26 % reduction in off-field population size were predicted after 10 years (even without spray drift).
2. The assessment based on a single spray would underestimate the long-term effects. This was demonstrated by the fact that, at high toxicities, the population decline was still ongoing after 20 years of pesticide use. Even standard toxicity scenarios required three years for populations to stabilise.
3. Landscape structure also clearly influenced the results as shown by differences between landscapes and between field boundary and unsprayed margin scenarios.

4. Specific protection goals in the risk assessment for non-target arthropods exposed to plant protection products

4.1. General considerations

The general protection goal, as defined in Regulation (EC) No 1107/2009, ultimately aims at protecting biodiversity and ecosystems. Therefore, the definition of specific protection goals in the risk assessment for NTAs has the aim of implementing this general protection into (i) explicit and viable mandates for risk assessors and (ii) practicable and effective suggestions for risk management measures.

A procedure to define specific protection goals has been developed by EFSA in consultation with stakeholders and has been published in the 'Scientific Opinion on the development of specific protection goal options' (EFSA PPR Panel, 2010).

As stated in this EFSA opinion, 'the Panel defines SPGs [specific protection goals] by the entities that need to be protected, the attributes and/or functions of those entities, as well as the magnitude, temporal and spatial scales of effects on these attributes and/or functions that can be tolerated without impacting the general protection goal' (EFSA PPR Panel, 2010).

Final decisions on the choice of specific protection goals involves risk management judgements, which are outside the remit of EFSA and the PPR Panel, and therefore need to be made in consultation with risk managers. If supported by scientific evidence, alternative options for specific protection goals might be developed in order to facilitate the consultation.

In the EFSA opinion (EFSA PPR Panel, 2010), several steps are proposed in order to identify and justify specific protection goals for aquatic and terrestrial organisms that may be affected as non-target organisms by PPP use.

The first step in the definition of a specific protection goal is the identification of important ecosystem services that are provided by agricultural ecosystems. By means of describing services that mankind receives from ecosystem performance, abstract ecological entities and processes become explicit (e.g. MEA, 2005a).

However, as reported in the Millennium Ecosystem Assessment (MEA, 2005a), 'modifications of ecosystems to enhance one service generally have come at a cost to other services due to trade-offs. Only 4 of the 24 ecosystem services examined in this assessment have been enhanced: crops, livestock, aquaculture, and (in recent decades) carbon sequestration. In contrast, 15 other services have been degraded [...]'. The impacts of these trade-offs should also be clearly described for ecosystem services in agricultural landscapes, so that risk managers can decide if and to what extent costs of trade-offs should be tolerated. In this respect, MEA (2005a) claims that 'many of the costs of changes in biodiversity have historically not been factored into decision-making'.

Despite the several criticisms that the concept of ecosystem services has been subjected to, namely its anthropocentrism (as it suggests that ecosystems should be protected in view of its value for mankind, as stated above), this concept is a valuable tool in characterising and communicating the protection goal. In view of this opinion, several ecosystem services were identified as being driven by NTAs in the agricultural landscape. These services are:

1. Biodiversity, genetic resources. NTAs are an extremely diverse organism group that contributes highly to biodiversity in agricultural landscapes.
2. Education, inspiration and aesthetic value. Several NTAs in agricultural landscapes are highly valuated for their cultural services (e.g. butterflies) and/or are listed as rare species with a species conservation status.
3. Regulation of arthropod pest populations. NTAs are valuable antagonists of pests affecting crop plant species. Moreover, NTAs are also antagonists of pests and parasites of mammals; for example, several ant species are known to be tick predators, and coleopterans are predators of cattle parasites.
4. Food provision. NTAs are the most significant and crucial part of the diet for organisms at higher trophic levels, e.g. amphibians, reptiles, birds and small mammals.
5. Pollination. NTAs are—as well as bees—effective pollinators of several plant species in- and off-crop.

The second step in the definition of a specific protection goal **is the characterisation of the main drivers** behind the ecosystem services deemed to be important in the agricultural landscape. In the sections dealing with the specific protection goals in the present opinion, NTA species and/or groups have been identified as having, through their activity (or presence), major influence on the service to be preserved.

The third step is the determination of the drivers' ecological entity to be considered in respect of the assessed ecosystem service. EFSA (PPR 2010) suggests to differentiate between the ecological entities 'individual', '(meta)population', 'functional group' and 'ecosystem'. The concept is based on the assumption that addressing organisms at one level of organisation will protect those at higher organisation level. For example, if the ecological entity to be protected is the 'individual', this means that the entities '(meta)population', 'functional group' and 'ecosystem' will implicitly be protected.

The ecological entity addressed in the assessment is specified in the definition of the single protection goals. In general, NTAs are not protected at individual level, and, considering the services mentioned above, where NTAs are the key drivers, the ecological entity to protect is, depending on the service, the (meta)population or the functional group (see details below). The term 'metapopulation' is not used in this opinion, as contradictory definitions regarding NTA species might lead to misunderstanding of the general concepts suggested. Instead, it is referred to as spatially structured populations of NTAs in the landscape. If it will be agreed by the risk managers that the level of protection should be different in in-field and off-field areas, different ecological entities might be selected for the same service.

The fourth step is the determination of the drivers' attribute to be measured in the assessment. Changes in behaviour, on survival and growth, in abundance/biomass, in a process rate or in biodiversity are suggested by EFSA (EFSA, 2010) as possible measurements to be done at the different drivers considered. In the case of NTAs, and according to the ecological entities considered in the previous step, the most reasonable attribute to measure will be abundance and/or biomass (see details below).

The fifth step is the determination of the magnitude of effect on the drivers that could be tolerated regarding the impact on the ecosystem service without affecting the general protection goal.

In the following, a partitioning of the magnitude of effects is proposed, which is derived from general effect classes in ecotoxicology. These effect classes are deemed to be pertinent for the assessment of effects on NTAs at local scale.

Scaling of the magnitude of effects on (meta)population/functional group/biodiversity (please also refer to the glossary):

- large effects: pronounced reduction above 65 %;
- medium effects: reduction between 35 % and 65 %;
- small effects: reduction above 10 % and below 35 %;
- negligible effects: reduction up to 10 % (comparable to non-detectable effects).

Effects on NTA adult survival and effects on reproduction should be considered separately but the overall effects on the (meta)population or functional group should be determined.

Regarding the magnitude of effects on NTAs arising from several years of PPP exposure in a landscape context, relevant measurement endpoints are still to be agreed upon in the scientific community. If the assessment of these effects is based on population models that address effects of PPPs on species, efforts should be made to identify those simulation endpoints that can be related to the magnitude of effects as defined above.

Experience is lacking on the consequence of effects on NTA species as predicted by population models, and no strict definition of effect classes can be given at present. Depending on the endpoint to be chosen in future for the assessment of PPP effects at landscape scale, negligible, small, medium and large effects will have to be defined (please also refer to the glossary). As modelling endpoints integrated several years of PPP application in a wider landscape context, it is anticipated that tolerable effects at landscape scale will be of lower magnitude than the effect classes given above regarding the assessment at local scale.

In particular, the definition of 'negligible effect' in a landscape context will be of great importance. For the time being, it can be stated that negligible effects should address the limits for the landscape-scale population. This means, on the one hand, that year-on-year decline in abundance should not be observed. On the other hand, negligible effects should also account for population range restrictions; here, not only individual abundance but also range of occupancy should not be reduced by more than a level that will be considered to be negligible.

As the landscape context in which effects occur deeply influences the degree of acceptability of the effects in the short and long term, the definition of a general acceptable percentage reduction compared with a control allocated in the same system is not possible. In terms of this opinion, the definition of a possible acceptable magnitude of effects as percentage reduction compared with a 'control' applies to a defined landscape context, i.e. landscapes supporting the highest NTA diversity that can be achieved in managed agricultural systems. In landscapes with lower NTA diversity, the acceptability of the same effects might be at a far lower magnitude level. This applies to all proposed specific protection goal options for NTAs.

For certain ecosystem services (e.g. minimum abundance of arthropods needed to support the survival of bird chicks), thresholds marking tipping points for the provision of the service might be defined. If no absolute threshold can be defined, maximum magnitudes of effects on NTA drivers are suggested marking the acceptable limits, in scientific terms, for the maintenance of the assessed service at a desired rate and ultimately for the general protection goal (EFSA PPR Panel, 2010). This means that, if such limits are breached, severe consequences for the ecosystem functioning and for stakeholders who rely on certain services can be expected. These limits mark the upper range of the magnitude of effects in the different specific protection goal options. The lower end of magnitude of effects in the specific protection goal options is set where no or negligible effects are observed in the NTA drivers, with no or negligible impact on the provision of the specific ecosystem service.

The sixth step is the determination of the temporal scale to be considered together with the magnitude of tolerable effects.

This step is of particular importance when addressing effects other than negligible effects, as it implies that some effects might be tolerable if ecological recovery occurs within a specified period. As stated in the EFSA Guidance on the risk assessment for aquatic organisms (EFSA PPR Panel, 2013), when including ‘recovery to identify (un)acceptable effects, all relevant processes that determine population viability and the propagation of effects to the community-, ecosystem- and landscape-level are to be considered’. In this respect, multiple applications of PPPs might pose a constraint to recovery processes in agricultural landscapes, in particular consecutive PPP uses throughout crop spraying schedules.

Considering the ecosystem services identified above in step 1, their timely provision might be of central importance. For example, as described for the ecosystem service ‘food web support’ (for details, please refer to section 4.2.4), effects occurring when birds raise their young have the highest implications which cannot be compensated for by recovery occurring several months later.

NTAs may display multi-, uni- or semivoltine life history strategies. For univoltine and semivoltine species, full recovery from chronic effects might be observed only one year after PPP use. Therefore, the Panel considers time lapses of one year or more as relevant for the demonstration of, for example, long-term effects on NTA species that may emerge after several years of PPP use or for the demonstration of recovery of species with long life cycles.

However, regarding the ecosystem services driven by NTA and having impact on other organisms, e.g. ‘pest control’, ‘pollination’ or ‘food web support’, full recovery in time ranges greater than the growing season are probably not adequate for the implementation of the protection goals.

Temporal scaling of effects on NTA drivers:

> 1 year: this temporal scale is not considered adequate for the implementation of protection goals except for effects other than negligible ones in terms of this opinion. Negligible effects are considered as no effect

Months: maximum of 6 months

Weeks: up to 4 weeks

Days: up to 7 days

4.2. Specific protection goals regarding effect assessment

4.2.1. Non-target arthropods as drivers in maintaining biodiversity and genetic resources in agricultural landscapes

Several reviews and books and numerous other publications have dealt in the last decades with the importance of biodiversity for supporting and delivering ecosystem services beneficial to mankind. For the purpose of this opinion, we do not give details of those evaluations but refer to the specific publications and the references therein (e.g. Altieri, 1999; Collins and Qualset, 1999; Hooper et al., 2005; MEA, 2005; Balvanera et al., 2006; Cardinale et al., 2012; Maier, 2012; EFSA, 2013a).

Essential statements on the current knowledge consensus on wider biodiversity are reported below from the work of Cardinale et al. (2012). These statements are essentially agreed upon in the scientific community and have fundamental implications for the assessment of biodiversity in agricultural landscapes:

1. There is unequivocal evidence that biodiversity loss reduces the efficiency by which ecological communities capture biologically essential resources, produce biomass, and decompose and recycle biologically essential nutrients.
2. There is strong evidence that biodiversity increases the stability of ecosystem functions over time.
3. The impact of biodiversity on any single ecosystem process is non-linear and saturating, such that change accelerates as biodiversity loss increases.
4. Diverse communities are more productive because they contain key species that have a large influence on productivity, and differences in functional traits among organisms increase total resource capture. Loss of diversity across trophic levels has the potential to influence ecosystem functions even more strongly than diversity loss within trophic levels.
5. Functional traits of organisms have large impacts on the magnitude of ecosystem functions, which give rise to a wide range of plausible impacts of extinction on ecosystem function.

Biodiversity as defined above is the ‘variety of life, including variation among genes, species and functional traits’. By contrast, in the discussion about definition of biodiversity as a protection goal in agricultural systems, EFSA (2013) states that the species diversity *per se* is often defined as ‘structural biodiversity’ versus a so-called ‘functional biodiversity’. It is agreed by the Panel that structural biodiversity delivers through the magnitude of the different species’ traits the ‘functional biodiversity’. Functional biodiversity focuses on the specific function that (a group of) species exerts in the performance of processes of interest. A clear assignment of a species to a functional group is, however, not possible. One single species might have different functions in different processes, as several traits can be assigned to every single species, and every species in turn is uniquely characterised by its trait configuration (e.g. Gardner et al., 2014). In addition, each ecosystem process involves differently assembled ‘functional groups’.

Given that the knowledge on the functions of the extremely diverse NTA species in ecosystem processes is far from being comprehensive, it is not recommended to address ‘biodiversity’ of NTAs at a merely functional level. The functional level is not considered adequate to implement the general protection goal defined in Regulation 1107/2009 in specific protection goals regarding ‘biodiversity’ and ‘genetic resources’.

Consideration of functional aspects is more pertinent when addressing single ecosystem processes that are the basis for ecosystem services beneficial for mankind, as is the case for further specific protection goals characterised in the next sections (e.g. ‘pollination’ or ‘pest control’). As several ecosystem services are simultaneously important in the agricultural landscape, it should be borne in mind that ‘maintaining multiple ecosystem processes at multiple places and times requires higher levels of biodiversity than does a single process at a single place and time’ (Cardinale et al., 2012).

The diversity of NTA species in agricultural landscapes is directly linked to the goal of protection in contrast to other ecosystem services performed by NTAs. This derives from the fact that the definition of biodiversity is based on the number of species present and/or on their individual densities. To assess the importance of NTAs in providing ‘biodiversity’ and ‘genetic resources’ it is therefore not necessary to quantitatively link the activity of NTA species to processes that deliver the services of interest, as is the case for, for example, ‘pest control’. The diversity of the NTA species present might therefore directly characterise the service. However, some open issues have to be addressed before specific protection goals for ‘biodiversity’ and ‘genetic resources’ can be defined:

- Appropriate biodiversity: biodiversity ‘normal operating ranges’ differ between different ecosystems.

- Evaluation of biodiversity is related to comparisons between areas or between time points.
- Landscape structure modulates biodiversity: implication for the definition of tolerable effects.

4.2.1.1. Appropriate biodiversity: biodiversity 'normal operating ranges' differ between different ecosystems

As a generic rule, no single biodiversity value—whatever measurement endpoint is chosen—can be defined as being appropriate for all ecosystems. In fact, apart from some ecosystems claimed to be 'highly diverse', an increase in species diversity owing to the additional presence of generalists on top of specialists might be an indication for the onset of disturbance (e.g. Begon et al., 2006). Regarding an appropriate 'normal operating range' especially for in-field areas in agricultural landscapes with high biodiversity values, extensively managed organic farmed fields with low PPP input could act as a reference system (e.g. Moreby et al., 1994; Hole et al., 2005; Tuomisto et al., 2012; literature review in Brühl et al., 2013).

As can be seen in Figure 9, 30 years of intensive farming can have dramatic impacts on the NTA diversity in-field. However, it is a sum of factors that is responsible for the losses of biodiversity: reduced crop rotation, changes in the landscape, intensive soil management but also intense use of PPPs. Meta-analyses of differences in biodiversity across agricultural landscapes in Europe have shown that PPP input has a decisive impact on species diversity in farmlands (e.g. Geiger et al., 2010). Even if it would be desirable to completely disentangle the effects of PPPs on the biodiversity of NTAs from the effects of other management drivers of biodiversity loss, there is enough evidence that links PPP use to losses in biodiversity in agricultural landscapes (e.g. Devine and Furlong, 2007; Frampton and Dorne, 2007; Brühl et al., 2013; van Lexmond et al., 2015).

In the assessment of the risk of NTAs being exposed to PPPs, it should be ensured that the exposure to the different products is not the limiting factor for colonisation of agricultural areas in terms of the magnitude of effects considered tolerable, and this should be done independently of other management practices.

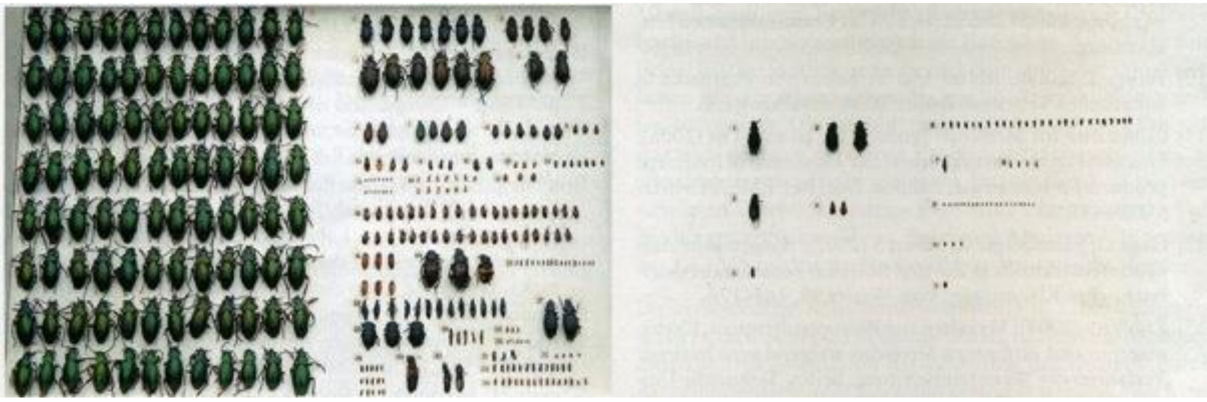


Figure 9: Coleoptera caught on a winter wheat field near Kiel (northern Germany). Left: sampling 1 July, 1951. Right: sampling 1 July, 1981. Adapted from Heydemann and Meyer, 1983.

A comprehensive literature review has been performed in view of the revision of the Guidance Document on Terrestrial Ecotoxicology (Brühl et al., 2013), with the aim of assessing the relationships between different NTA species and possible important factors affecting their diversity and population densities in the agricultural landscape. The relevance of different factors—as related to their impact on, for example, occurrence, abundance or species richness of terrestrial arthropods—may vary between the taxa, according to their life history traits. However, several investigated factors associated with unfavourable consequences (negative relationships) for many arthropod taxa are predominantly associated with agricultural practices. The most relevant environmental factors influencing arthropods in the agricultural landscape are shown in Table 3.

Table 3: Evaluation of published literature on the overall effects of investigated environmental and management factors on different arthropod taxa. Only factors for which data were available at least for five taxa are shown. Green: factors with predominantly positive relationships. Red: factors with predominantly negative relationships. Eight taxa considered in total (adapted from Brühl et al., 2013)

Factors assessed	**Studies reporting on the response of NTAs (No)**			**Investigated arthropod taxa (No)**
	Positive	**Neutral**	**Negative**	
Field margins/hedges	31	5	3	6
Plant species richness/flower abundance or area	26	5	0	6
Conservation headlands	24	5	0	7
Organic agriculture	21	10	3	5
Vegetation structure/height	14	3	2	6
Flower strips/beetle banks/grass strips	11	2	0	7
Percentage of semi-natural habitat in agricultural landscapes	11	3	0	5
Fertiliser	4	2	3	5
PPPs	3	9	32	8
Agricultural intensification	2	4	13	6
Mowing/grazing	1	2	9	5
Isolation/fragmentation	0	2	8	5

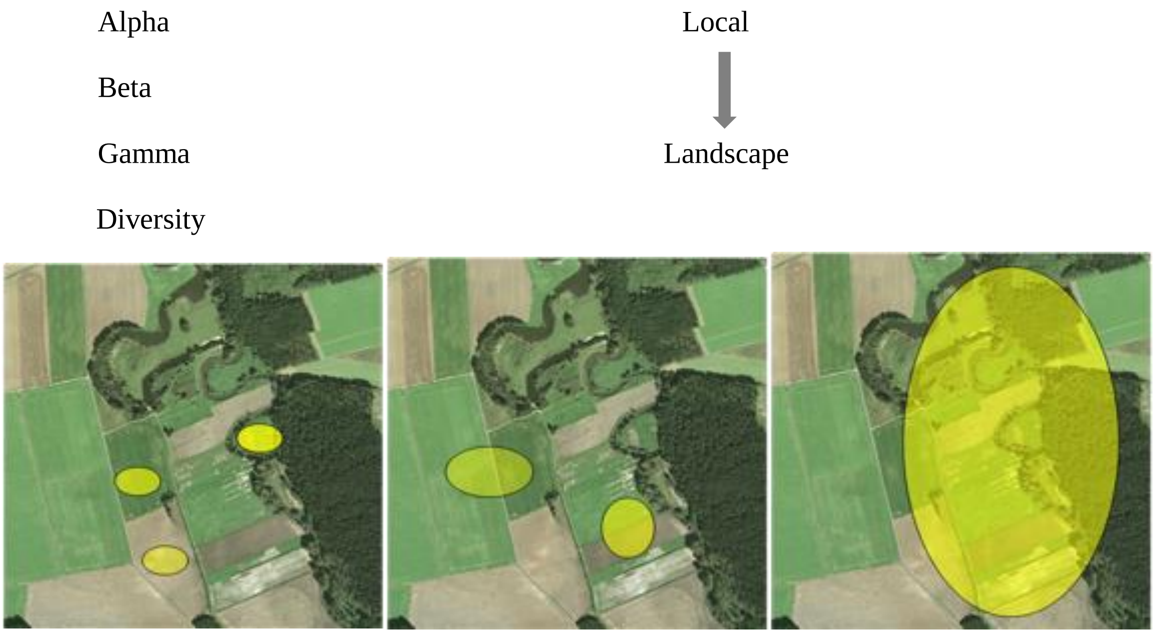
In particular, high pesticide input and agricultural intensification practices were often negatively related to the studied arthropod parameters. This tendency was consistent for pollinators (e.g. wild bees and other NTA pollinators), herbivores (e.g. cicadas) and predators (e.g. spiders).

When coming to 'high' or 'low' values of biodiversity in agricultural landscapes, the species number alone is not a good predictor of ecosystem integrity. On the one hand, it is required that the appropriate, typical suite of species for the ecosystem type is present. On the other hand, when assessing different biodiversity levels, single species traits often come into play. This means that the effects of increasing the species diversity by a value of 1 is not independent of the species added (e.g. sampling effects of ecosystem engineers). If these arguments suggest a pragmatic approach with a focus on presence/absence of key (indicator) species, there are several investigations showing that a high degree of diversity makes a system more resilient, productive and stable. Several theories support the 'better performance' expected from diverse systems (e.g. de Ruiter et al., 2005; Hooper et al., 2005; MAE, 2005; Cardinale et al., 2012; Tscharnke et al., 2012). Most prominent are the 'insurance hypothesis'—relying on functional redundancy of species that can buffer species losses or disturbance—and the 'facilitation hypothesis', where species interactions are not only driven by concurrence but also optimise resource exploitation.

The complexity of interaction between species does it make extremely difficult to determine the 'number' of species that a system can afford to loose. Some species loss can be compensated for, but, if the erosion process continues, a 'tipping point' for ecosystem functioning and ecosystem service provisioning is reached and the system may slip in a different status or definitively collapse (e.g. Lever et al., 2014).

4.2.1.2. Evaluation of biodiversity is related to comparisons between areas or between time points

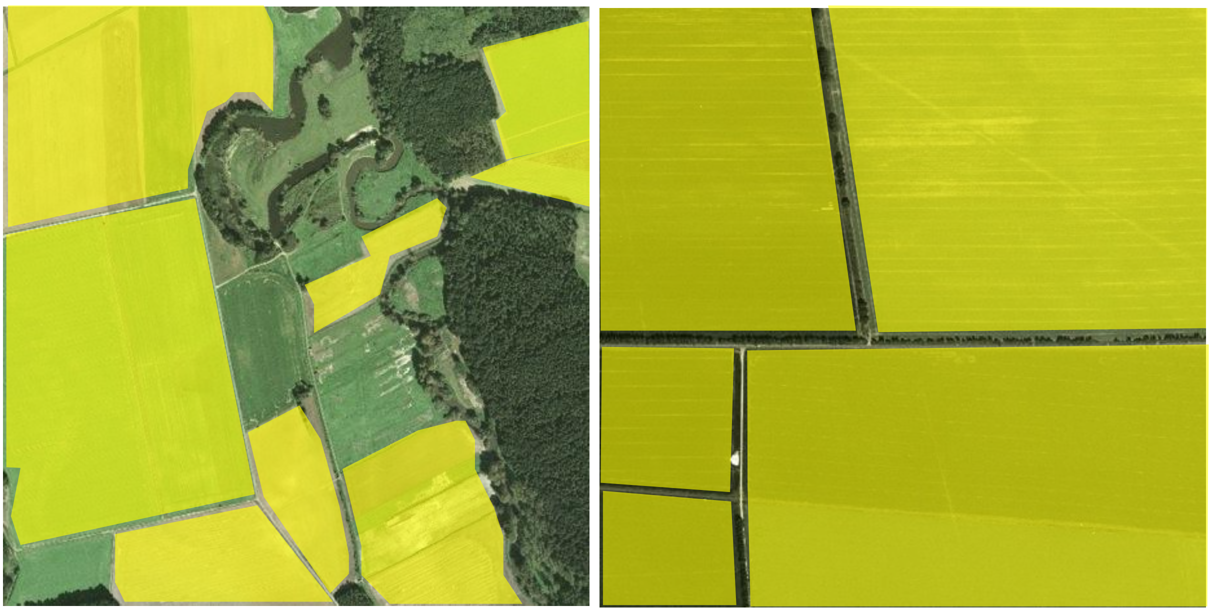
Measurement endpoints for 'biodiversity' are closely linked to landscape pattern and the definition approach is hierarchical. As shown below in figure 10, in one type of landscape, different biodiversity measurement endpoint might apply and deliver different metrics.



(©2009 Google – @2009 GeoBasis-DE/BKG (©2009 Vermessungsverwaltungen der Bundesländer und Bundesamt für Kartographie und Geodäsie)

Figure 10: Diversity according to different definitions. Left: Alpha diversity on a plot scale. Centre: Beta diversity between different areas. Gamma diversity on landscape scale. Landscape picture adapted from Brühl et al. (2013)

A controversial debate is ongoing on the relative contribution of local, alpha diversity to regional diversity (e.g. Gering and Crist, 2002; Clough et al., 2007; Karp et al., 2012). The relative dominance of the diverging diversity descriptors change at different spatial scales and the assessment of biodiversity only at local scale (alpha-diversity) might bias measures of beta diversity between different areas to higher values than are actually correct.



(©2009 Google – @2009 GeoBasis-DE/BKG (©2009 Vermessungsverwaltungen der Bundesländer und Bundesamt für Kartographie und Geodäsie)

Figure 11: Differently structured landscapes. Yellow-marked areas are in-field areas. Left: region in Rheinland-Pfalz. Right: region in Brandenburg (Germany). Picture from Brühl et al. (2013). Total area in both pictures is 4 hectares

A complex landscape structure—as shown on the left of Figure 11 above—is a best-case situation, as the off-field areas will support a high diversity and act as a donor for recolonisation (e.g. Topping et al., 2014). It is postulated that measurement of different biodiversity parameters (e.g. alpha, beta and gamma diversity) in a more realistic worst-case landscape (right of Figure 11) will not lead to biodiversity endpoint values diverging from each other.

As it is agreed by the Panel that several ecosystem services are important at the same time in agricultural landscapes and that different spatial assessment scales are relevant for different ecosystem services, not one single biodiversity measurement endpoint can be proposed as appropriate.

A description of realistic worst-case landscape scenarios in terms of support of biodiversity and ecosystem services linked to specific protection goals for NTAs should account for actual and possible future trends described for agricultural landscape structure. Considering a realistic worst-case landscape as on the right of Figure 11, it should be apparent that—irrespective of possible trade offs regarding food provisioning services—a certain degree of biodiversity has to be supported in the in-field areas in order to maintain an appropriate level of biodiversity of NTAs in the landscape and to maintain important ecosystem services provided by biodiversity at the local scale. This is translated in the setting of acceptable effects level for definition of specific protection goals.

4.2.1.3. Landscape structure modulates biodiversity: implication for the definition of tolerable magnitude of effects

Landscape structure and heterogeneity plays a pivotal role in modulating local and regional biodiversity (e.g. Tscharnke et al., 2012; Ekroos et al., 2013, 2014; Shackelford et al., 2013).

It is postulated that, in an agricultural landscape with a pertinent level of biodiversity, e.g. in a structured landscape with organic farming management, the acceptable magnitude of effect regarding the loss of NTA species and thus of biodiversity owing to the use of PPPs in the in-field area is higher than in a conventional managed agricultural landscape with large crop fields.

An assessment of PPP effects on NTA diversity at landscape level should ensure that no long-term effects from PPP use will emerge as a consequence of, for example, source–sink dynamics between off-field and in-field areas. This is particularly important for areas with simple structured agricultural fields and high PPP input (please also refer to sections 3.5, 3.6 and 5.6).

At the local scale, the spatial assessment in the agricultural landscape should be related to the individual range of the local NTA populations. The assessment should focus on an adequate spatial resolution so that multiple ecosystem services can be provided by NTA in an appropriate time scale. The potential recovery of NTA populations at local scale in longer time spans does not ensure that important services are provided in the in-field area when needed. As shown for NTAs as food web support for birds and mammals in agricultural landscapes, effects on NTAs occurring at the time of chick rearing have severe repercussions that cannot be compensated for by NTA recovery several months after breeding.

In order to implement the general protection goal of Regulation EU 1107/2009—no unacceptable effects of PPPs on biodiversity and the ecosystem—a certain degree of biodiversity has to be supported in the in-field areas in order to maintain an appropriate level of NTA biodiversity in the landscape and important ecosystem services provided by NTA biodiversity at the local scale. This goal should be accomplished, irrespective of possible trade offs regarding food provisioning services.

The magnitude of PPP effects on NTA biodiversity considered to be acceptable in the dialogue with risk managers should also relate to the most sensitive ecosystem service that has to be supported in-field. A sensitive service is understood as been driven by NTAs with high ecotoxicological and/or with high ecological sensitivity (e.g. low recovery potential) and/or a service that is highly susceptible to time constraints (e.g. NTAs as food web support for bird chicks in breeding season).

For off-field areas, only negligible effects on biodiversity are proposed to be acceptable. Regarding the assessment of effects at landscape level, an acceptable level of effects on specific parameters will have to be defined with risk managers once the endpoints are agreed in the scientific community.

Landscape-level assessment should ensure that the magnitude of effects on biodiversity in-field does not compromise the acceptable magnitude of effect agreed with risk managers for the off-field areas. At the landscape level, year-on-year decline in the abundance of species should not be observed. Negligible effects should also account for population range restrictions, meaning that not only individual abundance but also range of occupancy should not be reduced by more than a level that will be considered to be negligible (see also section 4.1).

For in-field as well as off-field areas, the tolerable magnitude of effects should take multiple PPP applications according to typical PPP 'spray schedules' into account. This will possibly implicate a lower level of tolerable effects for single PPP applications, especially in-field if the intended use fits in an application scheme that includes several other PPPs with potential effects on NTAs in the crop.

Multiple applications of several PPPs in typical schedules should also be taken in consideration when addressing the recovery of NTA at local scales (please refer to section 5.6.3).

Specific protection goal for non-target arthropods as drivers in maintaining biodiversity and genetic resources in agricultural landscapes.

In-field -

Ecological entity: (meta)population

Attribute: abundance

Magnitude /

Temporal scale: small effects on abundance and occupancy of NTA populations over months. This is possibly covered by the in-field SPGs for other ecosystem services. However it needs to be kept in mind that the ecological entities are the populations and not functional groups as in the other in-field SPGs.

Off-field

- Ecological entity: population
- Attribute: abundance
- Magnitude: negligible effects

→ at local scale < = 10% or comparable non-detectable effects on NTA species abundance that are directly caused by exposure in the off-field habitat

→ at landscape scale: negligible effects on NTA species abundance and spatial occupancy

4.2.2. Non-target arthropods as drivers in maintaining cultural services in agricultural landscapes (e.g. aesthetic value)

Agricultural landscapes are highly valued for the cultural services they provide. In detail, the conceptual framework of the MEA (2005) lists the following clusters of cultural services:

- cultural identity (that is, the current cultural linkage between humans and their environment);
- heritage values ('memories' in the landscape from past cultural ties);
- spiritual services (sacred, religious or other forms of spiritual inspiration derived from ecosystems);
- inspiration (the use of natural motives or artefacts in arts, folklore and so on);
- aesthetic appreciation of natural and cultivated landscapes;
- recreation and tourism.

Additionally, environmental education should be listed, as some types of agricultural landscapes are known by their educational interest and their nature trails (Vanderwalle et al., 2004).

Rural, agricultural landscapes are key cultural environments especially regarding heritage and inspiration values as well as aesthetic appreciation. Their scenic and recreation features are the basis for other businesses, such as tourism (e.g. de Groot et al., 2002). As Lefebvre et al. (2013) state, the role of farming in providing higher quality landscape as a public good is also recognised in the Common Agricultural Policy, where one sub-indicator of 'landscape state and diversity' (AEI28) is based on 'the interest and perception that society has for the rural-agrarian landscape (tourism, local products)' (COM, 2006).

The assessment of cultural services is *per se* difficult, as the perception of fulfilled values is very personal and/or dependent on the social context. Interestingly, if some contexts elicit aesthetic experiences that have traditionally been called 'scenic beauty', others may elicit different experiences, such as perceived care, attachment and identity (Gobster et al., 2007; Manachini et al., 2013). The latter differ between single individual observers but can sometimes be ascribed to generalised stakeholder groups. For example, significant differences in the perception of the cultural value of agricultural areas have been described for 'farmers', 'naturalists' and 'students', thus demonstrating that cultural services are not absolute values (e.g. Rogge et al., 2007; Natori and Chenoweth, 2008; Tempesta, 2010; Weyland and Laterra, 2014).

If a general rule can be set up, Weinstoerffer and Girardin (2000) see in humans an attraction for 'diversity, which is source of pleasure satisfaction, or happiness'. In this respect, Harrison et al. (2014) disentangle in their review the linkage between cultural services and explicit attributes regarding the diversity of species. They allocate more weight to species-level attributes (e.g. species abundance and richness) in the case of recreation services; in contrast, community attributes (e.g. community and habitat structure) are more relevant for the aesthetic perception of landscapes.

Some elements of NTA species diversity can be assumed to be shared by most stakeholders when it comes to the appreciation of the aesthetic values of agricultural landscapes. On the one hand, several species regarded as 'beautiful' can certainly be identified among NTAs, even if single perception might surely vary (see above). On the other hand, NTAs support a high degree of the biodiversity of agricultural environments and high biodiversity is key in perceiving the aesthetic value of a landscape as so-called visual or scenic quality (Clergue et al., 2005).

In general, 'beautiful' NTAs are deemed to be species that are big enough to be noticed, not aggressive in appearance and preferably colourful, best embodied by butterflies. The occurrence of other NTAs such as the golden ground beetle *Carabus auratus* might also possibly be judged as aesthetic in experience, especially for some stakeholders, but this leaves room for interpretation. It is proposed that a high concurrence exists between rare, endangered NTA species with a conservation status and NTAs in agricultural landscapes sharing the above-defined traits, especially regarding large arthropods with long life spans.

As a consequence, for the scope of this opinion, the definition of specific protection goal options for NTAs as drivers of cultural, especially aesthetic, values is coupled with the specific protection goal options for NTA biodiversity and genetic resources. We refer to the upcoming EFSA Scientific Committee Opinion on Endangered Species (EFSA Question No EFSA-Q-2013-00901) for specific matters regarding the risk assessment for endangered NTAs.

The demand of consumers on flows and values of aesthetic ecosystem services that originate from the agricultural landscape have changed in the past and will change in future (e.g. Bujis et al., 2006; van Zanten et al., 2013). However, the importance of conserving cultural service delivery in agricultural areas and their capacity of outdoor, recreation potential has been consistently demonstrated—sometimes against expectations (van Berkel and Verburg, 2014; Weyland and Laterra, 2014). Linking the specific protection goal for NTAs as drivers of cultural and aesthetic values to the specific protection goal for NTA biodiversity in agricultural landscapes will therefore help achieve the 'desirable complementary relationship between aesthetic pleasure and ecological health' (van Zanten et al., 2013).

4.2.3. Non-target arthropods as drivers of pest control in agricultural landscapes

As a rule, all species of animals are regulated by other living organisms (antagonists) which are not under manipulation by man but they are naturally occurring in crop surrounding environment.

Surrounding environment (e.g. field margins, see definition in section 3) provides an important agricultural habitat for a diversity of animal and insect groups (Haughton et al., 2001). They also play an important agricultural role in providing refuge for beneficial invertebrates (e.g. Aranea, some Carabidae, Staphylinidae, Heteroptera). Arthropods within agroecosystems provide numerous ecological services and economic benefits to land managers and for pest control. Predators, omnivores and parasitoids consume insect pests and weed seeds (Lundgren et al., 2006), detritivores aid in degrading crop residue and improve soil health and herbivores can reduce competition by non-crop plants and serve important roles as prey and hosts for natural enemies (Norris and Kogan, 2005). Predators and parasitoids (so-called natural enemies or antagonists) are important regulators of insect pest populations, playing a vital role in natural biological control. For example, whilst uncropped edges and margins act as refuge for flora and fauna, they can also promote movements of beneficial carabid beetle and spiders into the crop being a potential biological control agent against pest in the crop (Coombes and Sothertons, 1986; Thorbek and Topping, 2005).

Off-crop areas can provide natural enemies with food, especially in the case of parasitoids that need different flower sources as adults, refuge and alternative hosts/prey (Duelli, 1990a,b; Marino and Landis, 1996; Bianchi et al., 2005, 2006; Landis et al., 2008; Thomson and Hoffmann, 2010). Woody and herbaceous vegetation may also act as sources of pollen and nectar, which are essential prerequisites for many insects. In this manner these areas favour the presence of parasitoids and predators that can disperse into crops and contribute to pest control (Macfadeyen and Muller, 2013). These services and others in natural and managed habitats amount to an estimated USD 57 billion annually (Losey and Vaughan, 2006). Zhang and Swinton (2012) estimated the value of natural pest control ecosystem services at USD 84 million for the states of Illinois, Indiana, Iowa, Michigan and Minnesota in 2005.

Considerable effort has been devoted to examining the effects of PPPs on NTAs in laboratories or into the field. Field studies usually compare the abundance of non-target species in treated and untreated-fields; however, scant information is available regarding the antagonist present in the marginal areas of the field or in the surrounding environment and their role inside. Zhang and Swinton (2012) proposed a new bioeconomic optimisation model of the natural enemy-adjusted economic threshold

(NEET) for pest management; however, in this case, the population and densities of natural enemies were also considered only inside the crop.

Data regarding off-crop natural enemies can be taken from research done mainly for other purposes and few researchers have addressed this task specifically. While previous analyses contribute substantially to understanding the impacts on species and taxa, we have lacked a broader perspective of how PPPs may affect ecological functions of the complex insect communities associated with agroecosystems in the off-crop environments. Almost all field margin research in Europe has been done on cereal crops. In many studies, particularly in landscape studies with many sampling points, crops neighbouring the field margins are not specified. Using a systematic literature search, it is possible to obtain only a few studies with information concerning arthropod species in field margins of specific crops in Europe. Recently, a database concerning terrestrial NTA species in cultivated fields (i.e. maize, potato, sugar/fodder beet, oilseed rape, rice, soybean and cotton) and their margins in Europe has been released (Meissle et al., 2012). The database contains information for 3 030 arthropods species from arable crops and their field margins, based on literature and on personal communications. From field margins, 529 species have been collected in 926 records (i.e. a species recorded at one location with one sampling method). Data are available from field margins of maize, oilseed rape, beet and potato. There are records from 11 countries, although half of the records are from the UK. Studies of field margins mostly focus on ground beetles, spiders, bumblebees, moths and butterflies (Meissle et al., 2012).

However, it is important to consider that many species occurring in intensively managed cropping systems must be able to move between non-crop habitats (Duelli et al., 1990a) and fields, even at critical times such as harvest, and colonise fields at the start of the growing season in order to be effective control agents. Off-crop areas in this case are a real refuge for NTAs.

4.2.3.1. Parasitoids and predators

Insect parasitism is a vital component of herbivore population regulation; hence, the study of parasitism and parasitoid richness in off-field areas embedded in an agricultural matrix is relevant from conservation and management perspectives.

Parasitoid insects, whose larvae develop by feeding on other insects and ultimately cause their death, are involved in the regulation of herbivorous populations (Rauch and Weisser, 2007). Reduction of parasitoid populations could, in turn, trigger herbivory increases by releasing herbivores from top-down control (Kruess and Tscharntke, 1994).

Parasitism rates can also differ between the interior of the habitat and the artificial edges created by fragmentation, with positive (Roth et al., 2006; Valladares et al., 2006; Woodcock and Vanbergen, 2008), negative (Cronin, 2003) or no edge effects (McGeoch and Gaston, 2000; Elzinga et al., 2005) being reported in various agroecosystems.

Edge-driven effects could depend on intrinsic characteristics of parasitoid species (e.g. dispersion, foraging and searching capability) and also on the level of contrast between the original habitat and the matrix (Murcia, 1995; Ries and Sisk, 2010). Edge effects could strongly influence the parasitism patterns at the landscape level, particularly in highly fragmented systems with increased proportions of edge habitat (Ries and Sisk., 2004; Fletcher, 2005).

Several studies have highlighted the importance of natural antagonists that live in off-field areas (entirely or for part of their life cycle) in controlling the pest present in the filed. Landis and Menalled (1998) report that over 60 % of the alternative hosts of generalist parasitoids that control lepidopteran pests in maize, soybean, wheat and lucerne feed on trees and shrubs. However, a unequivocal quantification of this control is rather difficult as it is dependent not only on the host–pest relationship but also on the crop and the environment off-field (Genduso, 1979; Corbett and Rosenheim, 1996; Cãgan et al., 1999; Manachini, 2003; Kara et al., 2007). Other examples come from Rossetti et al.

(2013), who have explored, through experimental host exposure, remnant size and edge effects on parasitism levels, species richness and parasitoid community composition. Both per centage parasitism and the number of parasitoid species supported by the leafminer host were independent of forest size, but there was a higher per centage of parasitism and more of a tendency for larger parasitoid assemblages when the hosts were placed at the forest edge than when they were at the interior. The authors comparing the level of parasitism of the leafminer *Liriomyza commelinae* (Diptera: Agromyzidae) found that forests could play an important role as reservoirs of parasitoids with potential to control crop pests, a possibility heightened by the positive edge effects which could facilitate the transfer of this valuable ecosystem service to the adjacent cultivated land.

Generalist parasitoids species with irruptive population dynamics commonly use one or a few favoured hosts under low-density conditions and expand the host range to secondary hosts under high-density conditions. When the use of a secondary host persists in low density, however, it is of interest to know what nutritional, ecological and behavioural factors are involved. This is particular interesting if the secondary host becomes a primary host. Most studies of alternative host-use patterns focus on the abundance, acceptability and nutritional quality of the host, but a few studies are concerned with understanding if suppression of a pest could lead to a change of preferences in parasitoids.

However, further detailed studies on the biology and ecology of parasitoids can help to develop strategies to optimise their use as biological controls and understand their role in and off-crop.

Predatory arthropods constitute the main functional group of those collected in off-field areas of maize, beet and soybean, whereas herbivores dominate in potato, cotton and oilseed rape (Meissle et al., 2012). In most arable crops, the predators are largely dominated by beetles (Coleoptera) and spiders (Araneae) followed by syrphid flies (Diptera: Syrphidae) and predatory bugs (Hemiptera: Heteroptera). Among beetles, ground beetles (Coleoptera: Carabidae), rove beetles (Coleoptera: Staphylinidae) and ladybird beetles (Coleoptera: Coccinellidae) have been collected most frequently in all crops except rice, where aquatic beetles have been collected more frequently than other beetles. Ground beetles (Carabidae) are an invertebrate group of mainly generalist predators, abundant throughout most agroecosystems and thought to be an important group of beneficial insects contributing to restricting pest activity (Symondson et al., 2002). Östman et al. 2003) show that yield increases attributable to predators can be compared with yield increases from insecticide use for the evaluation of different management strategies. Bianchi et al. (2006) suggest that, under certain conditions and for certain sets of species, non-crop habitats in the direct vicinity of crops may attract generalist predators, leading to reduced pest control in arable fields.

Zhang and Swinton (2012) for the first time suggested that natural enemy predation contributes to optimal pesticide strategies. The population dynamics of ambient natural enemies are explicitly modelled along with their suppressive effect on the pest population and their mortality effect caused by the use of broad-spectrum PPPs. Their model was based on the species of natural enemies (mainly predators) present in the field. In fact, the quantification of the natural enemy presence was focused on major generalist predator species of the ladybeetle family, owing to their high abundance in both number and overall suppression effectiveness. The predators of aphid soybean pest belong mainly to ladybeetle species and were aggregated: *Harmonia axyridis* adult and larva, *Coccinella septempunctata* adult and larva, *Coleomegilla maculata* adult, *Cycloneda munda* adult and larva and *Hippodamia convergens* adult.

The predator populations were indexed to the equivalent predation rate of the Asian multi-coloured ladybeetle (*H. axyridis*). It was based on findings from the biological literature on the mean daily aphid consumption rate by multi-coloured Asian ladybeetle adults and the field observation of ladybeetle life stage composition proposing an approximate range of weighted average daily number of aphids eaten per ladybeetle. Based on the NEET model, it was evident that these predators boost producer net return most for moderately infested fields, and also sharply raise the pest density threshold for optimal pesticide use. These findings were based mainly on the effect of the predation action recorded in the field, and the variable to estimate from where these natural enemies arrive was not included. Of course several of them can migrate from the off-crop areas but this particular aspect was not included in the NEET model.

4.2.3.2. Landscape structure/diversity and reduction of pest

Natural pest control is an important ecosystem function that has been associated with biodiversity and landscape structure. Landscape composition affects the diversity and abundance of the natural enemy complex because different habitat types may favour different natural enemy species. A diversified agricultural landscape mosaic may therefore sustain a broad diversity of natural enemies. Non-crop habitats are often favourable habitats for natural enemies and act as source habitats from which the less favourable agricultural fields are invaded. Edge effects have mainly been examined as part of habitat fragmentation studies in agricultural landscapes. Many species in these ecosystems are well adapted to using ephemeral and spatially patchy resources and use a variety of habitats throughout their life (Bianchi et al., 2006). However, for other species, the landscape is more of a gradient of patches that span the full spectrum of suitable to unsuitable. They can utilise resources from both crop and non-crop patches, and the decision to move from one place to another is made depending on the risks associated with a particular matrix type (Macfadyen and Muller, 2013). However, insects are selective in their use of flowering plant species and specific plants may provide food for either pest species or natural enemies. The composition of the flora of non-crop habitats is therefore an important factor for the potential pest suppressive effect. Moreover, the moderate microclimate offered by off-crop areas in combination with presence of nectar sources, for example in wooded edges, result in higher parasitoid longevity, early season abundance and higher levels of parasitism than field centres. Natural enemies also use non-crop habitats for hibernation. Diversified landscapes hold the most potential for the conservation of biodiversity and sustaining pest control function (Bianchi et al., 2006). Enhanced activity of natural enemies is most frequently associated with herbaceous habitats (80 %) and somewhat less often with wooded habitats (71 %) and landscape patchiness (70 %). The differential habitat use and dispersal ability of natural enemies are likely to affect species composition, species interactions and pest control at the landscape level.

Adjacent habitats can have different effects on the density and presence of natural enemies. For example, maize-adjacent habitat did not significantly influence the number of *Macrocentrus grandii* Goidanich (Hymenoptera: Braconidae) captured, but did influence the capture of *Coleomegilla maculata* (De Geer) (Coleoptera: Coccinellidae), *Coccinella septempunctata* L. (Coleoptera: Coccinellidae) and *Chrysoperla cornea* Stephens (Neuroptera: Chrysopidae) with increased captures from fields with the herbaceous and intermediate habitats (Bruck and Lewis, 1998).

Reviewing of the literature, Bianchi et al. (2006) have shown that non-crop habitats act as reservoirs for biodiversity in agricultural landscapes and provide requisites for natural enemies that have the potential to control insect pests. Spatial scale and the distribution of crop and non-crop habitats in the landscape may influence the natural pest control function via multiple mechanisms. The diversity and density of natural enemy populations may decline with increasing distance from non-crop habitats, and the average distance between non-crop habitats and fields may affect the timing of field colonisation. Diversified small-scale landscapes therefore provide better conditions for effective pest control by natural enemies than do large-scale landscapes (Bianchi et al., 2006).

Moreover, higher parasitism levels and richer parasitoid assemblages at the edge of the forests also suggest a plausible extension of this service to the agricultural matrix, given the recognised role of edges as interfaces for the exchange of organisms between natural and cultivated systems (Blitzer et al., 2012).

Complex landscapes resulted in enhanced natural enemy populations in more than 70 % of the studies and included a variety of arthropod natural enemies and all types of enhancement effects. Landscape composition did not affect natural enemy populations in almost 20 % of the studies; for example, the oviposition rates of syrphid flies, parasitism rates in armyworms, activity density of carabid beetles and spider densities did not respond to landscape composition. The diversity and density of natural enemy populations may decline with increasing distance from non-crop habitats, and the average distance between non-crop habitats and fields may affect the timing of field colonisation. Diversified small-scale landscapes therefore provide better conditions for effective pest control by natural enemies than do large-scale landscapes (Bianchi et al., 2013). In addition, the presence of other kinds of crops can influence the antagonists present in the edges. For example, the effect of the presence of lucerne and maize fields near spring wheat crops on the presence of natural enemies of the cereal aphid was evaluated by Zi-Hua Zhao et al. (2013). The authors report that adjacent lucerne areas, as opposed to maize fields, can significantly increase the abundance and diversity of leaf-dwelling predators and parasitoids near the field edges. In addition, abundance and diversity were found to be significantly higher near the edges than in the centres of fields adjacent to lucerne areas. In contrast, no significant differences were found between edges and centres of fields adjacent to maize fields. Zi-Hua Zhao et al. (2013) conclude that the effect of within-field position and adjacent habitats on natural enemies of agricultural pests was species specific.

Bianchi et al. (2006, 2013) claim that spatial scale and the distribution of crop and non-crop habitats in the landscape may influence the natural pest control function via multiple mechanisms.

4.2.3.3. Off-crop areas as refuge for the pest

Non-crop habitats may act as reservoirs not only for natural enemies, but also for pest species that invade crops. Many polyphagous pests can utilise resources from both crop and non-crop. Aphids, phytophagous of canola, and also their parasitoids move frequently out of native vegetation (especially in the later season); in contrast, predators moved less commonly from native vegetation (Macfadyen and Muller, 2013). Off-crop areas can offer both nutritional and non-nutritional factors which can contribute to maintaining pest populations.

In fact, a number of agricultural pest species are associated with off-field habitats (e.g. field margins, road verges, fallows and meadows), such as aphids, herbivorous flies and beetles. Pest species also use non-crop habitats for hibernation, mating, resting locations for females and temporary escape from different stress (e.g. agricultural practices, natural enemies).

For example, it is known that European corn borer (ECB; *Ostrinia nubilalis* Hb., Lepidoptera: Crambidae) prefers to mate outside of maize crops. Female moths are found in taller vegetation during the day, for example in grassy edges of fields or wheat. In the evening, the females move into crops to lay several egg masses per night. ECB moths often rest during the day in grassy field edges, thus the weedy areas surrounding fields probably provide refuge for ECB, but also for natural enemies (Bruck and Lewis, 1998). Merrill et al. (2013) suggest strong correlations between ECB moth density and adjacent maize crops, prevailing wind direction and an edge effect. In addition, directional component effects suggest that more ECB moths were attracted to the south-western portion of the crop, which has the greatest insolation potential.

Most aggregations of *Euschistus servus* (Say) (Hemiptera: Pentatomidae) nymphs and adults were located on the edge of the maize, directly adjacent to the harvested wheat. Movement from wheat to maize was not consistent between the years and may have been influenced by factors such as variations in weather, timing of wheat harvest or other available alternative hosts (Reisig et al., 2013).

Additionally, in the case of tree pests, alternative hosts present in the off-field areas could provide benefits for the pest. Rossiter (1987) claim that the microhabitat of pitch pine (*Pinus rigida* Mil), a secondary host of gypsy moth (*Lymantria dispar* L.) which is phytophagous of oak, held less nuclear polyhedrosis virus (NPV), a major mortality agent of the gypsy moth; moreover, *L. dispar* individuals hatching from eggs laid on pitch pine were less infected with NPV and larvae dosed with a known amount of NPV survived longer when feeding on pitch pine foliage. The use of pitch pine by gypsy moth populations appeared to be beneficial and may have an important effect on its population dynamics. The mobility associated with host switching by late-instar larvae and with dispersal by first-instar larvae may represent an important mechanism for host range.

4.2.3.4. Changes from secondary pests to primary pests

Secondary pest outbreaks generally occur when pesticide applications kill natural enemies that were controlling a species that was not a pest before. These species can increase to densities that cause damage, because the natural enemies previously maintaining their populations at low densities are no longer present or abundant enough to control them (Hajek, 2004).

However, there is no evidence in any country that secondary pest outbreaks have emerged in crops as a consequence of insecticide effects on the off-crop areas.

The suppression of pest populations in crops by natural enemies provides environmental and economic benefits because it may reduce yield loss without the negative environmental consequences that result from chemical pesticide use (Östman et al., 2003; Bianchi et al., 2006). However, the role of natural enemies in maintaining natural pest control is controversial and needs more scientific support. Rodríguez and Hawkins (2000), Finke and Denno (2004) and Martin et al. (2013) showed that a simplified natural enemy community provides control of pest populations that is equal to or better than a complex natural enemy community. These findings are in line with observations from biological control programmes showing that effective control can in most cases be achieved by the introduction of one or few natural enemies (Myers et al., 1989). In contrast, there is also empirical evidence that diverse communities of natural enemies are more effective in regulating herbivore populations than poor communities (Schmidt et al., 2003; Snyder and Ives, 2003). At this point, general conclusions on the relationship between biodiversity and natural pest control function are uncertain.

It has been claimed that natural enemies of pests perform important ecosystem services in agricultural landscapes. Up to now, these services have rarely been evaluated in yield or monetary terms (Östman et al., 2003; Losey and Vaughan, 2006). Because these functional guilds interact differently with crop plants and environments, they are likely to be affected by pest management practices to varying degrees. Thus, a comprehensive examination that simultaneously accounts for different crop production systems and pest management practices is required to draw meaningful conclusions about the off-crop impact of PPPs. However, currently, such comparisons may miss important effects.

Integrating naturally occurring pest control services into decisions about pesticide-based control has the potential to significantly improve the economic efficiency of pesticide use, with socially desirable outcomes. While ecologically based approaches have long been promoted as alternatives to complement and partially replace current chemically based pest-management practices, there has been limited guidance on how to operationalise the concept also considering the natural enemies present in the off-crop areas.

Others factors can be detrimental to natural enemies present in the off-crop areas; for example, there are no indications on the extent to which field margin vegetation and hedgerow plants draw up PPPs from arable soils and what the effect of this could be on NTAs, especially predators or omnivores. In addition, the effect of herbicides could be evaluated because they can lead to a reduction of NTAs inhabiting the flora in edges. A few studies have addressed this area. The herbicide glyphosate had negative indirect effects on non-target spiders, *Lepthyphantes tenuis*, in field margins (Haughton et al., 2001). There are indications that margins which had the lowest rate of glyphosate supported significantly more spiders than margins with higher rates. The regression results suggested that there was a link between dead vegetation cover, height of vegetation, abundance of spiders and rate of herbicides. However, no investigations were performed to understand if this reduction in spider population densities also led to a reduction of pest predation.

Many authors claim that to conserve beneficial organisms in agroecosystems and to avoid repeated applications and excessive use of insecticides it is essential to preserve the service provided by parasitoids and predators.

The effects of PPPs on natural antagonists present off-field are sometimes contrasting. Desneux at al. (2005) suggest that the parasitoid *Diaeretiella rapae* (M'Intosh) (Hymenoptera: Braconidae) could limit populations of the green peach aphid *Myzus persicae* (Sulzer) (Homoptera: Aphididae) even after pyrethroid treatment because the presence of pesticide residues had little impact on the parasitoid. The authors claim that the pesticide treatment does not prevent recolonisation by off-crop parasitoids. On the other hand, continuous annual application of agrochemicals could gradually reduce the frequencies of certain species and cause further plant community shifts (Schmitz et al., 2014).

As described above, several attempts have been made to describe the relationship between predator presence and the reduction of specific pests. An alternative approach has been chosen by Freier et al. (2007) by defining so-called predator units—consisting of a range of different NTA pest predators—and to relate these units to rates of pest population development. Relative predator units (PU, Freier et al., 1998) are calculated for all target predator species or functional groups based on their surplus feeding rates determined in the laboratory. The PUs for all individual predators are then summed to obtain an overall predator community value. Investigations over 10 years in two contrasting regions in Germany showed that higher predator densities reduced the increase in aphid infestation (Freier et al., 2007). A tendency for stagnation of aphid infestation was observed between 3.9 and 4.2 PU depending on landscape settings. At higher predator densities, the change in aphid density became increasingly negative. Therefore, 4 PU/m^2 seemed to be a critical density threshold for active natural aphid control in the investigated wheat fields.

Pest control services driven by NTA species rely on the ability of natural enemies to move frequently into crop fields and, once there, attack and kill pest species. However, quantifying how species perceive habitat patches and matrices in agricultural landscapes is still a challenge. As biological control is a result not only of enemy diversity and abundance, but also of the trophic interactions occurring between enemies, the benefits of natural pest control are not self-evident, but depend on many factors and can easily be disrupted.

Therefore, it is not currently possible to determine the numeric impact that general changes in population size and community structures of natural enemies will have on functions and ecosystem services for natural pest regulation for different crops.

In order to be able, in the future, to quantify the magnitude of effect that can be tolerated in different crop/landscape combinations, the following points should be addressed:

- Improve knowledge about how much damage via pests would occur in the crop in the absence of NTAs as natural enemies.
- Description of typical plants and communities (habitat) of field margins in different European regions. No baseline should be considered when assessing potential effects of PPPs on the density of natural enemies in off-crop habitats.
- Identification of valued arthropods typically associated with plant communities in an agricultural environment. This can be done independently from specific crops, because the field margins as habitats rather than the neighbouring crops are the focus.
- Information on field margin types across Europe and information on arthropods associated with typical plant communities should be combined. This will allow conclusions to be drawn on which arthropods are likely to occur in the field margins in different European regions.
- Estimation of the impact of off-crop antagonists to in-crop populations of key pests and secondary pests. There is a need for the baseline to be defined to quantify the contribution of the antagonist present off-crop on pest populations in-crop.
- Valuation of toxic effects of PPPs on natural enemies that largely depend on the target pest, considering that such effects are common for all pest control methods.

As for the support of ‘biodiversity’ and the ecosystem service ‘genetic resources’, and for NTA species as natural pest enemies, the acceptable magnitude of effect regarding an impact of PPP use on the service ‘pest control’ is deemed to be higher in a structured landscape with general low PPP input than in a conventional managed agricultural landscape with large crop fields. The proposed tolerable magnitude of effects expressed as relative changes compared with controls relate to well-structured landscapes with high NTA biodiversity levels (see also section 4.1). It is of great interest to further develop indicators that attempt to describe the service of pest control in absolute terms (e.g. PUs, see above), as these can be set in specific protection goal options independently from the landscape context.

An assessment of PPP effects on NTA drivers of ‘pest control’ at landscape level should ensure that no long-term effects from PPP use will emerge as a consequence of, for example, source–sink dynamics between off-field and in-field areas. This is particularly important for areas with simple structured agricultural fields and high PPP input (please also refer to section 3).

Regarding the assessment of effects at landscape level, an acceptable level of effects on specific parameters will have to be defined together with risk managers once the endpoints are agreed upon in the scientific community. At the landscape level, negligible effects should exclude year-on-year decline in abundance of species, but also population range restrictions (see also section 4.1).

At the local scale, the assessment should focus on an adequate spatial resolution so that the ecosystem service ‘pest control’ can be provided by NTAs in an appropriate time scale. The potential recovery of NTA populations at local scale in longer time spans does not ensure that the service is provided in the in-field area when needed. However, the diversity of the NTA species as natural pest enemies implies a high diversification of spatial individual ranges that might be covered by individuals of local populations. It is important to focus the assessment of sensitive NTA drivers on ecotoxicological and ecological sensitivity.

For off-field areas, only negligible effects on NTAs are proposed to be tolerable without impacting the general protection goals (EFSA PPR Panel, 2010). Landscape-level assessment should ensure that the magnitude of effects on biodiversity in-field does not compromise the acceptable magnitude of effect agreed with risk managers for off-field areas.

For in-field as well as off-field areas, the tolerable magnitude of effects should take multiple PPP applications according to typical PPP ‘spray schedules’ into account. This will possibly implicate a lower level of tolerable effects for single PPP applications, especially in-field if the intended use fits in an application scheme that includes several other PPPs with potential effects on NTAs in the crop.

Multiple applications of several PPPs in typical schedules should also be taken in consideration when addressing the recovery of NTAs at local scales (please refer to section 5).

Specific protection goal options for non-target arthropods as drivers of pest control in agricultural landscapes (please refer to the text for justifications)

In-field

Ecological entity:	functional group
Attribute:	abundance
Magnitude:	medium effects
	at local scale: 35 % < effects < 65 %
	at landscape scale: to be defined
Temporal scale:	weeks

up to 4 weeks.

Off-field

Ecological entity:	population
Attribute:	abundance
Magnitude:	negligible effects
	at local scale: ≤ 10 % or comparable non-detectable effects on the abundance of NTA populations that is directly caused by exposure in the off-field habitat
	at landscape scale: negligible effects on abundance and spatial occupancy of NTA species for pest control
Temporal scale:	not relevant

Table 4: Key drivers for the ecosystem service pest control. Main taxa, examples of species, traits determining population range and exposure routes. For underlined species, test protocols are available

SPG	Key drivers	Examples of taxa	Main exposure routes	Population range			
				Large (larger than field size)		Small (smaller than field size)	
				Species of open habitats (including crops)	Species of other habitats (e.g. woody structures)	Species of open habitats (including crops)	Species of other habitats
Pest control/ antagonists	Predators Insectivorous Omnivorous	Coleoptera	Overspray Contact soil Oral	Ground beetles, e.g. *Bembidion lampros*, *Pterostichus melanarius*, *{Harpalus} rufipes*, *Poecilus cupreus*, *Carabus hortensis* Rove beetles: *Philonthus cognatus*, *Ocypus olens*	*Pterostichus oblongopunctatus*, *Carabus auratus* or *auronitens*		
			Overspray Contact leaves Oral	Ladybirds: *Coccinella septempunctata*, *Hippodamia* spp., *Rodolia cardinalis*		Ladybird, e.g. *Coleomegilla maculate*, *Chilocorus* spp., *Harmonia* spp.	Many species
		Diptera	Overspray Contact leaves Oral	Hoverflies Syrphidae (e.g. *Episyrphus balteatus*, *Syrphus* spp.)	Many species		Many species
		Araneae	Overspray Contact soil and leaves Oral	SPG main importance: money spiders, Linyphiidae (e.g. *Erigone atra*, *Oedothorax fuscus*, *Lepthyphantes tenuis*); Wolf spider; Lycosidae (e.g. *Pardosa palustris*, *Trochosa ruricola*)	Many species, e.g. *Pisaura mirabilis*		
		Heteroptera	Overspray Contact leaves oral	Damsel bugs, e.g. Nabidae (*Nabis rugosus*), Anthocoridae (*Orius laevigatus*), Miridae	Many species	*Deraeocoris* spp., Orius spp., Geocoris spp.	
		Neuroptera	Overspray Contact leaves oral	Green lacewing (e.g. *Chrysoperla carnea*) and brown lacewing (*Hemerobious* spp.)		Green lacewing (e.g. *Chrysoperla carnea*)	
		Acari	Overspray Contact soil	Acarina: Phytoseiidae (e.g. Euseius tularensis, Phytoseiulus persimilis)		Gamasidae, predatory mites, Phytoseiidae	Many species

SPG	Key drivers	Examples of taxa	Main exposure routes	Population range: Large (larger than field size): Species of open habitats (including crops)	Population range: Large (larger than field size): Species of other habitats (e.g. woody structures)	Population range: Small (smaller than field size): Species of open habitats (including crops)	Population range: Small (smaller than field size): Species of other habitats
			Oral				
	Parasitoids	Hymenoptera	Overspray Contact leaves (Contact) host	Parasitic wasps, Braconidae (e.g. *Aphidius*, *Cotesia* spp.), Trichogrammatidae (*Trichogramma* spp.) Calchoidea, Aphelinidae (*Encarsia* spp.), Mymaridae Ichneumonidae (*Campoletis sonorensis*, *Diadegma* sp.,)	Many species	Parasitic wasps (e.g. *Aphidius* spp., Cotesia spp.)	Many species
		Diptera	Overspray Contact leaves (Contact) host	Parasitic flies, Tachinidae (e.g. *Lydella thomposoni*, *Trichopoda* spp.), Diptera: Cecidomyiidae (*Aphidoletes aphidimyza*), Bee flies (Bombylidae)			

4.2.4. Non-target arthropods as food web support

Although the main focus of the current opinion is on the characterisation of effects of PPPs on NTAs, these effects cannot be discussed without relating them to the overall structure and function of communities. Arthropods constitute an important part of virtually all ecosystems, so any effect on their abundance, diversity or activity is reflected to some extent at the community level. Food web effects on higher trophic levels, especially on birds and small mammals, are of particular concern. This section presents an overview of the contemporary knowledge on links between abundance or biomass of NTAs—the food of many insectivorous and omnivorous species—and birds and small mammals. These secondary effects need to be considered in ecological risk assessment of PPPs, as their results can be sometimes as critical as direct toxic effects.

The decline of grassland birds (also called farmland or agricultural birds) is a well-known phenomenon. As stated in the review by Robillard et al. (2012), 'Over the last decades, farmland bird populations have declined rapidly in North America and Europe. Major declines have been observed in the guild of aerial insectivores, birds feeding mostly on flying insects. These major declines have been generally attributed to lack of food and/or scarcity of nesting sites. More specifically, increased use of pesticides in intensively managed farms has led to depleted food resources'. A recent comprehensive review of published literature has also been performed by DEFRA (2005) and Jahn et al. (2014), substantiating these conclusions.

Although the reasons of the decline could not be clarified in detail (and might be very different for particular species; see Jahn et al., 2014), it is commonly agreed, based on rich evidence, that agricultural intensification is the cause. Besides direct pesticide toxicity to birds (Mineau and Whiteside, 2013), which is out of the scope of this opinion paper, the indirect effects stemming from the loss of invertebrate food owing to insecticide use is at least in part responsible. However, effects of different agricultural practices on birds, such as habitat loss, direct toxicity of insecticides to birds and indirect trophic effects are usually highly correlated, making it hard to disentangle the impact of single factors (Mineau and Whiteside, 2013). The best cases in which indirect trophic-chain effects can be clearly identified are insecticides with low toxicities to vertebrates. The following example illustrates the effects of food shortage on bird chick survival. *Bacillus thuringiensis* (Bt) is used as an effective biological agent to control many insect pest species, and it is not toxic to vertebrates. Thus, any negative effects observed in birds after its application could be ascribed solely to indirect consequences of the damage to the food chain. Such effects have been shown recently by Poulin (2012) in Camargue, France, where Bt was experimentally used in 2006 to control mosquito populations. She found a significant decrease in insect food availability to birds in the three years (1997–1999) following the Bt application, accompanied by a significant decrease in chick feeding rates and breeding success of the house martin (*Delichon urbica*).

The author concluded that wide-scale Bt spraying 'can have deleterious effects on the demography of insectivorous birds through multitrophic interactions'. By identifying these indirect effects beyond any doubt, this work became the ultimate proof that a decline in insect abundance owing to pesticide use can indeed cause significant negative effects in bird populations.

Although with other classes of PPPs it is more difficult to prove indirect trophic-chain effects on insectivorous birds, a few studies showed more or less clear links between bird population numbers and abundance of prey invertebrates. The direct effect of PPPs and the resulting reduction of birds feeding mostly on flying insects was underlined by Robillard et al. (2012) in their work on swallows and sparrows. Although not studied by the authors, it may be worth mentioning that the guild of aerial insectivores also includes bats which may be affected by the decline in abundance of flying insects to the same extent as birds.

The relationship between effects on birds and pesticide use was also clearly identified by Rands (1986) who showed that the mean brood size of grey partridge and pheasant was significantly higher on plots where field edges were unsprayed than on fully sprayed control plots.

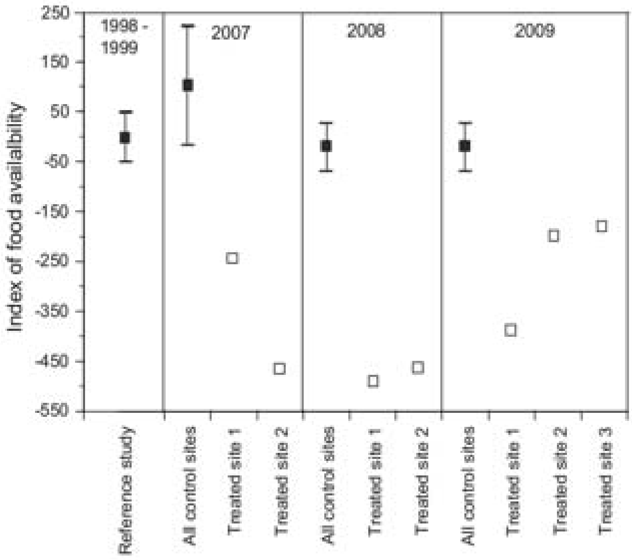
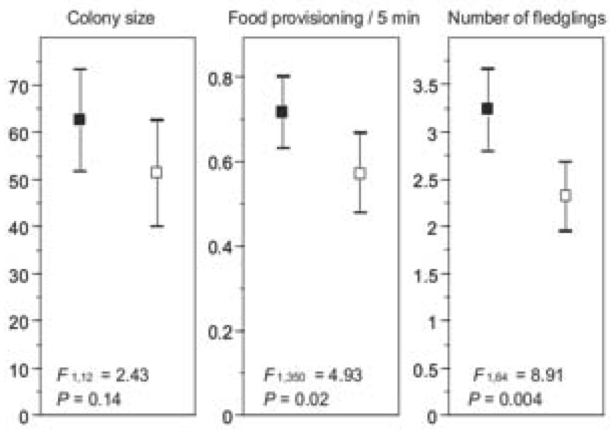


Figure 12: Insect food availability to birds in Camargue following the application of *Bacillus thuringiensis* (Bt) biocide in 2006 to control mosquito populations; Control sites filled squares. Open squares: sites treated with Bt. (means ± 95 % CI) From Poulin (2012)[9]



Food-chain effects of the Bt treatment in Camargue on house martin colonies. Control sites: filled squares. Treated sites: open squares (means ± 95 % CI). From Poulin (2012)[9]

Brickle and Harper (2000) proved, in turn, a whole series of relationships between pesticide use, abundance of chick-food invertebrates, nest survival probability and the weights of nestlings of corn buntings (*Miliaria calandra*). Invertebrate density was negatively correlated with the number of insecticide applications, the weights of nestlings were positively correlated with the abundance of chick-food invertebrates and the probability of nest survival was negatively correlated with the abundance of chick-food invertebrates close to the nest.

Benton et al. (2003), using multivariate analysis, showed that average densities of 15 common farmland birds in Scotland were significantly related to insect abundance and, independently, to measures of agriculture and climate. Although the authors were not able to separate pesticide effects from a range of agricultural practices, the most important result of this study in terms of possible indirect trophic-chain effects on bird populations is the proven relationship between insect abundance and densities of bird populations. Similarly, Evans et al. (2007) suggested that the reduced numbers of aerial invertebrates have contributed to barn swallow population declines, although in this particular study the decline in invertebrate numbers/biomass was not caused by pesticide use but rather by the reductions in the availability of pastures. The link between food (aerial invertebrates) availability and

[9] Reproduced from Poulin, B., 2012. Indirect effects of bioinsecticides on the nontarget fauna: The Camargue experiment calls for future research. Acta Oecologica 44, 28–32.

barn swallow population dynamics still holds, and negative effects of pesticide use on biomass and diversity of non-target invertebrates has been proved beyond any doubt. In 1975, Bryant (1975) showed that recruitment into house martin populations is significantly affected by food supply, which influences the clutch size, the occurrence of second clutches and nestling mortality.

A clear link between condition of yellowhammer (*Emberiza citrinella*) nestlings, insect food abundance and insecticide use was demonstrated by Hart et al. (2006). The mean chick body mass was significantly related to chick food abundance, which was, in turn, related to the probability of fledging. The observed depletion in insect food abundance correlated clearly with the pesticide use in the studied plots. As stated by the authors, 'the severity of the impact will depend upon the extent of applications around individual nests and their timing relative to the hatch date. Where isolated fields are sprayed during the nesting period, yellowhammers provisioning their young may simply select unsprayed fields and field boundary habitats nearby. However, as spraying becomes more extensive and the choice of unsprayed fields becomes more restricted, there will be a concomitant decrease in the mean abundance of chick food in the foraging range to levels where chick starvation and brood reduction are likely'.

Similar effects on the same species were shown also by Morris et al. (2005). They found that pesticide application during the yellowhammer breeding season had the most negative effects, but they also highlighted that multiple applications may have cumulative effects. The observed effects were especially pronounced in the case of broad-spectrum insecticides which, in the authors' opinion, should not be applied to crops in the breeding season 'unless there are serious implications for loss of yield'. The mortality among NTAs after application of such PPPs often approached 100 % a few days after the spray, and substantial effects lasted for up to two months.

Boatman et al. (2004) combined data available from published literature with their own experiments and concluded that 'indirect effects of pesticides indeed do occur, although, apart from the Grey Partridge, unequivocal evidence is only available for effects of insecticides. There is, however, strong circumstantial evidence for effects of herbicides, from a variety of sources; unfortunately practical difficulties in demonstrating indirect effects arising from herbicides have prevented their confirmation so far. […] However, it seems probable that indirect effects of pesticides form part of a suite of causal factors likely to be implicated in the declines of farmland bird species'. The authors also presented a clear quantitative relationship between food invertebrate abundance and yellowhammer chick mortality.

Despite the well-proven effects of PPP sprays on the abundance of bird-food invertebrates, it should be clearly stated that limiting the assessment to the effects of PPP use at the field scale is not sufficient to maintain healthy populations of birds (and other insectivores) and their prey invertebrates. As stated by Goulart et al. (2013), 'heterogeneity is the key'. The model constructed by the authors showed that 'farming practices have negative impact on the bird community supporting the idea that numerous and large areas should be conserved without human influences. […] Intensive farming leads to the decline of non-forest species because it directly influences the matrix' (Goulart et al., 2013). This is in line with the earlier statement by Benton et al. (2003) that recreating habitat heterogeneity is the key to restoring and sustaining biodiversity in temperate agricultural systems.

4.2.4.1. Proposals for the definition of tolerable magnitude of effects on non-target arthropods as food web support

As stated above, the relationship between the amount of invertebrates present in a given area and the reproductive success of birds in terms of, for example, number of fledging chicks has been quantitatively described for several species. The so-called 'chick–food index' describes the relationship between the probability of survival of a bird population and the amount of NTA species suitable for the diet of the bird species assessed. First applications of the 'chick–food index' were in the early-1980s. Potts (1980, in O'Connor and Shrubb, 1990) provided evidence that the population changes in the grey partridge were largely dependent on chick survival and this was in turn a function of the number of arthropods present in cereal crops. He could additionally determine that, in the investigated areas, the arthropods had to be present in sufficient numbers at a specific time, i.e. in the month of June. Potts and Aebischer (1991) conducted long-term monitoring and intensive investigations in several regions in the UK and found that chick survival rates were significantly correlated with the proportion of chick diets made up of caterpillars (Symphyta and Lepidoptera) and beetles (Chrysomelidae and Curculionidae). Potts and Aebischer (1991) also attempted to quantify the impact of total arthropod losses in the different landscapes (reduction of chick survival rates to levels below that necessary to replace adult losses) and the impact of doubling the arthropod numbers (restore chick survival rates and population density equilibrium to levels of the 'pre-pesticide era').

Preferred arthropod diet composition was also determined for other bird species and statistically significant relationships to population indices established, e.g. for the yellowhammer *Emberiza citrinella* (e.g. Boatman et al., 2004; DEFRA, 2005).

Odderskaer et al. (1997) published a project report on comparative studies of skylark (*Alauda arvensis*) breeding performance in sprayed and unsprayed spring barley fields. In all study years, the number of fledglings per nest was greater in unsprayed fields. Moreover, the probability for a clutch to survive in a sprayed and an unsprayed average field was 0.63 and 0.75, respectively. Reported nestling food consisted of beetles, Lepidoptera and Heteroptera, but in unsprayed fields the diet was more diverse and contained more herbivorous insects. Odderskaer et al. (1997), in addition, quantified the food abundance for the skylark as being on average three times higher in unsprayed than sprayed fields. The reduced reproductive output from sprayed fields was primarily a result of lower clutch survival, interestingly determined in this study by the lower number of successful breeding attempts. In fact, many skylark pairs continued breeding later in the season in the untreated fields in contrast to treated fields, apparently in better body conditions. The low availability of food was also a possible determinant of skylarks giving up nests in treated fields during the incubation period.

Next to the results of the investigations above, showing a strong decrease of NTAs as food web support in fields treated with PPPs, Holland et al. (2012) investigated the NTA diversity and calculated the chick–food index in 40 agricultural fields, cultivated with cereals, oil seed rape, peas, beans or potatoes. As could be expected, higher total numbers of invertebrates were to be found at the field edges than the mid-field. Strikingly, Holland et al. (2012) demonstrated that, in all 40 common crop fields investigated in their study, the chick–food index was only half or less the level required to ensure sufficient chick survival for the maintenance of the population of the grey partridge.

Frampton and Dorne (2007) quantified the impact of reduced PPP input on the abundance of several arthropod groups, including so-called 'chick-food insects'. This arthropod assemblage consisted of Lepidoptera and Symphita as well as some Carabid families. Excluding herbicide or jointly herbicides and other pesticides resulted in two to seven times higher abundances of 'chick-food insects', with, for example, doubled abundances of Lepidoptera.

The panel agrees with Boatman et al. (2004) in suggesting that endpoints regarding different 'chick–food indices' could be useful to assess food availability for birds in agricultural landscapes. Total numbers of NTA species delivering food web support or, if sufficiently justifiable, biomass of 'dietary functional groups' could be given as absolute minimum levels of service performance. In this case, the landscape context shaping the diversity of NTAs in different structured areas would still be of great importance, but an acceptable magnitude of effect on NTA drivers could be set in absolute values. Such an approach will support the definition of a tolerable magnitude of effects in addition to the characterisation of possibly acceptable percentage reductions of NTA drivers compared with controls, as these are proposed for well-structured landscapes with high NTA biodiversity levels (see also sections 4.1 and 4.2.1).

As for the previous ecosystem services driven by NTAs, an assessment of PPP effects on NTA drivers of 'food web support' at landscape level should ensure that no long-term effects from PPP use will emerge as a consequence of, for example, source–sink dynamics between off-field and in-field areas. This is particularly important for areas with simple structured agricultural fields and high PPP input.

At the local scale, the assessment should focus on an adequate spatial resolution so that the ecosystem service 'food web support' can be provided by NTAs in an appropriate time scale. The potential recovery of NTA populations at local scale in longer time spans does not ensure that the service is provided in the in-field area when needed. The 'chick–food index' should deliver a suite of NTAs as food web support for a number of birds and mammals in agricultural landscapes.

For off-field areas, only negligible effects on NTAs are proposed to be tolerable without impacting the general protection goals (EFSA PPR Panel, 2010). Landscape-level assessment should ensure that the magnitude of effects on biodiversity in-field does not compromise the acceptable magnitude of effect agreed with risk managers for off-field areas.

Regarding the assessment of effects at landscape level, an acceptable level of effects on specific parameters will have to be defined together with risk managers once the endpoints are agreed upon in the scientific community. At the landscape level, negligible effects should exclude year-on-year decline in abundance of species, but also population range restrictions (see also section 4.1).

For in-field as well as off-field areas, the tolerable magnitude of effects should take multiple PPP applications according to typical PPP 'spray schedules' into account. This will possibly implicate a lower level of tolerable effects for single PPP applications, especially in-field if the intended use fits in an application scheme that includes several other PPPs with potential effects on NTAs in the crop.

Multiple applications of several PPPs in typical schedules should also be taken in consideration when addressing the recovery of NTAs at local scales (please refer to section 5.6.3).

Specific protection goal options for non-target arthropods as <u>food web support</u> in agricultural landscapes (please refer to the text for justifications)

In-field

Ecological entity: functional group

Attribute: abundance/biomass

Magnitude /Temporal scale:

small effects up to months

at local scale: 10 % < effects < 35 % in breeding season; additionally, no shortfall below the limits given by chick–food indices

at landscape scale: to be defined; additionally, no shortfall below the limits given by chick–food indices

medium effects up to weeks

at local scale: 35 % < effects < 65 % up to four weeks outside the breeding season

at landscape scale: to be defined; additionally, no shortfall below the limits given by chick–food indices

Off-field

Ecological entity:	population
Attribute:	abundance/biomass
Magnitude:	negligible effects
	at local scale: $\leq 10\,\%$ or comparable non-detectable effects on the abundance of NTA populations that are directly caused by exposure in the off-field habitat
	at landscape scale: negligible effects on abundance and spatial occupancy of NTA species as food web support
Temporal scale:	not relevant

Table 5: NTAs providing food web support. Key drivers, main exposure routes and population ranges. For underlined species, test protocols are available

SPG	Key drivers	Examples of taxa	Main exposure routes	Population range: Large (larger than field size) Species of open habitats (including crops)	Population range: Large (larger than field size) Species of other habitats (e.g. woody structures)	Population range: Small (smaller than field size) Species of open habitats (including crops)	Population range: Small (smaller than field size) Species of other habitats
Food web support	Large NTA	Coleoptera	Overspray Contact soil **Oral**	Ground beetles, e.g. *Bembidion lampros*, *Pterostichus melanarius*, *{Harpalus} rufipes*, <u>*Poecilus cupreus*</u>, *Carabus hortensis*, Scarabaeidae, Elateridae Ladybirds: <u>*Coccinella septempunctata*</u> Rove beetles: *Philonthus cognatus*, *Ocypus olens* Weevils: Curculionidae Leaf beetles: Chrysomelidae	*Pterostichus oblongopunctatus*, *Carabus auratus* or *auronitens*		
		Diptera	Overspray Contact leaves	Many species Crane flies: Tipulidae Hoverflies: Syrphidae (e.g. *Episyrphus balteatus*), super Family Muscoidea	Many species		
		Araneae	Overspray Contact soil and leaves Oral	Main families of pest control importance: Linyphiidae (e.g. *Erigone atra*, *Oedothorax fuscus*, *Lepthyphantes tenuis*) Lycosidae (e.g. <u>*Pardosa*</u> *palustris*, *Trochosa ruricola*)	Many species, e.g. *Pisaura mirabilis*		
		Heteroptera	Overspray Contact leaves, oral	Shield bugs, e.g. Pentatoma			
		Auchenorrhyncha	Overspray Contact leaves, oral	Plant hoppers			
		Hymenoptera	Overspray Contact leaves Oral	Sawflies Symphyta			

SPG	Key drivers	Examples of taxa	Main exposure routes	Population range			
				Large (larger than field size)		Small (smaller than field size)	
				Species of open habitats (including crops)	Species of other habitats (e.g. woody structures)	Species of open habitats (including crops)	Species of other habitats
		Lepidoptera	Overspray Contact Oral water Oral leaves (larvae) Oral nectar	Butterflies (e.g. *Maniola jurtina*, *Vanessa io*, *Pieris rapae*) and moths			
		Orthoptera	Overspray Contact leaves Oral water Oral leaves	Grasshoppers			
		Neuroptera	Overspray Contact leaves Oral?	Lacewings (e.g. *Chrysoperla carnea*)			
	Small NTA	Collembola	Contact soil			Springtails (e.g. Entomobryidae)	
		Hemiptera	Contact leaves	Plant lice, Psyllidae			

4.2.5. Non-target arthropods as drivers of plant pollination in agricultural landscapes

This section deals with pollinating arthropods except bees which are considered in a separate Guidance Document (see EFSA 2013b).
Pollination is an essential process in natural communities, and is a prerequisite for a healthy ecosystem as many plants and animals depend on this process either directly (e.g. pollen as resource, plant reproduction) or indirectly (e.g. fruit as resource). While some plants are self-pollinated or wind-pollinated, most flowering plants require help from pollinators to produce fruit and seed. Pollinators play a significant role in the production of more than 150 food crops, from almonds, apples and lucerne, to melons, plums, and squash. Almost all fruit and grain crops require pollination to produce their crop.

Pollinators, over 100 000 invertebrate species, such as bees, moths, butterflies, beetles, and flies, serve as pollinators worldwide. There are nearly 20 000 known species of bees in nine recognised families, thus pollinators other than bees represent a huge number of species worldwide, almost 80 000 species. Only few, about 1 000 species, belong to vertebrates, as birds, mammals and reptiles might also pollinate many plant species. The conservation of this pollinator diversity is important because it contributes to maintaining a diverse community of floral species in agroecosystems (Fontaine et al., 2006).

Comparing the different groups, Calderone (2012) report that the values attributed to honey bees and non-bees pollinators reached USD 11.68 billion and USD 3.44 billion, respectively, by 2009; however, a decline in both of them was observed in respect to 2004. These data were recently confirmed; in fact, it was estimated that the benefit of all other pollinators, rather than bees, to US agriculture is between USD 4.1 and USD 6.7 billion annually (ESA, 2014); this means almost half of the contribution is from honey bees.

Trend analysis demonstrates that US producers have a continued and significant need for insect pollinators and that a diminution in managed or wild pollinator populations could seriously threaten the continued production of insect pollinated crops and crops grown from seeds resulting from insect pollination (Calderone, 2012).

Although some species of plants are visited only by one type of animal (i.e. they are functionally specialised), many plant species are visited by very different pollinators. In such cases, plants generalise on a wide range of pollinators, and such ecological generalisation is frequently found in nature. For example, a flower may be pollinated by bees, butterflies, and beetles or even birds. It is important to highlight that pollen is also food sources for many beneficial insects, including some pollinator as aphidophagous syrphid. Preference for certain nectar by pollinators has been examined in bees, butterflies birds and, less, in moths and other important insects (Tartaglia and Handel, 2014). Strict specialisation of plants relying on one species of pollinator is rather uncommon; however, there are new recorded of important plant genera and year by year this number increases. For example, all species of *Ficus* (Moraceae) and *Yucca* (Agavaceae) are pollinated exclusively by obligate seed-parasitic wasps and moths, respectively. In addition, the tree genus *Glochidion* (Euphorbiaceae) is pollinated exclusively by a moth of the genus *Epicephala* (Gracillariidae) (Kato et al., 2003).

The greatest variation of pollination methods is found among the plants that are fly pollinated (Faegri and van der Pijl, 1979). Several cases are reported also among Diptera. For example, consistent with their morphologies, *Leucospermum tottum* var. *tottum* (Proteaceae) is pollinated by long-proboscid flies (*Philoliche rostrata* and *Philoliche gulosa*), Cape sugarbirds (*Promerops cafer*), and, to a lesser extent, by Orange-breasted sunbirds (*Anthobaphes violacea*) (Johnson et al., 2014).

Many of the flies that feed on exposed fluids also eat small solid particles including pollen grains. Other plants that are fly pollinated include: Euphorbia, Potentilla, Trifolium, Tradescantia (personal observation), Sedum and various members of the Apiaceae, Brassicaceae and Orchidaceae families (Hagerup, 1951). Flies are especially important pollinators under certain climatic conditions, because they are present at all times of the year. Some plants flowering at odd times of the year may be completely dependent on flies for pollination (Hagerup, 1951). There are two types of fly pollination, myophily and sapromyophily.

Myophily flies that feed on nectar and pollen as adults, particularly bee flies (Bombyliidae), hoverflies (Syrphidae), and others regularly visit flowers. Hoverflies are also of particular interest because of their role as a beneficials in pest control.

Among the sapromyophily, the role of the common house fly in pollination under certain circumstance was highlighted recently. *Musca domestica* intensely forage pollen in male flowers and nectar in female flowers of *Ricinus communis* L. (Euphorbiaceae). In females flowers the foraging speed was 15.09 flowers/min; the foraging activity of *M. domestica* resulted in a significant increase in fruiting rate by 89.8 and 80.7 %; the number of seeds per fruit by 99.0 and 84.2 % and the normal seeds per fruit by 76.0 and 76.0 %, respectively, in 2010 and 2011. This improved performance is justified by the positive action of *M. domestica* on the pollination of flowers of R. *communis* (Douka et al., 2014).

Flies tend to be important pollinators in high-altitude and high-latitude systems, where they are numerous and other insect groups may be lacking (Larson et al., 2001).

Case of specialised pollination are recorded also among Lepidoptera, for example, for *Agrostemma githago* L., which is not visited by bees, bumblebees or hoverflies, but instead by butterflies according to the e-FLORA-sys database (Plantureux and Amiaud, 2010). Typical flowers pollinated by butterflies (Lepidoptera, psychophily) are generally open during the day and closed at night, have a light aroma and are vividly coloured including pure red. The flower is erect so that the butterfly can alight on the flower. The flowers have simple nectar guides with the nectaries usually hidden in narrow tubes or spurs. Flowers visited by butterflies include: *Silene*, *Rubus*, *Solidago*, *Salix*, *Lantana*, *Buddleia*, *Aster* and *Lonicera*. In addition, moths can pollinate flowers (phalaenophily) (Tartaglia and Handel, 2014) that are open at night and generally closed during the day, have a heavy sweet odour at night, are usually white or faintly coloured pale rose or pale yellow. Moth pollinated plants include: *Gaura*, *Yucca*, *Lilium*, *Salix*, *Centaurea*, *Cirsium* and night-blooming cactii.

Cantharophilous flowers (pollinated by Coleoptera) are usually large, single, dull in texture, greenish or off-white in colour and heavily scented. Scents include the spicy scent of many crab apples (*Malus* spp.) to the odour of decaying organic material.

Flowering plants that are beetle pollinated include *Nymphaea*, *Sambucus*, *Magnolia*, *Degeneria*, some species of *Rosa* and some species of the family Apiaceae (Faegri and van der Pijl, 1979). Beetle-pollinated flowers may be particularly important in some parts of the world such as semi-arid areas of southern Africa and southern California and the montane grasslands (Jones and Jones, 2001; Ollerton et al., 2003).

Importance of off-field areas for pollinators other than bees

Despite their importance, pollinators have been negatively affected by agricultural intensification, habitat losses and the decrease in crop diversity in Europe (Ricoua et al., 2014).

The relative abundances of floral resources can change throughout a pollinator’s life, necessitating seasonal switches in nectar diets. Many pollinators display these nectar diet shifts. Food preference may also change based on a pollinator’s capacity to learn and on seasonal resource availability. However, the insect capacity to adapt to different diet is not sufficient to insure their conservation. Field margins and off-field areas are important to sustain arthropods that play an important role in pollination of arable crops: honeybees, wild bees, bumblebees and hoverflies.

Aphidophagous syrphid abundance was higher in semi-natural habitats adjacent to oilseed rape fields than adjacent to wheat fields if the proportion of oilseed rape in the landscape was low, thus indicating local concentration. Haenke et al. (2014) highlight the potential of hedgerows to enhance the abundances of beneficial syrphid flies and their spill over to adjacent crop fields, especially when they are connected with forests. This local exchange was moderated by the extent of mass-flowering crops in the surrounding landscapes owing to local concentration.

Moreover, also unattractive flowers such as Poaceae, really common as weeds or in the marginal fields areas, can be a resource for pollinators in times of scarcity or if the given species is the only flowering species present during the foraging period (Ricoua et al., 2014).

Despite of their importance there is a lack of a predictive indicator at the species level to help different stakeholders, farm advisers and even farmers to gain insight into the impact of the floristic composition of semi-natural areas on pollinator groups in agroecosystems. Recently, a predictive indicator was identified that can be used at the field margin and floral levels to predict the pollination value of floral diversity and abundance of field margins on arable land. This approach takes into account also the contribution of hoverflies, which was half than the ones of bees and honeybees (Ricoua et al., 2014). However, the presence of hoverflies is considered also good indicator for the status of the overall landscape. In fact, functional analysis based on the hoverfly fauna proved to be a synthetic and informative tool to characterise and interpret a number of complex features in a standard and simple way (Sommaggio and Burgio, 2014).

The importance of the off-field areas for NTA as pollinators can be different according to the wider climatic conditions. In temperate and tropical environments, agricultural intensification has primarily negative consequences for pollinator conservation. However, in arid environments, agriculture is often highly dependent on irrigation and farms can offer higher availability of floral resources than the external environment. Floral visitation rates to wild plants inside and outside 40 agricultural gardens in South Sinai, Egypt, were compared (Norfolk et al., 2014). The mean number of flower visitors per plant during a 30-minute focal watch was significantly higher inside the gardens than outside, and this was true of orders Diptera, Hymenoptera and Lepidoptera.

4.2.5.1. Proposals for the definition of tolerable magnitude of effects on non-target arthropods as drivers of pollination

As described for other ecosystem services, in a structured agricultural landscape with low input of PPPs, the acceptable magnitude of effect regarding the loss of NTA species as drivers for pollination owing to the use of PPPs in the in-field area is deemed to be higher than in a conventional managed agricultural landscape with large crop fields. The proposed tolerable magnitude of effects expressed as relative changes of NTA drivers compared with controls relate to well-structured landscape with high NTA biodiversity levels (see also section 4.1).

An assessment of PPP effects on NTA at landscape level should assure that no long-term effects from PPP use will emerge as a consequence of, for example, source–sink dynamics between off-field and in-field areas. This is particularly important for areas with simple structured agricultural fields and high PPP input.

At the local scale, the spatial assessment in the agricultural landscape should be related to the individual range of the local NTA populations driving the ecosystem service ‘pollination’. The assessment should focus on an adequate spatial resolution, so that the service can be provided by NTA in an appropriate time scale. The potential recovery of NTA populations at local scale in longer time spans does not insure that the service ‘pollination’ is provided in the in-field area when needed.

For in-field areas, the magnitude of PPP effects on NTA as pollinators considered to be acceptable in agreement with risk managers should relate to the most sensitive functional group to be supported in-field. Most sensitive functional group is understood as including NTA with high ecotoxicological and/or with high ecological sensitivity (e.g. low recovery potential). Additionally, the pollination service might be susceptible to time constraints (e.g. flowering period).

For off-field areas, only negligible effects on NTA as pollinators are proposed to be tolerable without impacting the general protection goals (EFSA PPR Panel, 2010). Landscape-level assessment should ensure that the magnitude of effects on biodiversity in-field does not compromise the acceptable magnitude of effect agreed with risk managers for off-field areas. Regarding the assessment of effects at landscape level, acceptable level of effects on appropriated parameter will have to be defined together with risk managers once the endpoints are agreed in the scientific community. At the landscape level, negligible effects should exclude year-on-year decline in abundance of species, but also population range restrictions (see also section 4.1).

For in-field as well as off-field areas, the tolerable magnitude of effects should take multiple PPP applications according to typical PPP 'spray schedules' into account. This will possibly implicate a lower level of tolerable effects for single PPP applications especially in-field if the intended use fits in an application scheme that includes several other PPPs with potential effects on NTA in the crop. Multiple applications of several PPPs in typical schedules should also be taken in consideration when addressing the recovery of NTA at local scales (please refer to section 5).

Specific protection goal options for non-target arthropods as drivers of pollination in agricultural landscapes (please refer to the text for justifications)

In-field

Ecological entity: functional group

Attribute: abundance

Magnitude/Temporal scale:

small effect up to months

at local scale: 10 % < effects < 35 % during crop flowering

at landscape scale: to be defined

medium effects up to weeks

at local scale: 35 % < effects < 65 % up to four weeks outside flowering period

at landscape scale: to be defined

Off-field

Ecological entity: population

Attribute: abundance/biomass

Magnitude: negligible effects

at local scale: ≤ 10 % or comparable to non-detectable effects on the abundance of NTA populations that are directly caused by exposure in the off-field habitat

at landscape scale: negligible effects on abundance and spatial occupancy of NTA pollinator species

Temporal scale: not relevant

Table 6: NTAs as drivers for pollination. Key drivers, main exposure routes and population ranges. For underlined species, test protocols are available

SPG	Key drivers	Examples of taxa	Main exposure routes	Population range			
				Large (larger than field size)		Small (smaller than field size)	
				Species of open habitats (including crops)	Species of other habitats (e.g. woody structures)	Species of open habitats (including crops)	Species of other habitats (e.g. woody structures)
Pollination	Pollinators	Hymenoptera	Overspray Contact leaves Oral	Apoidea except *Apis mellifera* (e.g. Osmia sp. Apis sp.) Sawflies Symphyta			
		Diptera	Overspray Contact leaves Oral	Anthomydae, Muscidae, Calliphoridae, Hoverflies, Syrphidae (e.g. <u>*Episyrphus balteatus*</u>) soldier flies (Stratiomyiidae), Bombyliidae			
		Lepidoptera	Overspray Contact Oral water Oral leaves (larvae) Oral nectar	Butterflies (e.g. *Maniola jurtina*, *Vanessa io*, *Pieris rapae*) and Moths			
		Coleptera	Overspray Contact soil Oral	Cantharidae, Scarabeidae, Melyridae, Meloidae			

4.2.6. Summary of proposed specific protection goal options for non-target arthropods. Upper and lower limits of effects on non-target arthropod drivers

Table 7: Summary of the specific protection goal options for NTAs, and of the consequences of loss of, or reduction in, ecosystem services provided by NTAs

Ecosystem service	Specific protection goal options	Environmental consequences of loss/reduction of ecosystem service	Further consequences of loss/reduction of ecosystem service
Biodiversity and genetic resources (section 4.2.1)	In-field habitats: small effects on abundance and occupancy of NTA populations Off-field habitats: negligible effects	Reduction in species diversity reduces the efficiency with which ecological communities capture biologically essential resources, produce biomass, decompose	May lead to increased requirement for external inputs of nutrients to maintain crop yield. Soil structure may be adversely affected, reducing crop yield. General protection goal 'no unacceptable effect on biodiversity and

Ecosystem service	Specific protection goal options	Environmental consequences of loss/reduction of ecosystem service	Further consequences of loss/reduction of ecosystem service
	on individual densities of all NTA species occurring in the off-crop and on spatial abundance and occupancy of NTA species	and recycle biologically essential nutrients. Species loss above a tipping point may force ecosystems to move to a different (locally) stable state or to collapse. Loss of genetic resources removes the ability of an ecosystem to respond to external changes such as climate change (loss of resilience)	the ecosystem' set out in Regulation (EC) No. 1107/2009 is not achieved. The aim of halting of biodiversity loss by 2020 is not achieved: 'Halting biodiversity loss constitutes the absolute minimum level of ambition to be realised by 2020' (2009/2108(INI) and 2011/2307(INI)).[(a)] The aims of Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora not achieved
Cultural services (aesthetic value) (section 4.2.2)	Please refer to 'Biodiversity and genetic resources'	Loss of species that depend on high-value farmlands	The general public may value the agricultural landscape less if it becomes less attractive. The possibilities for education will be limited. Aesthetic values are important for mental health and human well being. Reduced human well being may lead to costs for the society. The loss of aesthetic values of agricultural landscapes may impact also tourism
Pest control (section 4.2.3)	In-field habitats: medium effects on abundance and occupancy of key driver functional groups (e.g. parasitoids, predators) Off-field habitats: negligible effects on abundance and occupancy of key driver populations	The 'balance of nature' may be destabilised, leading to outbreaks of fast-breeding pest species no longer regulated by predators and parasitoids. Ecosystem structure may change; for example, biodiversity is reduced when there are a few, numerically dominant species and the loss or reduction of less dominant species with a specific role in controlling particular pests can cause a shift in the structure of both pest and predatory or parasitoid species	Pests may increase both numerically and in geographical spread, leading to greater reliance on chemical pesticides and further reduction of biodiversity. Aims of Directive 2009/128/for achieving a sustainable use of pesticides are not implemented: 'Member States shall establish or support the establishment of necessary conditions for the implementation of integrated pest management. In protection and enhancement of important beneficial organisms, e.g. by adequate plant protection measures'. Incidence of certain diseases in livestock and humans may increase if vectors of disease (e.g. mosquitoes) are no longer controlled by their natural enemies
Food web support (section 4.2.4)	In-field habitats: small effects on abundance and occupancy of key driver functional groups (e.g. soil or leaf-dwelling NTAs). Generally, no shortfall below the limits given by chick food indices (see text in section 4.24) Off-field habitats: negligible effects on abundance and occupancy of key driver populations	Vulnerable species such as farmland birds that are highly dependent on invertebrates for chick growth and survival will decline further and may become extinct. Disruption of trophic networks can occur, impairing the ecological equilibrium of the system	Diverse income-earning activities such as game-bird shooting may disappear, leading to reduced financial viability of farms. Cultural services will be reduced if vulnerable species decline or disappear. General protection goal 'no unacceptable effect on biodiversity and the ecosystem' set out in Regulation (EC) No. 1107/2009 is not achieved. Aims of Council Directive 79/409/EEC on the conservation of wild birds and of Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora are not achieved.

Ecosystem service	Specific protection goal options	Environmental consequences of loss/reduction of ecosystem service	Further consequences of loss/reduction of ecosystem service
			The aim of halting of biodiversity loss by 2020 is not achieved: 'Whereas the disappearance of species may break the food chain that is key to the survival of other animal and plant species of vital importance for food production, adaptation to climatic conditions, resistance to external agents and the preservation of genetic values' (e.g. 2009/2108(INI) and 2011/2307(INI))
Pollination (section 4.2.5)	In-field habitats: small effects on abundance and occupancy of key driver functional groups (NTA pollinators) during flowering of the crop Off-field habitats: negligible effects on abundance and occupancy of key driver populations	Those flowering plants that depend on insect pollinators are not able to set seed if their pollinators are absent at flowering time. Without pollination, plant species would decline towards extinction, at least at a local scale	The production of more than 150 food crops, from almonds, apples and lucerne, to melons, plums, and squashes, depends on insect pollinators and production of such crops is reduced if there are insufficient pollinators at flowering time. General Protection Goal 'no unacceptable effect on biodiversity and the ecosystem' set out in Regulation (EC) No. 1107/2009 is not achieved. The aim of halting of biodiversity loss by 2020 is not achieved (e.g. 2009/2108(INI) and 2011/2307(INI)). Aims of Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora are not achieved

It is noted that the specific protection goals require assessments also at the landscape scale. Local-scale field experiments are dependent on the landscape context and the duration of field experiments is usually not sufficient to investigate whether there are year to year effects. The appropriate measurement endpoints still need to be determined. Please refer to the text for more information about the ecosystem services in the agricultural landscape in which NTA species are key drivers.

[(a)] Directive 2009/128/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides
EU legislation aiming at the conservation of biodiversity European Parliament resolution of 21 September 2010 on the implementation of EU legislation aiming at the conservation of biodiversity (2009/2108(INI))
Our life insurance, our natural capital: an EU biodiversity strategy to 2020 European Parliament resolution of 20 April 2012 on our life insurance, our natural capital: an EU biodiversity strategy to 2020 (2011/2307(INI))

4.3. Exposure routes of key drivers belonging to the non-target arthropods

Table 8: Exposure routes of key drivers belonging to the NTAs

Group affected (NTA key drivers)	Exposure route
All organisms—but different importance, e.g. butterfly adults and larvae, moths, flies, Hymenoptera	Overspray
Pollinators, e.g. butterfly and moths, etc.	Oral nectar
Predators, e.g. beetles	Oral prey
All organisms—but different importance, e.g. butterflies, flies, Hymenoptera	Oral water
Leaf-dwelling NTAs, e.g. butterfly larvae, grasshoppers	Contact leaves spray/dust
Ground-dwelling NTAs, e.g. beetles	Contact soil

5. General framework for risk assessment of non-target arthropods

As described in section 3, some NTAs are mobile species with individual ranges often larger than the local scale of a single treated agricultural field. Moreover, the range of a local population of NTA might cover an area that includes several fields as well as off-field habitats.

Currently, the assessment of the risk for NTA from the application of a PPP is performed by addressing ‘in-field’ and ‘off-field’ areas separately, assuming that there is no exchange between them. What is more, currently only PPP applications for one single year are considered in the assessment of the risk for NTA from PPP exposure, and no legacy of effects over the years is postulated. However, as was describe in detail in section 3, predicted effects of PPP on NTA in in-field habitats might also have impact on the NTA biocoenoses in off-field areas and the effects might emerge over the years.

As a consequence of the above, the possible effects of PPP applications at wider landscape level over several years of PPP use have to be considered already at lower assessment levels (lower tiers).

The suggested procedure is illustrated in Figure 13.

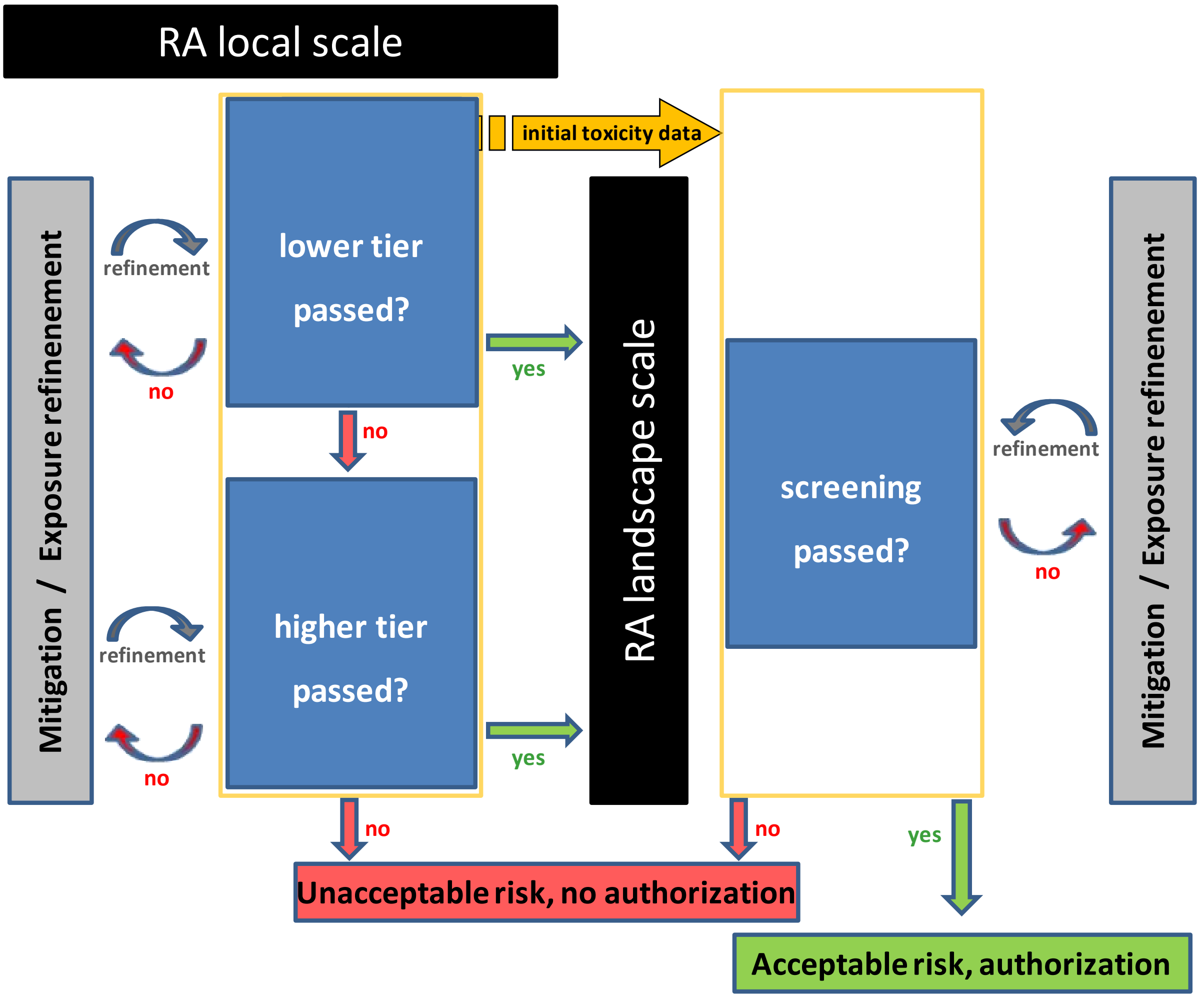
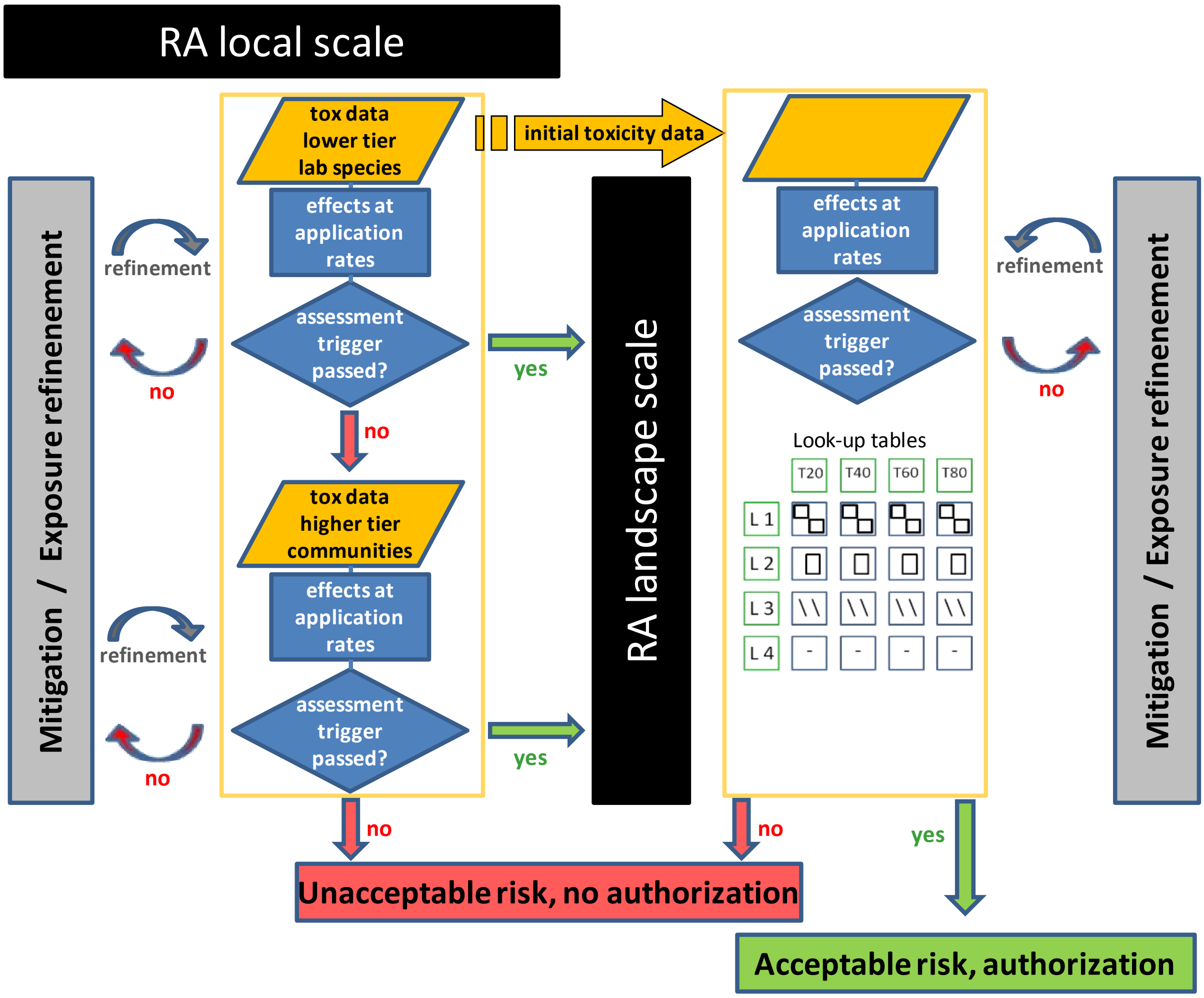


Figure 12: Draft scheme of the proposed procedure to assess the risk for NTAs exposed to active substances and their formulated PPPs. RA: risk assessment. For details, please refer to the text

It is thus proposed to assess the risk for NTA exposed to PPP at the local scale followed by an assessment at the landscape scale (see Figure 13). The need for an assessment at both scales arises from the different individual ranges of NTA species, the respective appropriate management options that can be implemented and the Specific Protection Goal options for the relevant ecosystem services in agricultural landscapes.

At the local scale, the risk for NTA from PPP use is assessed independently following the same proposed scheme for in-field as well as for off-field areas. If the risk as indicated by appropriated risk quotients is considered acceptable at the local scale, then the assessment of the possible effects at landscape scale is performed. Only when both assessment levels are passed (local and landscape scale), risk is considered acceptable and the approval of an active substance or the authorisation of a PPP is supported, respectively.

Figure 14 shows the single steps in more detail.



RA: risk assessment. T20…T80: toxicity descriptor of the active substance or PPP to be assessed. L1…L4: landscape scenarios with diverging field and off-field structures.
Legend of the single process layouts is:

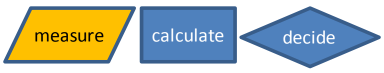


Figure 13: Detailed scheme of the proposed procedure to assess the risk for NTAs exposed to active substances and their formulated PPPs. Processes behind the different steps of the risk assessment are depicted.

5.1.1. Assessment at the local scale

Toxicity of the active substances or PPP is measured at lower tier in standard laboratory test with single species ('lower tier' rhomboid in the risk assessment scheme). Please refer to section 7 for details on the choice of the appropriate species and endpoints in lower tier assessment.

Effects at intended application rates are then calculated by comparing acute or chronic endpoints derived from the tests with the predicted exposure doses (rectangle in the risk assessment scheme). For the local the direct toxic effects from exposure in the off-field habitat are assessed, which should be negligible. Please refer to section 6 'Exposure assessment' for details on the exposure routes to be considered for NTA and for calculations of the predicted exposure doses.

If the agreed trigger (diamond in the risk assessment scheme) for the calculated risk quotient at lower tier is not passed, then exposure refinement options may apply to describe a more realistic exposure scenario for NTA. In addition, especially if the risk for NTA in the off-field is being addressed, several risk mitigation options may be implemented that will reduce the exposure of NTA to the PPP to be assessed. Both exposure and mitigation options will refine input parameters for the calculation of effects on NTA caused by PPP applied at intended rates.

If, after these refinement options, the trigger of the calculated risk quotient is still not met, then higher tier assessment of the risk for NTA can be performed. Section 7 lists requirements for higher tier test methods and describes which uncertainties of the prediction of the risk for NTA in field situations can be reduced by appropriate datasets. Finally, if after assessment of the risk at higher tier level unacceptable effects are predicted at local scale that can not be managed by implementing risk mitigation options, then no authorisation of the evaluated PPP use can be granted.

Conversely, if the trigger for the calculated risk quotient is passed at any level of the local-scale assessment, then the landscape-level risk assessment has to be performed.

5.1.2. Assessment at the landscape scale

Initial acute or chronic endpoints of NTA species serve as an input for the screening of predicted landscape scale effects in so called ‘look-up-tables’. The look-up-tables will flag landscape structures bearing an unacceptable risk for NTA population in association with the application of the PPP to be assessed. Look-up-tables are the results of pre-run model simulations with several landscape scenarios and a selection of physico-chemical active substance characteristic, both selections to be agreed at European level. Please refer to section 7.3 for specification of the requirements for models suitable for the simulation of NTA in agricultural landscapes over years of exposure to PPP. Appendix A gives more details on the derivation of look-up tables to assess the risk for NTA at landscape level. Given the uncertainties that results from the simulation of PPP effects on NTA species currently derived from initial lethal toxicity endpoints as input for single species models, section 7 addressed the description of acceptable changes in NTA population endpoints at landscape scales.

If the assessment of the effects for NTA at landscape scale indicates an acceptable risk, then the PPP under evaluation may be authorised. This means that the assessment of the risk for NTA exposed to PPP undergoes first an evaluation at local field scale and then at landscape scale. Only when both scales indicate an acceptable risk for NTA, authorisation of a PPP may be supported.

If the assessment of the risk for NTA from a PPP application indicates an unacceptable risk at landscape level, then several refinement options may apply as in the case of the local field scale assessment. It should be noted that refinement options for exposure scenarios at landscape scale might differ from possible refinement of the exposure at local field scale. At landscape scale, refinement might concern on the one hand the application patterns of the PPP to be assessed, as not all patterns may be expected to be covered by the pre-run model simulations. On the other hand, properties of the substance, for example parameters describing the degradation of the compound on different matrices or influencing the mode of action, may need to be specifically addressed in the simulations.

When doing a long-term landscape population-level risk assessment, the concept of recovery becomes subsumed under the evaluation of the landscape-scale population status. Therefore, no recovery option is needed, as, if the population is robust and shows no more than negligible effects over time, then recovery at local scale has already been assessed as integral part of this process.

5.1.3. Mitigation of identified risks

A novelty in the suggested risk assessment scheme is the implementation of management options to reduce the risk for NTA exposed to PPP in a landscape context over several years of PPP application. As NTA populations will be affected differently by PPP application in differently structured landscapes, effective management options at landscape scale will have to address the structure of off-field and/or unsprayed habitat in landscapes flagged to bear unacceptable risk for NTA from PPP use. Moreover, also in-field effects of PPP on non-target species are relevant in terms of the protection aims as laid down in Regulation EC 1107/2009, especially when it comes to the defined specific protection goal in this opinion.

Unacceptable effects of PPP use on biodiversity, the services of pollination, food web support and pest control driven by mobile and non-mobile NTA species might require risk mitigation measures that go beyond currently strategies aimed at reducing PPP input in off-field habitats. In this respect, especially cropped no-spray zones, fallow land and flowering margins are deemed to be suitable risk management measures. Necessary measures to ensure an acceptable risk of assessed PPP might even benefit from the implementation of landscape related measures outside the PPP risk regulation such as the greening measures under the Common Agricultural Policy (CAP) and National Action Plans for the Sustainable Use of Pesticides (NAP). To benefit from such synergies with the CAP or/and the NAP, minimum standards with respect to the necessary ecological quality and extent in the context of the risk management of PPP should be defined. However, a safe use of PPPs cannot depend on the implementation of such measures under other legal instruments, as effective risk mitigation measures are mandatory in PPP regulation

5.2. The principles of a tiered approach

For an overview of the principles of a tiered approach we refer to the guidance document for aquatic risk assessment (EFSA, 2013). In summary: According to Boesten et al. (2007) and Solomon et al. (2008) the general principles of tiered approaches are:

- lower tiers are more conservative than higher tiers;
- higher tiers aim at being more realistic than lower tiers;
- lower tiers usually require less effort than higher tiers;
- in each tier all available relevant scientific information is used;
- all tiers aim to assess the same protection goal.

In short, the tiered system as a whole needs to be (i) appropriately protective, (ii) internally consistent, (iii) cost-effective and (iv) address the problem with a higher accuracy and precision when going from lower to higher tiers (see Figure 15).

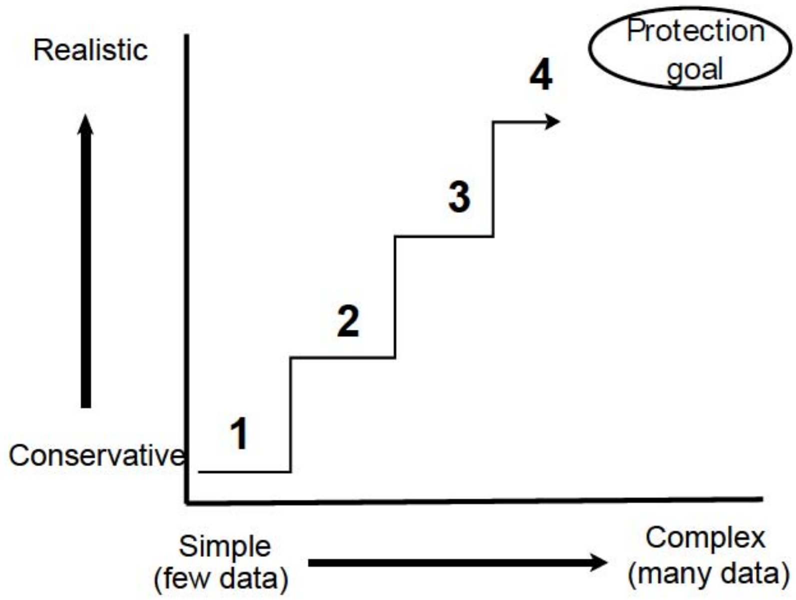


Figure 14: Tiers in the risk assessment process, showing the refinement of the process through the acquisition of additional data (redrafted after Solomon et al., 2008)

Figure 16 shows the relationship between the different tiers, the reference tier and the protection goal.

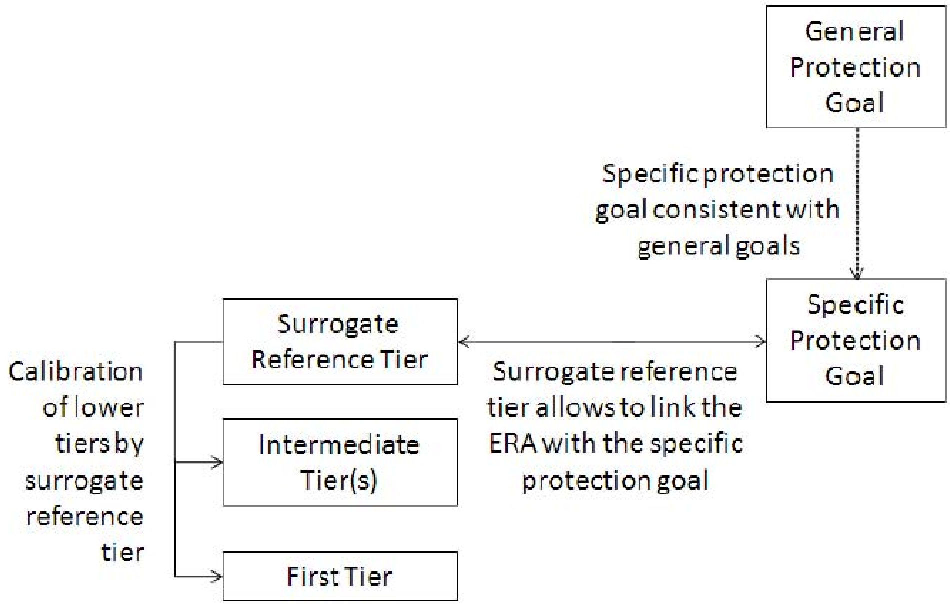


Figure 15: Illustration of the relationship between tiers of the risk assessment process and protection goals, in the approach used by the PPR Panel (EFSA PPR Panel, 2010)

5.3. Tiered approach in the risk assessment for non-target arthropods and definition of (surrogate) reference tiers

In order to assess the risk from PPP use on NTAs it is indicated in section 3 that effects on NTA should be assessed at the local and at the landscape level. The actual reference tier thus is the NTA community present in the field and influenced by processes at landscape scale.

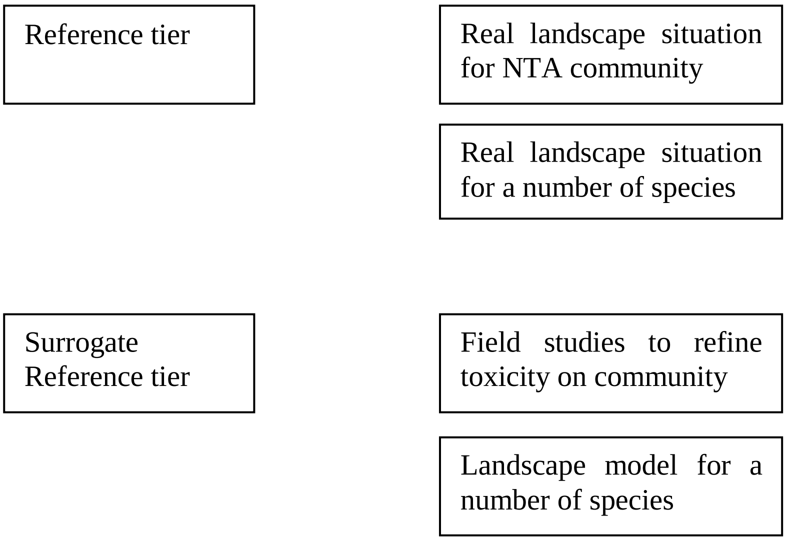


Figure 16: Reference tier versus surrogate reference tier in the risk assessment of NTAs

In the current risk assessment scheme, the highest available tier is represented by field studies performed at local scale. For a number of reasons it not possible to conduct field studies at the landscape level. These are practical reasons of finding comparable replicates, untreated controls, etc. In addition, monitoring of the arthropod community at the landscape level over time puts very high demands on the number of samples and the connected work for identification of taxa. Furthermore it can be questioned whether it is acceptable to test a new PPP before registration at the landscape level. The same aspect can play a role when conducting larger scale off-field field studies.

Therefore, a combination of both assessing the effects at local scale through testing the toxicity of PPP on communities and assessing the effects at landscape scale through modelling long time exposure of single species is proposed as a surrogate reference tier.

For the present opinion, the actual requirement is for a definition of surrogate reference tier. A surrogate reference tier is a compromise between what would be desirable and what is practical. A simple criterion for the latter is that it would need to be realisable within the time frame of production of the guidance document.

In principle one would like a single over-arching reference tier covering the combination of effects assessment, exposure assessment and population modelling. In practice, it seems that two versions are needed, distinguishing effects at local scale for all NTA species and additional aspects for mobile species at landscape level.

5.3.1. Surrogate reference tier for effects of plant protection product use on non-target arthropod species at local scale

Effects of PPP use on biodiversity and the services of pollination, food web support and pest control are driven by mobile and non-mobile NTA species (see section 4). As both mobile and non-mobile species contribute to the provision of the services above and the service provision is needed at local scale, a surrogate reference tier for the effects of PPP on all NTA at local scale is needed. However, a stated above, for mobile species not all aspects can be covered at local scale (see next section).

Non-mobile species (or better: non-mobile NTA life stages, in the following 'non-mobile species',) are defined in section 3 of this opinion. For these species, we propose that a suitable field study is apt to fully define the surrogate reference tier for the intended use. The size of the plots in the field study should be such that source–sink dynamics do not need to be addressed, and plot size and test design should prevent that organisms migrate from and to the untreated control, avoiding both false negatives as false positives. These restrictions are not valid for mobile species present in these field studies; the dynamics of populations of those species might be of influence on the populations of non-mobile species. This aspect should be taken into account when interpreting the results of such field studies.

For non-mobile species, it is proposed to use the field community studies (actual field situation) as a surrogate reference tier. These field studies address community composition, population dynamics, effects on different life stages, indirect effects (food loss), chronic exposure, eventually repeated exposure, interactions between and within species and exposure in the actual field situation.

In the description of the specific protection goal's a number of non-mobile taxa are mentioned that are important for especially food web support and pest control. In order to link the surrogate reference tier to the specific protection goals is of importance that in the field study representatives of these taxa are present.

Different types of field studies are available, reflecting the highest tier of effects assessment in the present NTA risk assessment (see section 7). Most available studies are on replicated plots, most of them were performed in-field, but recently also in off-field environments. Effects on 'natural' NTA population present at the time of the experiment are studied and, generally, crop fields with relatively rich NTA communities are analysed. Exposure is mostly reached according to the intended uses. A usual plot size in current field study designs is 24 × 24 m. Traditionally, a limited number of dosages is tested (1 or 2), focussing therefore on questions of in-field risk assessment. One single study is available in which effects of one single PPP use were studied on larger plots (2 ha) and in different countries. In this study, effect monitoring was intensive, with frequent and detailed sampling (see also section 7).

With increasing attention to effects of PPP use on the off-field environment, studies in grasslands with several test doses are currently performed investigated. These tests allows for an assessment of the consequences of risk mitigation measures on direct, local PPP effects in off-field habitats (see e.g. appendix in De Jong et al., 2010).

In order to determine whether the tested exposure is a realistic worst case and to possibly extrapolate the effects of field studies to other situations, a better insight in the exposure actually eliciting the observed effects on NTA species would be needed in field studies. As NTA species might, for example, move up and down in the canopy several times during the day, simply measuring exposure in the field studies would be insufficient.

As study plots are relatively small in most of the field study designs, recovery can potentially occur from other plots or from the off-crop environment (recolonisation). Therefore, processes demonstrating recovery of NTA species in this type of field studies are only meaningful for the present non-mobile species. For mobile species, e.g. species displaying individual ranges bigger than the size of the test plot, other options such as enclosure studies might be feasible (see section 7).

The available study design with small plot set-ups can, however, also deliver extremely useful information on PPP effects on mobile species:

- direct initial effects of PPP use on communities;
- direct long lasting effects of PPP use on communities in those cases in which internal recovery cannot be observed;
- direct PPP effects on reproduction of mobile species in those cases and time frames in which the presence of, for example, emerging juveniles cannot be ascribed to recolonising adults;
- direct PPP effects on mobile species in cases recolonisation does not take place because of toxic residue levels;
- indirect PPP effects on the habitat quality of mobile species, if these, for example, do not recolonise the test plot because of food scarcity.

5.3.2. Surrogate reference tier for effects of plant protection products on mobile species at landscape scale

For mobile species, modelling is needed because of concerns about landscape-level effects and unreliability of recovery information from field studies owing to source–sink dynamics, i.e. the local-scale assessment for these species is considered to be deficient.

For a surrogate reference tier, modelling would cover realistic landscapes with realistic agricultural practices against which to assess the new product. In the model, parts of a landscape to which the product was not being applied would have normal current agricultural applications of all PPPs and normal crop distributions, thus the baseline for the reference tier model would be current agricultural landscapes. An advantage of using real landscapes in modelling is that they would, in principle, allow subsequent comparison of model to reality, and thus reduce uncertainty regarding the model's realism.

Much of the data needed for modelling exists for most of the EU: realistic landscape structures, form of agriculture down to farm level, accurate crop usage patterns, PPP usage. Information to support this is collected for EU farming subsidy payments and pesticide usage and in theory could be available if local authority access can be granted. There are a number of EU mapping datasets available at a course scale (e.g. CORINE). In addition, the LTER-Europe network may provide a useful source of mapping where other detailed GIS mapping are unavailable. Therefore, in principle a surrogate reference tier model can be quite specific to a chosen scenario and should be possible to generate for most Member States.

Species models that could be used include the *Bembidion* model which is available now (Bilde and Topping, 2004) and is discussed in detail elsewhere in the opinion. A model for a linyphiid spider (*Erigone*) is also available now and in three years time it should be possible to have a butterfly model. In the same or a shorter time frame, it should be possible to extend the *Bembidion* and *Erigone* models to more species, for example spring breeding carabid beetles. All these models are part of the ALMaSS system (http://ccpforge.cse.rl.ac.uk/gf/project/almass/), which is open source and thus could be used to derive a common framework for standardising agreed models under some form of administrative control (e.g. using standard EFSA versions of the models). Development of these models following good modelling practices (EFSA, 2014c) would typically take one to three person years per species, but as well as standardisation, using a common framework for landscape simulation has the advantage that landscape simulation need not be developed newly for each new model. For all species models, new and old, it is recommended that they undergo a testing (validation) phase using data from real landscapes selected for the reference tier.

However, a model also needs a dose–response curve for the new product, linking effects to exposure. The problem with using data from laboratory studies is that the NTAs are often exposed to dried residues (e.g. the two standard first tier test species *T. pyri* and *A. rhopalosiphi*) whereas field studies combine exposure from overspray, contact to fresh and dried residues and oral exposure. However, in some laboratory studies, the animals and substrate are oversprayed (e.g. tests with *Poecilus cupreus*, *Pardosa* sp.). Therefore, one refinement which could be used to reduce uncertainty regarding exposure in the field for acute effects would be to use residue measurements in field studies to calibrate the calculated exposure to actual exposure. By doing this the uncertainty regarding the proportion of time organisms spend in different compartments is also reduced (e.g. leaf surface or soil for hunting spiders). This approach can only be used for rapidly acting stressors, as, if the period between application and exposure is too long, then ecological factors such as spatial dynamics, phenology, and changing reproductive rates may bias the effects estimate from the field. Achieving this requires a measure of exposure from the field experiments and a measure of immediate effect. A suitable field study would need to explore an appropriate range of application rates so as to ensure adequate knowledge about the dose–response. Because the actual exposure distribution is not known for a field study, exposure assessment calculations would also be needed in order to estimate exposure from the known application rate(s). The same exposure calculations are needed in the model and provide a way to extrapolate the measured effects via exposure to different fields in time and space using dynamic modelling (see section 6.10). This therefore assumes that effects are correlated to exposure using the dose–response curve.

Where the dose–response remained subject to considerable uncertainty, a 'reasonable worst case' dose response compatible with the field study data could be used instead. Reproductive effects would also need to be incorporated: in principle for the modelled species but in practice for a surrogate test species from a laboratory study, possibly with an adjustment for uncertainty.

The natural dose–response curve(s) to use in the model would be the curve for the modelled species; but, if another more sensitive species was of particular interest or concern, it might be acceptable to use the dose–response curve for that species even though the model would not apply directly to that species. However care must be taken, this is not a valid strategy to deal with most of causes of variation between different sensitive species because landscape-scale population sensitivity is also a function of life history strategy and behaviour. Therefore, it is strongly recommended that vulnerable species with different ecologies are identified and used in the modelling framework as soon as practicable.

5.3.3. Suggestion for a lower tier landscape-scale screening tool for non-target arthropods

One of the complexities of the real world that requires an innovative solution in regulatory pesticide ecological risk assessment (ERA) is the fact that the precise effect of a particular landscape configuration relies on complex spatial and temporal dynamics involved in animal behaviour and ecology. Of particular importance for arthropods is the 'action at a distance' or 'source–sink' phenomenon (e.g. pollinators flying to treated fields to forage). For example conventional wisdom would suggest that placing source habitats close to a treated area so that they receive over-spray would increase the impact of the chemical at the population level. However, this may not be so. In the case of field voles it has been demonstrated that the rescue effects of close proximity of source populations can over-ride the higher rate of pesticide induced impacts (Dalkvist et al., 2013). Similarly, this problem is not easily solved by using either simple landscape structures or small sections of landscape (Holland et al., 2007; EFSA, 2014), which will induce heavy and unpredictable bias in the assessments. This leads to the situation that a simple lower tier test, not including these dynamics and interactions, may be less conservative than a higher tier approach to modelling the system. The solution is clearly to test the new product in the relevant landscapes together with the relevant management for the intended product usage. Up to now this would not have been feasible, but recent developments in landscape-scale modelling and data acquisition now render this practicable.

Current simulation systems are capable of integrating not only large landscapes at high resolution but also complex management of farms together with the ecology and behaviour of non-target species to create highly realistic impact assessments (see e.g. Dalkvist et al., 2009; Topping, 2011). These simulation tools, although open-source, require considerable expertise to deploy effectively. Hence, it would not be practical to expect wide adoption of these methods without significant development of general frameworks to ensure standardised applications to novel pesticides. Whilst this might be a very useful long-term goal, perhaps under EFSA control, a practical short-term solution to the problem of spatial dynamics and NTAs is required.

One such shorter term workable solution to the problem of incorporation of complex modelling in lower tier assessments is to pre-run the scenarios to be used and provide the impacts in the form of a look-up table, matching product properties and usage to the closest pre-run scenario results. This removes the need for extra work and costs for the applicant requiring only that the requisite number of scenarios have been performed independently to populate the look-up table.

This proposal is based upon what is currently possible with existing models; at the end of the section an assessment of future development options is also provided.

The aim of this lower tier assessment, if adopted, would be to highlight combinations of landscape, pesticide, and NTA properties that would trigger causes for concern at the population level. As a lower tier the scenarios used to support this decision process would be developed as realistic worst cases, each applied to a range of representative landscapes from Europe. Inputs data required from an applicant for this process would be the same as that used for existing regulatory lower tiers, i.e. would not necessitate increased data requirements compared with present procedures and would be implemented in parallel with those procedures.

The steps for lower tier assessment using this system would be the following:

1. The applicant provides data on toxic thresholds, environmental fate in the form of a DT_{50}, and application timing, rates and frequency following good agricultural practice (GAP).
2. This data is used to select a model pesticide scenario with the closest match to the data provided by the applicant. This should be done as an automatic process based on well-defined rules defined as a standard process (possibly by EFSA).
3. The pesticide scenario would then be cross-checked against all landscapes in the database of pre-run scenarios and a profile of the impacts generated. This profile could be used to identify regions of Europe or landscape types where the use of the pesticide could be considered of concern.
4. If impact exceeds a trigger threshold in a fixed percentile of cases (e.g. 0 % or 95 %) then the pesticide has failed this lower tier and a higher tier assessment may be triggered.

The pre-run scenarios will need to fulfil the following requirements in order to provide a useful basis for assessment:

- represent a wide range of landscape structures found in Europe, ranging between heterogeneous and homogenous in both composition and structural organisation and including those considered as being suitable for GAP representative use;
- cover a wide range of potential PPP toxicity such that prospective new products can be encompassed by the scenarios available;
- cover a wide range of application timing and frequency options.

In order to ensure realistic worst-case scenarios a number of simplifications will also be needed. These will include the assumption that the exposure is equivalent to the field rate (i.e. 100 % exposure), that toxicity and DT_{50} (dissipation) is conservative, and that spray drift is incorporated following standard assumptions. It will also be necessary to assume a landscape with a realistic maximum proportion of the area covered by the crop and assume a 100 % market share, in addition to current agricultural management stressors, but excluding any other pesticide of the same type. The assumption being that the number of applications of the product being evaluated replaces any applications pesticides of the same type. It is there necessary to consider multiple applications in the scenarios, and for details of worst-case frequency of usage to be supplied.

Endpoints for each scenario would be simplified to be aggregate measures of impact, as population reduction during 10 years of pesticide use following 10 years of simulated pesticide usage. Three endpoints would be considered, a total population reduction, a population reduction in-crop, and a population reduction off-field. These endpoints would be measured relative to a baseline scenario where the product is not applied (and no substitute product is applied). Impacts would therefore proportional to population size and not based on raw numbers of affected individuals.

Adopting such a system has a number of advantages:

1. It allows the range of agricultural situations to be evaluated and avoids bias due to small or unrepresentative landscapes being used to address spatio-temporal dynamics issues such as source–sink dynamics.
2. It is easy to implement once the scenarios are developed, and from an applicant’s perspective is not more difficult than the current Hazard Quotient (HQ) approach.
3. Conservatism is built into the scenarios ensuring lower tiers are more conservative than subsequent tiers.
4. It provides a standard baseline for development of higher tier modelling approaches, effectively refining the conservative scenarios used at the first tier.
5. The issue of multiple stressors and multiple applications can readily be addressed.
6. The approach is transparent because standard results are used i.e. the same input data will always result in the same result.

Three basic elements are required to be developed before this approach could be fully implemented.

The first is the implementation of the specific protection goals in terms of trigger values from the simulations. What is the critical level for, for example, off-crop population reductions or total long-term population declines? (See section 2 on protection goals.)

The second is the development of the range of potential scenarios. Running these scenarios is relatively easy, but gathering the required number of agreed landscape maps and agricultural management systems would entail a coordinated effort and would need resources allocated to this.

The third component would be the development and testing (validation) of key species models. Some models may be available already, e.g. the ALMaSS beetle model used in section 3.5, but others would need to be developed and tested following EFSA Good Modelling Practice (2014) guidelines. For models that are not already available development resources will need to be allocated.

After implementation, the system could be expanded and improved in a number of directions. The main lack is the range of species available for modelling at the moment. This should be expanded to include a wider range of NTAs, including pollinators (see section 5.3).

If a wider range of species could be included, it should also be possible to include species from southern areas of Europe. This would be necessary to incorporate the range of climatic scenarios that would be necessary to represent the whole European region. Currently the scenarios would be restricted to species common in northern temperate Europe, which are not the same species dominant at more southerly latitudes.

It is suggested in this opinion that chronic effects should also be considered for NTAs (see section 7.4). If this is adopted then chronic effects would need to be incorporated into the simulations based on standard inputs. This would extend the range of scenarios needed.

Further options that could be considered are refinements of the landscape scenarios. These would include altering the crop compositions from mono-culture to realistic crop compositions for the corresponding landscapes. This would reduce the use of the test pesticide even assuming 100 % market share, as it would not be applied to all crops. One less drastic refinement might be to consider very simplified agricultural systems e.g. 100 % monoculture versus 50 % monoculture arable, 50 % grazing systems. How this should be done in practice would need to be addressed by future guidance documentation.

Alternatively, this proposal could be extended by refining model inputs to increase realism by inclusion of more realistic pesticide use and agricultural landscapes and management. An important concept here is that these landscapes should be agreed and standardised, but that particular agricultural management and pesticide use in the area of intended use can be modelled. In this case it would be necessary to reduce the number of scenarios to a manageable volume for highly detailed simulations and because of the need for maintenance of the scenarios to keep them up-to-date with current practice. The updating of the scenarios could be carried out at regular intervals to incorporate changes in landscapes and agricultural management through time. In order to deal effectively with the multiple stressor issues raised by developing a population approach, it will be necessary to change focus from assessment of single isolated products to systems level and accepting a replacement approach to assessment. In this case the endpoint of the simulations would be a systems assessment based on whether the use of the product would improve or worsen the current system state. The data required to parameterise such a system are now available at local and European levels (e.g. subsidy claim data and pesticide usage data). Hence, a part of the maintenance of models and scenarios would be a relatively automated process.

Mitigation could also be addressed (see sections 3.5 and 5.1.3) by allowing options for landscape management to reduce pesticide impacts, e.g. provision of uncropped field areas or boundaries or unsprayed margins.

In all cases, however, it is suggested that higher tier scenarios are still based on variations of standard scenarios such that a transparency of simulation settings and interpretation of outputs is maintained.

5.4. Relationship between effect and exposure assessment for non-target arthropods

EFSA (2010) described the framework for the risk assessment for aquatic and terrestrial organisms. The risk assessment requires two parallel tiered flow charts, one for the effect assessment and one for the exposure in the field. Considering in more detail the interactions between the flow charts for field exposure and effect, there are only arrows from field-exposure to effect tiers (Figure 18). The EFSA (2012) mentioned that all options for delivery of field-exposure assessments to effect tiers are possible (called the ‘criss-cross’ model).

This is possible provided that (i) the tiers in each of both charts address the same aspect (i.e. effect respectively exposure, (ii) the lower tiers of each type of assessment are calibrated with respect to the next higher tier and the reference tier for that type of assessment and (iii) the separation between effects and exposure is assured as one moves through the tiers of each type.

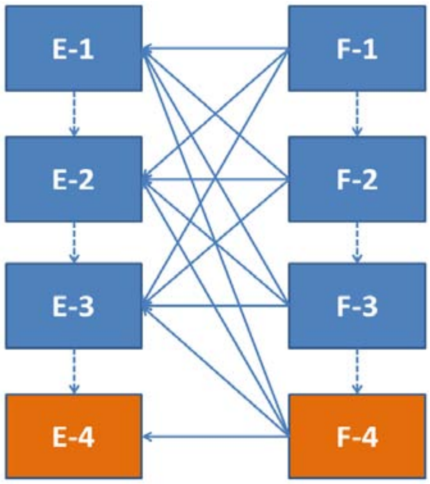


Figure 17: Possible routes through the combined effect and field-exposure flow charts for an NTA risk assessment. The boxes E-1 to E-4 are four effect tiers and the boxes F-1 to F-4 are four tiers for assessment of exposure in the field (‘F’ from ‘field’). Dashed arrows indicate movement to a higher tier. Arrows from right to left indicate delivery of field-exposure estimates for comparison with effect concentrations in the effect flow chart. Notice that in contrast to the original figure proposed by Boesten et al. (2007), there are no arrows from lower tier exposure tiers to effect tier E-4.

However, the criss-cross model has been developed for the local scale, i.e. for a single concentration representing the ERC. For the landscape level, approaches are proposed for higher tiers of the risk assessment. Such a higher tier assessment would involve simulating the intended use (i.e. GAP) at the landscape level and including fate and exposure, as well as the ecotoxicological model as directly simulated components. The appropriateness of the criss-cross model for use in risk assessment should be investigated. Maybe a separate criss-cross model is required for the landscape-level approach. Figure 19 shows in detail how the interaction between exposure and effect assessment works for an arbitrary combination of an effect and a field-exposure tier (by zooming in on an arbitrary combination of an effect and field-exposure tier from Figure 18). Up till now the standard procedure in ecotoxicological experiments for NTAs is to use a range of doses to derive a dose–response relationship. Assessment endpoints within effect tiers have to be expressed in terms of the same type of ERC as the endpoints of the field-exposure tiers. This implies that there are two equally important types of exposure assessments required for the risk-assessment procedure. The first assessment (in the field-exposure box in Figure 19) involves estimating the exposure (in terms of a certain type of ERC) that will occur <u>in the field</u> resulting from the use of the PPP in agriculture. This is part of the field-exposure flow chart (Figure 18) and is often referred to as the predicted environmental concentration (we use ‘predicted environmental dose’ because in the case of NTAs, effect studies are mostly based on dose–response relationships). The second exposure assessment (in the effect box in Figure 19) is a characterisation of the exposure (defined in terms of the same type of ERC) to which the organisms were exposed <u>in the ecotoxicological experiments</u>. This second exposure assessment is part of all tiers in the effect flow chart.

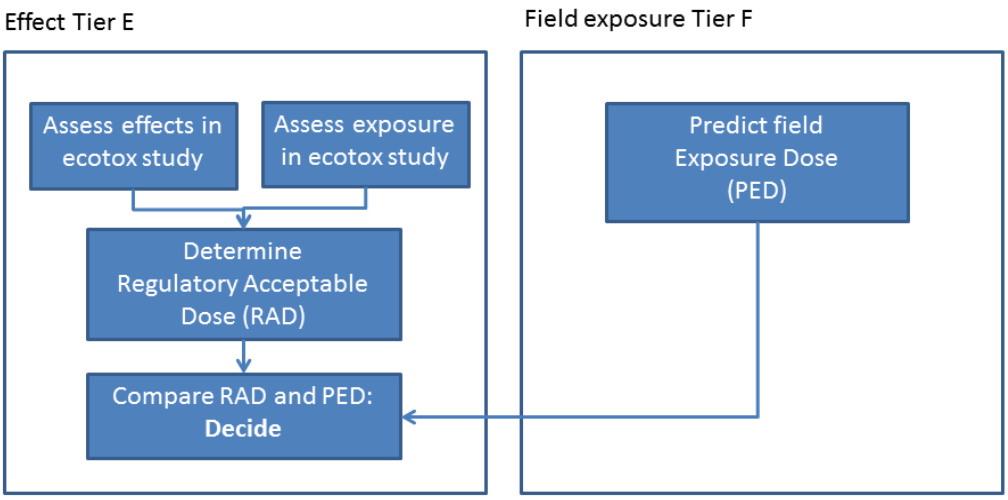


Figure 18: Schematic representation of the two types of exposure assessments which are needed in any combination of tiers of the effect and field-exposure flow charts (adapted from Boesten et al., 2007)

Moreover, for NTAs, there is a further obstacle in the way of applying the paradigm: (the distribution of) exposure is not known in the toxicity tests and field studies (section 7.2). Instead, the tests are based on known application rates for the PPP. In a study conducted on a glass plate, one might argue that the residue density on the surface is directly determined by the application rate although in fact residues are first permitted to dry before the arthropods are exposed. On other substrates or in 3D tests or field studies, there is no direct correspondence. The problem then is that any difference between observed effects, at the same application rate and for the same species, in (say) a glass plate study and a field study will be due in part to differences in exposure. Consider what happens, compared with an assessment based on the glass plate study and the application rate, if one tries to combine a refined exposure assessment with an effects field study. The difference in exposure will essentially be counted twice unless one is certain that the refinement in the exposure assessment does not overlap with the exposure differences between the studies. For example, if the exposure refinement is to do with leaf area of the crop, that will already have affected exposure in the field study, whereas if the refinement was about differences between the landscape where the field study was carried out and the landscape of regulatory interest, that would not have affected exposure in the field study and it might be reasonable to apply the refinement to the results of the field study.

It might seem that a pragmatic solution is to reduce the criss-cross model to one that combines lower tier effect studies (glass plates) with any tier of the exposure assessment described in section 6. However, that too is problematic. The calibration of the lower tier tests, which determines the assessment factor to be applied, is based on direct comparison against field studies. It would not be possible to combine with the lower tier tests any exposure refinement relative to the lower tier tests which is already effectively incorporated in those field studies as that refinement would then effectively be counted twice. This does not preclude combination of any refinement of exposure with lower tier testing but does mean that careful consideration would be needed.

In the future, it might become possible to make more cross-linking, even of higher tier effects studies (field studies) and higher tier exposure studies. This would require both the use of process-knowledge included in models and field measurement of relevant factors such as LAI and exposure concentrations. Note that field measurements alone would not determine the exposure distribution for a species; one would need at least to know the proportion of the time individuals were exposed in each compartment and how that varies between and within life-stages. In principle, it might then be possible to extrapolate effects to other exposure situations not yet field tested. This should be a focus area for future research.

5.5. Calibration of non-target arthropod risk assessment

5.5.1. The paradigm for tiered risk assessment based on specific protection goals

EFSA developed a methodology to define specific protection goals for the ecotoxicological risk assessment of pesticides (EFSA, 2010). An important step in the development of new risk assessment Guidance for NTAs is the definition of specific protection goals in consultation with risk managers. According to EFSA (2010), 'for each key driver (taxonomic group or other ecological entity), a reference tier should be identified, [..] based on the most sophisticated experimental or modelling risk assessment method currently available that addresses the specific protection goal. This reference tier will then be used to calibrate lower tiers using simpler methods that are practical for routine use'.

Typically, in a reference tier, quantification of risk through the calculation of a quotient and the determination of an acceptability criterion might take the following form:

$$\frac{\text{Exposure}}{\textbf{Toxicity}} \leq \textbf{Acceptability criterion}$$

where 'exposure' and 'toxicity' are the outcomes respectively of the exposure and toxicity assessment components of the reference tier.

5.5.2. Adapting the paradigm to risk assessment for non-target arthropods

The difficulties involved in applying the criss-cross model of tiered risk assessment to NTAs have been discussed in section 5.4. In particular, the existing effects tiers all report toxicity as an application rate and no measurements are made of actual exposure. Because of this issue, NTA risk assessment is based on calculating the hazard quotient:

$$\text{HQ} = \frac{\text{Application Rate}}{\text{Toxicity}}$$

where 'toxicity' is expressed also as an application rate. The HQ is then evaluated for acceptability by comparing it to a 'trigger value'. An equivalent formulation is to first divide the 'toxicity' value by an assessment factor equal to the trigger value and then compare the resulting HQ to 1. Calibration for the current risk assessment scheme was described in terms of trigger values (with some use of assessment factors). For describing calibration for the proposed risk assessment structure, there are some advantages to using the assessment factor formulation.

5.5.3. The basis of the calibration of the current risk assessment scheme

The current risk assessment scheme predates the specific protection goal and reference tier approach and has been described earlier in sections 5.1.1 and 5.1.2.

The current tier 1 risk assessment computes a HQ by dividing the intended application rate (g/ha) by the LR_{50}. For the in-field assessment, a higher tier assessment is triggered when the HQ > 2. For the off-field assessment a VDF is applied together with an assessment factor to allow for a wider range of sensitivity for off-field species. It is assumed that, when testing the current tier 1 'sensitive indicator species' *A. rhopalosiphi* and *T. pyri* (which is an EC data requirement), 'the risk of failing to detect significant adverse effects when testing only these two indicator species appears to be very limited' (Candolfi et al., 2001). The validity of the current tier 1 acceptability criterion based on toxicity data for those two standard species determines whether a substance requires further testing and heavily relies on the assumption of a high sensitivity. The tier 1 acceptability criterion has been calibrated using field/semi-field data for only the two standard species.

The calibration of the current scheme was based on two key steps: (i) that the two standard lower tier test species are sensitive species and are therefore likely to have lower LR_{50} values than most other species; and (ii) a calibration exercise, for a number of PPPs, linking laboratory test results for the standard test species to effects in field studies for the same species. Implicitly, it is assumed that the standard test species are not only amongst the most sensitive species in the laboratory but are also sensitive in the field and that the field studies for the standard species account also for interactions between species in the field. Currently, there is no assessment factor for inter-species sensitivity differences between the standard test species and other 'in-field species' considered in tier1 risk assessment (Candolfi et al., 2001).

The sources for considering *T. pyri* and *Aphidius* spp. as 'sensitive indicator species' are Candolfi et al. (1999) and Vogt (2000). Of these two articles, it is only the first that reaches a strong conclusion that the standard test species are sensitive; the conclusion in the second is much weaker. DEFRA (2007) examine the reasons for the differences in outcome between the two articles and conclude that the data used by Candolfi et al. (1999) ought to be re-analysed to try to understand the source of the difference as it might be due to the relatively small number of species tested for most pesticides and the frequent testing of a single highly insensitive species, *P. cupreus*. Unfortunately, the original data used by Candolfi et al. (1999) are no longer available.

The main difficulty with interpreting the evidence as presented by Candolfi et al. (1999) is that it is based on knowing only whether or not there were effects (lethal or sub-lethal ≥ 30 %) of a pesticide on a particular species or species group. As presented, the data do not record relative magnitudes of effect on different species/groups. From such data we can only make an inference about relative sensitivity to a particular pesticide when either (i) there are no effects for both test species and effects for another species or (ii) there are no effects for other species and effects for at least one of the two standard species. Consequently, there is only limited evidence to address the question of interest: how likely is it that one or both of the two standard test species is at the sensitive end of the distribution for any particular pesticide or, expressed in another way, for what percentage of pesticides would at least one of the two standard species be more sensitive than other species. The Candolfi et al. (1999) analysis does suggest that over the range of pesticides considered, the two standard test species tended to be as sensitive or more sensitive, according to the qualitative measure of sensitivity used in the article, than the other species/groups considered but that is in fact the answer to a different question. They quote 4.2 % (= 4/95) as the rate of 'false negatives'. By false negative, they mean a situation where there are no effects in testing a pesticide on *T. pyri* and *Aphidius* spp. but where effects are observed on another species. DEFRA (2007) point out that this rate is not really informative about the question of interest, as it includes cases in the denominator which have no capacity to provide information about the chance that other species are more sensitive: cases where there are effects for one or both standard test species. DEFRA (2007) argue that a rate of greater interest is the rate of false negatives conditional on finding no effects on the standard test species. In that situation, there were effects on other species for 44 % (= 4/9) of pesticides for which no effects were found in testing on the standard species. DEFRA (2007) call this the 'true' false negative rate: the proportion of negatives that are false. They also note that the number of species tested was often very small and that the proportion of negatives that are false could only increase if more species were tested. DEFRA (2007) make a Bayesian analysis of the summary data presented in Candolfi et al. (2009), which they find suggests that the true false negative rate 'is likely to exceed 80 % and could approach 100 %'. Their conclusion is that it is 'potentially misleading to refer to *T. pyri* and *Aphidius* spp.' as 'sensitive indicator species'. However, they note that this does not preclude their use provided that appropriate assessment factors are applied, i.e. that the trigger value for the HQ is appropriate.

To calibrate the current HQ trigger, HQ values of the form application rate/LR_{50} on glass, were plotted against the effects in semi-field/field studies for both *A. rhopalosiphi* and *T. pyri* (Campbell et al., 2000). The criterion according to which the trigger was set was that no effects above 40 % compared with the control should occur in the field (i.e. in field-/semi-field studies with both species). This validation resulted in the suggestion of an HQ of 12 or 8 as acceptability criterion. For commercial data protection reasons, neither names of active substances nor the mode of action of any PPP used by Campbell et al. (2000) in this calibration exercise were provided or could be made available. However, it is stated that pyrethroid, carbamate, organophosphate and neonicotinoid insecticides as well as strobilurin and azole fungicides were included. During the ESCORT 2 workshop, data provided by German authorities suggested that the HQ proposals of 12 and 8 could underestimate effects on *T. pyri* and *A. rhopalosiphi* in the field and the HQ trigger was set to 2. The Campbell et al. (2000) analysis is purely empirical and makes no attempt to account quantitatively for statistical uncertainties.

DEFRA (2007) attempted to calibrate more formally laboratory tests against field effects for the standard test species. Unfortunately, they were unable to obtain access to field study reports and were limited to a re-analysis of the data published by Campbell et al. (2000). DEFRA (2007) used a standard linear regression model for the dependence of field effect (transformed using logit) on the logarithm of HQ and carefully analysed the extent to which HQ predicts field effects and the possibility to use a trigger value to control the probable maximum level of field effects. Subject to their choice of transformations and some statistical modelling assumptions, they made a careful analysis of uncertainty and its consequences. They concluded that there was considerable uncertainty about the field effects expected for the current HQ trigger of 2. They also noted that, like Campbell et al. (2000), their analysis only addressed the field effects on the standard test species. It did not address effects on other species or on communities or in off-field environments. They called for further analysis to address those issues. Finally they presented a table of other sources of uncertainty which should be taken into account at least qualitatively. They emphasised that their model assumes that the relationship between effects and HQ is the same for all pesticides and that there is a high level of variability in effects, at the same or similar application rate, when more than one field study is available for the same PPP.

In the opinion of the panel, the outcome of the DEFRA (2007) Bayesian analysis of the data presented by Candolfi et al. (1999) is interesting but it is not definitive because it depends on a number of assumptions and because the DEFRA (2007) version of the false negative rate is also not a direct answer to the original question. However, even without the Bayesian analysis, the DEFRA (2007) discussion strongly undermines the conclusion of Candolfi et al. (1999). The reality is that interspecies variation in sensitivity was hardly explored at all at the level of an individual pesticide; in more than 80 % of cases analysed by Candolfi et al. (1999), only four species were tested of which two were the standard species and in more than 75 % of cases a third species was the generally insensitive *P. cupreus*. This does not mean that the two standard species do not have a tendency to be sensitive in the laboratory; we simply do not know how far into the tail of the distribution of sensitivity for any particular pesticide (for all the species that have not been tested) they tend to lie, and we cannot know this without much more test data to establish species sensitivity distributions (SSDs), as are commonly used in aquatic systems (EFSA, 2013).

In the opinion of the panel, the detailed conclusions of the DEFRA (2007) quantitative analysis of uncertainty for the data presented by Campbell et al. (2000) depend on the choice of a particular statistical model and it is clear that there is uncertainty about the right form of model to use. However, that uncertainty might not contribute much additional uncertainty to the assessment of field effects for the standard species to be expected at the current trigger level for the HQ.

5.5.4. Fundamentals of calibration for the proposed risk assessment structure

Refinements in the exposure assessments described in section 6 are by construction in order of decreasing exposure. Consequently, the lower tiers are necessarily more conservative and further calibration is not needed.

For the modelling used for risk assessment for mobile species, the model for tier 1 is the same model as for the surrogate reference tier. The difference is that the inputs to the model such as dose–response, landscape, weather, etc. are specific to the scenario of interest in the surrogate reference tier whereas in tier 1 a number of standard scenarios would be used. The tier 1 assessment is constructed to be more conservative than the reference tier.

One approach would be to seek directly to calibrate separately each tier against the reference tier. However, that would require a lot of relevant data for each such calibration and such data are not available. Alternatively, one might seek to establish an assessment factor between each pair of adjacent tiers with the expectation that the individual assessment factors would then be multiplied. This would be more achievable because of the greater availability of datasets suitable for calibrating some of the individual steps. However, the multiplication of assessment factors does not have solid theoretical basis (Cooke, 2010) and is likely to result in an overall factor which would be unnecessarily large unless some of the individual assessment factors are chosen not to be very protective. An example of this approach in human risk assessment is considered by Kodell and Gaylor (1999) and Gaylor and Kodell (2000).

5.5.5. Obtaining assessment factors from a statistical model

For NTA risk assessment, an assessment factor expresses and addresses uncertainty about the ratio between a suitable summary of toxicity measurements and the application rate for the product leading to acceptable effects in the reference tier. The assessment factor will depend on which toxicity summary is being used and should change as more species are tested or when one moves to a higher tier. A transparent and rational way to derive an assessment factor for a particular toxicity summary is to express uncertainty about the ratio using a probability distribution. The assessment factor is then obtained as an appropriate (low) percentile of that distribution. This is the approach taken for example in the widely used SSD-based HC_5 calculation proposed by Aldenberg and Jaworska (2000) and also advocated by Cooke (2010) as a general principle and by Slob and Pieters (1998), Kodell and Gaylor (1999), and WHO/IPCS (2014) for hazard assessment for chemicals.

However, because there are multiple tiers of assessment and the possibility of testing additional species, there is more than one toxicity summary we need to consider and there is more than one such ratio of interest. Moreover, it would not usually be straightforward simply to write down a suitable probability distribution to describe uncertainty about a particular ratio. Consequently, it is better to build a statistical model which links together the various possible test results, field study outcomes and reference tier effects for relevant species. Relevant species would include those tested but also those covered by specific protection goals. The statistical model can then be used to infer the probability distribution for any particular ratio of interest in the light of available test and field study data for the PPP being considered. In this framework, reduction of the assessment factor comes about because additional tests or field studies reduce uncertainty and that will usually lead to a probability distribution for the ratio which covers a narrower range of values for the same coverage probability. This is the general approach advocated by Cooke (2010); examples of methodology for somewhat simpler contexts than NTAs are in Aldenberg and Jaworska (2000), EFSA (2006), Craig et al. (2012) and Hickey et al. (2012).

In this approach, the statistical model would directly address the reasons to apply an assessment factor:

1. to account for inter- and intraspecies variation in sensitivity;
2. to account for potential difference between the application rate causing acute 50 % mortality (LR_{50}) and the application rate causing other adverse consequences of concern (chronic, reproductive,...);
3. to account for potential difference between the application rate causing adverse effects in the laboratory and the application rate causing adverse effects in the field;
4. to account for effects under field conditions, e.g. effect from exposure of different life stages, indirect effects, effects on interactions, behavioural aspects, etc.

In a tiered approach it should be possible to add information through the tiers, refining the risk assessment. As uncertainty is reduced by providing more data to the statistical model, this should result in a lower assessment factor (but not per definition in a higher endpoint). In aquatic risk assessment, it is common to use the SSD approach (Forbes and Calow, 2002) to model the reduced uncertainty resulting from testing additional species. The statistical model proposed below implicitly uses the SSD concept although the resulting calculations would be quite different from standard SSD calculations, for example those of Aldenberg and Jaworska (2000). When the DEFRA (2007) report stated that SSD methodology would not be helpful, they did not mean that it was either unhelpful or unnecessary to think about interspecies variation but that the standard tools require a lot of data. It may help to contrast the use of SSD methodology and standard assessment-factor thinking:

A. Traditional SSD methodology treats each chemical (here a PPP) without reference to data from others. It is assumed that one has sufficient data that one can determine an HC_5 or PNEC with an acceptable level of uncertainty and formal statistical arguments are often used to calculate the required concentration and its uncertainty.
B. Traditional AF methodology is based on having relatively few data for the chemical of interest but on having data for a reasonably large number of chemicals about the relative sensitivity of test species, i.e. how much one has to reduce the concentration/rate from the value causing a specified effect in the test species in order to be quite confident of acceptable effects in most species. However, AFs have often been determined without transparent detailed quantitative reasoning.

In the NTA context, DEFRA (2007) took the view that it was unlikely that there would be test data for enough species to follow the A path. The statistical modelling approach proposed below brings the benefits of transparent statistical reasoning to the B path and also addresses two further DEFRA (2007) concerns: (i) that standard test species are unlikely to be representative in the sense required for traditional SSD methodology as they were chosen with the intent of being likely to be sensitive; and (ii) that variation in sensitivity between species as measured in current laboratory tests is unlikely to be representative of variation between species' sensitivities in the field.

5.5.6. The structure of a Bayesian statistical model for use in non-target arthropod risk assessment

The key to building such a statistical model is to consider the 'dose–response' (really application rate-effect) relationship between effects and application rate (logarithm base 10) for each test or field study which is relevant to a particular species. We need dose–response because the results of different tests and field studies are not comparable. Laboratory tests usually deliver an estimate of the application rate causing a particular effect in 50 % of individuals whereas field studies tend to deliver an estimate of effects caused by a particular application rate and these two kinds of result can only easily be related through the dose–response relationship. There is no theoretical reason why the shape of the dose–response curve should be the same in each case but the reality is that we have insufficient data for NTAs to enable us to make a universal choice of shape or to inform us about how much variation in shape there might be. It therefore seems pragmatic to use a single dose–response shape, i.e. a family of dose–response functions having two parameters: (i) the $\log_{10} ER_{50}$ (logarithm of the application rate causing 50 % effect); and (ii) the slope of the response at the $\log_{10} ER_{50}$. We shall write θ to denote the pair of dose–response parameters. Obvious candidates for a standard shape are the logistic and probit families.

Conceptually, we now make links between different testing procedures, including field studies, for a species by considering θ and how it varies between endpoints and test/study methods. We can also address interspecies variation by considering how θ varies between species for a single endpoint and test/study method. We shall distinguish different values of θ by using a subscript to indicate the species to which the value refers and a superscript which denotes the test type and endpoint. For example, for *T. pyri* there is a tier 1 glass plate test which could provide estimates of

$\theta_{T.pyri}^{GP,L}$ (for acute lethality) and $\theta_{T.pyri}^{GP,R}$ (effects on reproduction) there are (semi-)field studies in vineyards and orchards which consider only *T. pyri* which could provide an estimate of

$\theta^{FS,P}_{T.pyri}$ (population effect)

and there is the potential to observe effects on mite populations in multi-fauna field studies of the type which are envisaged as the reference tier, for assessing effects on non-mobile species, which could provide an estimate of

$$\theta^{RT,P}_{T.pyri}$$

For several other species, there are also test procedures of various kinds and the possibility to measure effects in field studies and for each of those we could write down a list of the types of dose–response which we wish to consider. For others, there are no specific tests and we are concerned only with

$$\theta^{RT,P}_{\mathrm{species}}$$

for each of those species.

The paradigm is that each relevant specific protection goal should be translated into a statement about acceptable outcomes in the reference tier. This in turn should be translated into a statement, for the species involved, about the values of

$$\theta^{RT,P}_{\mathrm{species}1}, \theta^{RT,P}_{\mathrm{species}2}, \ldots, \theta^{RT,P}_{\mathrm{species}n}$$

Where a single species is the focus, this should be phrased as a requirement relating to

$$\theta^{RT,P}_{\mathrm{the species}}$$

for example that there is 95 % probability effect is less than some specified percentage at the planned application rate. For general protection of most species, one might require high probability that, at the planned application rate, the great majority of species have population effects less than some specified percentage. That requirement would then be translated into a requirement about the collection of values

$$\theta^{RT,P}_{\mathrm{species}1}, \theta^{RT,P}_{\mathrm{species}2}, \ldots, \theta^{RT,P}_{\mathrm{species}n}$$

Figure 20 shows an idealised representation of the relationships between the dose–response parameters for some of the various dose–responses of interest. Specifically it is a graphical representation of a Bayesian network (for example, Gelman et al., 2013) modelling the structure of probabilistic judgements about those relationships. An arrow means that the quantity at the end of the arrow is related to the quantity at the start of the arrow but involves an extra component of uncertainty. This effectively equates to a loss of precision of information about the quantity at the start of the arrow. Thus, we can see, for example, that knowing the dose–response for *T. pyri* for the glass plate test lethality endpoint is a loss of information relative to knowing the field study dose–response and that knowing the latter is a, not necessarily large, loss of information relative to knowing the dose–response in a reference tier study. For *P. cupreus*, the lowest tier study is carried out on a glass plate and that dose–response gives less information about the reference tier dose–response than the extended laboratory study which uses quartz sand as the substrate.

A fully specified version of the Bayesian network would provide a complete probability model of uncertainty relating to the quantities shown in the network for the PPP being considered. The information required in order to fully specify that model would derive from two sources: (i) prior knowledge from other products about inter-chemical variability and (ii) test and/or field study results for the product under consideration. In a somewhat simpler context, this is the approach used by a number of the risk procedures in EFSA (2006).

The role of prior knowledge is to provide an initial probabilistic description for each arrow in the network. For example, we would need to provide a probability distribution for the difference between

$\theta_{T.pyri}^{GP,L}$ and $\theta_{T.pyri}^{FS,P}$

Following what Cooke (2010) calls the 'random chemicals' approach, that distribution would derive from knowledge about inter-chemical variation of the difference between

$\theta_{T.pyri}^{GP,L}$ and $\theta_{T.pyri}^{FS,P}$

Such knowledge could be obtained directly by analysis of data when there exists a suitable dataset, in this case of pairs of values of glass plate test result and field study effects for a number of PPPs. The analysis would provide estimates of parameters for the distribution and also statistical uncertainty due to the size of the dataset. The process is illustrated in the example in Appendix I. Where a suitable dataset was not available, one would elicit quantitative expert judgements (EFSA, 2014) from which to derive the required distribution.

When a particular test (or field study) is carried out, say the glass plate lethality test for T. pyri, that test provides information about the corresponding dose–response parameters θ, in this case

$\theta_{T.pyri}^{GP,L}$

Thus, we have reduced uncertainty about

$\theta_{T.pyri}^{GP,L}$

Consequently, we have reduced uncertainty about all the other dose–response parameters to which

$\theta_{T.pyri}^{GP,L}$

is linked, and in particular reduced uncertainty about the reference tier version:

$\theta_{T.pyri}^{RT,P}$

The mechanism by which the reduction in uncertainty propagates is the standard mathematics of probability, known in this case as Bayesian inference.

It is important to recognise that the reported result of a test or field study is only an estimate of the quantity of real interest and one should distinguish the two. All the tests/studies are subject to inter-test variation: were the test repeated the answer would not be exactly the same. For example, the *T. pyri* glass plate lethality test delivers an estimate of the LR_{50} and the true value of the LR_{50} could be determined to high accuracy only by very large-scale testing: multiple labs and/or much large numbers of mites. Allowing for inter-test variation is potentially very important in risk assessment inference (Hickey et al., 2012), especially when the magnitude of the variation is large. This can be done by extending the Bayesian network to include nodes representing results of tests and field studies. For example, one would add a node representing the outcome of the *T. pyri* glass plate lethality test with an arrow from the existing

$$\theta_{T.pyri}^{GP,L}$$

node to the new node. The probability distribution corresponding to that particular arrow might in principle be estimated from the results of ring-testing although it is not straightforward do so. In any case, the example in Appendix I shows that we may be able to finesse this issue.

The description of the model above does not distinguish between mobile and non-mobile species nor between in-field and off-field assessments. In practice, the model would need to include both off-field and in-field reference tier dose–responses. It would also need to address short-term effects for all species and field-study full time-period effects for non-mobile species.

5.5.7. Illustrative example

Appendix I provides details of an example showing how to build part of the statistical model required. Specifically, it links together the glass plate lethality test for *T. pyri* with the field test for the same species and with the reference tier and shows in principle how such a model might be used to derive an assessment factor to be applied to the glass plate LR_{50} in order to protect a specified percentage of species in the reference tier. Part of the model is built using data, of a similar nature to those in Campbell et al. (2000) and DEFRA (2007), on the relationship between glass plate LR_{50} values and effects in *T. pyri* field tests for a number of PPPs. Other parts of the model require expert judgements which have not yet been elicited and hypothetical judgements are used in their place in order to show what judgements would be needed and how they would be incorporated.

5.5.8. Evaluation of additional uncertainties

Important categories of additional uncertainty include: choice of particular statistical model; representativeness of data used for the statistical modelling in relation to the substance under consideration; relevance of expert judgements used in the statistical modelling to the substance under consideration. Each of these categories would involve a number of sources of uncertainty. The statistical modelling proposed would address some but not all sources of uncertainty.

Once the detailed model was built and used to calculate assessment factors, an evaluation should then be made of other sources of uncertainty and their potential to influence risk assessment outcomes (EFSA, 2007).

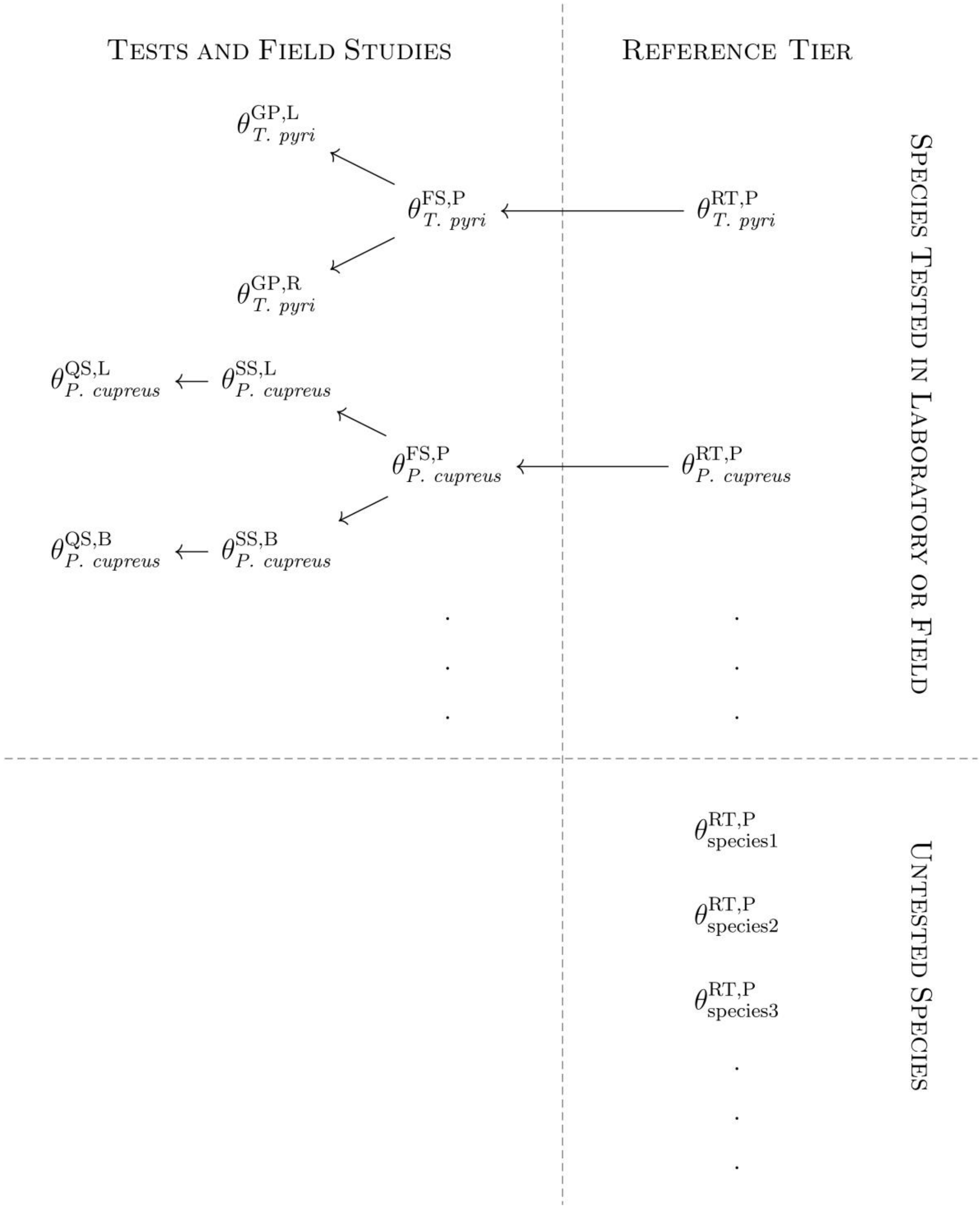


Figure 19: Bayesian network for relationships between species' dose–response parameters for tests, field studies and reference tiers

5.6. Recovery

5.6.1. Definition of recovery in the context of non-target arthropod risk assessment

Organisms are exposed to varying degrees of stressors such as drought, flooding, avalanches, depletion of food resources, parasites and also toxins which can be of natural or anthropogenic origin. Mechanisms evolved which enable organisms, populations and ecosystems to resist and recover from random fluctuations in environmental conditions.

ESCORT 3 defines recovery as follows: The return of populations, communities or functional groups to levels that would be reached without the specific stressor (Brühl et al., 2012.

The upcoming opinion of the Scientific Committee gives the following definition:

'We define ecological recovery as the return of an ecological entity to a defined reference state after a disturbance (e.g. return to its pre-disturbance state). Ecological recovery can thus be defined at all levels of biological organisation from populations to ecosystems, and including both structural and functional attributes.'

The same definition is followed in the current opinion.

Pesticides can pose a pulse (limited duration) or a press disturbance (prolonged stress period) depending on their properties (degradation rates and frequency or application). A pulse disturbance may cause a relatively instantaneous but short-term alteration of the densities of certain sensitive species. However, in case that impacted organisms are not able to recolonise the stressed habitat then a pulse disturbance could also result in long-term effects. If short-term exposures to stressors repeatedly occur (e.g. by multiple applications) and the period between exposure events is shorter than the recovery time of impacted populations then the impact is similar to a press disturbance. A system can shift to a new configuration if populations of key species are eliminated from the system. Therefore, it is important to ensure that those species do not get lost. This needs to be considered when choosing focal species for the risk assessment and assessment of recovery.

A distinction can be made between actual recovery and potential recovery (Brock and Budde, 1994). Potential recovery is defined as the disappearance of the stressor to a level (concentration) at which it no longer has adverse effects.

5.6.2. Protection goals and recovery

In order to integrate recovery in the risk assessment it is important that this is in line with the specific protection goals. It needs to be defined what needs to recover, where it needs to recover and over what time period. ESCORT 3 workshop suggests as relevant endpoints to be assessed recovery of communities in the off-field area and recovery of ecosystem functions (e.g. pollination, pest control).

It makes a difference if the aim is to maintain biodiversity at the landscape scale or if the aim is to ensure the provision of a certain level of the ecosystem service (e.g. pest control, pollination, food web support) during a certain period of time. For example, for maintenance of biodiversity at larger landscape scale the assessment could focus on whether recovery of populations takes place at landscape level. However, if the assessment aims at the protection of a certain level of ecosystem function (e.g. food web support, pollination, pest control) then impacts may be unacceptable even if the NTA community returns to its pre-disturbed state. For example a certain abundance of insects is needed during the breeding season of farmland birds. If the abundance of insects falls below a certain threshold then the chick survival is impacted (see section 4.2.4). Depending on landscape structure, even if recovery can be shown, exceeding certain effect thresholds might not be an option in line with the specific protection goals.

In the current opinion it is assumed that the specific protection goal would focus on populations of NTA species and their long-term survival in the landscape as well as provision of a certain (not yet defined) level of ecosystem function. The level of ecosystem function needs to be defined in the context of the agricultural practices.

5.6.3. Recovery and multiple stress

The EFSA 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, 2010) states, '[..]Multiple stress from pesticides should also be considered to prevent additive impact on the abundance and diversity of non-target species. This appears to be required by point 5 in the General Principles on Decision Making, which states 'Member States shall ensure that use of plant protection products does not have any long-term repercussions for the abundance and diversity of non-target species' (note this refers to plant protection products in the plural and 'does' in the singular, which implies that their use is being referred to collectively). This may imply that the risk assessment of individual plant protection products needs to be more conservative for products used in crops with an intensive multiple plant protection product use than for crops with low plant protection product input [..]'. Consequently, risk assessments developed under the umbrella of both the PPP regulation (1107/2009) and the Sustainable Use of Pesticides Directive (2009/128/EC) are required to take into account multiple stress.

During the EFSA workshop on 'Biodiversity as Protection Goal in Environmental Risk Assessment for EU agro-ecosystems' it has been recognised that effects of single regulated products are assessed in the current regulatory practice while in real agro-ecosystems multiple stress will be relevant when (and if) implementing the recovery concept in ERA. As a way forward for considering multiple stress in the assessment, it was discussed to define realistic 'packages' of multiple stressors on which the risk assessment could be based (please refer to EFSA, 2013a).

It is concluded that any assessment of recovery of NTAs from use of a single PPP needs to account for multiple stress caused by normal agricultural practice (e.g. sequential use of different pesticides) that might hinder recovery.

5.6.4. Important aspects for assessment of recovery

Recovery of populations after pesticide impact can come from surviving individuals or a reservoir of life stages not affected (e.g. eggs in the soil) or from immigration of individuals from neighbouring areas (= recolonisation). The rates of both, internal and external recovery of populations depend on life cycle characteristics of affected species such as the number of generations per year, life cycle strategies (r- versus K- strategies), the presence of insensitive life stages and the migration capacity (Barnthouse, 2004; Solomon et al., 2008; Kattwinkel et al., 2012). Important species traits for recovery can be related to demography (related to population growth rates) and their ability for recolonisation. Rubach et al. (2011) lists as important demographic traits life span, survival to reproduction, generation time, number of offspring and number of reproductive events and as important recolonisation traits the dispersal capacity, distribution patchiness, territorial behaviour, trophic levels, diet specialisation, reproduction mode and dispersal mode. Arthropod species with a poor dispersal ability, only one generation per year, a low number of offspring and which spend their whole life cycle in the treated area will have a low capacity to recover. This should be considered when choosing appropriate 'focal species' for the assessment of recovery in addition to the criteria on selection of species as outlined in section 4 (specific protection goals) and section 7 (tiered approach in the effects assessment of NTA).

The life cycle characteristics of arthropods may be different depending on the temperature. In cooler climates they may have only one generation per year while under warmer climates two or more. Such differences can significantly change the recovery capacity and hence need to be considered for example when extrapolating results from a study to different climatic conditions.

In a recent review of the literature on the ecological recovery of populations of vulnerable species driving the risk assessment of pesticides, Kattwinkel et al. (2012) highlighted the following points with regard to NTA studies:

- The size of areas treated with PPPs and distance from 'refuge' habitats influences recovery rates of NTAs in cereal crops after broad-spectrum insecticide applications—illustrated by dimethoate effects on Carabidae, Staphylinidae and Linyphiidae (Jepson and Thacker, 1990) and deltamethrin effects on aphid parasitoids and hyper-parasitoids (Hymenoptera) (Longley et al., 1997).
- It is very difficult to clearly distinguish recovery that occurs *in situ* and that which occurs as a result of recolonisation from a refuge habitat. For highly mobile flying insects such as the aphid parasitoid *Aphidius ervi*, recolonisation is more important than vertical recruitment, even in cases where the mummy is more protected than the adult stage (Desneux et al., 2006).
- Characteristics of refuge habitats such as hedgerows influence NTA recovery—e.g. Wratten et al. (2003) who found that syrphid Diptera movements into crops differed between four types of field boundaries (fences, lines of poplars with or without gaps, and controls (i.e. no potential barriers).
- Sub-lethal effects of PPPs could influence recovery: Toft and Jensen (1998) found that wolf spiders (*Pardosa amentata*) exposed to sub-lethal concentrations of cypermethrin in a laboratory showed enhanced performance (e.g. in females killing and feeding rate increased compared with the control), meaning that population recovery could be stimulated while other studies with other species and pesticides show adverse effects on lomotor activity (e.g. Everts et al 1991).
- Duffield et al. (1996) demonstrated a relationship between predator recovery and predation rate.
- Rates of dispersal of some NTAs have been published but were not included in the review by Kattwinkel et al. (2012).

These findings suggest that apart from compound properties and species traits (such as mobility, dispersal and reproduction) also landscape structure with source–sink dynamics and direct and indirect effects may be important factors to be considered in the assessment of recovery.

For populations which rely on external recovery (recolonisation) the landscape structure and the distribution of source and sink populations are of great importance. It can be expected that recolonisation will take place within a much shorter period of time if source populations are close to the impacted area. ESCORT 3: Return of population densities after disturbance (e.g. the application of a PPP) to levels similar to those in undisturbed controls can be observed under field conditions in many situations (e.g. one hectare field experimental plots). However, especially for mobile taxa, observed return to the control levels or its absence is not considered to be a robust predictive indicator for the likelihood of recovery under larger scale use of pesticides: it does not consider e.g. applications of different products or different ecological conditions such as the size and distribution of refugia/reservoirs or life cycle parameters of species.

5.6.5. Current methods to assess recovery

The risk assessment conducted according to the Guidance Document on Terrestrial Ecotoxicology under Council Directive 91/414/EEC. SANCO/10329/2002 rev.2 final, 17 October 2002) takes recovery of in-field populations into account. The potential of recovery needs to be demonstrated if effects of greater than 50 % are indicated in first tier tests. This can be done either via aged residues studies or by field studies. Hence the ultimate protection goal is in-field recovery of abundance of NTA species within one year. Usually this is investigated in field studies.

Field studies are useful in order to investigate impacts on populations of non-target organisms. Field studies can address effects on communities, effects on interactions and behaviour, indirect effects, effects on different life stages. The exposure conditions are more realistic but more variable than under laboratory conditions. However, field studies allow only limited extrapolation to other landscapes and climatic conditions (for details, see section 7).

As concluded by ESCORT 3 (Brühl et al., 2012) and also Topping et al. (2014), the 'potential for recovery' or 'actual recovery' as assessed in current risk assessment using small scale plot experiments are 'not a robust predictive indicator for the probable effect of landscape-scale usage of the pesticide'. These limitations could be partly overcome by monitoring exposure over time in field experiments as well as climatic conditions and providing a detailed description of the landscape setting. Using process-based knowledge, the results of a certain field study could then be brought into a wider context in the risk assessment.

Another shortcoming of NTA field studies used in regulatory practice is the fact that landscape specific features and source–sink dynamics are not considered. This could be elaborated using modelling approaches (see section 3.5). However modelling has also some limitations. Usually it is only one population of a species that is investigated. It is difficult to include in a model all different kind of interactions, behaviour and indirect effects which may play a role for recovery. Often field data are not available to build and validate models appropriately. Field studies and modelling approaches could be used complementary in order to address the effects and recovery potential of affected NTA populations.

5.6.6. Conclusions and recommendations for assessing recovery of non-target arthropods

Consideration of specific protection goals outline what needs to be protected, where does it need to be protected and over which period of time. If the provision of a certain level of ecosystem function (e.g. food web support, pollination, pest control) needs to be maintained then impacts may be unacceptable even if the NTA community returns to its pre-disturbed state.

All (experimental or modelling) approaches for assessing the recovery of NTAs from use of a single PPP need to account for multiple stress caused by normal agricultural practice (e.g. sequential use of different pesticides) that might hinder recovery.

Key species for the ecosystem functioning and ecosystem service need to be identified. These species should not get lost otherwise the ecosystem may not return to the original state and/or important ecosystem services may get lost.

Distinction should be made between recovery from within the affected area and recolonisation from source habitats.

Direct and indirect effects are of importance. Both need to be considered for the assessment of recovery of populations.

In higher tier assessments, recovery should be assessed on the basis of species whose traits make them vulnerable (slower/reduced recovery and recolonisation).

Current field study designs do not take into account source–sink dynamics and may not be appropriate to assess recovery.

Depending on the risk assessment question it may be more appropriate to address recovery with field studies or with modelling. Usually both approaches are complementary. For example field studies can provide information on magnitude of effects on an in-field community including indirect effects while modelling can be used to investigate effects in different landscape context and climatic conditions for some species.

It would be beneficial to identify and evaluate risk management measures which facilitate recovery of NTAs.

5.7. Identification of vulnerable non-target arthropod species to be addressed in the risk assessment

The regulatory framework (EC No 1107/2009; EU No 283/2013) requires the consideration of the impacts of active substances and of formulated PPPs on non-target species, on their ongoing behaviour and impacts on the biodiversity and the ecosystem, including potential indirect effects via alteration of the food web.

In order to meet these requirements for an extremely diverse group as the NTAs, the working group identified firstly those NTAs that are key drivers of essential ecosystem services in agricultural landscapes. Then, for NTA acting as key drivers of the ecosystem services pollination, food web support and pest control and for NTA species supporting biodiversity and genetic resources, specific protection goals were derived (see section 4).

The NTA identified as key drivers belong to several classes of organisms and display a multitude of traits regarding their life history, behaviour and food and habitat requirements. It should be kept in mind that NTA comprise, e.g. species with complete versus incomplete metamorphosis, possibly entirely changing life forms, habitat and feeding requirements in different life stages. All these species traits determine the degree of NTA vulnerability following PPP use, as they influence species susceptibility and the extent and duration of exposure and the potential for recovery.

Table 8 of section 4 (specific protection goals) lists taxa and examples of species driving the identified ecosystem services. As expected, traits of key drivers particularly important for one ecosystem services seldom correspond to the traits of the key drivers providing another ecosystem service. For example, the traits of NTA species driving the service ‘pest control’ (e.g. being a predator present at a particular time of the year) are not essential when addressing the service ‘pollination’. This means that different combinations of traits are to be defined for different assessment purposes.

In addition to the fact that species’ traits are the base for the delivery of particular ecosystem services, the working group has defined several steps of the risk assessment procedure in which particular attention should be paid to species’ traits:

1. The assessment of effects. The intrinsic toxicity of PPP, particularly of those with specific modes of action, is the consequence of traits determining the ecotoxicological sensitivity of NTA species.
2. The assessment of exposure. The correct description of exposure routes needs to rely on the characterisation of traits influencing the internal residue doses after PPP use, be it by approximation.
3. The assessment of recovery. For example, life history traits determine the ecological sensitivity of NTA species, hampering or facilitating recovery after initial PPP effects.
4. The assessment of effects at landscape scale over longer time periods. The consideration of the landscape structure and management modulating the impact of PPP on NTA requires particular attention to those traits that affect the species behaviour in in- and off-field habitats over the year.

From the listing above it appears that the definition of trait combinations, possibly pooled in few species, which are relevant for every aspect of the risk assessment, might be challenging. A specific combination of traits that defines vulnerable species for, for example, exposure assessment might not describe vulnerable species in terms of recovery. Moreover, a vulnerable species in terms of exposure and recovery might not necessarily be a key driver in more than one ecosystem service covered by a specific protection goal. This implies that an approach aimed at defining vulnerable species might require the identification of several groups of species for different specific protection goals. It should be stressed that it is not a matter of defining unrealistic worst-case trait assemblages but to explore the possibilities of a reductionist approach in the selection of representative vulnerable species disregarding emerging properties of trait combinations.

It is acknowledged that the description of (a group of) vulnerable, representative species might be a prerequisite for a practicable assessment of the risk for NTA arising from PPP use. However, the characterisation of vulnerable NTA species might not be equivalent to the selection of 'focal species' within a relatively homogeneous taxon such as, for example, birds (EFSA 2009). The assessment of the risk for birds follows a stepwise approach from an 'indicator species' to a 'generic focal species'—both being merely defined by trait combinations—and finally to a 'focal species' –which is a real bird species with explicit traits. The single bird 'focal species' that aims to cover all species present in the crop are determined by field work and are representative for wider regions in Europe.

The diversity of NTA species and the diversity of NTA communities between different regions in Europe are deemed to be so eminent that an approach based on 'focal species' seems unrealistic. Regional aspects of species recruitment in community composition would impede extrapolation of selected 'focal species' to other situation to be assessed.

More promising seems an approach based on, for example, 'indicator groups' that are representative for a set of several species with common trait combinations. This would acknowledge the recruitment of different species in different regions—i.e. sets of species that share similar traits, occupy similar ecological niches and possibly perform similarly in the ecosystem services covered by the Specific Protection Goals. However, special care should be taken when refining the assessment of the risk for such 'indicator taxa', as a refinement addressing traits at single species level would be inappropriate, given that one 'indicator group' possibly covers several hundred other NTA species.

It is suggested that a comprehensive inventory of traits driving the risk of NTA exposed to PPP should be developed in the Guidance Document on the risk assessment for NTA and, subsequently, several realistic combinations of traits that are realised in NTA species in nature should be described in order to best address vulnerable key drivers of the proposed specific protection goals.

Step 1:

- list of traits determinant of intrinsic toxicity;
- list of traits determinant of exposure;
- list of traits driving recovery;
- list of additional traits to be taken into account for landscape-level approaches.

Step 2:

- realistic combination of traits describing vulnerable key drivers of the different ecosystem services;
- pest control: indicator group predators (e.g. relatively large animals with long generation time, present in field and off-field) and indicator group parasitoid (e.g. soft-bodied animals living on top of the vegetation, exposure also in-field);
- e.g. pollination: indicator group pollinator (e.g. butterflies with herbivorous larvae, adult feeding on nectar also in field).

In Figure 21 below, a first description of traits that are considered to be important for assessing the effect of PPPs on NTAs at the landscape scale is shown.

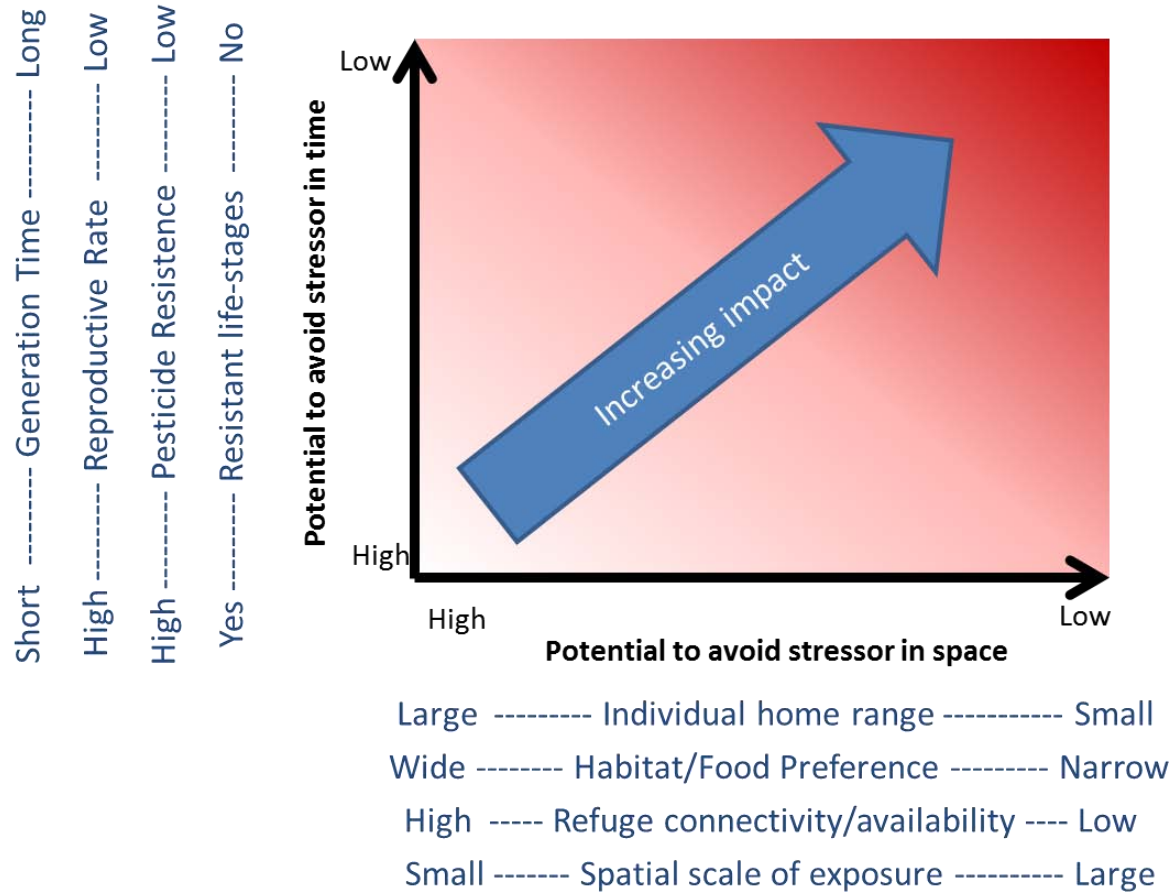


Figure 20: Some aspects of life history traits that alter the likelihood of a landscape-scale population to avoid a stressor in space and time, and therefore to reduce or increase population-level impacts. Note that any one species may exhibit contrasting traits along the same axis, thus making the identification of focal species difficult, as the precise contribution to avoidance of each trait and their interactions is difficult to quantify. In addition, particular behaviours (e.g. attraction to crops) may further alter the probable impact, as can spatial structure of landscapes (e.g. are source habitats near to treated fields). (Figure adapted from Recovery SO, EFSA in prep)

6. Exposure assessment for NTAs

6.1. Introduction

As explained in section 1 of the current opinion, NTAs are pragmatically defined as (life stages of) terrestrial invertebrates that dwell primarily on the soil surface and/or the vegetation, whereas (life stages of) terrestrial invertebrates that move primarily in the soil are called ‘in-soil organisms’. This distinction was made because the main exposure routes for these two groups of organisms differ and hence also the risk assessment differs:

1. For (life stages of) species dwelling in the soil and the soil litter layer, the exposure to PPPs will derive from dermal and oral uptake routes taking place in the surrounding soil compartment;
2. Terrestrial invertebrates moving most of the time on the soil surface or in the vegetation will experience different exposure routes and contamination levels, as the surrounding medium is mostly atmosphere and contamination takes place upon local contact and/oral exposure to residues or direct overspray.

This section describes the field exposure assessment for NTAs (NTAs), whereas exposure in the effects experiments is described in section 7. The field exposure assessment for in-soil organisms is described in EFSA (2012). The risk assessment (i.e. the combination of the effect assessment and the field exposure assessment) for in-soil organisms will be described in a later EFSA opinion.

6.2. Environmental compartments to be considered in the exposure assessment of non-target arthropods

Terrestrial invertebrates that are considered to be NTAs can generally be divided into three categories:

1. soil-surface dwelling animals;
2. herbivores, spending most of their life time on plants;
3. arthropod species such as pollinators, living on the top layer of the plant canopy.

Many NTAs fall into two or even all three of these categories during their lifetime. For example, many butterflies and most moths are herbivorous in larval stage, pupate in soil, and become pollinators as adults. Larvae of carabid beetles live in topsoil, while adults are typical surface dwelling hunters. Other insectivores, however, can spend all their life on plants, feeding on herbivorous insects, like ladybirds do. Many dipterans are herbivorous as larvae and live on leaves, but become pollinators as adults, being exposed mostly through floral nectar and pollen.

From ecological and ecotoxicological points of view, the above three categories, and therefore the compartments in which they reside, should be considered in the exposure assessment of NTAs.

6.3. Dominant exposure routes for non-target arthropods

While the relevance of the exposure routes ‘overspray’ and ‘contact to leaves and soil’ are unquestioned in the risk assessment of NTA exposed to PPP, the relevance of the exposure of NTAs via contaminated food items is subject to debate. It was discussed whether the assessment of the risk for NTA exposed via contact to residues on plant and soil surface would possibly cover also the risk that can arise from exposure of animals to food items contaminated with PPP. The Panel concluded that this is not the case and hence it is essential to address also oral exposure of NTAs to PPP residues in food. The most important arguments for addressing oral exposure are provided below.

1. A potentially higher susceptibility of arthropod species due to oral exposure than due to contact exposure cannot be excluded.

   Using data compiled from Draft Assessment Reports published by EU Member States for PPP active substance evaluation, doses causing 50 % mortality for honey bees (*Apis mellifera*) via oral or contact exposure towards different PPPs were compared. Bees did not show a systematically higher susceptibility being exposed via either of the two paths. This indicates that effects via oral exposure cannot be predicted by only testing contact exposure and vice versa. It is acknowledged that the comparison of bee oral and contact toxicity (OECD 213/214) probably represents the comparison of an overspray versus an oral uptake of fresh residues scenario. This means that for herbivorous NTA there will be larger differences in bioavailability of dried and fresh residues on plant material taken up orally or via contact. Bundschuh et al. (2012) assessed acute toxicity of different insecticides on grasshopper nymphs and compared acute mortality on a plastic surface and on grass mixtures sprayed with the same application rate. The authors found no higher toxicity of purely contact versus grasshoppers exposed by both oral and contact. However, no comparison of the doses received was possible. Moreover, different bioavailability and surface area concentrations of the tested substances on plastic surfaces resp. grass do not allow drawing a conclusion whether oral exposure would be covered by a contact scenario in risk assessment. Regarding *Lepidoptera*, a comparison of LD_{50} values for different compounds compiled by Schmitz (2008) indicates that larvae of *Pieris brassicae* showed higher sensitivity to toxicants when compared with *Apis* *mellifera*, indicating that toxicity of PPP towards bees does not well predict toxicity for, for example, *Lepidoptera*.

2. For hydrophobic compounds, sorption to substrates will reduce the availability via contact exposure. Relative importance of oral exposure versus contact exposure may be increased.

   When addressing internal concentrations of toxic compounds in animals—being the concentrations finally responsible for lethal or sub-lethal effects—the accumulation via the body surface will predominate for hydrophilic compounds bioavailable from surfaces in contact with the animal. By contrast, for compounds with high affinity to organic matter, the greater relative contribution of oral exposure versus contact exposure to total body burdens has been demonstrated (biomagnification versus bioconcentration; e.g. Ma et al., 1995; Green et al., 2006; Laskowski and Hopkin, 1996; Heddle et al., 2012; Saenz-de-Cabezon et al., 2006).

3. Considering the time course of exposure in different substrates, oral exposure to PPP residues will gain importance over time.

   It is postulated that the relative contribution of oral exposure and contact exposure for PPP toxicity will increase over time, as PPP residues that are not available on plant surface any more owing to sorption to organic substrate and uptake by the plant will be available in the plant tissue for a longer period of time when these are consumed as food.

4. For active substances in PPP with systemic properties, oral exposure is the relevant exposure path when carry-over from one growing season to the other is concerned. Oral exposure is the main exposure path for systemic substances available in pollen, nectar and plant material. This is especially true for systemic substances applied as seed treatment.

6.4. Selection of the ecotoxicologically relevant types of concentrations

As described by EFSA (2010), any assessment of the risk to organisms has to be based on concentrations that are most relevant for the effect (called the ecotoxicologically relevant concentrations, abbreviated to ERCs). The relevant concentration can be the peak, a time weighted average or concentration profile over time and these may vary over space. Two extreme cases are (i) the highest peak concentration in a 3D structure and (ii) a time and space weighted average over a volume (an above ground-level vegetation). Currently it is unclear what is the ERC. As a pragmatic approach it is proposed to take initial concentrations on top of the plant canopy. It is acknowledged that this is a worst-case scenario and once detailed information on the ERC becomes available it may be refined.

When linking exposure to effects for risk assessment, the same ERC should be used for both field exposure estimates and effect estimates. However, in the case of contact exposure[10] and exposure by direct overspray, the relevant test endpoints for NTAs are traditionally expressed in mass per area rather than concentrations. Consequently, field exposure has to be expressed in the same unit (mass per area). As a reasonable balance between precision and manageability we found estimating the exposure as follows:

1. Overspray—amount of a substance per area relevant animal surface. Given appropriate data, this exposure can be calculated as a dose expressed as amount of substance per animal (EFSA, 2013).
2. Contact exposure
   a. Soil surface—amount of a pesticide substance per soil surface area after the spray. Studies on pesticide toxicity with NTA living on soil surfaces are usually performed

[10] In lower tier effect assessments, NTAs are generally exposed to residues on dry surfaces. This fact is disregarded in this chapter. Exposure estimates for moments shortly after application will be for wet residues. This may influence the toxicity and internal exposure; this influence is, however, unknown.

using sprayed artificial or natural substrates. Experimentally tested amounts of active substance per area are estimated based on a range of pertinent application rates.

b. Plant surface—amount of a substance on flowers and leaf surface area, as these are habitats for a variety of arthropods, including herbivores, predators and pollinators.

3. Oral exposure

Concentration in food items—residues in food items given as mg substance/kg food normalised for kg substance applied/ha (RUD). Pollinators, such as bees, flies, butterflies and moths, are exposed significantly to PPP through nectar or pollen on tops of plants. In addition, predators such as ladybirds may be exposed significantly by oral exposure, as these feed on (pest) organisms such as aphids.

Concentration of PPPs in leaves (crop for in-crop estimates; natural plant communities in hedges). Plant leaves are consumed as food by herbivores. The concentration in upper parts of plants can be much higher than in lower parts, especially for pesticides sprayed over larger plants. As the movement of herbivorous invertebrates throughout the canopy is rather unpredictable, these animals may be exposed to these high concentrations as well. The concentration in the upper part of the canopy is therefore a relevant concentration, in want of a properly defined ERC.

In all cases, initial concentrations on soil and plant surfaces are needed, as this is the highest concentration right after spraying, which currently is assumed most relevant to most NTAs. Especially with modern, non-persistent insecticides, the first few days or even hours after the spray are decisive when addressing lethal exposure and possible onset of effects on reproduction, and animals surviving this initial impact are usually relatively safe.

6.5. Specification of the exposure assessment goal

As described in EFSA (2014b), specific protection goals should preferably be defined in terms of percentiles of effects distributions. This 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 direct identification of percentiles of effects distributions. However, an agreed methodology for developing such environmental scenarios is not yet available. For this reason, the Panel proposes the exposure assessment goals to be defined separately until better alternatives have become available.

So the field exposure estimate should apply to a given percentile of the concentration distribution (usually the 90th percentile) of the treated fields. Following a consultation of risk managers, EFSA (2012) chose the exposure assessment goal to be the 90th percentile exposure concentration in the intended area of use in each of the three regulatory geographical zones (North, Centre, South). In line with this, the Panel proposes to base the field exposure assessment of NTAs on the 90th percentile concentration in each of the three regulatory geographical zones (North, Centre, South).

Developing an exposure scenario for a given percentile requires simulating the concentration distribution in the entire target area (e.g. EFSA, 2012). The model for simulating this concentration distribution should preferably include all relevant exposure routes (e.g. spray drift deposition, dust drift deposition, overspray and oral exposure). As such models are not yet available for regulatory purposes at the European level, the simplifying assumption is made that the individual exposure routes can be assessed separately.

6.6. Overall assessment of the exposure based on the different routes

In the following sections, the assessment of exposure by contact, direct overspray and oral uptake are described. For each of the routes, a tiered approach is suggested that needs to be further elaborated when developing the guidance document. As combination into one overall risk assessment is currently not possible, it seems necessary to assess all exposure routes and evaluate the routes separately. A safe use is then possible when each of the assessments results in an acceptable risk (Figure 22).

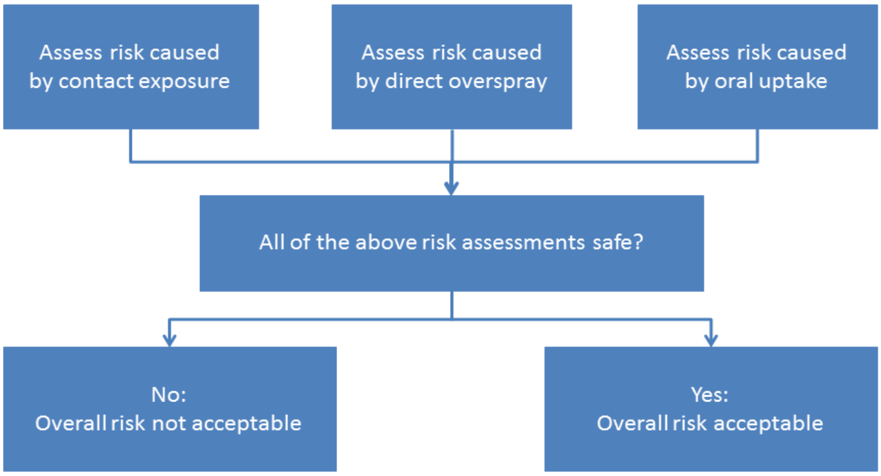


Figure 21: Scheme for the exposure assessment of NTAs

Currently available exposure data, for example data on RUDs, probably do not distinguish between the various exposure routes and the results are then to be seen as lumped over the three routes. Available data should therefore be carefully checked and attributed to one of the three exposure routes or marked 'lumped'. Appropriate values should be used for estimating exposure according to three specified routes. Notice that the use of RUD values based on lumped exposure routes might lead to conservative assessments of the risk if exposure is overestimated.

RUD data are contents of PPP residues in food and feed items, harmonised to a PPP treatment with a unit dose, i.e. a typical application with an amount of 1 kg/ha. It is assumed that actual contents can be calculated by multiplying the RUD with the used application rate.

RUD values are usually determined immediately after the application event on species collected from the treated field, except for food/feed items that need time to establish a concentration. In these latter cases, the highest concentration over time may be taken. RUD data usually concern the edible part, which for NTA usually is the whole body, and do usually not distinguished between exposure routes. Data concerning NTA may be biased to higher concentrations owing to collection of the organisms; dead organisms may have a higher chance of being included in the sample, and these dead organisms may have experienced relatively high exposure.

The use of RUD data is the determination of exposure of NTA is might be conservative, especially when lumped values are used as proxies for specific exposure routes.

6.7. Overview of the assessment of contact exposure

The exposure endpoint in this section is the load of a surface (mass per area, areic mass). This is a surrogate for the real exposure, which is generally unknown and not estimated in currently available and accepted test methods (section 7). Four situations, which require different exposure calculations, are distinguished:

1. in-field, on crop
2. in-field, on soil
3. off-field, on plants
4. off-field, on soil.

The exposure assessment for each of these four situations is described in the sections below. Each section starts with an overview of the most relevant processes and concludes with a proposal for a tiered exposure assessment approach. The highest tier in each of these assessment schemes is a spatially distributed approach, which may be used separately or be incorporated as a component in landscape-level effects models.

Many NTAs live in more than one of the above environmental compartments. A landscape-level approach to risk assessment would be needed for such moving animals as described in section 5. As these approaches are not yet available for all species, a lower tier conservative approach could be to evaluate the risk in all relevant compartments separately and to accept a PPP only if all underlying risk assessments result in a safe use (Figure 23).

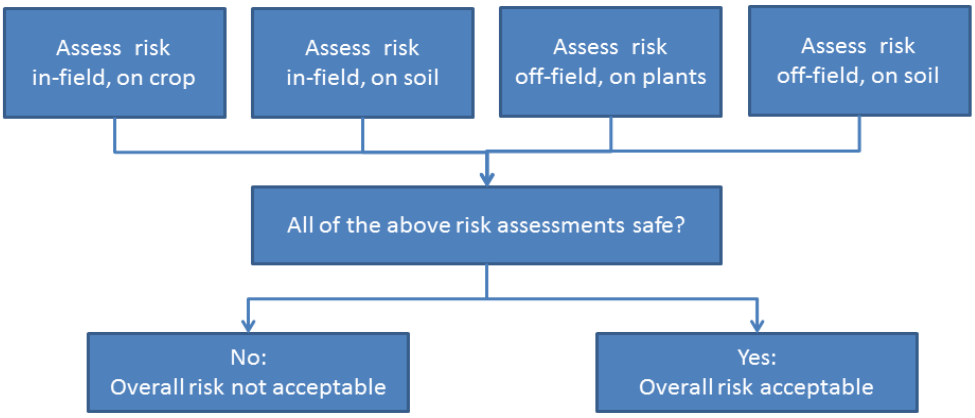


Figure 22: Scheme for the assessment of contact exposure for (life stages of) NTAs that reside in more than one of the above environmental compartments

6.7.1. Assessment of in-field, on-crop contact exposure

As described in section 6.4, the maximum areic mass on the leaf surface is needed[11]. It is assumed that only applications within one growing season are relevant and that there is no carry-over from previous growing seasons. This is rather obvious for crops that are removed from the field upon harvest, but it is also in line with the default dissipation rate (*DisT50* of 10 days) of PPP from leaf surfaces (EFSA, 2014). As the lower tiers of the exposure assessment should be protective for all types of animals, in these lower tiers a maximum of 1 is used for the leaf area index (LAI, ratio between single sided leaf area and surface area of the field).

Using a LAI of 1 implicitly implies that NTAs are exposed to the dose on the surface of the upper vegetation layer. This is obvious for populations of NTA species that live habitually on flowers and leaves on the top of the canopy (e.g. butterflies and hoverflies). The top vegetation layer is also highly attractive for several NTA species, not only for foraging but also for thermoregulation (sun basking butterflies), for mating (especially in those case species where males show a territorial behaviour) or as landmark for orientation while species migrate (Dover, 1997; Longley and Sotherton, 1997).

The LAI value may be refined in higher tiers based on ecological considerations (e.g. populations of species dwelling deeper in the canopy or living on the soil are less exposed, even if daily migration patterns to higher strata of canopies have been observed (e.g. Abraham and Vas, 1999; O'Neill and Rostol, 2012)) or exposure considerations (e.g. the use of spraying equipment that effectively distributes the substance over the total leaf area).

[11] In current NTA risk assessment the peak value (PIED) is required. It is possible that time-weighted averages are required in future as well. For the interpretations of higher tier effects experiments, it may be necessary to have information on the time course of the exposure concentration as well (e.g. to quantify the effect of recovery; see the Aquatic Guidance Document for examples).

6.7.1.1. Theoretical considerations

The exposure of NTAs is expressed as the areic mass of substance on the crop canopy, expressed as mass per unit area single sided leaf surface. The initial amount can be calculated from the nominal dosage using the fraction intercepted by the crop canopy (see also Figure 24):

$$A_{d,p} = \frac{1}{LAI} f_i A_{d,f} \tag{1}$$

where $A_{d,f}$ (kg m^{-2}) is the areic mass of pesticide applied to the field, LAI (m^2 m^{-2}) is the Leaf Area Index, $A_{d,p}$ (kg m^{-2}) is the areic mass of pesticide arriving on the crop canopy and f_i (-) is the fraction of the dose intercepted by the crop canopy. The fraction intercepted (f_i) depends on the crop development stage and should be obtained from the improved FOCUS interception tables as published by EFSA (2014; http://www.efsa.europa.eu/en/efsajournal/doc/3662.pdf).

PPPs can be dissipated at the plant surface by volatilisation into the air, penetration into the plant and by degradation at the plant surface. In current exposure models (e.g. Tiktak et al., 2000), these three processes are generally lumped into one process, based on first-order kinetics:

$$R_{dsp} = k_{d,p} A_p \tag{2}$$

where R_{dsp} (kg m^{-2} d^{-1}) is the areic mass rate of dissipation of pesticide at the plant surface, $k_{d,p}$ (d^{-1}) is lumped rate coefficient for dissipation from the plant canopy, and A_p (kg m^{-2}) is the areic mass of pesticide at the crop canopy. The dissipation rate is calculated as follows:

$$k_{d,p} = \frac{\ln(2)}{DisT50_p} \tag{3}$$

where $DisT50_p$ (d) is the dissipation half-life. The default value of the dissipation half-life is 10 days EFSA (2012). Notice that the dissipation half-life at the crop canopy is not corrected for temperature, as insufficient data is available to establish such a relationship. Furthermore, the dissipation from the canopy is a result of various processes, with each different dependency on temperature.

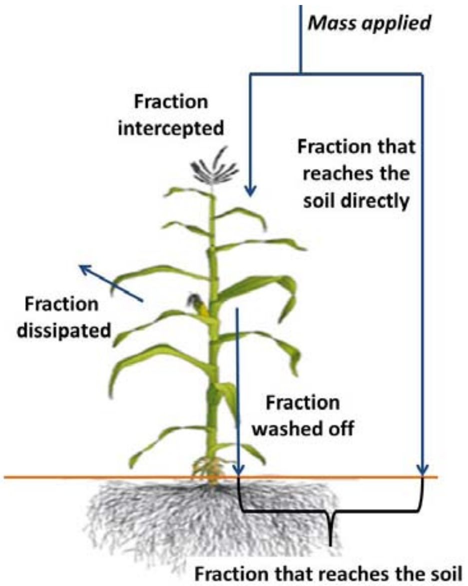


Figure 23: Schematic overview of the processes occurring at the crop canopy. The fraction of the dose reaching the soil is the sum of wash-off from the canopy and the fraction of the dose that reaches the soil directly

Pesticides can further be washed-off from the plant canopy. The areic mass rate of pesticide wash-off is taken proportional to the throughfall flux:

$$R_{w,p} = (1 - \exp(-P\, w_p)(SC\, P - P_i)A_p \tag{4}$$

where $R_{w,p}$ (kg m^{-2} d^{-}-^{1}) is the areic mass rate of pesticide wash-off from the crop canopy, w_p (m^{-1}) is an empirical wash-off factor (set to 100 m^{-}-^{1} according to EFSA, 2012), P (m d^{-1}) is precipitation, P_i (m d^{-1}) is intercepted rainfall, and SC (-) is the soil cover fraction. In current exposure models used at EU-level (e.g. Tiktak et al., 2000), the soil cover fraction is estimated from the LAI using Beer's law:

$$SC = 1 - e^{-\kappa LAI} \tag{5}$$

in which κ is the extinction coefficient for diffuse solar radiation (set to 0.6 based on Kroes et al., 2008). Interception of rainfall by the crop canopy can be calculated from the empirical equation (Braden, 1985):

$$P_i = a\, LAI\left[1 - \frac{1}{1 + (SC\, P)/(aLAI)}\right] \tag{6}$$

where P_i (m^3 m^{-}-^{2} d^{-1}) is intercepted precipitation, P (m^3 m^{-}-^{2} d^{-1}) is precipitation and a is an empirical parameter with a default value of 0.25 for ordinary agricultural crops (Braden, 1985).

Please note that the wash-off factor in Equation 4 (100 m^{-}-^{1}) is the maximum value found in a series of wash-off studies as reported by Leistra, 2005. This value therefore delivers a conservative estimate for the fraction of the dose reaching the soil, but may not be sufficiently conservative for the residue at the plant canopy. The Panel recommends also deriving realistic worst-case values for wash-off factor with respect to the residue left at the canopy.

Redistribution of substance over the plant canopy may occur because of plant growth and uptake of substance by the plants. These processes may become relevant if other ERC than peak concentrations are required and in case of repeated application. Insufficient knowledge and data are available to account for these processes at the moment. As a first approach, in case of multiple applications redistribution could be accounted for by assuming the total dose (residues from earlier applications and the last application) being uniformly distributed over the relevant LAI.

6.7.1.2. Proposed tiered assessment scheme

The proposed tiered approach uses the theory above and consists of four tiers with increasing level of complexity. The first two tiers are simple conservative assessments based on a single scenario. The third tier involves calculations for realistic worst-case scenarios for each of the three regulatory zones North, Central and South. The highest tier is a spatially distributed exposure modelling approach.

Tier	Description
1	*Conservative assessment assuming complete interception by the crop* The total amount (100 %) of the annual application dose is intercepted by the crop. The areic mass (amount per surface area leaves) is calculated based on the LAI of the crop at the earliest growth stages included in the GAP, with a maximum LAI of 1
2	*Conservative assessment assuming partial crop interception* Interception is determined based on the growing stage(s) of the crop using the interception tables published in EFSA (2013). In the case of multiple applications within a growing season, interception is calculated for each application separately. The areic mass for the assessment is calculated from the total intercepted amount and the LAI at the time of the last application, with a maximum LAI of 1 for downwards spraying techniques and 2 for sideways/upwards spraying techniques
3	*Scenarios for the regulatory zones North, Central and South* Interception is calculated as in tier 2 for each application in a growing season, but dissipation of substances due to wash-off and dissipation from leaf surfaces in between applications is taken into account. As the processes depend on climatic conditions and crop development, different scenarios need to be derived for the three regulatory zones North, Central, South. The relevant end-point is the 90th percentile of the maximum areic mass on the leaves in the area of use in a regulatory zone. Note that the maximum amount is not necessarily the amount after the last application. As in earlier tiers, a maximum LAI of 1 respectively 2 is used for downwards and sideways/upwards spraying techniques. However, if it can be demonstrated that the application method effectively distributes the substance over the total leaf area, the real LAI may be used in the calculations. The use of a dilution factor based on the real LAI may also be justified based on ecological considerations (e.g. the primary habitat of a species is in the lower part of the canopy) and demonstrated distributions of the sprayed substance over the canopy (see also section 6.7.6)
4	*Spatially distributed assessment for each of the three regulatory zones North, Central and South* As tier 3, but the relevant end-points are simulated using a spatially distributed approach. This allows direct quantification of the 90th percentile from the simulated frequency distribution. It also allows integrating the exposure and effects assessment in landscape-level effects models as described in section 5

Mitigation options are possible and may be taken into account, where appropriate, in tier 3 and 4. One mitigation option, for example, could be the use of special spraying machines (with lower to zero interception, or more or less even distribution over the total (two-sided) leaf surface).

The relative deposition distribution factor may be different for in-field and off-field as the influence of the spraying technique on the deposition pattern over the leaves will be greater for the in-field exposure assessment.

6.7.2. Assessment of in-field, on-soil exposure

As described in section 6.4, the maximum areic mass on the soil surface is needed. In contrast to the situation at the crop canopy, PPPs may accumulate in soil following multi-year applications, which implies that carry-over from previous growing seasons cannot be ignored. In support to the exposure assessment of soil organisms, EFSA (2012) developed a simple analytical model, which accounts for these effects. It is proposed to base the on-soil exposure assessment of NTAs on the same model.

Theoretical considerations
The areic mass reaching the soil surface ($A_{d,s}$) is the sum of the dose that directly reaches the soil surface and the amount washed-off from the crop canopy (Figure 24):

$$A_{d,s} = f_s A_f = ((1 - f_i) + f_i f_w) A_f \tag{7}$$

where A_f (kg m^{-2}) is the areic mass applied to the field, f_s (-) is the fraction of the dose reaching the soil, f_i (-) is the fraction of the dose intercepted, f_w (-) is the fraction of the dose washed off from the canopy. Both the interception fraction and the fraction of the dose washed off from the canopy depend on the crop development stage. EFSA (2014) developed tables of the fraction of the dose reaching the soil based on many simulations with the pesticide fate models PEARL and PELMO. These tables can be used for deriving f_s; however, for dynamic simulations the use of the mathematical equation above is indispensable.

The simple analytical model describing the fate of an active substance and its metabolites is described in section 3 of EFSA (2012). This model first calculates the amount of residue in soil ($A_{plateau}$) just before the next application and after an infinite number of annual applications:

$$A_{plateau} = \frac{z_{avg}}{z_{Til}} A_{d,s} \frac{X}{1 - X} \tag{8}$$

where z_{avg} (m) is the averaging depth (set to 0.01 m, as only on-soil exposure is considered here), z_{til} (m) is the plough depth (for arable crops this is fixed at 0.2 m according to EFSA, 2012) and X is defined as:

$$X = \exp(-t_{cycle} f_T k_{ref}) \tag{9}$$

where t_{cycle} (d) is the time between annual applications (usually fixed at 365 days, i.e. the crop is treated each year), f_T (-) is a factor describing the effect of soil temperature on the degradation rate coefficient, and k_{ref} (d^{-1}) is the first-order degradation rate coefficient at reference temperature.

The dimensionless factor f_T describing the effect of temperature on degradation is described by the Arrhenius equation:

$$T > 0^o C \quad f_T = \exp\left(\frac{-E}{R}\left[\frac{1}{T} - \frac{1}{T_{ref}}\right]\right) \tag{10a}$$

$$T < 0^o C \quad f_T = 0 \tag{10b}$$

where E (kJ mol^{-1}) is the Arrhenius activation energy, R is the gas constant (0.008314 kJ mol^{-1} K^{-1}), T (K) is the soil temperature, and T_{ref} (K) is the temperature at reference conditions (20 ºC so 293 K). The Arrhenius activation energy was set to 65.4 kJ mol^{-1} (EFSA, 2007b). The degradation rate coefficient k_{ref} is calculated from the degradation half-life in soil by:

$$k_{ref} = \frac{\ln(2)}{DegT_{50}} \tag{11}$$

where $DegT_{50}$ (d) is the degradation half-life at the reference temperature. The maximum exposure will occur immediately after the next application:

$$A_{peak} = A_{plateau} + A_{d,s} \tag{12}$$

where A_{peak} (kg m^{-}-^{2}) is the maximum areic mass in total soil. Degradation after application may be calculated using first-order kinetics using the same degradation half-life as in Equation 11.

Proposed tiered assessment scheme

The proposed tiered approach uses the theory above and consists of four tiers with increasing level of complexity. The first three tiers are assessments based on realistic worst-case scenarios for each of the three regulatory zones North, Central and South. Parameter values of these scenarios are those of the concentration in total soil scenarios as described by EFSA (2012). The highest tier is a spatially distributed modelling approach. Within tier 3 and tier 4, models with a different level of complexity could be used ranging from a simple one-layer model to more comprehensive pesticide fate models such as PEARL (Tiktak et al., 2000) and PELMO (Klein, 1990).

Tier	Description
1	*Conservative scenarios for the regulatory zones North, Centre and South assessment assuming no crop interception* It is assumed that all the pesticide reaches the soil. The total dose is calculated by summing up the doses of all applications within a growing season. The maximum amount expressed per surface area of soil is calculated by summing this annual dose and the plateau value obtained with the simple analytical model based on the amount applied annually. Parameter values for the three scenarios are those of the concentration in total soil scenarios as described in EFSA (2012)
2	*Conservative scenarios for the regulatory zones North, Centre and South assessment assuming crop interception* Interception by the crop is considered. The amount reaching the soil should be derived from Equation 7 or using the tables published by EFSA (2014). The total dose is calculated by summing up the amount reaching the soil for all applications within a growing season. The maximum amount expressed per surface area of soil is calculated by summing this annual dose and the plateau value obtained with the simple analytical model. Parameter values for the three scenarios are those of the concentration in total soil scenarios as described in EFSA (2012)
3	*Scenarios for the regulatory zones North, Central and South* The background concentration is calculated with the simple analytical model. All applications within a growing season are added individually and degradation in between these applications is simulated using first-order kinetics. Alternatively, more sophisticated models such as PEARL and PELMO can be used, as these models also take other dissipation processes such as leaching, plant-uptake and volatilisation into account. Parameter values for the three scenarios are those of the concentration in total soil scenarios as described in EFSA (2012)
4	*Spatially distributed assessment for each of the three regulatory zones North, Central and South* As tier 3, but the relevant end-points are simulated using a spatially distributed approach. This allows direct quantification of the 90th percentile from the simulated frequency distribution. It also allows integrating the exposure and effects assessment in land-scape level effects models as described in section 7

Mitigation options are possible and should be taken into account, where appropriate, in the higher tiers. One mitigation option, for example, could be the use of special spraying machines (with higher interception).

6.7.3. Exposure pathways for off-field exposure

The presence of PPPs on off-field non non-target areas is a combination of two processes during and after the application of the compounds in the field: (i) the emission of the applied product out of the field edges and (ii) the deposition of the emitted amounts onto the off-field areas. Drift is currently considered to be the most important factor for off-field emissions to non-target terrestrial areas. Drift is defined as droplet drift but also vapour drift and dust drift are considered to be important emissions in some particular cases. Deposition on non-target areas is defined as the entry path for transport of airborne substances from the air as an environmental compartment to the non-target area, i.e. to an aquatic or terrestrial compartment or to non-target plants, arthropods, bees, etc. In some cases off-field exposure is considered to be negligible and not further assessed, e.g. in the case of rodenticides, substances used for wound protection or in the case of substances used in stored products or in greenhouses.

As droplet drift is generally the most important exposure pathway for off-field areas, this is elaborated in the paragraphs below. For a discussion on vapour drift and dust drift, the reader is referred to EFSA (2014b).

Droplet drift

Spray drift is defined as the part of the applied product that leaves the treated field through the air because of air currents during the application of the PPP. These droplet drift emissions do not include emissions by volatilisation. Droplet drift is considered to be a short-distance process (0–30 m) and occurs only during and shortly after application (i.e. within a few minutes actually defined by the time between spraying and collection of samples during drift experiments).

Droplet drift is not compound specific but is mainly dependent on droplet size, wind speed, wind direction and crop and spray-boom height during spraying. The spray drift is calculated on the basis of spray drift tables, which give the deposition as a percentage of pesticide application rate deposited at a given distance from the last crop row as a function of crop type (arable crops, fruits, grapes, hops and vegetables), crop stage (early or late) and spraying technique. Different spray drift curves are available (Southcombe et al., 1997; Rautmann et al., 2001; van de Zande et al., 2012, 2014). These spray drift curves were obtained from deposit measurements on artificial receptors (e.g. filter paper strips) on soil level. Most off-field emissions are calculated for deposition on surface water or soil. However, interception by non-target plants can be influenced in a different way because droplets have less contact with leaf surface owing to lower velocity and because of the presence of a laminar air layer on the leaf surface which influences contact. Moreover, the height and structure of the canopy is different from that of bare soil. For example, Kjær et al. (2014) demonstrated that spray drift deposition decreased with height in the plant canopy and that the effect of height is different at larger distances from the field. The PPR Panel did not review datasets to quantify these effects and assumes that the current methodology to assess spray drift deposition (FOCUS, 2001) will continue to be used for exposure assessment at EU level until better alternatives become available.

Currently, estimation of spray drift deposition is based on the values given by Rautmann (2001). These values apply to 90th percentile conditions. However, in a workshop on harmonisation of European drift values (Huijsmans and van de Zande, 2011; van de Zande et al., 2014), it was concluded that spray drift values for the reference situation in field crops originating from recent research were considerably higher than the values used by FOCUS (2001). These differences were particularly important at short distances (0–3 m) of the treated crop (Table 9) and were caused by differences in the selection of datasets to fit the reference deposition curves upon. Crop and spray-boom height during application of the pesticide are other important reasons for differences in spray deposition. For this reason, van de Zande et al. (2012, 2014) suggest using a different spray drift curve for developed crops than for short crops.

Table 9: Estimated spray drift deposition for field crops (% of in-field target deposition) downwind of a sprayed (downwards) bare soil surface and a crop situation based on joined spray drift data from Germany and the Netherlands (van de Zande et al., 2014) and FOCUS (2001)

	Distance from the treated area of the crop (m)			
	1	**3**	**5**	**10**
van de Zande et al. (2014): crop	42.5	6.7	2.8	0.88
van de Zande et al. (2014): bare soil	7.77	1.93	1.01	0.42
FOCUS (2001): crop and bare soil	2.77	0.95	0.57	0.29

Because of the use of machinery, there will always be a certain distance between the off-field target area and the last treated row in the field (Figure 25). This minimum distance is called the minimum agronomic crop-free zone and because it results from agronomic practices its width cannot be changed by risk managers (Van de Zande et al., 2012). The consequence of the presence of an agronomic crop-free zone is that spray-drift deposition at the edge of the off-field target area is less than 100 % of the sprayed dose rate. (Notice that this differs from EFSA (2014 NTTP), who stated that drift deposition at the edge of the non-target area is 100 %.) So the risk assessment for the off-field area could consist of two steps. In the first step, the exposure would be carried out using the deposition rate at the distance of the minimum agronomic crop-free zone. If the protection goal for the off-field area would not be met in this step, risk mitigation options would have to be assessed in a follow-up step. Options to mitigate spray drift deposition to off-field areas include (i) the use of spray drift reducing techniques, and (ii) the establishment of non-spray buffer strips, with or without crop. As spray drift deposition decreases with distance and drift-reducing technique class, spray drift mitigation options could be evaluated using a matrix approach (Van de Zande et al., 2012). Spray drift deposition could, for example, be evaluated first for the standard spraying technique, second for drift-reducing techniques and measures and third for all spray techniques with stepwise wider buffer strip.

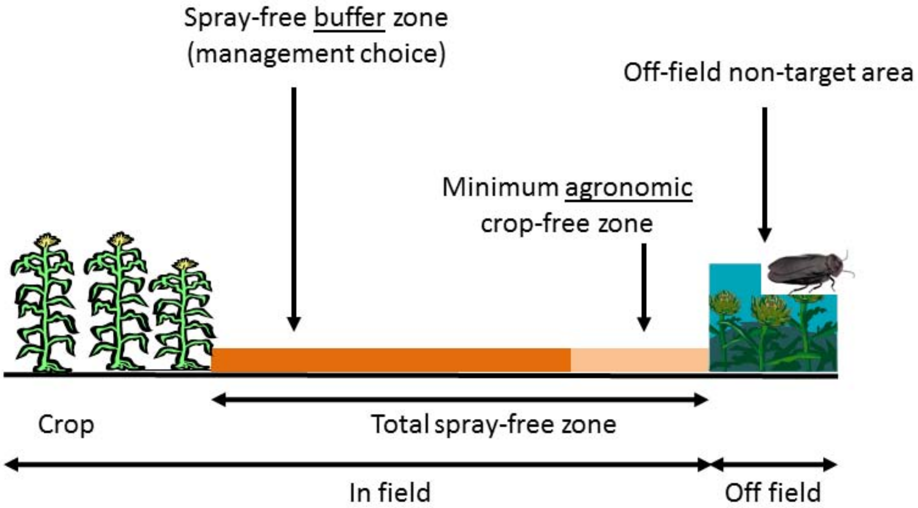


Figure 24: Schematic representation of a crop with total crop-free zone, minimum agronomic crop-free zone and crop-free or spray-free buffer zone. The distance of the minimum agronomic crop-free zone is 0.25 m for grass and cereals, 0.75 m for crops grown on ridges, 3 m for orchards, 2 m for avenue trees and 0.5 m for all other arable crops (Van de Zande et al., 2012)

Spray drift deposition differs between crop types (grass and bare soil, field crops, fruit crops, vines and hops) and crop development stage. For this reason, a spray drift deposition curve and hence an evaluation matrix is needed for each combination of crop type and crop development stage, or classes of these. For estimating spray drift deposition onto surface waters, spray drift deposition curves were developed by the FOCUS Surface Water working group (FOCUS, 2001) for many major crops. Harmonised European drift curves are currently only available for bare soil, grass and fully developed arable field crops (van de Zande et al., 2014); spray drift curves for fruit crops are expected to become available by autumn, 2014. However, for vine and hops, no updated values are foreseen in the near future. In this situation, the PPR Panel recommends the evaluation of new spray drift curves when they become available and to start revising the spray drift assessment methodology accordingly. For the time being, the PPR panel recommends the use of the current assessment based on FOCUS (2001). However, please note that the exposure assessment for all environmental compartments in which spray drift is relevant would benefit from this revision. A summary of the current spray drift assessment methodology and guidance for how to calculate the maximum exposure rate is given in Appendix D of EFSA (2014).

6.7.4. Assessment of off-field exposure, on-plant exposure

As described in section 6.4, the maximum areic mass on the leaf surface is needed. As it is *a priori* unknown where the organisms reside on the plant, a maximum LAI of 1 is used in the lower tiers of the assessment. It is further assumed that only applications within one growing season are relevant and that there is no carry-over from previous growing seasons. The risk assessment starts at the edge of the field, taking into account all relevant exposure routes as described in the previous section. (Note that taking the edge of the field does not imply that spray drift deposition is 100 %.)

Theoretical considerations
The exposure of NTAs is expressed as the areic mass of substance on the canopy, expressed as mass per unit area single sided leaf surface. Canopy processes are calculated in the same way as in the in-field, on-crop assessment (section 6.7.1). The only difference is that in the calculation of the initial amount the fraction deposited by spray drift deposition is taken into account:

$$A_{d,p} = \frac{1}{LAI} f_d \, f_i A_{d,f} \tag{13}$$

where f_d (-) is the fraction of the dose deposited by the spray drift deposition.

Proposed tiered assessment scheme
The proposed tiered approach can use the same four tiers as in section 6.7.1. Refinements at tiers 3 and 4 are, however, based on different considerations:

1. Spray drift to the off-field target area may be reduced by the use of spray drift reducing techniques, and by the establishment of non-spray buffer strips; evaluation of these mitigation options can be systematically done using a matrix approach (see *Section 6.7.3*).
2. Refinement of the LAI is possible based on properties of the vegetation and ecological considerations.

Scenario properties (e.g. type of vegetation in the off-field area, climate properties, etc.) have to be established during the development of the guidance document and should be based on distributions in real landscapes.

6.7.5. Off-field, on soil

As described in section 6.4, the maximum areic mass on the soil surface is needed. The risk assessment starts at the edge of the field, taking into account all relevant exposure routes as described in section 6.7.3. (Note that taking the edge of the field does not imply that spray drift deposition is 100 %.)

Theoretical considerations
The areic mass reaching the soil surface ($A_{d,s}$) is the sum of the dose that directly reaches the soil surface and the amount washed-off from the canopy, taking into account the fraction deposited by spray drift deposition (Figure 25):

$$A_{d,s} = f_d \, f_s A_f = f_d ((1 - f_i) + f_i \, f_w) A_f \tag{14}$$

Proposed tiered assessment scheme
The proposed tiered approach uses the same four tiers as in section 6.7.2. Refinements at tiers 3 and 4 are as in section 6.7.4.

6.7.6. Accounting for distribution of the substance over the leaves

A VDF has been used to account for distribution of the substance over the leaves. The use of the VDF originates from current risk assessment for NTAs in off-field areas according to ESCORT 2 (Candolfi et al., 2001). Its purpose is to relate results of lower tier effect studies to the 'real' off-field environment where (leaf-dwelling) organisms are assumed to be less exposed because of a different vegetation structure and a larger dilution of exposure than in the field. In current practice, the VDF is set to 10 by default. There have been several reviews of this figure and attempts for deriving an appropriate default figure for the VDF (please refer to Appendix E for an overview). Based on this, the PPR Panel considers the scientific basis and the number of experimental measurements too small to recommend a (conservative) default value at the moment.

In the proposed tiered assessment scheme, the VDF is not used. Instead, the exposure is calculated using Equations 1 and 13. At lower tiers of the assessment, the LAI is set to a conservative value of 1. At higher tiers, the exposure could be adjusted for the distribution of the substance over the leaves. If the substance is applied with conventional spraying equipment, one could expect that there is a non-even distribution of the areic mass of the substance on the leaves. This could be accounted for by using a so-called deposition distribution factor (DDF instead of VDF in order to avoid confusion between current practice and best practice). The value of this DDF is dependent on the spray distribution over the leaves and (most) relevant areic mass (for example the maximum areic mass on the various leaves of the plant or the average areic mass on the leaves). The DDF will be dependent on the spraying equipment, including for example nozzle types and operational parameters as tank pressure. The DDF should therefore be based on experimental evidence. The maximum of this DDF is 2 LAI, i.e. the two-sided relative surface area of the leaves. This maximum of the DDF is only reached when the spraying equipment is capable of distributing homogeneously over the total leaf area (double sided). By definition LAI indicates the single sided leaf area in relation to the field area.

The relative DDF may be different for in-field and off-field as the influence of the spraying technique on the deposition pattern over the leaves will be greater in-field.

6.8. Overspray

Deposition of substance on the NTA includes direct overspray and deposition of drift droplets,. This is relevant for each of the four situations discussed in section 6.7. The amount deposited on the NTA is directly related to the effective dose on the surface and the relevant projected 2D surface area of the NTA:

$$D_{NTA} = O_{NTA} A_{d,x} \tag{15}$$

where D_{NTA} (kg) is the amount deposited on the NTA, O_{NTA} (m^2) is the projected relevant surface area of the NTA and $A_{d,x}$ (kg ha^{-1}) is the dose as calculated for the four cases in the previous section.

The number of NTA is large. Moreover, as it might be necessary to consider life stages separately, it will be impossible to consider all species and life stages. One way to handle this could be the approach taken in the birds and mammals risk assessment methodology (EFSA, 2008) where (generic) focal species are defined as the risk assessment is performed using these species. This principle could be developed further in future guidance.

It seems that at least two generic focal species need to be defined, one for leaf-dwelling NTA and one for soil surface-dwelling NTA. The requirements for such species include at least the relevant (projected) body surface area to body volume ratio is large, for example the 90^{th} percentile of this ratio for species for which this ratio is known. Typically such a selection would point to a relatively small NTA. Focal species could then be selected from the species mentioned in the section on specific protection goals (section 4); for example a small mite or wasp as a leaf dwelling organism and a tiny spider as a soil surface dwelling organism.

In practice, in higher tier experiments, in might be difficult to distinguish between overspray, contact and possible other routes of exposure. There are, however, estimates of the exposure of NTA as RUD data available for different groups and sizes of insects (see Appendix 14 in EFSA (2008)).

The RUD is a measure of residues in insects after spray applications. It is expressed in mg/kg insect. It combines all different exposure routes but mainly it will be from direct overspray and contact on surfaces, if the estimates are for a moment in time shortly after application (thus excluding for example oral exposure). It is unknown however whether the RUD data are biased, i.e. whether the exposure of the individuals on which the RUD is based is representative of the exposure the population is the field. For example, the samples on which the RUD is based may have a relatively high abundance of dead NTA. It would need additional experiments with ranges of application rates to examine this in more detail.

In principle, it would be possible to compare RUD values with the corresponding endpoints from laboratory studies. The problem for the assessment is that the corresponding endpoints from the studies with NTAs are lacking. The endpoints are measured as application rates (EC_{50} expressed as g/ha) and not as a dose (mg/kg insect). At least the amount taken up by the NTA (the in-body concentration) should become available for establishing such relationship. The test also only measures toxicity from exposure to dry residues, which might influence the uptake and internal concentrations.

It could be investigated whether in-field RUD tests with (selected) species of NTA could form the basis of an assessment tier in future. This then could cover both contact and overspray exposure. As a consequence, the (exposure) schemes for contact and overspray should then be combined into one assessment scheme.

6.9. Oral exposure

Current NTA risk assessment does not include oral uptake explicitly. However, oral uptake was identified as a relevant exposure route for NTA (see section 6.5). As standardised tests for addressing this issue are missing, the relevant exposure endpoint has not been defined. The relevant exposure endpoint could be the total amount (per unit body weight) to which an organism is exposed, which can be translated into a concentration in food items that are consumed by the NTA if the total food intake is known. The assessment will be different for carnivores and herbivores as the food intake is totally different, but differentiation may also be required within these two main categories as feeding habits may be different for different life stages. Stages should be chosen with respect to specific feeding modes (e.g. leaves, nectar or predator) at the specific life stage of an NTA.

As with the overspray assessment, the PPR Panel recommends developing assessment methodologies for generic focal species and for focal species, as a first approach and to refine this methodology later, depending on experience and available data.

Exposure of the NTA will depend on the situation as described in section 6.7, i.e. whether the NTA resides in-field or off-field, on-crop or on-soil. The various exposure assessments, including the tiered approaches, are relevant for oral intake as well, depending on the type of NTA and its life stage. The following sections therefore assume that an areic mass, as obtained from the application regime for the various situations, can be related to food uptake.

The RUD values for different food items (see database for birds and mammals) could be used to estimate residues in food items in order to conduct an assessment for oral uptake. This could be useful for risk assessment for both leaf consuming and predatory arthropods. The problem here is that we need data on consumption patterns and rates of different NTA taxa.

6.9.1. Herbivorous non-target arthropods

Both leaf-dwelling and soil-dwelling NTA may be exposed via eating plant leaves or plant residues. Separate assessments are required for these organisms (or life stages of these organisms). It is not necessary to account for seasonal carry-over of PPP residues on plant leaves or plant residues (see section 6.7 for justification). It is envisaged that dissipation is in general fast enough to lead to negligible residues on leaf surfaces after a period of several months. However, it may be necessary to account for carry over of pesticide residues in soil and subsequent uptake by plants in the following year/growing season, for example for systemic substances with sufficiently long half-life in soil (see Bee Guidance, EFSA 2013b). Soil residues and plant uptake in crops in the following season could be calculated with the currently available soil exposure models (EFSA, 2012). In general, the amount taken up will be lower than 5 % of the remaining amount in the plough layer of the soil, dependent primarily on the sorption constant of the substance, but it may be the single source for the substance uptake by the NTA in certain conditions. As a first approach, the approach taken in the Bee Guidance Document (EFSA 2013b) could be followed. This approach includes not only exposure via leaf eating but also via feeding on nectar and pollen.

The total intake of a substance by an NTA is related to the food intake and the substance contents of the various food items (similar equations can be given for the both acute and the chronic situation:

$$intake_{NTA} = \sum_{i=1}^{n} ff_i content_i DFI_{NTA} \quad (16)$$

where *intake* (g) is the total amount of substance taken up by the NTA, NTA is the NTA under consideration, *i* a specific food item, *n* the total number of food items to be taken into account for the respective NTA, ff_i (-) the fraction of food item *i* in the diet, and DFI_{NTA} (g) the (daily) food intake by the NTA (i.e. the food intake over the time period that is relevant for the acute or chronic situation). Intake should be expressed in units that correspond with units used in relevant toxicity tests. Notice that the DFI values should correspond with the growth stage under consideration.

The number of food items to be taken into account may be dependent on the NTA and should be specified for the (generic) focal species. A conservative approach may be that (for the acute situation) the NTA feeds upon a single contaminated food item with a content derived from the appropriate tier (see section 6.7 on contact assessment).

The daily food intake is dependent on many variables, amongst others the NTA considered, the growth stage considered and environmental conditions (e.g. temperature). The *DFI* may be expressed in terms of amount of fresh food, dry food or energy and be based on the metabolic rate of the NTA, characteristic of the life stage. It is assumed that the assessment will be conservative if the food intake is calculated for optimal growing conditions. The necessary *DFI* may be obtained from specific allometric equations that relate energy or food requirement to the body mass of the NTA. *DFI* expressed in terms of energy is the most appropriate when more than one food item is taken into consideration and energy efficiencies vary over the various food items. Substance contents in the various food items should then be given in terms of amount per unit energy, taking into account the efficiency with which the NTA uses the energy content of the food item. So *DFI* may also be given as:

$$DFI_{NTA} = \sum_{i=1}^{n} EF_{NTA,i} DFI_{NTA,i} \quad (17)$$

where $EF_{NTA,i}$ is the energy efficiency for NTA and food item *i* and $DFI_{NTA,i}$ is the food intake of item *i* by the NTA in terms of energy content of the food item. Data on *DFI* and energy efficiency are scarce and should therefore be made available through target research. For the purpose of a guidance document, realistic worst-case values could be derived from the literature and used until better data have become available.

The amount of substance taken up with the food could be determined from RUD values for these food items and the appropriate dose. Appendices 14, 17 and 18 of the Birds and Mammals Opinion (EFSA, 2008) give information on the contents of substances in crop residues upon time after application, dependent on crop growth stage, per unit dose rate. These values could be used to derive appropriate input for the calculations, upon assumptions for mass of crop (fresh/dry) at the specific growth stage, LAI, etc.

The generic focal species to be defined would be a relatively small NTA as small animals usually have a relatively high energy requirement per unit body weight. Furthermore, the specific life stage of the generic focal species is a stage with a high growth rate, maximising the energy requirement and therewith the exposure. A focal species might probably be a relatively small caterpillar as caterpillars consume a high amount of plant material in a short period of time. It is expected that assessments for caterpillars will cover also other herbivorous NTAs.

6.9.2. Carnivorous non-target arthropods

Carnivorous NTA feed on other animals, for example other NTA. Exposure of these NTA is via contaminated food items. Oral uptake is then related to the exposure of the prey organisms and the number of prey organisms that are consumed over the relevant period of time. A formula similar to that in section 6.9.1 can be derived, taking into account various prey animals and their substance contents (including energy efficiencies if necessary). It will be possible if the source of contamination of the prey animal is by overspray or oral intake, more difficult if possible at all for prey exposed via contact (as in general this route of exposure is not taken into account in experiments (not in the data requirements).

Both, larvae and adult carnivorous NTAs are exposed through their food and contact. Both life stages therefore need to be addressed.

6.9.3. Other non-target arthropods

Some NTA use plant fluids as source of energy (for example aphids) or have quite specific ways of taking up energy (e.g. Lepidoptera feeding on pollen and nectar). If considered necessary, the exposure of other NTA can be given in formulae such as the one in section 6.9.1, with assumptions about concentrations of the substance in the plant fluids, nectar and pollen. These assumptions could be based on data available in the bee guidance document (EFSA 2013b).

6.10. Landscape scale dynamic exposure modelling

For mobile species and species which live in different exposure compartments during their life cycles, spatio-temporal patterns of exposure can be critical in determining population-level impacts. This will particularly important when combined with animal behaviour, e.g. a pollinator choosing a sprayed field for foraging, or a carabid beetle spending larval and pupal stages in soil in in-field, but as an adult moving between in field and off-field. In these cases exposure needs to be calculated dynamically in space and time and linked directly to simulation of phenology and spatial dynamics of the NTA. In general this can be done for exposure using the same approaches described for the field exposure assessments described above (sections 6.1 to 6.9). However, this is a non-trivial exercise and currently there are few models available that can integrate these factors.

Developing a dynamic simulation of exposure therefore accomplishes two goals. The first is the provision of spatio-temporally varying stressor as part of a dynamic environmental scenario for the effect models. The second is that the dynamic simulation could be used to generate the range of exposure profiles predicted for real landscapes, and thus be able to place any field experiment in the context of this distribution (e.g. identify if the field experiment represents the 90^{th} percentile of exposure concentrations).

To link to the effects models the dynamically simulated exposure must be able to represent the pesticide applications in space and time at a resolution commensurate with the detail required to run the NTA model, but also must be capable of doing this at the scale necessary to represent landscapes of sufficient size to avoid spatial structure bias. This means that the calculations for exposure must be handled very efficiently. For instance a 10×10 km^2 landscape at a 1-m resolution has 100 000 000 cells that require independent calculations for pesticide fate typically at a daily time step.

Capabilities required are:

1. That the scale of landscape considered is large enough to avoid spatial bias and large enough to adequately represent source–sink dynamics and spatial interactions between sub-populations (typically this is expected to be at scales of 100 km^2 or greater).
2. The representation of differential application timings, frequencies and rates dependent upon field or farm and region.
3. There needs to be the facility to provide rainfall and temperature varying at the same time-step as the model.
4. Vegetation growth and LAI needs to be dynamically simulated in order to drive deposition and wash-off fate equations. This should take into account different crop and off-crop vegetation structures.
5. The following need to be simulated at a high spatial resolution (e.g. 1 m^2) to ensure field boundary vegetation can be particularly important off-crop habitat for NTAs to be accurately represented:
    a. The representation of drift from in-field to off-crop areas based on drift calculations from section 6.7.3, but including wind direction to prevent simultaneous drift in opposite directions. It is important that the wind direction varies realistically for the landscape under consideration. This is because the combination of spraying on days with different wind directions gives the possibility for double-drift to off-crop areas. Without this the worst-case tails of the exposure distribution will be missing.
    b. Dynamic simulation of interception of the pesticide spray by crop canopy (and off-crop vegetation) to produce a vegetation concentration following section 6.7.1. This presupposes the capability to independently model vegetation growth within fields and off-crop areas, i.e. different fields with the same crop will have different crop vegetation profiles at the same instant in time, perhaps owing to different sowing dates, or fertiliser applications. Note here that this method does not use the look-up tables described in section 6.7.1, but utilises the LAI generated for each vegetation structure for each time-step. This is necessary to correctly generate the variation in exposure in time and space.
    c. Simulation of environmental decay of vegetation residues following section 6.7.1.
    d. Simulation of wash-off of pesticide from vegetation to soil surface following section 6.7.1.
    e. Simulation of deposition of the pesticide to the on-soil compartment following section 6.7.2.
    f. Temperature related environmental decay of pesticide residues in the soil compartment following section 6.7.2.
    g. Pesticide residues at both on-soil and vegetation compartments must be available for modelled NTAs to link effects with exposure modelling.
    h. The order in which the fate processes are considered is important, as decay rates differ between compartments. To prevent bias, it is suggested that chemical processes are considered first (i.e. decay), then physical processes, i.e. wash-off in this case. No decay or wash-off should be carried out on the day of application, unless smaller time-steps than one day are being considered.

6.11. Review of the vegetation distribution factor for use in non-target arthropod risk assessment

The areic mass of pesticide spray on a canopy can be expressed as mass per unit area leaf surface. For a defined vegetation structure, the surface concentration of substance per unit area leaf surface as a proportion of the areic mass of substance on the canopy is called VDF. The current risk assessment for

NTAs in off-field areas according to ESCORT 2 (Candolfi et al., 2001) refines tier 1 exposure using a so-called VDF to relate the results of the in-field risk assessment to the off-field environment where (leaf-dwelling) organisms are assumed to be exposed less because of a different vegetation structure and a larger dilution of exposure than in the field. By diluting foliar exposure in lower tier effect studies, this factor can be omitted.

Current guidance indicates that this figure of the VDF of 10 is considered unreliable. Therefore, 'more appropriate data should be used as soon as they become available' (please refer to SANCO/10329/2002 rev 2 final) and the question whether the current VDF can be considered appropriate has been also raised by Member State in the public consultation on the existing Guidance Documents on Aquatic and Terrestrial Ecotoxicology under Directive 91/414/EC1 (refer to EFSA 2009). There have been several reviews of this figure and attempts for deriving an appropriate figure for the VDF (please refer to Appendix E) in NTA risk assessment. The derived figures for a VDF all express as the surface concentration of substance per unit area leaf surface as a proportion of the areic mass of substance on the canopy.

A stated previously, in current higher tier effects studies both the exposure assessment and the effects assessment are refined, see section 7, section 7.2.4. As exposure is not measured in these studies, implementation of the criss-cross model in full should currently be done with due care because this would lead to possible double counting of refinements (e.g. refining exposure in both the exposure assessment and implicitly in such legacy field studies). Such a combination of lower tier effects assessment and higher tier exposure assessment according to the criss-cross model is possible. However, this requires that exposure in the reference tier used for calibration of risk assessment is considered, i.e. relevant factors such as LAI measured are in the field and taken into account when linking exposure and effects. As this was not considered when calibrating current risk assessment using (semi-)field experiments, current tier 1 risk assessment cannot be used to extrapolate effects to other exposure situations. Consequently, none of the current figures for the VDF (expressing the difference between the areic mass of substance on the canopy and the surface concentration of substance per unit area leaf surface) can be used on the basis of the calibration of current risk assessment. When developing the future risk assessment for NTAs, process-knowledge should be used to link lower tier exposure and effects to higher tier exposure and effects (field studies) assuming relevant factors are measured in the field.

7. Effects assessment for non-target arthropods

7.1. Introduction

The new demands of Regulation (EC) No 1107/2009 need to be implemented in the existing risk assessment. It needs to be ensured that the current methodology for effect assessment is sufficient to determine whether following the intended use of a PPP the specific protection goals defined according to the legislative demands are met. The group identified two primary aspects of the current effect assessment of NTAs that need to be revised considering new information that has become available since publication of the (draft) guidance document for terrestrial ecotoxicology was in 2002 (see e.g. section 2).

1. As specific protection goals require maintenance of particular ecosystem services within cropped fields as well as in off-crop areas, it needs to be made sure that the current (higher tier) methodology is able to detect relevant effects on ecosystem services and lower tier methods are available that allow a robust prediction of field effects.
2. The present risk assessment scheme does not take into account the effects of configuration of agricultural fields and spatial dynamics on NTA populations. The working group proposes a way forward how this essential aspect necessary for realising the specific protection goals could be addressed using individual based modelling based on what is currently possible with existing models as described in section 5.3. It needs to be made sure that appropriate data are available to describe the toxicity on NTAs as input for the proposed modelling approach.

For this reason, the following section reviews and gives recommendation on methodology for effect assessment at the local scale and characterises the appropriate toxicity input that would be necessary to realise the proposed landscape scale assessment.

7.2. Assessment of effects on non-target arthropods at the local scale

7.2.1. Introduction

The toxicity input for the tiered risk assessment approach for assessing effects at the local scale are discussed here in two main parts:

- Laboratory studies for first tier risk assessment. A limited number of species should be tested. Tests should be carried out for every substance and should be relatively simple. Chronic effects should be assessed as part of the first tier in order to detect the occurrence of delayed effects of acute exposure. The tests should allow a robust prediction of adverse effects in the field.
- (Semi-) field studies for use in higher tier risk assessment. Field studies ideally represent the most sophisticated testing approach available for assessing the effects of PPPs on NTAs at the local scale (reference tier for immobile species at the local scale, see also section 5.3) but also can deliver important information on mobile species. Depending on the quality of field studies, their results may be used as part of the toxicity summary in risk assessment and/or to justify lowering assessment factors.

For NTA risk assessment, an assessment factor expresses and addresses uncertainty about the ratio between a suitable summary of toxicity measurements and the application rate for the product leading to acceptable effects in the reference tier and may also include quantitative expert judgements about additional uncertainties that are not addressed by the reference tier. The assessment factor will depend on which toxicity summary is being used and should change as more species are tested or when one moves to a higher tier. How such an assessment factor is obtained and under which conditions it can be reduced is explained in section 5.5. How exactly the assessment factor changes if additional toxicity measurements become available will be elaborated when developing the guidance document.

The approach adopted to revising the current test procedures was as follows:

- An overview of the existing methods to test the effects of PPP and their active ingredients on NTA was compiled. According to specific criteria, suitable laboratory toxicity tests for the first tier of risk assessment were identified.
- Furthermore, the evaluation of (semi-)field methods for higher tier risk assessment including the assessment of uncertainties is described. For both laboratory and (semi-)field methodology gaps and recommendations for further development are identified.

7.2.2. Overview on available test systems

Laboratory studies for first tier risk assessment
As a first step in the choice of suitable test methods, a list of available laboratory and so-called extended laboratory test methods was compiled (see Appendix G). It should be noted that the information presented has been compiled by the working group experts from public literature but has no claim of being exhaustive.

Appendix G lists IOBC published standard test methods 'to evaluate side-effects of plant protection products to non-target arthropods', as recommended by current guidance. Furthermore, Appendix G provides test protocols that have been submitted during peer-reviewed PPP active substance evaluation in the EU or identified from open literature and considered to address traits, taxonomic groups, endpoints or exposure paths missing in the standard IOBC test methods.

The existing test systems do not follow a uniform design for the different NTA species. The substrate on which the organisms are exposed, the life stages exposed, the duration of exposure and the determined endpoints vary greatly between the different study protocols. In order to facilitate an overview, test systems are listed by species tested. Details on the test design, assessment endpoints and exposure routes (e.g. oral, contact, duration, substrate type) are reported. Some general comments on the tests tabulated follow.

As can be seen in Appendix G, tests with NTA are performed on a variety of substrates. Moreover, the routes and intensity of exposure also vary depending on the mode of application of the PPP at the start of the test. In some tests, NTAs are exposed to dried PPP residues, in others the animals are over-sprayed during PPP application. The definition of 'extended' laboratory tests vs. standard laboratory tests is based on the use of more natural substrate (often leaves) compared with sprayed glass plates.

The endpoints to be assessed in the tests are also very variable. It is important that both acute and chronic endpoints can be assessed in different studies. Several tests do address the reproduction of adults following exposure to PPP for a short time although not to exposure to PPP through the period of reproduction.

The life stages of the species selected to be exposed are also very variable. The larvae and adults of holometabolous arthropods generally prefer different habitats and have divergent feeding modes, so the choice of the life stage to be tested can strongly influence the outcome of the test.

(Semi-)field studies for use in higher tier risk assessment

The available test systems for semi-field and field tests are listed in Appendix H. This overview is adapted from Brown et al. (2009). In Appendix H, a distinction is made between semi-field and field studies. Semi-field studies are in this overview tests where the mobility of the test species is artificially limited using, for example, enclosures. In contrast, field studies are conducted on plots, mostly in-field, where species can freely move within the plot and out of the plot. These can be targeted studies in which one species is studied as well as studies of the whole NTA community present in the plots. For reasons of completeness, some modelling studies are also included.

Appendix H briefly summarises for the available test systems: test design, endpoints studied, exposure regime, eventual caveats and a reference to the guidelines. An overview of studies that have been submitted during peer-reviewed PPP evaluation is not provided, but the study types are represented by the studies in Appendix H.

Appendix H shows that there is high variability in testing methods. Single species field studies are submitted for the peer review process in a standardised way. In semi-field experiments, leaf or soil dwelling NTAs are often exposed in a similar manner as in (extended) laboratory experiments with the difference that a part of the experiment takes place in the field (on the soil in the treated field or on crop plants). The exposure in experiments is not influenced by migration of organisms, as organisms are usually enclosed in the treated area after application of the PPP but can be influenced by climatic conditions. Species tested are standard laboratory test species or species with similar traits. The life stages of the species selected to be exposed are also very variable. The larvae and adults of holometabolous arthropods generally prefer different habitats and have diverging feeding modes, so the choice of the life stage to be tested can strongly influence the outcome of the test.

Appendix H shows that also for field studies a high variability exist concerning: aim of the study, duration, community studied, scale, etc. Guidance for evaluating these types of field studies is developed by De Jong et al. (2010). This guidance might be of help to enhance not only the way these studies are evaluated, but also to increase uniformity in the test design.

Exposure in the field studies is mostly the rate according to the intended use, or in the case of drift studies some of the predicted drift rates. Exposure is expressed as dosage per surface area. Applied rates are often checked by determining the amount sprayed or by determining the deposition, e.g. using water sensitive paper. The actual exposure is not measured in any of the field experiments.

7.2.3. Laboratory studies for first tier risk assessment

Criteria for choice of first tier test systems

As specific protection goals require maintenance of particular ecosystem services within cropped fields as well as in off-crop areas, it needs to made sure that the current (higher tier) methodology is able to detect relevant effects on ecosystem services and tier 1 test systems are available that allow a good prediction of field effects. Tier 1 test systems should therefore be designed to prevent the missing of unacceptable effects of intended uses of a PPP on ecosystem services defined to be important in the agricultural landscape. Unacceptable effects could be missed if the required standard test protocols do not test (i) susceptible taxonomic groups, (ii) susceptible life stages or developmental processes targeted by the PPP or (iii) relevant routes of exposure.

In order to increase the reliability of tier 1 risk prediction, the working group recommends that the following selection criteria for tier 1 test systems should be applied:

- Test species should be *spread over taxonomic groups* expected to show different sensitivity and *toxicologically sensitive indicators* should be chosen.
- Test systems should address *sensitive life stages* of NTAs.
- Test systems should cover *all exposure routes* that are most relevant for the species tested (oral, contact, overspray). Test systems should also cover those exposure routes that are particular to the active substance/PPP to be evaluated, e.g. spray versus solid formulations or soil fumigants.
- Test systems should allow the detection of effects resulting from *specific/novel modes of action* (e.g. insect growth regulators, insect feeding inhibitors).

As a basis for selection of tier 1 test species, Table 10 lists available tier 1 test systems classified according to the main selection criteria listed above (for detailed description of laboratory and extended laboratory methods please refer to Appendix G). It is neither feasible nor necessary to test all existing taxonomic groups of NTAs to achieve a reliable prediction of risk for NTAs from the use of PPP. However, if from the sensitivity analysis of available test outcomes it is not possible to decide whether a species in a standard test system can serve as a sensitive surrogate for several others, then test species representing key drivers performing specific ecosystem services (pollination, food web support and pest control, aesthetic value) should be preferred for testing. Hence, Table 10 additionally includes key drivers for specific protection goals identified in section 4. Note that the selective table only lists groups that were identified as key drivers for the specific protection goals pollination, food web support and pest control. Biodiversity as a specific protection goal concerns the complete variety of species and is therefore not considered as criterion for selection of tier 1 test systems.

Under which circumstances toxicity as measured in (semi-)field experiments can serve as input for the landscape scale assessment will be discussed in section 7.3.

Do the available test systems match the criteria set?
On the basis of the proposed criteria for the choice of tier 1 test species and the crosscheck performed with key NTA in Table 10, the working group concluded that the important parameters are covered by current test methods as described below.

Testing different taxonomic groups and sensitive indicators of these groups:

Current standard test protocols include tests with parasitic wasps (Hymenoptera), lacewings (Neuroptera), flower bugs (Heteroptera), hoverflies (Diptera), ladybird beetles, ground beetles and rove beetles (Coleoptera), predatory mites (Acari) and wolf spiders (Araneae).

An example of a major group that is missing and yet considered to be a key driver for important ecosystem services as described in section 4 and summarised in Table 10 above is money spiders (Linyphiidae). This group contributes to the ecosystems service pest control as well as food web support and could be more susceptible to pesticides than wolf spiders because of their smaller size (larger surface area to volume ratio).

Currently, both a leaf-dwelling ladybird beetle and a ground beetle (*Poecilus cupreus*) are tested, whereas test systems for other groups of beetles such as rove beetles, weevils and leaf beetles are missing.

An important group for which standard test protocols are missing is Lepidoptera. The working group considers Lepidoptera to be important drivers for the ecosystem services pollination and food web support (see section 4). For Lepidoptera larvae, there are test guidelines available in the open literature that could be adapted for oral and contact exposure (see Appendix G). Grasshoppers, plant hoppers and several groups of flies and wasps are also missing from the standard test systems.

The choice of *Poecilus cupreus* as a representative test species for NTAs has been criticised, as it is a burrowing beetle that also has a relatively thick cuticle (Brown et al., 2008). Consequently, owing to its behaviour and structure, this species will be less exposed than species that remain on the surface and/or have a different physiology. There is a similar criticism of the tested group of spiders of the genus *Pardosa* (wolf spiders) as they are larger and more robust than other families and also hide in soil.

Testing chronic effects, sensitive life stages and developmental processes on which the effects could differ systematically

Some stages of arthropods are more sensitive to PPPs than others for a variety of reasons that include physiology, metabolic rate, activity and feeding preferences. Current test methods with Hymenoptera, Coleoptera and Araneae are performed with adults exposed to PPPs, whereas for the other available tests, larvae or nymphs are exposed. This is the case for, for example, *O. laevigatus*, *T. pyri*, *C. carnea*, *C. septempunctata* and *Syrphus corollae*. Tests with Lepidoptera species are performed with larvae. In tests with *C. septempunctata* and *S. corollae*, effects of an initial exposure on pupation/edcysis are assessed. In the test with *C. carnea*, the reproduction of the emerged adults is also assessed, making it the only test species for which both acute effects on adults and reproductive effects are assessed. None of the current tests, however, takes into account chronic exposures and allows for assessment of PPPs on the whole life history of NTAs.

Tests using long-term exposure have not traditionally been part of lower tier tests. The working group believes that this extension is needed. While the major goal of ERA is to provide means for a long-term protection of populations and ecosystems against toxic chemicals, most current ecotoxicological bioassays actually do not address that question properly. The most common method is estimating acute effects, such as LC_{50} or LT_{50}, in short-term tests. The tests that measure some more chronic effects, e.g. on fecundity, body growth rate, etc. are still scarce. Among 44 ecotoxicological tests reviewed by Léon and Van Gestel (1994), 25 were on invertebrates and are thus, at least in part, relevant to this opinion (although the majority was on soil invertebrates). The duration of these tests vary from 2 to 63 days, most representing the lower end of the interval. Comparing this to the life span of tested invertebrates, it is apparent that the majority of tests cover only a minute part of the life cycle of a test organisms (Table 10; Laskowski, 2001). Especially noteworthy in this table are such invertebrates as earthworms, honey bees, rove beetles, ground beetles and isopods, for which the discrepancy between test duration and their life span is particularly striking. Note that most of them are typical non-target terrestrial arthropods.

Table 10: Organisms used in ecotoxicological tests, test duration and approximate life span (after Laskowski, 2001)

Organisms	Test duration	Life span
Parasitic wasps	2–18 days	3–4 weeks
Honey bees	2 days	Few weeks–few months
Earthworms	2–8 weeks	Few months–few years
Spiders	2–14 days	ca. 1 year
Enchytraeids	4–9 weeks	ca. 10 weeks
Isopods	8 weeks	1–2 years
Springtails	4–9 weeks	Few months
Ground beetles	6 days	1 –2 years
Rove beetles	15 days	ca. 1 year

As shown by Laskowski (2001), too short ecotoxicological tests may result in serious underestimation of some effects, while overestimating others. In the cited work, the author tested toxicity of two chemicals of different modes of action, namely the insecticide imidacloprid and a heavy metal (cadmium). Briefly, the results showed that short-term tests on the pea aphid (*Acyrthosiphon pisum*) overestimated population-level effects of the insecticide and heavily underestimated chronic effects of cadmium. Laskowski (2001) concluded that short-term assays neglect the accumulative nature of some chemicals, and possible accumulation of their toxic effects; they usually also neglect any effects other than on mortality and fecundity (e.g. growth rate, aestivation, etc.), only take into account a small part of the whole life history of an organism and, as a result, do not allow for drawing meaningful conclusions about population dynamics under toxic stress. Underestimating toxic effects, should always be avoided because this leads to underestimation of risk. Although the underestimation of population-level effects was shown by Laskowski (2001) on cadmium, which might seem of no interest in the opinion on effects of PPPs on NTAs, it should be remembered that PPPs is an extremely diversified group of chemicals, including, besides organic insecticides, a vast number of chemicals of different modes of action. For example, the fungicides most commonly used in vineyards are copper compounds, chemicals whose toxicity is based on metal ions similar to cadmium. Organic chemicals may accumulate in their bodies and trigger damage that also accumulates in time.

Moreover, some effects may be secondary in nature, connected with changes in food quality. Such effects are not well documented although recent work by Stark et al. (2012) on Behr's metalmark butterfly (*Apodemia virgulti*) showed that they are not only possible but actually happen in the field and can cause significant damage to NTAs. The authors showed that herbicides used to remove invasive weeds from the dunes reduced the number of adult butterflies that emerged from pupation by as much as 24–36 %.

Recently it has been shown that also some insecticides may exhibit delayed and time-cumulative effects in NTAs. Rondeau et al. (2014) estimated that imidacloprid at a concentration of 0.25 μg/kg would be lethal to a large proportion of overwintering bees, and similar delayed effects are expected also in ants and termites. The authors concluded that 'chronic tests for pesticide toxicity to pollinators should be extended to 30 days or more and use time-to-effect measurements'. It should be stressed in this respect that Rondeau et al. (2014) considered only effects on mortality, and not on fecundity. Thus, the overall delayed effects on population dynamics are expected to be even more serious. Such studies support the need to implement long-term tests in ERA, based on chronic exposure of invertebrates to potentially toxic chemicals. Only such studies will make possible predicting population-level effects of PPPs under realistic exposure scenarios. The NTA working group stands on the position that development of such chronic tests, covering a significant part of a species life history and allowing for assessment of reproductive effects of PPPs, should be one of the most crucial and urgent tasks in modern ERA. Recognising that this may result in more time-consuming and costly tests, the working group suggests that chronic and reproductive effects of PPPs could be performed on species with relatively short life cycle and easy to maintain in a laboratory. A good candidate would be a multi-generation test with *Drosophila* species, such as *D. melanogaster*. Both the acute and chronic effects can be studied with such a test in short time. The additional advantage of *D. melanogaster* is that with such a well studied species, with known genetics, it would be possible to select appropriately sensitive strains.

Testing relevant exposure routes

Standard test methods after IOBC guidelines (Candolfi et al., 2000) are so called contact-toxicity tests. Tests are either performed by spraying glass plates or by spraying artificial/natural soil with PPP. For tests performed on glass plates, organisms are exposed to freshly dried residues after application. In tests performed on moist artificial or natural soil, the spiders of the genus *Pardosa* or the ground beetle *P. cupreus* and their food are oversprayed, whereas the rove beetle *Aleochara bilineata* is exposed to freshly sprayed residues.

In regulatory practice, procedures from IOBC guidelines (Candolfi et al., 2000) testing contact exposure on glass plates are often adapted by spraying the test item on a natural substrate (i.e. leaf disks) or whole plants to refine exposure conditions (so-called 'extended laboratory tests'). These extended laboratory tests only differ from IOBC standard test protocols in the substrate on which the exposure phase takes place. Duration of exposure and endpoints assessed are usually the same. According to Grimm et al. (2001), tests with *Typhlodromus pyri* on sprayed leaf disks are often performed according to a method adapted from Oomen (1988). Tests with *Aphidius rhopalosiphi* on sprayed plants are mostly performed according to an unpublished draft guideline by Mead-Briggs and Longley (1997). This guideline was edited and republished in a more standardised form by Mead-Briggs et al. (2010).

Further (extended) laboratory test protocols have been submitted for registration of PPP active substances and are evaluated in Draft Assessment Reports (DARs) prepared by member state authorities. However, except for Syrphidae, taxonomic groups and traits, endpoints as well as exposure paths represented by those species are already represented by the species shown in Appendix H for which standard guidelines are available. Syrphidae represent Diptera, which constitute an important order that is not covered by IOBC standard test protocols.

Standard test protocols performed on glass plates or leaf material only test contact exposure towards freshly dried residues

For Lepidoptera larvae, there are test guidelines available that could be adapted for testing oral exposure to PPP residues on food items. Test protocols are from open literature.

Bundschuh et al. (2012) tested the sensitivity of grasshoppers (*Chorthippus* sp.) in plastic boxes where either the inner surface of the test vessels, grass provided as food or the surface + food was sprayed with plant-protection products. The results indicated that exposure *via* contact only (surface) yielded the most sensitive endpoints. A comparison of the LR_{50} values from standard toxicity studies with those of *A. rhopalosiphi* and *T. pyri* on glass plates covered the toxicity towards *Chorthippus* spp. for 4 tested insecticides with a different mode of action. However, if the exposure conditions in toxicity studies are refined (tests performed on natural substrate such as plant material), toxicity towards species mainly exposed via contact might underestimate the acute toxicity towards herbivores as *Chorthippus* spp.

For the available standard protocols, the reproductive phase of adults always takes place in an uncontaminated environment. For *C. septempunctata* and *Syrphus corollae*, individuals are still exposed until the end of pupation/edcysis. *Chrysoperla carnea* is also exposed until it reaches the pupal stage.

Testing specific modes of action

Vogt et al. (2000) pointed out that ‘for testing side effects of insect growth regulators (IGRs) or PPPs with similar mode of action and for other novel type PPPs other species than *Aphidius* spp. and *T. pyri* have to be chosen. As an example, coccinellids (*C. septempunctata*) or the green lacewing (*C. carnea*) proved to be suitable test organisms for testing IGRs (Hassan et al., 1991, 1994; Stark et al., 1999; Vogt, 1998; Vogt et al., 2000).’ The available test systems with soil dwelling wolf spiders and ground beetles allow assessing the effects on the feeding behaviour that could be particularly relevant for feeding inhibitors.

Conclusion
Applying the selection criteria for tier 1 test systems (take in to account spread over taxonomic groups, toxicologically sensitive indicators, sensitive life stages/developmental processes, relevant exposure routes) to Appendix G, based on the test systems that are currently available or could be practicable in the near future, the working group recommends available tests with the following species to be carried out already at tier 1:

Aphidius rhopalosiphi	leaf-dwelling parasitoid (Hymenoptera)
Typhlodromus pyri	leaf-dwelling predator (Acari)
Coccinella septempunctata	mobile leaf-dwelling predator (Coleoptera)
Lepidopteran larvae	leaf-dwelling herbivore (Lepidoptera)

This includes an oral toxicity study with Lepidopteran larvae as representative of herbivorous species, which is considered of high importance by the working group. Test protocols could be adapted for this purpose from existing test protocols available in literature.

Additional species might need to be chosen, depending on the mode of action, application method or crop type.

We note that exposure to dried residues on glass plates reduces variability compared with a substrate such as leaves (which are both variable and tend to dry out over time). As a general aspect regarding test design, it would be also desirable that exposure in the test systems is controlled especially during the reproductive phase of the test organisms.

The working group strongly recommends that effects on reproduction should be tested at tier 1; if this is possible according to the test design, reproductive effects should be assessed regardless of the substrate on which exposure takes place.

In the list of recommended tests, no exposure via overspray is included because the available systems to test this exposure path were not considered appropriate owing to shortcomings regarding, for example, test design or species/life stage tested. The working group considers the toxicity endpoints derived from the tests with bees (fresh residues) to be a possible surrogate for the overspray exposure route. Please refer to section 6 for calculation of exposure for NTA species.

7.2.4. (Semi-)field studies for use in higher tier risk assessment

Specific protection goals require maintenance of particular ecosystem services within both in field and off-field. When interpreting field studies it needs to be made sure that they are able to detect relevant effects on ecosystem services. Depending on the quality of field studies, their results may be used as part of the toxicity summary in risk assessment and/or to justify lowering assessment factors. The two main uncertainties that a field study could address for immobile species at the local scale are:

1. Uncertainties with regard to laboratory data to field impact, extrapolation intra- and inter-species variations (biological variance).
2. Uncertainties with regard to short-term to long-term toxicity extrapolation and assessing recovery.

What aspects are to be considered when interpreting the results of field studies in order to address these uncertainties is described in detail in the below sections.

Furthermore, based on current regulatory field studies, it is difficult to separate the influence of exposure and different biological aspects on the outcome. As discussed in section 5.3.1, to characterise uncertainties with regard to exposure in field studies and determine the worst-case character of exposure, a measurement of the ‘ERCs’, i.e. the concentration(s) actually eliciting the observed effects in field experiments would be required. It still needs to be elaborated what the ERCs are. Measurement of ERCs in combination with a dose–response design are needed to determine the consequences of risk mitigation measures on direct, local PPP effects in off-field habitats or help extrapolating to different in-field situations.

As described in section 5.3.1, field studies represent the surrogate reference tier for addressing different uncertainties with regard to effects of PPP use on NTA species at local scale. It was recognised that current best practice field studies could address uncertainties with regard to the extrapolation of laboratory data to field impact, for example:

- inter-species variations in sensitivity
- effects on different life stages
- exposure in the actual field situation
- interactions between and within species.

As summarised in section 7.2.2, current regulatory field test protocols are mostly performed in-field (recently also in off-field environments) on replicated plots and study effects on ‘natural’ NTA populations present at the time of the experiment. Plot size is relatively small and for studies focusing on in-field risk assessment, usually a limited number of dosages is tested (1 or 2). To determine the consequences of risk mitigation measures on direct, local PPP effects in off-field habitats, off-field experiments usually test a dose–response (see e.g. Appendix in De Jong et al., 2010). When a valid field study is conducted, the uncertainty of its outcome with regard to the potential outcome under actual field situation should be assessed. Uncertainties that should be checked could be connected to, for example:

- product
- dosage
- method of application
- time, frequency and interval of application
- type of ecosystem (depends on abiotic factors as soil, climate and on composition of non-target groups)
- location and isolation of the test system
- region
- history of the test system
- crop and crop-stage
- in-field and off-field.

These aspects are discussed in detail in De Jong et al. (2010). The guidance document of De Jong et al. (2010) indicates which taxa should be present in arable crops, orchards and off-crop field study as representative community of these types of agro-ecosystems (see Appendix F). The possibility to extrapolate results of field studies between regions is also discussed extensively in ESCORT 3 (Alix et al., 2010). A study by Aldershof and Bakker (2010) for insecticides shows that the differences in the effects of insecticides in different regions might be limited.

It is particularly important that species that are considered particularly vulnerable according to their traits are present in field experiments (see also section 5). Vulnerability of the NTA species, as well as how representative the NTA species is, should be carefully checked, in order to be sure that the specific protection goals are covered in the specific field study (see section 4). It should be considered if other endpoints can be derived from field studies (e.g. chick-food index; DEFRA, 2007).

It is important to understand the power of various field study designs to detect effects at magnitudes relevant to the specific protection goals. Effects from multiple sampling times in field studies are summarised using either the Henderson-Tilton (1955) or the Abbott (1925) calculation. Consequently, the final effect measurement is a complex quantity which is influenced by the population dynamics within the control and treatment groups. Without in-depth theoretical understanding of the population dynamics, it is extremely challenging to make theoretical power calculations.

If raw data for sub-plots are available for a field study and one assumes independence of sub-plots, it is possible to assess the statistical significance of the difference between treatment and control at individual time points. By making additional assumptions about dynamics, such as those underlying the Henderson-Tilton calculation, one might arrive at an overall assessment of statistical significance for a field study. However, unlike aquatic mesocosm studies for example, the sub-plots are not truly independent in a NTA field study. This is due to movement between sub-plots by mobile species and holds even if such species are not the focus of the study.

Given raw data for a substantial number of representative field studies, it may be possible to arrive at an empirical assessment of the overall power of a particular field study design. However, power for any particular study will depend on various factors, including initial abundances of species and environmental conditions during the study, and it may be difficult to assess the influences of such factors in an empirical power assessment.

For the *T. pyri* field study protocol, Blümel et al. (2000b) made an empirical study of the pooled results of a number of studies. They examined the relationship between magnitude of overall effect, as summarised by Henderson-Tilton or Abbott, and the P-values from t-tests for difference between treatment and control groups at individual time-points. Although of interest, their study is not a power calculation. Moreover, any wider interpretation would rely heavily on the representativeness of the studies they considered in relation to the assessment of a new PPP use. As mentioned above, it is probable that power is dependent on the initial mite density and this was not considered in their analysis. The empirical evidence in Figures I3 and I4 of Appendix I, for a substantial number of PPPs, is that there is considerable variation in measured effect for studies on the same product at similar application rates.

The use and the magnitude of an assessment factor on the outcome of field studies will be considered further in the guidance document.

Note that the available study design with small plot set-ups can also deliver useful information on PPP effects on mobile species (please refer to section 5.3.1)

As discussed in section 5.3.1, for non-mobile species, field studies could also uncertainties with regard to short-term to long-term toxicity extrapolation and assessing recovery, for example:

- effects of chronic (repeated) exposure

- indirect effects (e.g. due to food loss)
- population dynamics
- interactions between and within species.

As study plots are relatively small in most of the field study designs, recovery can potentially occur from other plots or from the off-crop environment (recolonisation). Therefore, processes demonstrating recovery of NTA species in this type of field studies are only meaningful for the present non-mobile species. As discussed in the previous paragraph, it is important that species present in field experiments are representing the characteristic fauna and traits that make species more vulnerable (please refer to section 5).

When assessing long-term effects, compound properties related to the dissipation of the substance under the respective conditions are considered to become increasingly important.

As recognised in section 5.6.4, field experiments for assessing the long-term effects on NTAs from use of a single PPP need to account for multiple stress caused by normal agricultural practice (e.g. sequential use of different pesticides) that might hinder recovery. The duration required to assess whether recovery of non-mobile NTAs occurs also depends on the life cycle of the NTA studied.

As discussed in section 5.6.4, there is evidence that in cooler climates NTAs tend have fewer generations per year than under warmer conditions. This could significantly change the recovery capacity and hence need to be considered for example when extrapolating results from a study to different climatic conditions.

General recommendations for further development of existing field study methods

- To characterise uncertainties with regard to exposure in field studies and determine the worst-case character of exposure would require the measurement of the 'ERCs', which still need to be elaborated.
- Measurement of ERCs in combination with a dose–response design this could prove helpful to determine the consequences of risk mitigation measures on direct, local PPP effects in off-field habitats or help extrapolating to different in-field situations.
- As performing large plot experiments at the desirable local scale to exclude the influence of recolonisation for most of the less mobile species is desirable but might be not practicable, a possible compromise would be to conduct a field study with a limited number of large plots, which is combined with a number of smaller plots.
- It is clearly important to understand the power of various field study designs to detect effects at magnitudes relevant to the specific protection goals. Without in-depth theoretical understanding of the population dynamics, it is extremely challenging to make theoretical power calculations for the Henderson-Tilton or Abbott summaries of effects, as both use data from multiple sampling times.
- For interpretation of field experiments, not only representativeness but also vulnerability of the species present is important.
- It should be considered whether additional measures can be derived from field studies (e.g. chick-food index; DEFRA, 2007).
- De Jong et al. (2010) describe in detail the assessment of higher tier studies, and the requirement for a sound study. These requirements can be used for the design of field studies as well and will be developed further in the guidance document.

Other higher tier methods

Apart from field studies, there are other methods for refinement of lower tier uncertainties. One useful possibility is monitoring of actual NTA populations in the field. In practice, for new substances, this is not possible, as the product is not used. For re-registration of existing product, where concern exists for NTAs, this is an option.

Another option is to apply modelling approaches. Models can be used to address specific questions and uncertainties. Models need to be validated, however, before they can be used as a decision tool in risk assessment.

7.3. Assessment of effects of plant protection products on non-target arthropods at the landscape scale

As described in section 5, the present risk assessment scheme does not take into account the effects of configuration of agricultural fields and spatial dynamics on NTA populations. Hence, the working group proposes a way forward how this essential aspect necessary for realising the specific protection goals could be addressed using individual based modelling based on what is currently possible with existing models as described in section 5.3 and it needs to be made sure that appropriate data are available to describe the toxicity on NTAs as input for the proposed modelling approach.

The following aspects should be considered when defining an appropriate toxicity input for the models:

Dose responses based on either field studies or glass plate toxicity tests need to be incorporated into the modelling. There are different ways this can be done, but each has a consequence for the risk assessment. Some examples could be:

1. Effect probability above a threshold. This might be a daily probability calculated to give a certain effect (e.g. LC_{90} would give 90 % mortality) over the period that the test was measured in the laboratory. The disadvantage of this is that for long-period of exposure effects are virtually certain as probabilities combine each day, a result of multiple double jeopardy probability tests. This does not represent the case where individuals have differential sensitivities and could result in local extinction which could have important consequences for impact and recovery.
2. Using an individual sensitivity distribution whereby individuals have different threshold levels for effects. This prevents very high mortality with long exposure, but equally prevents long exposure having any impact above the instant the highest dose is experienced. This may prevent high impacts at local scale because there will always be a group of individuals that are completely safe because they have high individual threshold, or 100 % mortality if field rates are higher than the highest individual threshold, again with effects on impacts and recovery.
3. Dose–response relationships—these are more complicated because each time-step the organisms may have different doses and the probability of effects will change daily. This is also subject to double jeopardy effects.

In all cases the effects of multiple applications, long-term exposure and internal accumulation need to be considered, whilst avoiding ‘double jeopardy’ effects. It is also possible that previous exposure predisposes individuals to effects rendering them more sensitive to the same dose experienced later. One useful facet of laboratory toxicity testing is that the effect rate may change with time, typically highest in the first period and declining with time e.g. LC_{50} of cadmium (Ardestani and Van Gestel, 2013). This could be used if individuals carry a memory of past exposure and effect probabilities be reduced with time. However, declining probabilities will not provide a full solution and further consideration of how to address this problem will be needed for any future guidance document.

7.4. Conclusions and recommendations on non-target arthropod testing and effect assessment

It is concluded that the present first tier risk assessment is not protective for the occurrence of effects in the field situations in all cases. It is therefore recommended to carry out tier 1 toxicity tests on at least four species.

It is recommended that existing glass-plate test protocols should be performed testing effects on mortality and reproduction. The working group strongly recommends that effects on reproduction should be tested at tier 1.

The working group strongly recommends performing an oral toxicity study with Lepidopteran larvae as representative of herbivorous species.

Field studies are important to address a number of uncertainties connected to the limitations of the lower tier studies. The limitations of field studies, especially in relation to the mobility of species, should be considered while conducting and assessing a field study.

More uniform field study methods would enhance the correct use of field studies in relation to the uncertainties that can be addressed, and it would enhance the more uniform assessment of the field study results in the risk assessment.

Exposure in field studies should be addressed in such a way that extrapolation of results is possible.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

The focus of risk assessment for NTAs has been on species that are beneficial in integrated pest management for more than 23 years. This is true of both the ESCORT 1 (1994) and the ESCORT 2 (2002) protocols. The assessment of effects on biodiversity is not explicitly addressed under the existing guidance documents, even though EC91/414 specifies that there should be 'No unacceptable impact on non-target species'. Appropriate risk assessment methodology therefore needs to be developed for protection of biodiversity and a range of ecosystem processes, including biological control of pests, food web support and pollination. EFSA (2010) now places a new emphasis on ecosystem processes, general protection goals and specific protection goals. New risk assessment methodology should be based on achieving specific protection goals that protect important ecosystem processes both within cropped agricultural fields ('in-field') and off-field. A key issue is assessment of effects of PPPs both in-field and off-field. Some PPPs (especially insecticides) are expected to affect in-field NTAs adversely. It is essential to protect NTA biodiversity off-field in order to maintain *donor areas for recovery* of in-field NTAs.

It is first necessary to define the temporal and spatial boundaries in the context of risk assessment. These boundaries relate to the protection goal (e.g. 'where is the community of interest for this specific protection goal?') in relation to the route and distance covered of the emission coming from the in-field. It is also necessary to distinguish between the area which is for cultivation with crops (= in-field) and the area surrounding a field (= off-field). The off-field can be either (semi-)natural habitats or simple structures (fence or a bare strip of land). In most cases, the off-field should not be influenced by the farmer's use of PPPs. Another off-field category comprises man-made structures, e.g. an adjacent field, roads, etc. The actual off-field is not known for every field. It is therefore proposed that a generic protection goal for the off-field area be defined. Another important spatial element is the buffer strip. It is a cropped or non-cropped zone of a defined width at the edge of a field, which is influenced by the farmer's action (e.g. spray drift). The buffer strip is normally enforced by authorities and underlies prescribed actions in order to meet the specific protection goal for the off-field. In addition, buffer strips may provide a potential source of NTA species for recovery from impacts in the cropped area (see section 3).

It is important to consider the distribution and the mobility of individuals of NTA species. The current risk assessment addresses the risk to NTAs at a local (within-field) scale. The local-scale risk assessment is appropriate to assess impacts of pesticide application on certain in-field ecosystem services (e.g. pest control). The local-scale risk assessment is also considered sufficient to address impacts on species with a very limited mobility. The overall population-level impact may, however, be underestimated for highly mobile species. Even if there is no exposure of individuals in the off-field area, the off-field population can be affected if the treated field acts as a 'sink'. It is therefore also recommended that the risk to mobile species at larger spatial scales than the fields treated be addressed (see section 3.5).

There is a general protection goal concerning biodiversity under Regulation EU 1107/2009: *no unacceptable effects of PPP on biodiversity and the ecosystem*. A certain degree of biodiversity therefore has to be supported in the in-field areas in order to maintain both an appropriate level of NTA biodiversity in the landscape and the important ecosystem services provided by NTA diversity at the local scale. For in-field areas, the magnitudes of effects of PPPs on NTA biodiversity that are considered to be acceptable relate to the most sensitive ecosystem service to be supported in-field. It is understood that a sensitive service is driven by NTAs with high ecotoxicological or ecological sensitivity (e.g. low recovery potential) and/or a service that is highly susceptible to windows of opportunity in time (e.g. NTAs as food web support for bird chicks in the breeding season). Indirect effects of the use of PPPs on food web structure and stability are considered to be especially important for NTAs, as these animals deliver a substantial part of the diet of several farmland birds and small mammals. In this respect, the determination of required dietary demands for bird and mammals would deliver acceptable magnitude of effect on NTAs as food web support in absolute values (see section 4)

Landscape-level assessments should be used to ensure that the magnitude of effects on biodiversity in-field does not exceed the acceptable magnitude of effects in off-field areas. Effects of PPPs on NTAs in off-field areas should be negligible at most. For in-field as well as off-field areas, assessment of the magnitude of effects should take into account multiple PPP applications according to typical PPP spray schedules. This assessment may indicate a lower level of tolerable effects for single PPP applications, especially in-field; for example, if the intended use is part of an application scheme that includes several other PPPs with potential effects on NTAs in the crop (see section 4).

In order to include recovery in the risk assessment, it is necessary to first identify those NTA species that are key for ecosystem functions to ensure that they are not lost. It is essential to take into consideration both direct and indirect effects from multiple stresses (e.g. repeated application of different pesticides). The investigation of recovery should focus on sensitive species with traits making recovery more difficult for them (e.g. low number of offspring, low dispersal capacity). Provision of a certain level of some ecosystem functions (e.g. food web support, pollination, pest control) may need to be maintained continuously, in which case impacts of PPPs that reduce provision of the ecosystem service for a time may be unacceptable, even if the NTA community returns to its pre-disturbed state. Modelling and field studies are complementary for assessment of recovery. Field studies can provide information on the magnitude of effects on an in-field community, including indirect effects. Modelling can be used to investigate effects of PPPs on some species in different landscape contexts, including source–sink dynamics and different climatic conditions. It would be beneficial to identify and to evaluate risk mitigation options in order to facilitate recovery of NTAs (see section 5.6).

Exposure assessment needs to consider NTAs living primarily in each of the following four environmental compartments: in-field on crop, in-field on soil, off-field on vegetation and off-field on soil. Section 6 provides tiered approaches designed to assess exposure of NTAs that are not very mobile (e.g. home range less than field size) but these approaches may overestimate exposure for more mobile species, for which development of a landscape approach is recommended. Worst-case effects for the off-field are expected to occur immediately adjacent to the treated field. For these situations, drift deposition is considered to be the major source of exposure. It is recommended to base drift-deposition values on drift measurements that have become available recently.

A number of drift mitigation measures have recently been developed and it is therefore recommended that drift mitigation be accounted for in higher tiers of the assessment. One approach to accounting for drift mitigation is the so-called matrix approach, in which crop or crop-group drift deposition is related to growth stage, classes of drift-reducing technology and the distance between the crop and the location for which the risk assessment is performed. General risk management measures can be accounted for in the matrix approach as well (see section 6).

The exposure assessments take three NTA exposure routes into account: contact, overspray and oral consumption. All three exposure routes must be considered, although, in practice, it may be difficult to distinguish between them because of lack of suitable data. The ERC (the exposure concentration that is linked to ecological effects) is in general unknown. The application-dose rate is used as a surrogate to link exposure to effects in the proposed risk assessment schemes. In order to improve the risk assessment, it needs to be established what the ERCs are. This requires exposure concentrations in future to be measured in ecotoxicological effect experiments (see section 6).

The current risk assessment for NTAs uses a VDF in order to calculate off-field exposure to the applied dose. It might be more appropriate to use a lower concentration in calculating exposure because distribution of the sprayed PPP is uneven and the total leaf surface is larger than the area on which the vegetation stands. It is, however, not possible to recommend default values in the absence of proper ERCs and exposure concentrations. For example, it should be known whether the average areic mass on a plant or the maximum on one of its leaves is causing the observed effects. In the case of multiple applications in one growing season, accumulation of a substance on crop leaves may occur. The currently available default values for wash-off and other dissipation routes from leaves are based on limited numbers of experimental values. It is recommended that more independent (experimental) data are gathered to better underpin default values. Furthermore, it should be investigated whether scenario conditions, such as temperature, may be taken into account when calculating the dissipation (see section 6).

RUD values for insects could potentially provide a good estimate for exposure from overspray and contact to fresh residues on plants and soil surface. A simple and quick screening step assessment could be conducted together with endpoints expressed in the same units as the RUD values. It is recommended that it be investigated further whether or not the underlying residue data justify the use of RUDs as a conservative estimate of contact exposure. It also needs to be decided whether the acute contact endpoint from honeybees (LD_{50} µg/bee, recalculated to mg/kg insect) should be used in the assessment or new studies with NTAs should be proposed where the toxicity from contact to fresh residues (including overspray) is investigated (see section 6).

The current (ESCORT) risk assessment scheme for NTAs does not take into account the ecological effects of configuration of agricultural fields and spatial dynamics of NTA populations. Therefore, the risk assessment procedure needs to contain two elements: a tiered approach for assessing effects at the local (field-scale) level and an additional approach that relates to landscape-scale effects. It is necessary to assess local-scale effects in addition to landscape-scale effects because specific protection goals often require maintenance of particular ecosystem services within cropped fields as well as in off-crop areas.

A tiered approach to assessing the effects of PPPs on non-target organisms should include a relatively simple, robust set of tests as the lowest tier. Selection of species for testing ecotoxicological effects should be based on the specific protection goals, not just on practical convenience, although of course practicality also needs to be considered. Tier 1, local-scale test systems should therefore be designed to prevent missing unacceptable effects of intended uses of a PPP on ecosystem services defined to be important in the agricultural landscape. After reviewing the existing tests available, the Panel recommends carrying out tier 1 toxicity tests on four species (minimum), chosen to represent different lifestyles and taxonomic groups. Existing glass plate protocols should be updated in order to provide robust tests for effects on *reproduction* as well as mortality. Tests should also include an oral toxicity study with lepidopteran larvae to represent herbivorous NTAs. The Panel considers that the toxicity endpoints derived from tests with bees (fresh residues) could provide a possible surrogate for the overspray exposure route.

The Panel recommends that assessment factors be derived on the basis of statistical modelling of the relationships between effects for different species in the various possible lower tier tests, higher tier field studies and the surrogate reference tier. In particular, a Bayesian network model can exploit information from both experimental data and expert judgement and provides a relatively transparent method for deriving assessment factors in order to ensure high probability of acceptable effects for uses that pass the risk assessment. In this context, the panel considers that the SSD conceptual model is very useful at the reference tier level but that standard SSD methodology cannot yet be applied to NTAs because of lack of data.

Higher tier studies currently use semi-field and field studies based on multiple in-field plots (typically 24 × 24 m). These can be misleading for mobile species that move into and out of plots during the course of a study. Replicated landscape-scale studies are desirable but usually impractical. A possible compromise is to carry out a field study with a limited number of large plots in combination with a larger number of smaller plots.

The panel recommends that landscape-scale effects should be studied as part of NTA risk assessment. This can be achieved using landscape-scale modelling that follows the recommendations of the EFSA

Scientific Opinion on Good Modelling Practice (2014). The available choices of both landscape modelling systems and species modelled are currently quite limited. The Panel recommends that the range of species modelled should be expanded and that it should include a wider range of NTAs and species from southern areas of Europe.

RECOMMENDATIONS

The new regulatory framework for PPPs requires consideration of impacts on non-target species, on their ongoing behaviour and impacts on the biodiversity and the ecosystem, including potential indirect effects via alteration of the food web. It is recommended to translate these general protection goals into specific protection goals in consultation with risk managers. Clearly defined specific protection goals are needed to develop risk assessment guidance which ensures that the protection goals are achieved.

The Panel proposes specific protection goals which relate to NTAs in their role as drivers of biodiversity, aesthetic value, pest control, food web support and pollination. More information is needed to be able to predict quantitatively effects of changes in NTA populations on related ecosystem services. Research is needed to generate data which allow a quantification of impacts on key driver species and the related ecosystem services they provide.

The effect magnitudes need to be considered in the landscape context. To ensure that effects in-field do not have unacceptable effects on NTA biodiversity, it is suggested that a landscape-level risk assessment is conducted. A landscape-level assessment is also needed because off-field areas can be negatively affected by treated fields acting as a sink for mobile arthropods in the off-field.

Monitoring studies will be needed to verify in which agricultural landscape and management settings the specific protection goals as outlined in this opinion can be implemented and under which circumstances. The effectiveness of the different management options to mitigate identified risks at landscape scale (as suggested in section 5) should be investigated in view of the defined specific protection goals.

With regard to effect modelling it is recommended that future risk assessment consider populations in landscape with a diverse range of structures and the agricultural practice (including the use of other pesticides). Long-term (multi-generation, more than 10 years) impacts of exposure to stressors can be demonstrated as long-term NTA population declines. Therefore, future risk assessment should take account of long-term population exposure.

For landscape-scale population-level risk assessment, the entity of interest is the whole population and its distribution in space and time. This approach automatically incorporates recovery, as it is the long-term population status which is the ultimate endpoint. It is recommended that relevant measurement endpoints (e.g. abundance and/or distribution) to assess this should be identified.

Direct and indirect effects from multiple stressors (e.g. repeated application of different pesticides) need to be considered in the assessment. Modelling and field studies are complementary for assessment of recovery. In order to protect vulnerable species, the investigation of recovery should include species with traits indicating a low recovery potential (e.g. low number of offspring, low dispersal capacity).

The Panel recommends that assessment factors be derived on the basis of statistical modelling of the relationships between effects for different species in the various possible lower tier tests, higher tier field studies and the surrogate reference tier. In this context, the panel considers that the SSD conceptual model is very useful at the reference tier level. The standard SSD methodology, however, may be difficult to apply at tier 1 because of lack of data.

The environmentally relevant concentrations need to be defined and determined in field studies. The Panel recommends that field study protocols are developed with measurement of relevant concentrations.

Rather high uncertainties still exist about processes influencing exposure, both in-field and off-field, for example wash-off. Different sets of factors may need to be developed for the risk assessment of soil-dwelling and leaf-dwelling organisms. The uptake process on leaves need to be understood better for more realistic exposure estimates.

For deposition of drift and dust and exchange of air-borne substances with receptor surfaces, several datasets exists. These datasets should be combined in order to produce harmonised approaches (e.g. drift curves).

RUD values could be used to estimate oral exposure of NTAs. The list of RUD values for plant (leaves), nectar, pollen, insects, etc. on which NTAs feed should be extended.

It should be investigated whether RUD values can be used to verify estimates of exposure modelling. It is recommended to investigate whether the underlying residue data justify the use of RUDs as a conservative estimate of oral, contact and overspray exposure of NTAs

Furthermore, RUD values for insects could provide a first tier estimate for exposure of arthropods from overspray and contact to fresh residues on plants and soil surface. A quick screening step assessment together with endpoints expressed in the same units as the RUD values should be developed.

It also needs to be decided whether the acute contact endpoint from honeybees (LD_{50} µg/bee and recalculated to mg/kg insect) should be used in the assessment. If not, new studies with NTAs should be proposed where the toxicity from contact to fresh residues (including overspray) is investigated.

As neither the implicit dilution of exposure in field studies via vegetation distribution nor the actual ERCs of tested NTAs was considered when calibrating current risk assessment, it is recommended to stop using the VDF as a refinement of off-field exposure.

It is recommended that dynamic exposure modelling is used to link effects assessment of mobile NTAs to changing patterns of exposure in space and time at the landscape scale. This should include exposure distribution and more realistically should address temporal issues related to co-occurrence of NTAs and stressor. Dose–response information should be integrated in such modelling.

There is not enough evidence that the present first tier effect assessment is sufficient for a robust prediction of effects on all key driver species in the field. This is due to differences in species sensitivity, differences in exposure and lifestyles. To reduce this uncertainty, it is recommended to carry out tier 1 toxicity tests on at least four species out of which one should be a Lepidoptera larvae as a representative of herbivorous species.

The Panel recommends developing new test methods that would allow assessment of effects from chronic exposure and delayed effects in NTAs in the lower tiers. They should allow for estimating effects on the most crucial life history parameters, such as longevity and reproduction rate.

Field studies are important to investigate direct and indirect effects on communities under realistic field exposure situations. The panel recommends development of new field study protocols in order to address uncertainties and to aid the consistent evaluation of field studies. Exposure should be measured in field studies in order to link exposure and effects.

It is important to understand the power of various field study designs to detect effects at magnitudes relevant to the specific protection goals. Without in-depth theoretical understanding of the population dynamics, it is extremely challenging to make theoretical power calculations. There is empirical evidence of high variability in field studies. It is recommended to re-analyse raw data from field studies in order to evaluate the statistical power.

The Panel recommends further development of landscape simulation models. Development of new models would be beneficial for risk assessment and therefore should be prioritised, even though it is time and resource consuming. Development should include production of simulation data for look-up tables and modelling new species and landscape scenarios. A suite of standard models could be developed for NTA risk assessment.

Effect modelling, including TK/TD modelling, is a research area which could result in future application in risk assessment. It could be beneficial to develop such models for NTAs in order to reduce testing. In combination with exposure modelling, it may be possible to develop new risk assessment approaches.

For the development of a guidance document, it is recommended that easy to use software tools are produced and made available together with the guidance document.

REFERENCES

Abbott WS, 1925. A method of computing the effectiveness of an insecticide. Journal of Economic Entomology, 18, 265-267.

Abraham L and Vas J, 1999. Preliminary report on study of the daily activity pattern of neuroptera in Hungary. Acta Phytopathologica et Entomologica Hungarica 34, 153–164.

Aldenberg T and Jaworska JS, 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotoxicology and Environmental Safety, 46, 1–18.

Aldershof S and Bakker F, 2010. Comparison of Arthropod Community Responses to an Insecticidal Active in Different Geographic Regions. In: ESCORT 3, 2010.

Alix A, Bakker F, Barrett K, Brühl CA, Coulson M, Hoy S, Jansen JP, Jepson P, Lewis G, Neumann P, Süßenbach D and van Vliet P (Eds), 2012. ESCORT 3 – linking non-target arthropod testing and risk assessment with protection goals. CRC SETAC Press, 1–151.

Altieri MA, 1999. The ecological role of biodiversity in agroecosystems. Agriculture, Ecosystems and Environment, 74, 19–31.

Ardestani MM and Gestel CAM. van, 2013. Toxicodynamics of copper and cadmium in Folsomia candida exposed to simulated soil solutions. Environmental Toxicology and Chemistry, 32, 2746–2754.

Bakker F, Grove A, Blumel S, Calis J and Oomen P, 1992. Side-effect tests for phytoseiids and their rearing methods. Bulletin OILB/SROP, 15, 61–81.

Balvanera P, Pfisterer AB, Buchmann N, He J-S, Nakashizuka T, Raffaelli D and Schmid B, 2006. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. Ecology Letters, 9, 1146–1156.

Barnthouse LW, 2004. Quantifying population recovery rates for ecological risk assessment. Environmental Toxicology and Chemistry, 23, 500-508.

Barrett KL, Grandy N, Harrison EG, Hassan S and Oomen PA, 1994. Guidance document on regulatory testing procedures for pesticides with non-target arthropods. Report of the SETAC/ESCORT Workshop 51. Wageningen, the Netherlands, and SETAC-Europe, Brussels, Belgium.

Begon M, Townsend CR and Harper JL, 2006. Ecology: from individuals to ecosystems, 4th edition. Wiley-Blackwell, Oxford, 752 pp.

Benton TG, Vickery J and Wilson JD, 2003. Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecology and Evolution, 18, 182–188.

Bianchi FJJA, van Wingerden WKRE, Griffioen AJ, van der Veen M, van der Straten MJJ, Wegman RMA and Meeuwsen HAM, 2005. Landscape factors affecting the control of *Mamestra brassicae* by natural enemies in brussels sprout. Agriculture Ecosystems and Environment, 107, 145–150.

Bianchi F, Booij CJH and Tscharntke T, 2006. Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control. Proceedings of the Royal Society B Biological Sciences, 273, 1715–1727.

Bianchi FJJA, Ives AR, Schellhorn NA. 2013. Interactions between conventional and organic farming for biocontrol services across the landscape. Ecol. Appl. 23:1531–43

Bilde T and Topping C, 2004. Life history traits interact with landscape composition to influence population dynamics of a terrestrial arthropod: A simulation study. Ecoscience, 11, 64–73.

Blitzer EJ, Dormann CF, Holzschuh A, Klein A-M, Rand TA and Tscharntke T, 2012. Spillover of functionally important organisms between managed and natural habitats. Agriculture Ecosystems and Environment, 146, 34–43.

Boatman ND, Brickle NW, Hart JD, Milsom TP, Morris AJ, Murray AW, Murray K and Robertson P, 2004. Evidence for the indirect effects of pesticides on farmland birds. Ibis, 146, 131–143.

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.

Blümel S, Bakker FM, Baier B, Brown K, Candolfi MP, Gossmann A, Grimm C, Jäckel B, Nienstedt K, Schirra KJ, Ufer A and Waltersdorfer A, 2000a. Laboratory residual contact test with the predatory mite *Typhlodromus pyri* Scheuten (Acari: Phytoseiidae) for regulatory testing of plant protection products. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative, IOBC/WPRS, Gent, Belgium, 128–150.

Blümel S, Aldershof S, Bakker FM, Baier B, Boller E, Brown K, Bylemans D, Candolfi MP, Huber B, Linder C, Louis F, Müther J, Nienstedt KM, Oberwalder C, Reber B, Schirra KJ, Sterk G, Ufer A and Vogt H, 2000b. Guidance document to detect side effects of plant protection products on predatory mites (Acari: Phytoseiidae) under field conditions: vineyards and orchards. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative, IOBC/WPRS, Gent, Belgium, 145–158.

Brickle N and Harper D, 2000. Effects of agricultural intensification on the breeding success of corn buntings *Miliaria calandra*. Journal of Animal Ecology, 37, 742–755.

Braden H, 1985. Ein Energiehaushalts- und Verdunstugsmodell für Wasser und Stofhaushaltsuntersuchungen landwirtschaftlich genutzter Einzugsgebiete. Mitteilungen Deutsche Bodenkundliche Geselschaft 42, 294-299.

Brock TCM and Budde BJ, 1994. On the choice of structural parameters to indicate responses of freshwater ecosystems to pesticide stress. In: Hill IA, Heimbach F, Leeuwangh P, Matthiesen P (Eds), Freshwater Field Tests for Hazard Asses¬sment of Chemicals Lewis Publishers, Michigan, 19 56.

Brown K, Tomlinson J, Duncan J and Hinchcliffe A, 2009. Critical comparison of available and potential higher tier testing approaches for the risk assessment of plant protection products, considering at least field and semi-field experimental designs, extrapolation from dose–response relationship, and increased dosages (aquatic and terrestrial). Literature reviews on ecotoxicology of chemicals with special focus on plant protection products. CFT/EFSA/PPR/2008/01. Exponent, UK, 192 pp.

Bruck DJ and Lewis LC, 1998. Influence of adjacent cornfield habitat, trap location, and trap height on capture numbers of predators and a parasitoid of the European corn borer (Lepidoptera: Pyralidae) in central Iowa. Environmental Entomology, 27, 1557–1562.

Brühl CA, Neumann P, Aldershof S, Bohan D, Brown K, Candolfi M, Geiger F, Kovalkovicová N, Kula C, Nienstedt K, Pestanudo S, Roembke J, Schmidt T, Topping CJ and van Vliet P, 2012. Recovery. In: ESCORT 3: Linking non-target arthropod testing and risk assessment with protection goals. Eds Alix A, Bakker F, Barrett K, Brühl CA, Coulson M, Hoy S, Jansen J, Jepson P, Lewis G, Neumann P, Süssenbach D, van Vliet P, eds.. Society of Environmental Toxicology and Chemistry, 39–41.

Brühl CA, Alscher A, Berger G, Bethwell C, Graef F, Hahn M, Schmidt T and Weber B, 2013. Protection of biodiversity in the risk assessment and risk management of pesticides with a focus on arthropods, soil organisms, and amphibians. Report of the Research and Development Project Nr 3709 65 421. German Federal Environment Agency (Umweltbundesamt), 244 pp.

Bryant DM, 1975. Breeding biology of house martins *Delichon urbica* in relation to aerial insect abundance. Ibis, 117, 180–216.

Buijs AE, Pedroli B and Luginbühl, 2006. From hiking through farmland to farming in a leisure landscape: changing social perceptions of the European landscape. Landscape and Ecology, 21, 375–389.

Bundschuh R, Schmitz J, Bundschuh M and Brühl CA, 2012. Does drift of insecticide adversely affect grasshoppers (Orthoptera: Saltatoria) in field margins? A case study combining laboratory acute toxicity with field monitoring data. Environmental Toxicology and Chemistry, 1874–1879.

Cãgan L, Turlings T, Bokor P and Dorn S, 1999. *Lydella thompsoni* Herting (Dipt., Tachinidae), a parasitoid of the European corn borer, *Ostrinia nubilalis* Hbn. (Lep., Pyralidae) in Slovakia, Czech Republic and south-western Poland. Journal of Applied Entomology, 123, 577–583.

Calderone NW, 2012. Insect pollinated crops, insect pollinators and us agriculture: trend analysis of aggregate data for the period 1992–2009. PLOS ONE 7(5), e37235. doi:10.1371/journal.pone.0037235

Campbell PJ, Brown KC, Harrison EG, Bakker F, Barrett KL, Candolfi MP, Canez V, Dinter A, Lewis G, Mead-Briggs M, Miles M, Neumann P, Romijn K, Schmuck R, Shires S, Ufer A and Waltersdorfer A, 2000. A hazard quotient approach for assessing the risk to non-target arthropods from plant protection products under 91/414/EEC: hazard quotient trigger value proposal and validation, Journal of Pest Science, 73, 117–124.

Candolfi MP, Bakker F, Canez V, Miles M, Neumann Ch, Pilling E, Primiani M, Romijn K, Schmuck R, Storck-Weyhermuller S, Ufer A and Waltersdorfer A, 1999. Sensitivity of non-target arthropods to plant protection products: could *Typhlodromus pyri* and *Aphidius* spp. be used as indicator species? Chemosphere, 39, 1357–1370.

Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H, 2000a. Guidelines to evaluate side-effects of plant protection products to non-target arthropods. IOBC, BART and EPPO Joint Initiative, IOBC/WPRS, Gent Belgium, 158 pp.

Candolfi M, Bigler F, Campbell P, Heimbach U, Schmuck R, Angeli G, Bakker F, Brown K, Carli G, Dinter A, Forti D, Forster R, Gathmann A, Hassan S, Mead-Briggs M, Melandri M, Neumann P, Pasqualini E, Powell W, Reboulet J-N, Romijn K, Sechser B, Thieme T, Ufer A, Vergnet C and Vogt H, 2000b. Principles for regulatory testing and interpretation of semi-field and field studies with non-target arthropods. Journal of Pest Science, 73, 141–147.

Candolfi MP, Barrett KL, Campbell P, Forster R, Grandy N, Huet M-C, Lewis G, Oomen PA, Schmuck R and Vogt H, 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, and SETAC-Europe, Brussels, Belgium.

Cardinale BJ, Duffy JE, A. Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS and Naeem S, 2012. Biodiversity loss and its impact on humanity. Nature, 486, 59–67. doi:10.1038/nature11148

Clergue B, Amiaud B, Pervanchon F, Lasserre-Joulin F and Plantureux S, 2005. Biodiversity: function and assessment in agricultural areas. A review. Agronomy for Sustainable Development, 25, 1–15.

Clough Y, Holzschuh A, Gabriel D, Purtauf T, Kleijn D, Kruess A, Steffan-Dewenter I and Tscharntke T, 2007. Alpha and beta diversity of arthropods and plants in organically and conventionally managed wheat fields. Journal of Applied Ecology, 44, 804–812. doi: 10.1111/j.1365-2664.2007.01294.x

Collins WW and Qualset CO (eds), 1999. Biodiversity in agroecosystems. CRC Press, Boca Raton, FL, 334 pp.

Cooke R, 2010. Conundrums with uncertainty factors (with discussion). Risk Analysis, 30, 330–338.

Coombes DS and Sotherton NW, 1986. The dispersal and distribution of polyphagous predatory Coleoptera in cereals. Annals of Applied Biology, 8, 461–474.

Corbett A and Rosenheim JA, 1996. Impact of a natural enemy overwintering refuge and its interaction with the surrounding landscape. Ecological Entomology 21, 155–164.

Craig PS, Hickey GL, Luttik R and Hart A, 2012. Species non-exchangeability in probabilistic ecotoxicological risk assessment. Journal of the Royal Statistical Society Series A: Statistics in Society 175, 243–262.

Cronin, JT, 2003 Patch structure, oviposition behaviour, and the distribution of parasitism risk. Ecological Monographs, 73, 283–300

Dalkvist T, Topping CJ and Forbes VE, 2009. Population-level impacts of pesticide-induced chronic effects on individuals depend more on ecology than toxicology. Ecotoxicology and Environmental Safety, 72, 1663–1672.

Dalkvist T, Sibly R and Topping C, 2013. Landscape structure mediates the effects of a stressor on field vole populations. Landscape Ecology, 28 (10), 1–14.

de Groot RS, Wilson MA and Boumans RMJ, 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Special Issue: The dynamics and value of ecosystem services: integrating economic and ecological perspectives. Ecological Economics, 41, 393–408.

De Jong FMW, Bakker FM, Brown K, Jilesen CJTJ, Posthuma-Doodeman CJAM, Smit CE, Van der Steen JJM and van Eekelen GMA. 2010. Guidance for summarising and evaluating field studies with non-target arthropods. RIVM report 601712006. RIVM, Bilthoven, the Netherlands, 73 pp.

de Ruiter PC, Wolters V, Moore JC and Winemiller KO, 2005. Food web ecology: playing Jenga and beyond. Science, 309, 68–71. doi: 10.1126/science.1096112

DEFRA (Department for Environment, Food and Rural Affairs), 2001. Assessing the validity of current approaches for calculating off-crop exposure to pesticides. Defra Project code PS2001, 18 pp.

DEFRA (Department of Environment, Food and Rural Affairs), 2005. Assessing the indirect effects of pesticides on birds. Final report PN0925 under Project CTD9804, 40 pp.

DEFRA (Department for Environment, Food and Rural Affairs), 2007. Methods of addressing variability and uncertainty for improved pesticide risk assessments for non-target arthropods. Defra Project code PS2307, 27 pp.

Desneux N, Denoyelle R and Kaiser L, 2006. A multi-step bioassay to assess the effect of the deltamethrin on the parasitic wasp Aphidius ervi. Chemosphere, 65, 1697-1706.

Desneux N, Fauvergue X, Dechaume-Moncharmont, X, Kerhoas L, Ballanger Y and Kaiser L, 2005. *Diaeretiella rapae* limits *Myzus persicae* populations after applications of deltamethrin in oilseed rape. Journal of Economic Entomology, 98, 9–17.

Devine, G.J. and M.J. Furlong (2007): Insecticide use: Contexts and ecological consequences. Agr. Human Values 24, 281–306. doi 10.1007/s10460-007-9067-z

Douka C and Tchuenguem F-N, 2014 Foraging and pollination activity of *Musca domestica* L. (Diptera: Muscidae) on flowers of *Ricinus communis* L. (Euphorbiaceae) at Maroua, Cameroon. Journal of Biodiversity and Environmental Sciences, 4, 63–76. Dover, J. W. (1997).

Dover JW, 1997. Conservation headlands: Effects on butterfly distribution and behaviour. Agriculture Ecosystems & Environment 63(1): 31-49.

Duelli P, Studer M, Marchand I and Jakob S, 1990a. Population movements of arthropods between natural and cultivated areas. Biological Conservation, 54, 193–207.

Duelli P, Studer M and Marchand I, 1990b. The influence of the surroundings on Arthropod diversity in maize fields. Acta Phytopathologica et Entomologica Hungarica, 24, 73–76.

Duffield SJ, Jepson PC, Wratten Sd and Sotherton NW, 1996. Spatial changes in invertebratepredation rate in winter wheat following treatment with dimethoate. Entomologia Experimentalis et Applicata, 78, 9-17.

EC (European Commission), 2002. Guidance document on terrestrial ecotoxicology under Council Directive 91/414/EEC (SANCO/10329/2002) revision 2, final, 17.10.2002. 1–39.

Edwards CA, Sunderland KD and George KS, 1979. Studies on polyphagous predators of cereal aphids. Journal of Applied Ecology, 16, 811–823.

EFSA (European Food Safety Authority), 2006. Opinion of the Scientific Panel on Plant Health, Plant Protection Products and their Residues (PPR) on a request from EFSA related to the assessment of the acute and chronic risk to aquatic organisms with regard to the possibility of lowering the uncertainty factor if additional species were tested. The EFSA Journal 2006, 301, 1–45. doi:10.2903/j.efsa.2006.301

EFSA (European Food Safety Authority), 2007a. Opinion of the Scientific Committee related to Uncertainties in Dietary Exposure Assessment. The EFSA Journal 2007, 438, 1–54. doi:10.2903/j.efsa.2007.438

EFSA (European Food Safety Authority), 2007b. Opinion on a request from EFSA related to the default Q10 value used to describe the temperature effect on transformation rates of pesticides in soil. The EFSA Journal 622, 1-32. doi:10.2903/j.efsa.2007.622

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 734, 181. doi:10.2903/j.efsa.2008.734

EFSA (European Food Safety Authority), 2009a. Guidance Document on Risk Assessment for Birds & Mammals on request from EFSA. EFSA Journal 2009; 7(12):1438. doi:10.2903/j.efsa.2009.1438

EFSA (European Food Safety Authority), 2009b. Outcome of the Public Consultation on the existing Guidance Documents on Aquatic and Terrestrial Ecotoxicology under Directive 91/414/EC. EFSA Journal 2009; 7(11):1375. 129 pp. doi:10.2903/j.efsa.2009.1375

EFSA (European Food Safety Authority), 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 (European Food Safety Authority), 2013a. Biodiversity as protection goal in environmental risk assessment for EU agro-ecosystems. Summary report of the 19th EFSA Scientific Colloquium, 72 pp. Available online: http://www.efsa.europa.eu/en/supporting/doc/583e.pdf

EFSA (European Food Safety Authority), 2013b. Guidance on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). EFSA Journal 2013;11(7):3295, 268 pp. doi:10.2903/j.efsa.2013.3295

EFSA (European Food Safety Authority), 2014a. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734, 278 pp. doi:10.2903/j.efsa.2014.3734

EFSA (European Food Safety Authority), 2014b. EFSA Guidance Document for evaluating laboratory and field dissipation studies to obtain DegT50 values of active substances of plant protection products and transformation products of these active substances in soil. EFSA Journal 2014;12(5):3662, 37 pp. doi:10.2903/j.efsa.2014.3662

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2010. Scientific Opinion on outline proposals for assessment of exposure of organisms to substances in soil. EFSA Journal 2010;8(1):1442, 38 pp. doi:10.2903/j.efsa.2010.1442

EFSA PPR Panel (EFSA 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 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), 2013c. 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

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2014c. 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

Ekbom BS, Wiktelius S and Chiverton PA, 1992. Can polyphagous predators control the bird cherry-oat aphid (*Rhopalosiphum-padi*) in spring cereals – a simulation study. Entomologia Experimentalis et Applicata, 65, 215–223.

Ekroos J, Olsson O, Rundlöf M, Wätzold F and Smith HG, 2014. Optimizing agri-environment schemes for biodiversity, ecosystem services or both? Biological Conservation, 172, 65–71.

Ekroos J, Kuussaari M, Tiainen J, Heliölä J, Seimola T and Helenius J, 2013. Correlations in species richness between taxa depend on habitat, scale and landscape context. Ecologial Indicators, 34, 528–535.

EPPO (European and Mediterranean Plant Protection Organization)/Council of Europe Risk, 1994. Decision making scheme for the environmental risk assessment of plant protection products. Chapter 9, Arthropod Natural Enemies. Bulletin OEPPIEPPO. Bulletin, 24, 17–35.

EPPO (European and Mediterranean Plant Protection Organization), 2003. Decision making scheme for the environmental risk assessment of plant protection products. EPPO Bulletin, 33, 131–139.

Elzinga JA, Turin H, van Damme, JMM & Biere A, 2005. Plant population size and isolation affect herbivory of Silene latifolia by the specialist herbivore Hadena bicruris and parasitism of the herbivore by parasitoids. Oecologia, 144, 416–426

ESA (Ecological Society of America), 2014. Revealing secrets about... the birds and the bees. Pollination: An Essential Ecosystem Service. In: Communicating Ecosystem Services Project.

EU (European Union), 1997. Doc. 397L0057. Council Directive 97/57/EC of 22 September 1997 establishing Annex VI to Directive 91/414/EEC concerning the placing of plant protection products on the market. Official Journal L265, 0087-0109.

European Commission, 2006. Development of agri-environmental indicators for monitoring the integration of environmental concerns into the common agricultural policy. Communication COM 508 final.

Evans KL, Wilson JD and Bradbury RB, 2007. Effects of crop type and aerial invertebrate abundance on foraging barn swallows *Hirundo rustica*. Agriculture, Ecosystems & Environment, 122, 267–273.

Everts JW, Willemsen I, Stulp M, Simons L, Aukema B and Kammenga J, 1991. The toxic effect of deltamethrin on linyphiid and erigonid spiders in connection with ambient temperature, humidity and predation. Archives of environmental contamination and toxicology, 20 (1), 20-24

Faegri K and van der Pijl L, 1979. The principles of pollination ecology, 3rd edition. Pergamon Press, Oxford, UK, 248pp.

Fahrig L and Nuttle WK, 2005. Population ecology in spatially heterogeneous environments. In: Ecosystem function in heterogeneous landscapes. Eds Lovett GM, Jones CG, Turner MG and Weathers KC. Springer-Verlag, New York, USA, 95–118.

Finke DL and Denno RF, 2004. Predator diversity dampens trophic cascades. Nature, 429, 407–410.

Fletcher R J, 2005. Multiple edge effects and their implications in fragmented landscapes. J. Anim. Ecol. 74:342–352.

FOCUS. 2000. FOCUS groundwater scenarios in the EU review of active substances.EU.

FOCUS. 2001. FOCUS surface water scenarios in the EU evaluation process under 91/414/EEC. Report of the FOCUS surface water scenarios workgroup on surface water scenarios.SANCO.

Forbes VE and Calow P, 2002. Species sensitivity distributions revisited: a critical appraisal. Human and Ecological Risk Assessment, 8, 473–492.

Frampton GK and Dorne JLCM, 2007. The effects on terrestrial invertebrates of reducing pesticide inputs in arable crop edges: a meta-analysis. Journal of Applied Ecology, 44, 362–373.

Freier B, Triltsch H, Moewes M and Rappaport V, 1998. Predator units – an approach to evaluate coccinellids within the aphid predator community. IOBC-WPRS Bulletins, 21, 103–111.

Freier B, Triltsch H, Moewes M and Moll E, 2007. The potential of predators in natural control of aphids in wheat: Results of a ten-year field study in two German landscapes. BioControl, 52, 775–788. doi 10.1007/s10526-007-9081-5

Gardner RH, Engelhardt KAM, Elmore AJ and Cadol D, 2014. A traits-based model of species diversity. Ecological Modelling, 228, 178–194.

Gaylor DW and Kodell RL, 2000. Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Analysis, 20, 245–250.

Geiger F, Bengtsson J, Berendse F, Weisser WW, Emmersond M, Morales MB, Ceryngier P, Liira J, Tscharntke T, Winqvist C, Eggers S, Bommarco R, Pärt T, Bretagnolle V, Plantegenest M, Clement LW, Dennis C, Palmerd C, Oñate JJ, Guerrero I, Hawro V, Aavik T, Thies C, Flohre A, Hänke S, Fischer C, Goedhart PW and Inchausti P, 2010. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. Basic and Applied Ecology, 11, 97–105.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB, 2013. Bayesian data analysis, 3rd edition. Chapman and Hall/CRC Press, London, UK.

Genduso P, 1979. Grain pentatomids in Italy and investigations on the biocoenoses in the newly sown fields and shelter vegetation. I pentatomidi del grano in Italia e ricerche sulla biocenosi nei seminativi e nella vegetazione rifugio. Bollettino dell'Istituto di Entomologia Agraria e dell'Osservatorio di Fitopatologia di Palermo, 9, 59–74.

Gering JC and Crist TO, 2002. The alpha–beta–regional relationship: providing new insights into local–regional patterns of species richness and scale dependence of diversity components. Ecology Letters, 5, 433–444. doi: 10.1046/j.1461-0248.2002.00335.x

German Federal Environment Agency (UBA), 2006. Exposure calculation for arthropods in field border structures – selection of an appropriate 'vegetation distribution factor'.

Gobster PH, Nassauer JI, Daniel TC and Fry G, 2007. The shared landscape: What does aesthetics have to do with ecology? Landscape Ecology, 22, 959–972.Gonzalez-Valero J, Campbell P, Fritsch H, Grau R and Romijn K, 2000. Exposure assessment for terrestrial non-target arthropods. Anzeiger fuer Schaedlingskunde, 73, 163–168.

Goulart FF, Salles P, Saito CH and Machado RB, 2013. How do different agricultural management strategies affect bird communities inhabiting a savanna-forest mosaic? A qualitative reasoning approach. Agriculture, Ecosystems & Environment, 164, 114–130.

Green ID, Jeffries C, Diaz A and Tibbett M, 2006. Contrasting behaviour of cadmium and zinc in a soil–plant–arthropod system. Chemosphere 64, 1115–1121

Grimm C, Reber B, Barth M, Candolfi M, Drexler A, Maus C, Moreth L, Ufer A and Waltersdorfer A, 2000. A test for evaluating the chronic effects of plant protection products on the rove beetle aleochara bilineata gyll. (coleoptera: Staphylinidae) under laboratory and extended laboratory conditions. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative – IOBC/WPRS, Gent, Belgium, 1–12.

Grimm C, Schmidli H, Bakker F, Brown K, Campell P, Candolfi M, Chapman P, Harrison EG, Mead-Briggs M, Schmuck R and Ufer A, 2001. Use of standard toxicity tests with Typhlodromus pyri and Aphidius rhopalosiphi to establish a dose response relationship. Anzeiger für Schädlingskunde-Journal of pest science, 74 (3), 72-84.

Haenke S, Kovacs-Hostyanszki A, Fruend, J, Batáry P, Jauker B, Tscarntke T and Holzschuh A, 2014. Landscape configuration of crops and hedgerows drives local syrphid fly abundance. Journal of Applied Ecology, 51, 505–513.

Hagerup O, 1951. Pollination in the Faroes – in spite of rain and poverty of insects. Biologiske Meddelelser Copenhagen, 18, 1–48.

Hagler JR and Naranjo SE, 2005. Use of gut content ELISA to detect whitefly predator feeding activity after field exposure to different insecticide treatments. Biocontrol Science Technology, 15, 321–339.

Hajek AE, 2004. Natural enemies: an introduction to biological control. Cambridge University Press, Cambridge, UK.

Harrison PA, Berry PM, Simpson G, Haslett JR, Blicharska M, Bucur M, Dunford R, Egoh B, Garcia-Llorente M, Geamănă N, Geertsema W, Lommelen E, Meiresonne L and Turkelboom F, 2014. Linkages between biodiversity attributes and ecosystem services: A systematic review. Ecosystem Services, 9, 191–203. doi:10.1016/j.ecoser.2014.05.006.

Hart JD, Milsom TP, Fisher G, Wilkins V, Moreby SJ, Murray AWA and Robertson PA, 2006. The relationship between yellowhammer breeding performance, arthropod abundance and insecticide applications on arable farmland. Journal of Applied Ecology, 43, 81–91.

Hassan SA, Albert R, Bigler F, Blaisinger P, BogenschUtz H, Boller E, Brun J, Cheverton P, Edwards P, Englert WD, Huang P, lnglesfield C, Naton E, Domen PA, Overmeer WP, Rieckmann W, Samsoe-Petersen L, Staubli A, Tuset JJ, Viggiani G and Vanwetswinkel O, 1987. Results of the third joint pesticide testing programme by the IOBC/WPRS-Working Group 'Pesticides and Beneficial Organisms'. Applied Entomology, 103, 92–107.

Hassan S, Bigler F, Bogenschütz H, Boller E, Brun J, Calis J, Chiverton P, Coremans-Pelseneer J, Duso C, Lewis G, Mansour F, Moreth L, Oomen P, Overmeer W, Polgar L, Rieckmann W, Samsoe-Petersen L, Stäubli A, Sterk G, Tavares K, Tuset J and Viggiani G, 1991. Results of the fifth joint pesticide testing programme carried out by the IOBC/WPRS-working group 'pesticides and beneficial organisms'. Biocontrol, 36, 55–67.

Hassan S, Bigler F, Bogenschütz H, Boller E, Brun J, Calis J, Coremans-Pelseneer J, Duso C, Grove A, Heimbach U, Helyer N, Hokkanen H, Lewis G, Mansour F, Moreth L, Polgar L, SamsÃ¸e-Petersen L, Sauphanor B, Stäubli A, Sterk G, Vainio A, van de Veire M, Viggiani G and Vogt H, 1994. Results of the sixth joint pesticide testing programme of the IOBC/WPRS-working group on pesticides and beneficial organisms. Biocontrol, 39, 107–119.

Hassan SA, Halsall N, Gray AP, Kuehner C, Moll M, Bakker FM, Roembke J, Yousef A, Nasr F and Abdelgader H, 2000. A laboratory method to evaluate the side effects of plant protection products on Trichogramma cacoeciae Marchal (Hymenoptera: Trichogrammatidae). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative – IOBC/WPRS, Gent, Belgium, 107–119.

Haughton AJ, Bell JR, Boatman ND and Wilcox A, 2001. The effect of the herbicide glyphosate on non-target spiders: Part II. Indirect effects on *Lepthyphantes tenuis* in field margins. Pest Management Science, 57, 1037–1042.

Hedde M, van Oort F and Lamy I, 2012. Functional traits of soil invertebrates as indicators for exposure to soil disturbance. Environmental Pollution 164, 59-65

Henderson, C.F. and E. W. Tilton, 1955. Tests with acaricides against the brow wheat mite, J. Econ. Entomol. 48:157-161.

Heydemann B and Meyer H, 1983. Auswirkungen der lntensivkultur auf die Fauna in den Agrarbiotopen. SCHRIFTENREIHE DES DEUTSCHEN RATES FÜR LANDESPFLEGE (eds.) Leopold Bonn Verlagsdruckereigesellschaft, Landespflege und Landwirtschaft 42/12, 174-191

Heimbach U, Wehling A, Barrett KL, Candolfi MP, Jäckel B, Kennedy PJ, Mead-Briggs M, Nienstedt KM, Römbke J, Schmitzer S, Schmuck R, Ufer A and Wilhelmy H, 2000. A method for testing effects of plant protection products on spiders of the genus Pardosa (Araneae: Lycosidae) under laboratory conditions. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative – IOBC/WPRS, Gent, Belgium, 71–86.

Hickey GL, Craig PS, Luttik R and De Zwart D, 2012. On the quantification of intertest variability in ecotoxicity data with application to species sensitivity distributions. Environmental Toxicology and Chemistry, 31, 1903–1910.

Hole DG, Perkins AJ, Wilson JD, Alexander IH, Grice PV and Evans AD, 2005. Does organic farming benefit biodiversity? Biological Conservation, 122, 113–130.

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

Holland JM, Smith BM, Birkett TC and Southway S, 2012. Farmland bird invertebrate food provision in arable crops. Annals of Applied Biology, 160, 66–75.

Hooper DU, Chapin FS III, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setälä H, Symstad AJ, Vandermeer J and Wardle DA, 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecology Monographs, 75, 3–35. http://dx.doi.org/10.1890/04-0922

Hoye TT, Skov F and Topping CJ, 2012. Interpreting outputs of agent-based models using abundance–occupancy relationships. Ecological Indicators, 20, 221–227.

Huijsmans JFM and van de Zande JC, 2011. Workshop Harmonisation of drift and drift reducting methodologies for evaluation and authorization of plant protection products. Wageningen, The Netherlands 1-2 December 2010. Foundation Stichting Dienst Landbouwkundig Onderzoek (DLO)

Humphreys IC and Mowat DJ, 1994. Effects of some organic treatments on predators (Coleoptera, Carabidae) of cabbage root fly, Delia readicum (L) (Diptera, Anthomyiidae), and on alternative prey species. Pedobiologia, 38, 513–518.

IOBC (International Organization for Biological Control of noxious animals and plants), 1988. Guidelines for testing the effects of pesticides on beneficials: short description of test methods. West Palaartic Regional Section Working Group: Pesticides and beneficial organisms. Bulletin Section Regionale Ouest Palearctique, XI, 1–143.

Jahn T, Hötker H, Oppermann R, Bleil R and Vele L, 2014. Protection of biodiversity of free living birds and mammals in respect of the effects of pesticides. Final Report Federal Environement Agency (Umweltbundesamt) Project No 371063411. UBA Texte 30/2014, http://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/texte_30_2014_protection_of_biodiversity.pdf

Jepson PC and Thacker JRM, 1990. Analysis of the spatial component of pesticide side-effects onnon-target invertebrate populations and its relevance to hazard analysis. Functional Ecology 4,349-355.

Johnson CM, He T and Pauw A, 2014. Floral divergence in closely related *Leucospermum tottum* (Proteaceae) varieties pollinated by birds and long-proboscid flies. Evolutionary Ecology, 28, 849–868.

Joisten M, 2000. The 'island method' – a new laboratory test method to determine side-effects of plant protection products on the predatory mite Typhlodromus pyri Scheuten (Acari, Phytoseiidae). Journal of Applied Entomology – Zeitschrift Fuer Angewandte Entomologie, 124, 267–268.

Jones GD and Jones SD, 2001. The uses of pollen and its implication for Entomology. Neotropical Entomology, 30, 314–349. doi:10.1590/S1519-566X2001000300001

Kara K, Gözüaçik C and Mart C, 2007. Tachinid Parasitoids (Diptera: Tachinidae) of Mythimna (Acantholeucania) loreyi in the Southeast Anatolian Region of Turkey. Phytoparasitica 35: 136-139.

Karp DS, Rominger AJ, Zook J, Ranganathan J, Ehrlich PR and Daily GC, 2012. Intensive agriculture erodes β-diversity at large scales. Ecology Letters, 15, 963–970. doi: 10.1111/j.1461-0248.2012.01815.x

Kattwinkel M, Römbke J and Liess M, 2012. Ecological recovery of populations of vulnerable species driving the risk assessment of pesticides. EFSA Supporting Publications 2012, EN-338. 98 pp.

Kjaer C, Bruus M, Bossi R, Lofstrom P, Andersen HV, Nuyttens D and Larsen SE, 2014. Pesticide drift deposition in hedgerows from multiple spray swaths. J.Pestic. Sci 39, 14-2

Klein M, 1990. Pesticide Leaching Model (PELMO).Schmallenberg.Fraunhofer Institut für Umweltchemie und Ökotoxikologie.

Koch H and Weisser P, 2001. Spray deposits of crop protection products on plants – the potential exposure of non-target arthropods. Chemosphere, 44, 307–312.

Koch H and Weisser P, 2004. Die Gesamtoberflaeche in Saumstrukturen als potentielle Retentionsflaeche fuer Driftpartikel, Retention Area Index (RAI). Nachrichtenblatt des Deutschen Pflanzenschutzdienstes, 56, 65–69.

Koch H, Weisser P and Landfried M, 2003. Effect of drift potential on drift exposure in terrestrial habitats. Nachrichtenblatt des Deutschen Pflanzenschutzdienstes, 55, 181–188.

Kodell RL and Gaylor DW, 1999. Combining uncertainty factors in deriving human exposure levels of noncarcinogenic toxicants. Annals of the New York Academy of Sciences, 895, 188–195.

Kroes JG, Van Dam JC, Groenendijk P, Hendriks RFA and Jacobs CMJ, 2008. Swap version 3.2 : theory description and user manual.Wageningen.Alterra. Rep. No. 1649.

Kruess A and Tscharntke T, 1994. Habitat fragmentation, species loss, and biological control. Science 264:1581–1584

Landis DA, Gardiner MM, van der Werf W and Swinton SM, 2008. Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes. Proceedings of the National Academy of Sciences USA, 105, 20552–20557.

Larson BMH, Kevan PG, Inouye DW, 2001. Flies and flowers: taxonomic diversity fo anthophiles and pollinators. Canadian Entomologist, 133 (4), 439-465.

Laskowski R, 2001. Why short-term bioassays are not meaningful – effects of a pesticide (imidacloprid) and a metal (cadmium) on pea aphids (*Acyrthosiphon pisum* Harris). Ecotoxicology, 10, 177–183.

Laskowski R and Hopkin SP, 1996. Accumulation of Zn, Cu, Pb and cadmium in the garden snail (Helix aspersa): implications for predators. Environ Pollut 91, 289-297.

Laskowski R, Pyza E, Maryanski M and Niklinska M, 1998. Test on the Centipede Lithobius mutabilis L. Koch 1852 (Chilopoda:Lithobiidae). In: Handbook of soil invertebrate toxicity. Eds Løkke H and Van Gestel CAM. John Wiley & Sons, Chichester, UK, 181–196.

Lefebvre M, Espinosa M, Gomez-y-Paloma S, Piorr A and Zasada I, 2013. Agricultural landscapes as multi-scale public good and the role of the Common Agricultural Policy. Proceedings of the AIEAA Conference ‘Between crisis and development: which role for the bio-economy’, 6–7 June, Parma, Italy.

Leistra, M. 2005. Estimating input data for computations on the volatilisation of pesticides from plant canopies and competing processes.Wageningen.Alterra. Rep. No. 1256.

Leon CD and Van Gestel CAM, 1994. Selection of a set of laboratory ecotoxicity tests for the effects assessment of chemicals in terrestrial ecosystems. Report no D94004, Department of Ecology and Ecotoxicology, Vrije Universiteit, Amsterdam, the Netherlands.

Lever JJ, van Nes EH, Scheffer M and Bascompte J, 2014. The sudden collapse of pollinator communities. Ecology Letters, 17, 350–359. doi: 10.1111/ele.12236

Linden AMA, Boesten JJTI, Brock TCM, Eekelen GMA, ter Horst MMS, de Jong FMW, Montforts MHMM and Pol JW, 2008. Revised proposal for the risk assessment of persistence of plant protection products in soil. RIVM report no 601712003/2008.

Longley M, Jepson PC, Izquierdo J and Sotherton N, 1997. Temporal and spatial changes in aphid and parasitoid populations following applications of deltamethrin in winter wheat. Entomologia Experimentalis et Applicata 83, 41-52.

Longley M and Sotherton W, 1997. Factors determining the effects of pesticides upon butterflies inhabiting arable farmland. Agriculture, Ecosystems & Environment 61, 1-12.

Losey JE and Vaughan M, 2006. The economic value of ecological services provided by insects. BioScience, 56, 311–323.

Louis F and Ufer A, 1995. Methodical improvements of standard laboratory tests for determining the side effects of agrochemicals on predatory mites (acari: Phytoseiidae). Anzeiger Fuer Schaedlingskunde Pflanzenschutz Umweltschutz, 68, 153–154.

Lundgren JG, Shaw JT, Zaborski ER and Eastman CE, 2006. The influence of organic transition systems on beneficial ground-dwelling arthropods and predation of insects and weed seeds. Renewable Agriculture and Food Systems, 21, 227–237.

Ma WC, Immerzeel J and Bodt J, 1995. Earthworm and food interactions on bioaccumulation and disappearance in soil of polycyclic aromatic hydrocarbons: studies on phenanthrene and fluoranthene. Ecotoxicology and Environmental Safety 32, 226-232.

Macfadeyen S and Muller W, 2013. Edges in agricultural landscapes: species interactions and movement of natural enemies. PLoS ONE, 8, e59659. doi:10.1371/journal.pone.0059659

Maier DS, 2012. What's so good about biodiversity? A call for better reasoning about nature's value. Springer, Heidelberg, 568 pp.

Makoto K, Takimura A and Kawakita A, 2003. An obligate pollination mutualism and reciprocal diversification in the tree genus *Glochidion* (Euphorbiaceae). Proceedings of the National Academy of Sciences of the United States of America, 100, 5264–5267.

Manachini B, 2003. Effects of transgenic corn on Lydella thompsoni Herting (*Diptera:Tachinidae*) parasitoid of Ostrinia nubilalis Hb. (*Lepidoptera Crambidae*). Bollettino di Zoologia Agraria e di Bachicoltura, 35, 111–125.

Manachini B, Billeci N and Palla F, 2013. Exotic insect pests: the impact of the red palm weevil on natural and cultural heritage in Palermo (Italy). Journal of Cultural Heritage, 14, Supplement, e177–e182.

Marino PC and Landis DA, 1996. Effect of landscape structure on parasitoid diversity and parasitism in agroecosystems. Ecological Applications, 6, 276–284.

Martin EA, Reineking B, Seo B and Steffan-Dewenter I, 2013. Natural enemy interactions constrain pest control in complex agricultural landscapes. Proceedings of the National Academy of Sciences USA, 110, 5534–5539. doi: 10.1073/pnas.1215725110

MEA (Millennium Ecosystem Assessment), 2005a. Ecosystems and human well-being: biodiversity synthesis. World Resources Institute, Washington, DC, 86pp.

MEA (Millennium Ecosystem Assessment), 2005b. Cultural and amenity services. Chapter 17. World Resources Institute, Washington, DC, 455–476.

Mead-Briggs M and Longley M, 1997. A standard 'extended laboratory' test to evaluate the effects of plant protection products on adults of the parasitoid, aphidius rhopalosiphi (hymenoptera, braconidae) (unpublished).

Mead-Briggs MA, Brown K, Candolfi MP, Coulson MJM, Miles M, Moll M, Nienstedt K, Schuld M, Ufer A and McIndoe E, 2000. A laboratory test for evaluating the effects of plant protection products on the parasitic wasp, *Aphidius rhopalosiphi* (DeStephani-Perez) (Hymenoptera: Braconidae). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative – IOBC/WPRS, Gent, Belgium, 13–35.

Mead-Briggs MA, Moll M, Grimm C, Schuld M, Ufer A and Walker H, 2010. An extended laboratory test for evaluating the effects of plant protection products on the parasitic wasp, Aphidius rhopalosiphi (Hymenoptera, Braconidae). Biocontrol, 55, 329–338.

Meissle M, Álvarez-Alfageme F, Malone LA and Romeis J, 2012. Establishing a database of bio-ecological information on non-target arthropod species to support the environmental risk assessment of genetically modified crops in the EU. EFSA Supporting Publications 2012, EN-334, 170 pp.

Merrill SC, Walter SM and Peairs FB, 2013. The distribution of European corn borer (Lepidoptera: Crambidae) moths in pivot-irrigated corn. Journal of Economic Entomology, 106, 2084–2092.

Metge K and Heimbach U, 1998. Tests on Staphylinid *Philonthus cognatus* Steph. 1832. In: Handbook of soil invertebrate toxicity. Eds Løkke H and Van Gestel CAM. John Wiley & Sons, Chichester, UK, 157–179.

McGeoch MA and Gaston KJ, 2000. Edge effects on the prevalence and mortality factors of Phytomyza ilicis (Diptera, Agromyzidae) in a suburban woodland. Ecology Letters 3, 23-29.

Mineau P and Whiteside M, 2013. Pesticide acute toxicity is a better correlate of U.S. grassland bird declines than agricultural intensification. PLoS ONE, 8, e57457.

Moreby SJ, Aebischer NJ, Southway SE and Sotherton NW, 1994. A comparison of the flora and arthropod fauna of organically and conventionally grown winter wheat in southern England. Annals of Applied Biology, 125, 13–27.

Morris AJ, Wilson JD, Whittingham MJ and Bradbury RB, 2005. Indirect effects of pesticides on breeding yellowhammer (*Emberiza citrinella*). Agriculture, Ecosystems & Environment, 106, 1–16.

Murcia C, 1995. Edge effects in fragmented forests: implications for conservation. Trends Ecol. Evol. 10:58–62

Myers JH, Higgins C, and Kovacs E, 1989. How many insect species are necessary for the biological control of insects? Environ. Entomol. 18: 541-547.

Naranjo SE, 2005. Long-term assessment of the effects of transgenic Bt cotton on the function of the natural enemy community. Environmental Entomology, 34, 1211–1223.

Natori Y and Chenoweth R, 2008. Differences in rural landscape perceptions and preferences between farmers and naturalists. Journal of Environmental Psychology, 28, 250–267.

Norfolk O and Gilbert F, 2014. Insect visitation rates to wild flowers increase in the presence of arid agriculture in South Sinai, Egypt. Journal of Arid Environments, 109, 83–87.

Norris RF and Kogan M, 2005. Ecology of interactions between weeds and arthropods. Annual Review of Entomology, 50, 479–503.

O'Connor RJ and Shrubb M, 1990. Farming and birds. Cambridge University Press, Cambridge, 290pp.

O'Neill K and Rolston M, 2012. Short-term dynamics of behavioral thermoregulation by adults of the grasshopper Melanoplus sanguinipes. Journal of Insect Science, Article 27 1

Odderskaer P, Prang A, Elmegaard N and Andersen PN, 1997. Skylark reproduction in pesticide treated and untreated fields. Comparative studies of skylark *Alauda arvensis* breeding performance in sprayed and unsprayed spring barley fields. National Environmental Research Insititute, Ministry of Environment and Energy, Danish Environmental Protection Agency. Bekaempelsesmiddelforskning fra Miljoestyrelsen, 32, 72 pp.

Ollerton J, Johnson SD, Cranmer L and Kellie S, 2003. The pollination ecology of an assemblage of grassland asclepiads in South Africa. Annals of Botany, 92, 807–834.

Oomen PA, 1988. Guideline for the evaluation of sideeffects of pesticides *Phytoseiulus persimilis*. Bulletin OILB/SROP, 11, 51–64.

Östman Ö, Ekbom B and Bengtsson J, 2003. Yield increase attributable to aphid predation by ground-living polyphagous natural enemies in spring barley in Sweden. Ecological Economics, 45, 149–158.

Plantureux S and Amiaud B, 2010. e-Flora-sys, a website tool to evaluate the agronomical and environmental value of grasslands. In: 23th General meeting of the European Grassland Federation, Kiel (Allemagne).

Potts GR and Aebischer NJ, 1991. Modelling the population dynamics of the grey partridge: conservation and management. In: Bird population studies. Eds Perrins CM, Lebreton JD and Hirons GJM. Oxford University Press, Oxford, UK, 373–390.

Poulin B, 2012. Indirect effects of bioinsecticides on the nontarget fauna: The Camargue experiment calls for future research. Acta Oecologica, 44, 28–32.

Pullen AJ, Jepson PC and Sotherton NW, 1992. Terrestrial non-target invertebrates and the autumn application of synthetic pyrethroids: Experimental methodology and the trade-off between replication and plot size. Archives of Environmental Contamination and Toxicology, 23, 246–258.

Pulliam HR, 1988. Sources, sinks, and population regulation. American Naturalist, 132, 652–661.

Rands MRW, 1986. The survival of gamebird (Galliformes) chicks in relation to pesticide use on cereals. Ibis, 128(1), 57-64.

Rauch G and Weisser WW, 2007. Local and spatial dynamics of a host–parasitoid system in a field experiment. Basic and Applied Ecology, 8, 89–95.

Rautmann D, Streloke M and Winkler R, 2001. New basic drift values in the authorisation procedure for plant protection products. In Workshop on Risk Assessment and Risk Mitigation Measures in the Context of the Authorisation of Plant Protection Products (WORMM). . Mitteilungen aus der Biologischen Bundesanstalt fur Land und Forstwirtschaft 381, 133-141.

Reisig DD, Roe M and Dhammi A, 2013. Dispersal pattern and dispersion of adult and nymph stink bugs (Hemiptera: Pentatomidae) in wheat and corn. Environmental Entomology, 42, 1184–1192.

Ricketts TH, 2001. The matrix matters: Effective isolation in fragmented landscapes. American Naturalist, 158, 87–99.

Ricoua C, Schnellera C, Amiaudc B, Plantureuxa S and Bockstaller C, 2014. A vegetation-based indicator to assess the pollination value of field margin flora. Ecological Indicators, 45, 320–331.

Rieckmann W, 1989. Auswirkungen von pflanzenschutzmittel auf die schwebfliege *Syrphus corollae* (fabr.) im laboratorium. Guideline part vi, 23-2.1.7 of the Federal German Biological Agency for Agriculture And Forestry.

Ries L and Sisk TD, 2004. A predictive model for edge effects. Ecology 85:2917–26

Ries L and Sisk TD, 2010. What is an edge species? The implications of sensitivity to habitat edges. Oikos 119:1636-1642.

Robillard A, Garant D and Bélisle M, 2012. The Swallow and the Sparrow: how agricultural intensification affects abundance, nest site selection and competitive interactions. Landscape Ecology, 28, 201–215.

Rodrìguez MÁ and Hawkins BA, 2000. Diversity, function and stability in parasitoid communities. Ecology Letters, 3, 35–40.

Rogge E, Nevens F and Gulinck H, 2007. Perception of rural landscapes in Flanders: Looking beyond aesthetics. Landscape and Urban Planning, 82, 159–174.

Rondeau G, Sanchez-Bayo F, Tennekes HA, Decourtye A, Ramirez-Romero R and Desneux N, 2014. Delayed and time-cumulative toxicity of imidacloprid in bees, ants and termites. Scientific Reports, 4, 5566. doi: 10.1038/srep05566.

Rossetti RM, Salvo A and Videla M, 2013. Forest remnants contribute to parasitoid conservation: experimental evaluation of parasitism on a leafminer host. Journal of Insect Conservation, 17, 1137–1144.

Rossiter MC, 1987. Host by non-outbreak populations of the gypsy moth. Ecology, 68, 857–868.

Roth D, Roland J and Roslin T, 2006. Parasitoids on the loose - experimental lack of support of the parasitoid movement hypothesis. Oikos 115:277 - 285

Rubach MN, Ashauer R, Buchwalter DB, De Lange HJ, Hamer M, Preuss, TG, Töpke K and Maund SJ, 2011. Framework for traits-based assessment in ecotoxicology. Integrated Environmental Assessment and Management, 7, 172-186.

Sáenz-de-Cabezón Irigaray FJ and Zalom FG, 2006. Side Effects of Five New Acaricides on the Predator Galendromus occidentalis (Acari, Phytoseiidae). Experimental & Applied Acarology 38, 299-305

Sáenz-de-Cabezón Irigaray FJ, Zalom FG, Moreno-Grijalba F, Marco V, Pérez-Moreno I and Carvajal Montoya L, 2008. Terrestrial bioaccumulation: experimental approaches on an arthropod prey-predatory mite system. Proceeding Endure Conference, Diversifying crop protection, O-59

SANCO (Directorate General for Health and Consumer Affairs), 2002. Guidance document on terrestrial ecotoxicology under Council Directive 91/414. Draft Working Document. EU (DG Health and Consumer Protection), Brussels, Belgium. SANCO/1039/2002 rev 2 final.

Schmidt MH, Lauer A, Purtauf T, Thies C, Schaefer M and Tscharntke T, 2003. Relative importance of predators and parasitoids for cereal aphid control. Proceedings of the Royal Society B, 270, 1905–1909.

Schmitz J, 2008. Auswirkungen der Abdrift von Pflanzenschutzmitteln auf phytophage Insekten am Beispiel von Heuschrecken und Schmetterlingen. Diploma thesis, University Koblenz-Landau, Campus Landau, 103 pp.

Schmitz J, Hahn M and Brühl CA, 2014. Agrochemicals in field margins – An experimental field study to assess the impacts of pesticides and fertilizers on a natural plant community. Agriculture Ecosystems & Environment 193: 60-69.

Schmuck R, Candolfi MP, Kleiner R, Jans D, Waltersdorfer A, Mead-Briggs M, Moll M and Wilhelmy H, 2000. A laboratory test system for assessing effects of plant protection products on the plant dwelling insect Coccinella septempunctata L. (Coleoptera: Coccinellidae). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative – IOBC/WPRS, Gent, Belgium, 45–70.

Shackelford G, Steward PR, Benton TG, Kunin WE, Potts SG, Biesmeijer JC and Sait SM, 2013. Comparison of pollinators and natural enemies: a meta-analysis of landscape and local effects on abundance and richness in crops. Biological Reviews, 88, 1002–1021. doi: 10.1111/brv.12040

Slob W and Pieters MN, 1998. A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Risk Analysis, 18, 787–798.

Smart LE, Stevenson JH and Walters JHH, 1989. Development of field trial methodology to assess short-term effects of pesticides on beneficial arthropods in arable crops. Crop Protection, 8, 169–180.

Snyder WE and Ives AR, 2003. Interactions between specialist and generalist natural enemies: parasitoids, predators, and pea aphid biocontrol. Ecology, 84, 91–107.

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.

Southcombe ESE, Miller PCH, Ganzelmeier H, Van de Zande JC, Miralles A and Hewitt AJ, 1997. The international (BCPC) spray classification system including a drift potential factor. In 'Brighton Crop Protection Conference - Weeds', pp. 371-380, Brighton.

Sommaggio D and Burgio G, 2014. The use of *Syrphidae* as functional bioindicator to compare vineyards with different managements. Bulletin of Insectology, 67, 147–156.

Stark JD, Chen XD and Johnson CS, 2012. Effects of herbicides on Behr's metalmark butterfly, a surrogate species for the endangered butterfly, Lange's metalmark. Environmental Pollution, 164, 24–27.

Stark JD, JEPSON PC; MAYER DF, 1995. Limitations to Use of Topical Toxicity Data for Predictions of Pesticide Side Effects in the Field. Journal of Economic Entomology, Volume 88, Number 5, pp. 1081-1088(8)

Swarowsky K, Süßenbach D and Streissl F, 2013. How protective is NTA risk assessment? Poster presentation on the SETAC Europe 23rd Annual Meeting, 12–16 May, Glasgow, United Kingdom.

Swift TL and Hannon SJ, 2010. Critical thresholds associated with habitat loss: a review of the concepts, evidence, and applications. Biological Reviews, 85, 35–53.

Symondson WOC, Sunderland KD and Geenstone MH, 2002. Can generalist predators be effective biocontrol agents? Annual Review of Entomology, 47, 561–594.

Tajovsky K, 1998. Test on the Millipede Brachydesmus superus Latzel 1884. In: Handbook of soil invertebrate toxicity. Eds Løkke H and Van Gestel CAM. John Wiley & Sons, Chichester, UK, 197–205.

Tartaglia ES and Handel NS, 2014. Nectar plant preferences of Hemaris (Sphingidae) on co-occurring native Cirsium and non-native Centaurea (Asteraceae) inflorescences. Journal of Pollination Ecology, 5, 2011, 86–89.

Tempesta T, 2010. The perception of agrarian historical landscapes: A study of the Veneto plain in Italy. Landscape and Urban Planning, 97, 258–272.

Thacker JRM and Jepson PC, 1993. Pesticide risk assessment and non-target invertebrates: Integrating population depletion, population recovery, and experimental design. Bulletin of Environmental Contamination and Toxicology, 51, 523–531.

Thomson LJ and Hoffmann AA, 2010. Natural enemy responses and pest control: importance of local vegetation. Biological Control 52:160-166.

Thorbek P and Topping CJ, 2005. The influence of landscape diversity and heterogeneity on spatial dynamics of agrobiont linyphiid spiders: an individual-based model. Biocontrol, 50, 1–33.

Tiktak A, van den Berg F, Boesten JJTI, van Kraalingen D, Leistra M, and van der Linden A. M. A. 2000. Manual of FOCUS PEARL version 1.1.1.Bilthoven.RIVM. Rep. No. report 711401008.

Tilman D, May RM, Lehman CL and Nowak MA, 1994. Habitat destruction and the extinction debt. Nature, 371, 65–66.

Toft S and Jensen AP, 1998. No negative sublethal effects of two insecticides on prey capture and development of a spider. Pesticide Science, 52, 223-228.

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, Kjær LJ, Hommen U, Høye TT, Preuss TG, Sibly RM and van Vliet P, 2014. Recovery based on plot experiments is a poor predictor of landscape-level population impacts of agricultural pesticides. Environmental Toxicology and Chemistry, 33, 1499–1507.

Tscharntke T, Tylianakis JM, Rand TA, Didham RK, Fahrig L, Batáry P, Bengtsson J, Clough Y, Crist TO, Dormann CF, Ewers RM, Fründ J, Holt RD, Holzschuh A, Klein AM, Kleijn DA, Kremen C, Landis DA, Laurance W, Lindenmayer D, Scherber C, Sodhi N, Steffan-Dewenter I, Thies C, van der Putten WH and Westphal C, 2012. Landscape moderation of biodiversity patterns and processes – eight hypotheses. Biological Reviews, 87, 661–685. doi: 10.1111/j.1469-185X.2011.00216.x

Tuomisto HL, Hodge ID, Riordan P and Macdonald DW, 2012. Does organic farming reduce environmental impacts? – A meta-analysis of European research. Journal of Environmental Management, 112, 309–320.

Valladares F, Sanchez D and Zavala MA, 2006. Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. Journal of Ecology 94:1103–1116.

van Berkel DB and Verburg PH, 2014. Spatial quantification and valuation of cultural ecosystem services in an agricultural landscape. Ecological Indicators, 37, 163–174.

van Lexmond MB, Bonmatin J-M, Goulson D and Noome DA, 2015. Worldwide integrated assessment on systemic pesticides. Global collapse of the entomofauna: exploring the role of systemic insecticides. Environmental Science and Pollution Research, 22, 1–4. doi 10.1007/s11356-014-3220-1

van de Zande JC, Holterman HJ and Huijsmans JFM, 2012. Spray drift for the assessment of the exposure of aquatic organisms to plant protection products in the Netherlands. Part 1: Field crops and downward spraying. .Wageningen.WUR-PRI Rep. No. 419.

van de Zande JC, Rautmann D, Holterman HJ and Huijsmans JFM, 2014. Joined spray drift curves for boom sprayers in The Netherlands and Germany.Wageningen.Plant Research International. Rep. No. 526.

van Zanten BT, Verburg PH, Espinosa M, Gomez-y-Paloma S, Galimberti G, Kantelhardt J, Kapfer M, Lefebvre M, Manrique R, Piorr A, Raggi M, Schaller L, Targetti S, Zasada I and Viaggi D, 2013. European agricultural landscapes, common agricultural policy and ecosystem services: a review. Agronomy for Sustainable Development, 34, 309–325.

Verboom J, Schippers P, Cormont A, Sterk M, Vos CC and Opdam PFM, 2010. Population dynamics under increasing environmental variability: implications of climate change for ecological network design criteria. Landscape Ecology, 25, 1289–1298.

Vogt H, 2000. Sensitivity of non-target arthropods species to plant protection products according to laboratory results of the IOBC WG 'Pesticides and beneficial organisms'. IOBC Bulletin, 23, 3–15.

Vogt H, Bigler F, Brown K, Candolfi MP, Kemmeter F, Kühner ChH, Moll M, Travis A, Ufer A, Viñuela E, Waldburger M and Waltersdorfer A, 2000. Laboratory method to test effects of plant protection products on larvae of *Chrysoperla carnea* (Neuroptera: Chrysopidae). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative – IOBC/WPRS, Gent, Belgium, 27–44.

Wahlberg N, Klemetti T and Hanski I, 2002a. Dynamic populations in a dynamic landscape: the metapopulation structure of the marsh fritillary butterfly. Ecography, 25, 224–232.

Wahlberg N, Klemetti T, Selonen V and Hanski I, 2002b. Metapopulation structure and movements in five species of checkerspot butterflies. Oecologia, 130, 33–43.

Weinstoerffer J and Girardin P, 2000. Assessment of the contribution of land use pattern and intensity to landscape quality: use of a landscape indicator. Ecological Modelling, 130, 95–109.

Weisser P, Koch H, Laun N and Kleisinger S, 2001. Belagsqualität an Gemüsekulturen bei der Applikation von Pflanzenschutzmitteln mit luftunterstützten Feldspritzgeräten. Gesunde Pflanzen, 53, 15–23.

Weyland F and Laterra P, 2014. Recreation potential assessment at large spatial scales: A method based in the ecosystem services approach and landscape metrics. Ecological Indicators, 39, 34–43.

WHO (World Health Organization), 2014. Guidance document on evaluating and expressing uncertainty in hazard characterization. Available online: http://www.who.int/ipcs/methods/harmonization/uncertainty_in_hazard_characterization.pdf

Woodcock BA and Vanbergen AJ, 2008. Parasitism of the beech leaf-miner weevil in a woodland: patch size, edge effects and parasitoid species identity Insect Conservation & Diversity 1 (3) 180-188

Wratten SD., Bowie MH, Hickman JM, Evans AM, Sedcole JR, and Tylianakis JM, 2003. Field boundaries as barriers to movement of hover flies (Diptera: Syrphidae) in cultivated land. Oecologia 134, 605-611.

Zhang W and Swinton SM, 2012. Optimal control of soybean aphid in the presence of natural enemies and the implied value of their ecosystem services. Journal of Environmental Management, 96, 7–16.

Zhao Z-H, Hui C, He D-H, Ge F, 2013. Effects of position within wheat field and adjacent habitats on the density and diversity of cereal aphids and their natural enemies. Biocontrol, 58, 765–776.

APPENDICES

Appendix A. *Bembidion lampros* model

The simulations were run using the ALMaSS system (Topping et al., 2003), a model system designed to provide answers to policy level questions related to changing land-use or management and the resultant impacts on animal wildlife.

The model system

The model used is part of ALMaSS. The ALMaSS project is an open source project hosted on CCPForge (http://ccpforge.cse.rl.ac.uk), where program code can be downloaded. ALMaSS itself is a large system comprising many interacting agent-based models and hence a detailed description cannot be provided here. The reader is therefore directed to the online documentation found at Anon (2014). This documentation follows ODdox format (Topping et al., 2010), combining model description with doxygen (van Heesch, 1997) code documentation. The animal models included in ALMaSS have been tested using a pattern-oriented approach (Grimm et al., 2005; Topping et al., 2010) to maximise confidence in their structure and function. The models are quite detailed in their behaviour and hence run times for ALMaSS can be long, usually measured in hours or even days. This is particularly the case for invertebrate model simulations, which have been recorded as having up to 46 million concurrent agents. The model used for the simulations presented in this report was the *Bembidon lampros* model; however, a short introduction to the whole ALMaSS system is presented below for the purpose of understanding the structure of the overall model.

ALMaSS – short overview

ALMaSS comprises two main components: the environment and its associated classes, and the animal representations (classes). The environment interface is provided by the ‘Landscape’ class. This class contains a map of the landscape to be simulated together with individual landscape elements such as fields, hedges, roads and woodlands. Fields are a special case. Fields are linked in groups to form farms. These groups are typically based on ownership or management information from municipal or EU-farming subsidy sources. Each farm is an instance of the ‘Farm’ class, which simulates the detailed management of its fields, dependent upon its farm type, the weather, soil type and past history of management. There is a degree of stochasticity in farmer decisions, and hence the result is a dynamic pattern of farm management across the landscape, with farmers with the same farm types, growing the same crops, making similar but not identical decisions.

All vegetated landscape elements (crops and non-crops) undergo type-specific daily vegetation developed based on weather and fertiliser inputs as drivers. Farm management events (e.g. harvest or ploughing) directly interact with vegetation height and biomass, providing a dynamic picture of changing landscape conditions as a result of both environmental and anthropogenic processes and factors.

The second main ALMaSS component is the simulation of animals, represented by specific species classes all derived from a common base class. All animals are agents and are affected by environmental variables, vegetation structure, and by direct interaction with other agents or farm management. Each animal represents an individual of a particular species, with its own behavioural rules and interactions with its environment. Animals can sense the characteristics of their environment (habitat type, vegetation structure, temperature, etc.), management events, and their own physiological condition. Hence, animals exposed to management will choose behaviour suitable for that management, their current location, and physiological state. Animals can interact with each other in a variety of ways ranging from simple local density-dependent interactions to complex behavioural messaging, depending upon animal type and current activity. All animals share a common basic form of control simulated as a state machine. This means that they exhibit behaviour associated with a specific state, and make transitions to other behavioural states as a result of internal or external cues.

The *Bembidion lampros* model

B. lampros is a Palearctic species that has been studied intensively as one of the most common beetles in European agroecosystems, and therefore has a well-described biology and natural history. The parameters are based on field and laboratory data on *B. lampros* from the literature. When necessary data were not available, parameters were estimated based on published data on other ground beetles with similar life histories. It is considered to be a useful natural enemy of pests in agricultural fields (e.g. Edwards et al., 1979; Ekbom et al., 1992; Humphreys and Mowat, 1994).

In common with all other ALMaSS animal models, the *Bembidion* model's individuals are agents designed to simulate the ecology and behaviour of individual beetles. Each agent moves around in its virtual world in much the same way that a real animal would, picking up information from its surroundings as it goes and acting upon this to achieve the twin goals of survival and reproduction. Since the environment is dynamic, the resultant response of the sum of the agents' interactions with each other and their environment, through space and time, produces an emergent population response. The original model was described in Bilde et al. (2004), and full documentation is available in ODdox format from http://www2.dmu.dk/ALMaSS/ODDox/Bembidion/index.html.

Bembidion behaviour is characterised by annual dispersal and aggregation phases with aggregation linked to non-cultivated habitats, and dispersal and breeding largely occurring in open fields. Primary drivers in the model are temperature-controlled developmental rates of eggs, larvae and pupae, together with adult beetle interactions with the landscape. Each individual beetle reacts to the local environmental drivers of beetle density within a 2 m radius, and to landscape management and global weather drivers.

Model Procedures

Day-degree calculations are driven by actual weather data stored as daily records of mean temperature, mean wind speed and total precipitation.

Reproduction

Only beetles present in a preferred habitat (such as open agricultural fields) in spring may initiate reproduction. The rate of egg production, $f_{(egg)}$, is temperature dependent and was calculated as:

(eq.1) $f_{(egg)}(T) = (T - T_0) * C$

T, temperature; T_0, lower threshold for egg production; C, egg production slope.

Development

Temperature-dependent development was calculated daily for each developmental stage: egg, first, second and third instar and pupa with a transition to the next stage when the $\sum f_{(dev)}$ for that stage was greater or equal to 1.0. Development $f_{(dev)}$ was calculated as:

(eq.2) $f_{(dev)}(T) = T - T_0/L$

T, temperature; T_0, lower threshold for development; L, duration of stage in day-degrees.

Movement

Movement patterns were based on behavioural decisions and preference for habitat. Movement was determined by five parameters:

1. a directional vector that indicates the preferred direction;
2. a weight indicating the strength of the bias towards the directional vector;
3. a maximum allowed distance per time step;

4. a quality assessment in terms of attractiveness for the moving beetle (this is dynamic and may change e.g. between reproduction and migration behaviours);

5. the probability of a beetle accepting a sub-optimal habitat e.g. to cross a road.

For each possible direction, the habitat was assessed and the beetle was moved to one of the locations based on the quality assessment (or chance of accepting a sub-optimal step). If more than one location of similar quality was available, movement to a particular location was chosen at random.

Mortality

Beetles could die because of external events, density-dependent factors (i.e. competition) or because they reached the end of their life span. External factors causing mortality were either farm operations, mainly soil cultivation operations, or temperature-dependent winter mortality. The different cultivation methods were assigned different probabilities of beetle mortality (Thorbek and Bilde, 2004), which were implemented when the farm operation in question occurred.

Density dependence was introduced by limiting the number of beetles present in the area surrounding any beetle. Both the size of the area tested and the number of beetles allowed in the area are system parameters and can be used to control overall population size. Excess beetles were removed from the simulation. Tests indicated that, at the densities usually used for beetles, altering these parameters does not impact on results (Bilde and Topping, 2004). Background mortality was implemented for the egg stage and the three larval instars. Beetles were assessed for density-dependent and background mortality once per simulation day. In order to avoid bias resulting from concurrency problems, the order of assessment of beetles was randomised once per time step.

Habitat

Habitat preference depended on life history requirements. In autumn, beetles migrated to vegetated field boundaries for hibernation. Autumn migration was determined by a probability distribution starting on 1 October, which is consistent with empirical data, although it is not known what triggers this behaviour. In spring, beetles exhibited directional movement into agricultural fields to reproduce. Spring migration was temperature dependent.

Scenario set-up

Pesticide

The pesticide properties were chosen both to highlight the issues to be addressed, but also to be realistic in terms of action. No drift to off-field areas was assumed in order to completely isolate source-sink dynamics as drivers of change in off-field areas. Other insecticides, applied to winter wheat, and normal herbicide and fungicide applications were assumed to have no impact on beetles.

An 80 % field mortality rate for a foliar insecticide spray measured over seven days was chosen. Available regulatory field data indicate that this should be considered a realistic value rather than a realistic worst-case. We assumed a DT_{50} of 10 days and an application rate of twice the LR80 to all winter wheat fields, applied twice during the activity time of the adult beetles, the first on 31 May, the second 20 days later. No interception by the vegetation was assumed, so the available concentration for the beetle on the day of spraying was twice the LD_{80}, in this case a value of 50 units (the actual units are immaterial here).

For a subset of scenarios, toxicity was assumed to be increased by factors of 2, 5 and 10. These settings were used with scenarios with zero field boundaries only (FB0, see below). Thus, the trigger value for mortality chance was reduced from 25, to 12.5, to 5 and, finally, to 2.5 units. Decay rates and mortality rates were unchanged, hence the result of increasing toxicity was, in this case, to extend the period over which the pesticide was toxic, but still with an 80 % chance of a beetle dying over a seven day period above the toxic threshold.

A simplifying assumption that eggs, larvae and pupae were in soil and not exposed was used. Thus, in all cases, mortality because of pesticidal action was implemented only for adults.

Crops

All scenarios were run assuming that the landscape contained only a monoculture of winter wheat, which was either treated with the insecticide to be evaluated, or untreated in the baseline scenarios. In the case of the baseline scenarios where no treatment with the product occurred, this was not replaced by another insecticide. This crop was, however, managed following otherwise normal Danish agricultural practice, including herbicides, fungicides, soil tillage and harvest. These agricultural management practices were included as resulting in mortalities of 27 % for soil tillage operations and 36 % for harvest, as measured by Thorbek and Bilde (2004) for ploughing and grass cutting. Herbicides and fungicides were assumed to be non-toxic.

Landscapes

Two different landscapes were chosen for the simulations. These landscapes differed in both composition and arrangement of landscape elements. The Herning landscape has a mean field size of 3.32 ha, a maximum field size of 33.9 ha and a total of 3 022 arable fields. Præstø has a mean field size of 7.77 ha, a maximum field size of 136.6 ha and 915 arable fields in total. The structure of the off-field habitats also differs, with large wooded areas in the Prætø landscape, and heathland and small woodlots in the Herning landscape (Figure A1, Table A1).

Table A1: Percentage cover by area of each landscape element type in the two landscapes used in this study

Landscape Element Type	Herning	Præstø
Bushes/scrub	0.8	0.3
Fields (rotation) **	70.5	66.1
Heath *	3.4	0.0
Linear features (excl. hedge banks) *	3.9	2.5
Hedge bank *	0.9	0.3
Permanent Pasture *	1.2	0.0
Unmanaged grassland *	2.6	2.5
Urban	4.6	6.4
Water	0.6	0.7
Wetland	2.1	1.2
Woodland	8.6	19.6
Woodland plantation	1.7	0.6
Total	100.0	100.0

* Indicates suitable breeding habitat for beetles.
** Optimal breeding habitat.

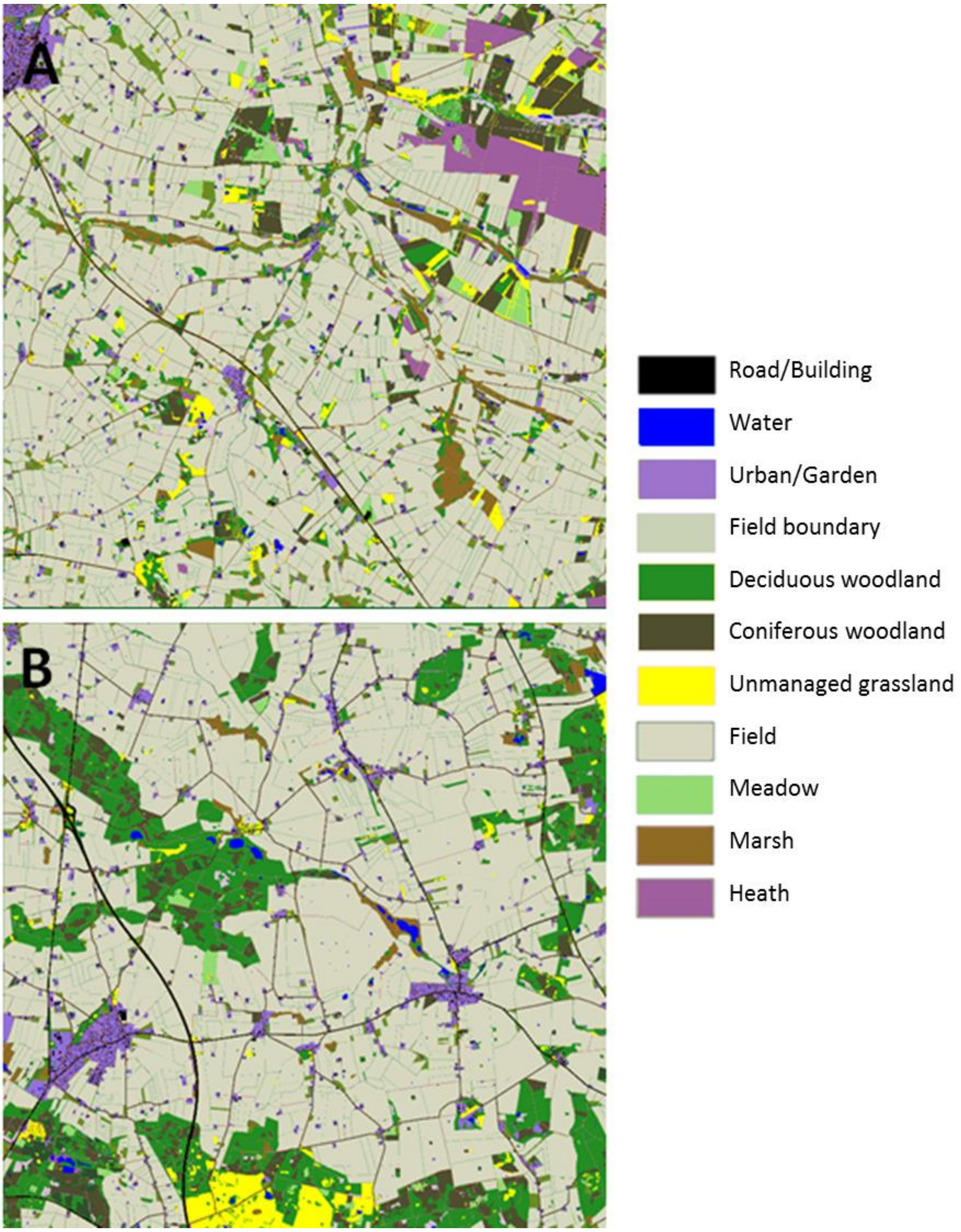


Figure A1: (A) Herning landscape; (B) Præsto landscape. Key denotes major landscape elements

Each landscape was used in two artificially manipulated forms. The first was with all grassy field boundaries removed (designated FB0), the second was with a grassy field boundary inserted around all fields (designated FB100, which was derived from FB0). These boundaries were applied to all fields in three widths, 1 m, 5 m and 10 m (FB100X1, FB100X5 and FB100X10, respectively). The resulting area cover for field boundaries in the two landscapes was markedly different, ranging from 1.0 % to 17.5 % cover as a proportion of field area (Table A2).

A further manipulation was to use FB100X1 versions of both landscapes but to add 2 m, 5 m or 10 m unsprayed cropped margins (USM) to all fields (designated as _USM2, _USM5 and _USM10).

Unsprayed margins contained crop and were treated identically to the crop in every way, except that pesticides were not applied to the margin.

Table A2: Percentage by area of arable field for non-cropped field boundaries for the Herning and Præstø landscapes

Field Boundary Width (m)	**Landscape**	
	Herning (He)	**Præstø (Pr)**
1	1.8%	1.0%
5	8.8%	5.2%
10	17.5%	10.4%

Other settings and replicates

All landscapes described above were simulated with beetles for 30 years with both baseline and product runs. Baseline conditions were identical to the product run except that no insecticide was applied to the winter wheat fields. Data was extracted from the simulations only after the first 10 years of simulation to allow the populations to equilibrate with the landscape (burn-in period). 10 years was determined to be adequate in initial tests.

Weather conditions were selected to represent the decade 1990–1999 from central Denmark. Each simulation was run for a for a total of 30 simulation years, looping the 1990–1999 weather data three times. Note that replicates were identical in terms of initial settings and weather inputs. Variation between replicate outputs was therefore entirely caused by different decisions or stochastic processes e.g. individual mortality chance.

Simulation data extraction

Two main sets of data were extracted from all the simulation runs. The AOR index information and spatial data on the numbers of adult female beetles extant.

Overall population impact

To compute the overall impact, statistical analysis was based on the mean differences between baseline and product simulations runs with time, resulting in an estimate of mean population depression during the second 10 years of pesticide application at the full landscape scale. The procedure was to use the raw adult abundance output from ALMaSS and average for each month over all replicates, then to average these values within each simulation year. The ratio of ‘with pesticide’ to ‘appropriate baseline’ was computed for each year, and the average over the final 10 years was taken. This was then converted to percentage loss. This method provided an estimate of impact relative to baseline, and controls for year-to-year variation caused by weather driven processes within the simulation.

The AOR index

Results from a comprehensive ABM are often themselves complicated and difficult to handle in a management or policy context. To alleviate this problem, ALMaSS output was used to create an index developed from the Abundance:Occupancy Relationship (AOR) (Gaston et al., 2000), often studied in macroecology. The AOR index has the advantage that it provides a clear picture of the changes in range and density of animals relative to a baseline condition (Hoye et al., 2012). Previously, ALMaSS results have been expressed as changes in local abundance and spatial distribution as described by the univariate Ripleys $K(r)$ (Jepsen et al., 2005). However, this approach is both statistically difficult and results in relatively complex outputs. The AOR index was designed to ease the calculation and communication, and works by comparing changes in occupancy and abundance to a baseline scenario.

The baseline acts as a reference against which the impacts of scenario changes can be evaluated, and hence, in order to provide good resolution for both positive and negative changes to the index, should not be based on extreme conditions.

Occupancy is quantified by overlaying the landscape with a regular grid and quantifying the proportion of grid cells containing individuals. The grid cell size should be large enough to allow more than one individual to be present in each grid cell, but small enough to avoid occupancy and abundance being identical. Three rules were used to identify the grid cell size for each model species as follows: (1) there should be a minimum of 250 cells in total; (2) in the baseline scenario no more than 50 % of the cells should be occupied; and (3) if possible within the above constraints, the grid size chosen should result in a mean occupancy of more than five. The resulting grid size for all simulations in this study was 50×50 m or 0.25 ha. Occupancy was quantified by the proportion of grid cells occupied by at least one adult female for each annual recording (in this case on 1 January) of the locations of individuals of a species averaged across a 10-year simulation period from year 20 to year 30. Abundance was calculated as the mean number of females in grid cells where individuals were present. The result can be recorded and translated into plots of AOR, indicating changes in abundance and occurrence relative to the baseline condition, and is typically expressed as a 2-D plot.

Temporal effects

To determine the extent to which year-on-year application of the pesticide resulted in instantaneous or long-term effects, the impact relative to baseline was used and compared over time. However, because of large annual fluctuations caused by the weather it was necessary to eliminate weather effects for this analysis. Since the weather cycle was repeated after 10 years, comparing the impacts between like weather years was necessary. Therefore, the ratio of impact relative to baseline from each year following pesticide application to the corresponding year 10 years later was taken. The analysis was carried out for Herning and Præstø FB0 landscapes with increased toxicity by factors of 2, 5 and 10.

REFERENCES

Anon, 2014 ALMaSS Model Documentation. Available online: http://bios.au.dk/om-instituttet/organisation/faunaoekologi/projekter/almass/documentation/

Bilde T and Topping C, 2004. Life history traits interact with landscape composition to influence population dynamics of a terrestrial arthropod: A simulation study. Ecoscience, 11, 64–73.

Edwards CA, Sunderland KD and George KS, 1979. Studies on polyphagous predators of cereal aphids. Journal of Applied Ecology, 16, 811–823.

Ekbom BS, Wiktelius S and Chiverton PA, 1992. Can polyphagous predators control the bird cherry-oat aphid (rhopalosiphum-padi) in spring cereals - a simulation study. Entomologia Experimentalis Et Applicata, 65, 215–223.

Gaston K, Blackburn T, Greenwood J, Gregory R, Quinn R and Lawton J, 2000. Abundance–occupancy relationships. Journal of Applied Ecology 37, 39–59.

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

Hoye TT, Skov F and Topping CJ, 2012. Interpreting outputs of agent-based models using abundance–occupancy relationships. Ecological Indicators 20, 221–227.

Humphreys IC and Mowat DJ, 1994. Effects of some organic treatments on predators (Coleoptera, Carabidae) of cabbage root fly, Delia readicum (L) (Diptera, Anthomyiidae), and on alternative prey species. Pedobiologia, 38, 513–518.

Jepsen JU, Topping CJ, Odderskaer P and Andersen PN, 2005. Evaluating consequences of land-use strategies on wildlife populations using multiple-species predictive scenarios. Agriculture Ecosystems & Environment, 105, 581–594.

Thorbek P and Bilde T, 2004. Reduced numbers of generalist arthropod predators after crop management. Journal of Applied Ecology, 41, 526–538.

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, Hoye 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.

van Heesch D, 1997. Doxygen. Available online: www.doxygen.org

Appendix B. Qualitative assessment of uncertainty in the ecological modelling of *Bembidion lampros* exercise

The purpose of the modelling exercise was to investigate:

1. effects in the off-field from in-field mortality;
2. effects of spatial scales;
3. underestimation of long-term effects from one year studies.

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 risk 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 effects		**Potential to overestimate the real effects**	
		Score	**Explanation**	**Score**	**Explanation**
Model uncertainty (see examples below)					
Model assumptions					
Any other model uncertainty?					
Parameter uncertainty (see examples below)					
Input variables (toxicity/effect data)	Variability/uncertainty of toxicity endpoints	+	Intraspecies sensitivity distribution not considered	+	Intraspecies sensitivity distribution not considered
	Intra – lab, inter- lab and inter species variations		Not relevant because toxicology data are from field studies		Not relevant because tox data are from field studies
	Whole dose-response curve: single endpoint derived	++	Threshold was applied		
	Type of data e.g. acute or chronic standard tests, semi-field/field data	++	Chronic effects are not considered		
	Estimation of values (e.g. literature, measured, estimated by calibration)	+	No dose response tests but low uncertainty around the outcome – good field trials	+	No dose response tests but low uncertainty around the outcome – good field trials
	Uptake, elimination rates		Mortality was measured in field studies therefore uptake and elimination rates are realistic		

		Potential to underestimate the real effects		Potential to overestimate the real effects	
		Score	**Explanation**	**Score**	**Explanation**
Input variables (other than toxicity/effect data)	Variability of most relevant parameters (tested; expected in the field)			+	Landscape structure in model simpler than in reality
		+++	No spray drift		
		++	Timing of spray in relation to life history	++	Timing of spray in relation to life history
				++	Number of sprays (2 times applied while in DK; usually there are 1–2 applications of insecticides)
		+	Monoculture (there could be crops which are worse for beetles than winter wheat)	++	Monoculture (usually more diverse crops)
	Gaps in measurement/measurement errors		Not relevant		
	Estimation of values (e.g. from literature, directly measured, estimated)	+	Life history data and ecological data are well established	+	Life history data and ecological data are well established
Life history characteristics (e.g. duration of life cycle) and migration/movement pattern			Good and reliable data are available		Good and reliable data are available
Life stage sensitivity and size		+	Only adult beetles are considered to be exposed		
Sensitivity to the chemical (e.g. data from species other than modelled species)			Not relevant		
Presence at time of exposure or when substance accumulated in the environment (e.g. exposed to PEC maximum)			Yes		Yes
Potential chronic, delayed, cumulative and carry-over of effects		++	Chronic effects not considered		

Model outputs (examples see below)					
Not relevant, model specifically designed to produce the required outputs					
Assessment + Environmental scenario (see examples below)					
Biological considerations	Abiotic stressors (context dependency) (e.g. agricultural management, resources, drying, pH, low oxygen, isolation vs. connectivity)	+	Weather cycles are from the past (weather could be more or less favourable in future)	+	Weather cycles are from the past (weather could be more or less favourable in future)
	Biotic stressors, depending on level of organisation (e.g. intra- or inter-specific competition, predation)	+	No other stressors are included	+	No other stressors are included
Exposure considerations	Exposure routes		Mortality measured in field studies therefore all exposure routes are covered		
	Exposure regime (e.g. pattern, concentration, duration in relation to type of effects)		The exposure is simulated very precisely		
	Exposure scale (currently exposure scenarios are not designed for landscape level)		Model is run at landscape level		
Spatial scale	Spatial scale (e.g. edge-of-field ditch for individual effects; larger scale for effects on populations or communities)		Large spatial scale used		
	Landscape structure (e.g. connectivity, off-field size as sources of recolonisation)		Explicitly incorporated		
Temporal scale	Temporal scale e.g. seasonality of effects, carry-over effects		Explicitly investigated		
Integration of exposure and effects (e.g. animal behaviour altering exposure)			Explicitly incorporated		
Any other uncertainties related to assessment and environmental scenario?				+	All fields are treated at the same day
Multiple PPP exposure (examples see below)					
Different applications of multiple PPP (e.g. combined or successive)		++	Not considered		
Output of model validation (examples see below)					
Comparison of model outputs to suitable independent datasets (e.g. baseline data and toxicity data)?			Few data are available to compare model outputs with field data at landscape scale.		
Comparison of model outputs to data of sufficiently contrasting scenario?			?		?
Fit of the model predictions to observed data patterns			Field studies are available which show similar effects at smaller		

Model outputs (examples see below)	
	scales
Any other model validation related uncertainties?	Not aware of any
Domain of applicability	
Extrapolations (i.e. list all extrapolations which have been made)	No extrapolation to other scenarios
Any other type of uncertainties?	Not aware of any
Overall assessment	The real effects are more likely to be underestimated than overestimated because spray drift, larvae mortality and dose–response mortality were not included

Appendix C. Articles SPG Pest Control

Table C1: Summary of selected articles that quantified pest pressure and natural enemy stimulation in crops in relation to landscape composition (Modified from Bianchi et al., 2006)

Off-crop areas	Crop	Pest	Natural enemy species/group	References
Field margins, hedges, field size	Sugar beet	Aphids	Aphid predator complex	Basedow (1990)
Forest, tree lines, grassland, channels	Brussels sprout	—	Predators, egg parasitoids	Bianchi et al. (2005)
Field area-to-perimeter ratio	Cereals, rapeseed, legumes	—	Carabid beetles	Bommarco (1998)
Wooded field edges, field size	Maize	—	Armyworm parasitoids	Costamagna et al. (2004)
Forest	Leek	Thrips	—	Den Belder et al. (2002)
Forest, grassland, CRPa, patchiness	Wheat	—	Aphid predator complex	Elliott et al. (1998)
Forest, CRP, grassland, wetlands, patchiness	Maize	—	Coccinellids	Elliott et al. (2002a)
Forest, CRP, wetlands, patchiness	Lucerne	—	Aphid predator complex	Elliott et al. (2002b)
Forest	Potato	Aphids	Coccinellids	Galecka (1966)
Wooded field edges	Lucerne	Weevils	—	Holland and Fahrig (2000)
Forest	Spinach	Lepidoptera	—	Klug et al. (2003)
Hedges	Winter cereals	—	Syrphids	Krause and Poehling (1996)
Wooded field edges, field size	Maize	Lepidoptera borer	Armyworm parasitoids	Marino and Landis (1996)
Wooded field edges, field size	Maize	—	Armyworm parasitoids	Menalled et al. (1999b)
Wooded field edges, field size	Maize	—	Armyworm parasitoids	Menalled et al. (2003)b
Field area–to-perimeter ratio, forest	Spring barley	Aphids	—	Östman et al. (2001a)
Field area–to-perimeter ratio	Cereal	—	Carabid beetles	Östman et al. (2001b)
Uncultivated areas	Cotton	—	Cotton natural enemy complex	Prasifka et al. (2004)
Forest, fallow, hedgerows, grassland	Winter wheat	—	Carabid beetles	Purtauf et al. (2005a)
Forest, fallow, hedgerows, grassland	Winter wheat	—	Carabid beetles	Purtauf et al. (2005b)
Forest, fallow, hedgerows, grassland	Winter wheat	Aphids	Aphid parasitoids	Roschewitz et al. (2005)
Ecological corridors	Wheat	—	Leaf beetle parasitoids	Sedivý (1995)

Off-crop areas	Crop	Pest	Natural enemy species/group	References
Forest, fallow, hedgerows, grassland	Winter wheat	—	Aphid parasitoids	Schmidt et al. (2003)
Forest, fallow, hedgerows, grassland	Winter wheat	—	Spiders	Schmidt and Tscharntke (2005a)
Forest, fallow, hedgerows, grassland	Winter wheat	—	Spiders	Schmidt et al. (2005)
Forest, fallow, hedgerows, grassland	Oilseed rape	rape pollen beetle	Rape pollen beetle parasitoids	Thies and Tscharntke (1999)
Forest, fallow, hedgerows, grassland	Oilseed rape	rape pollen beetle	Rape pollen beetle parasitoids	Thies et al. (2003)
Forest, fallow, hedgerows, grassland	Winter wheat	Aphids	Aphid parasitoids	Thies et al. (2005)
	Soybean	Aphids	Aphid natural enemies	Zhang and Swinton (2012)
Pasture, Native perennial vegetation, Fallow	Canola, wheat	Lepidopteran herbivores (larvae)	Parasitoids	Macfadyen and Muller (2013)
Other crops (lucerne, maize)	Spring wheat	Cereal aphid	Natural enemies: ground-dwelling predators, leaf predators, aphid mummies for parasitoids	Zi-Hua Zhao et al. (2013)

Appendix D. Approximation of the 'hedge' (off-field)

1. Flat (vertical) area with a lineic area of X times crop height per metre field edge; X ranging, for example, from 0.1 to 1.

2. As 1. but with more realistic (horizontal) LAI.

3. As 2. but with more realistic interception.

Appendix E. The use of a vegetation distribution factor in the non-target arthropod risk assessment

There have been several reviews of this figure and attempts at deriving an appropriate figure for the VDF in NTA risk assessment. As stated in section 6, these different figures all express the VDF as the ratio of areic mass of substance on the canopy and the surface concentration of substance per unit area of leaf surface for a defined vegetation structure. If using one of these figures for risk assessment this requires that exposure in the reference tier used for calibration of risk assessment is considered, i.e. relevant factors such as LAI measured are in the field and taken into account when linking exposure and effects.

The VDF value of 10 proposed in the report of ESCORT 2 (Candolfi et al., 2001) and used in the EU concept was derived by considering ‘Leaf Area Indices’ and ‘plant interception’ (see e.g. Gonzalez-Valero et al., 2000; Koch and Weisser, 2001; Weisser et al., 2003). The report of ESCORT 2 (Candolfi et al., 2001) and the Guidance Document on Terrestrial Ecotoxicology (EC, 2002) state that the current basis from which the VDF values are derived is not sufficient and further research is needed to calculate a more reliable value (e.g. validation of the VDF using field data).

The German Federal Environmental Agency (‘UBA’) proposed a VDF value of 5 instead of 10 (please refer to UBA, 2006). This estimation of the exposure was based mainly on the ‘Retention Area Index’ (RAI) characterising the total area for retention of sprayed PPPs in a canopy per base area (Koch and Weisser, 2004). This was done especially for meadow canopies of less than 20 cm in height, a scenario seen as a ‘realistic worst case’(see UBA, 2006). The derived VDF of 5 agrees well with field data by Koch et al. (2003), who compared measured residues of PPPs on two dimensional surfaces to the measured residues on meadows with a canopy height less than 20 cm, next to a treated area (factor of 4.4 to 6.5 between median spray residues on leaves when a standard nozzle was used for spray application, see also Table E1).

Table E1: Comparison of measured ground deposition on petri dishes (2D) and deposition on plant surfaces in a meadow (canopy height < 20 cm); table adapted from UBA (2006) and Koch et al. (2003)

	1 m	3 m	5 m	10 m
Drift percentages				
BBA, 50th percentile	0.970 %	0.340 %	0.210 %	0.110 %
Nozzle specification				
XR 110 03	8.814 %	1.644 %	0.692 %	0.252 %
AI 110 025	1.101 %	0.250 %	0.178 %	0.076 %
ID 120 05	0.884 %	0.096 %	0.045 %	0.015 %
Calculated residues (drift rates)				
Application rate	10 ng/cm^2			
BBA, 50th percentile	0.0970 ng/cm^2	0.0340 ng/cm^2	0.0210 ng/cm^2	0.0110 ng/cm^2
Nozzle specification				
XR 110 03	0.8814 ng/cm^2	0.1644 ng/cm^2	0.069^2 ng/cm^2	0.0252 ng/cm^2
AI 110 025	0.1101 ng/cm^2	0.0250 ng/cm^2	0.0178 ng/cm^2	0.0076 ng/cm^2
ID 120 05	0.0884 ng/cm^2	0.0096 ng/cm^2	0.0045 ng/cm^2	0.0015 ng/cm^2
Measured residues (median)				
Nozzle specification				
XR 110 03	0.1423 ng/cm^2	0.0307 ng/cm^2	0.0158 ng/cm^2	0.0039 ng/cm^2
AI 110 025	0.0068 ng/cm^2	0.0019 ng/cm^2	0.0011 ng/cm^2	0.0000 ng/cm^2
ID 120 05	0.0078 ng/cm^2	0.0008 ng/cm^2	0.0006 ng/cm^2	0.0004 ng/cm^2

	1 m	3 m	5 m	10 m
Quotient between calculated and measured residues (with nozzle reference value)				
Nozzle specification				
XR 110 03	6.2	5.4	4.4	6.5
AI 110 025	16.2	13.2	16.2	–/–
ID 120 05	11.3	12.0	7.5	3.8

A report by the UK Department for Environment, Food and Rural Affairs (DEFRA, 2001) reviews methods to calculate off-field exposure for non-target organisms. Based on data by Miller et al. 2000, the authors suggest that where there is significant vegetation as tall or taller than the crop, a VDF factor no higher than 3 should be used to extrapolate from a 2D to a 3D exposure situation (see Table E2).

Table E2: Drift deposition of plant protection products determined in strips above different structures in the off-field. Data from Miller et al. 2000.

Distance from the field edge (m)	Lines at 1.6 m height above tall grass plot (% of application rate)	Lines at 0.7 m height above cut plot (% of application rate)	Quotient of deposits on lines above tall grass and cut plot
2.25	3.17	10.00	3.15
4.25	1.46	6.25	4.28
6.25	0.70	3.44	4.91

In regulatory practice, ‘2D’ and extended laboratory tests (exposure on sprayed glass plates or leaf disks), as well as ‘3D’ extended laboratory tests (sprayed whole plants), are performed. Assuming direct proportionality between the LR_{50} values of, and the exposure to, the plant PPPs in tests performed in regulatory studies, the factor between the toxicity in those studies could give additional information on the numeric value of a VDF.

To derive an estimate for a VDF, LR_{50} values published in EFSA DARs and EFSA Conclusions for active substances were compared between test types for the same species.

Table E3: Comparison of different test types for *A. rhopalosiphi* and *T. pyri*.

Compared test types	LR_{50} ratio					
	Min	10th percentile	Median	90th percentile	Max	n
2D leaf/2D glass (*T. pyri* & *A. rhopalosiphi*)	0.351	0.789	1.943	5.891	8.166	14
3D plant/2D glass (*A. rhopalosiphi*)	1.068	1.362	5.576	61.782	73.708	12

Comparisons were based on the datasets given in **table E4** and **table E5**. For the comparison of 2D leaf/2D glass tests the data for *T. pyri* and *A. rhopalosiphi* were pooled because the sample size for *A. rhopalosiphi* was very low (n = 4).

As it can be seen in Table E3, there is a median factor of ca. 6 between ‘3D’ tests on sprayed whole plants and ‘2D’ tests on sprayed glass plates. It is not clear how much of the difference of toxicity between 2D and 3D laboratory and extended laboratory studies can be attributed to the difference in bioavailability between plant material and glass plates and how much can be attributed to vegetation distribution. There were no comparisons available between 2D studies on sprayed leaves and 3D studies on sprayed plants for any of the species. So how much of the difference in bioavailability that contributes to the difference in toxicity between glass plate studies and studies on plant material could be estimated by comparing 2D studies on glass plates and plant material? The median factor between LR_{50} values in 2D tests on glass plates and leaf disks was ca. 3. So, considering this, the resulting difference between 3D plant and 2D leaf tests attributable to vegetation distribution could be 6/2 = ca. 3.

Because of the high variability, the uncertainty around a VDF derived from this test data seems very high. However, the presented data could be used as additional information when discussing the VDF.

Table E 4: LR_{50} ratios between 2d leaf and 2d glass tests for *A. rhopalosiphi* and *T. pyri*.

Test item	LR_{50} ratio [2d leaf / 2d glass]			
	A. rhopalosiphi		*T. pyri*	
	Qualifier	Value	Qualifier	Value
Isopyrazam (formulation)			>	2.3176
Topramezone (formulation)			>	2.8754
Tolylfluanid WG 50 (formulation)				
Imidacloprid SL 200 (formulation)				4.5225
Carbendazim 125 g/L (Formulation)				1.3507
Tebuconazole EW 250 (formulation)		0.59		3.6379
Emamectin benzoate (formulation) /				1.4052
Acrinathrin 75 g/L EW (formulation)		0.35		
Fluquinconazole (formulation) /				6.3725
Etridiazole (formulation) /				1.2576
Acetochlor - Dow (formulation)				2.0349
Flufenoxuron (formulation)				1.7884
Spirotetramat OD 150 EU (formulation)				4.7688
Methomyl 20 SL (formulation) /				2.6953
Fenamiphos 240 CS (formulation) /				8.1657
Cyanamide (formulation) /				4.6454
Prosulfocarb 800g/L EC				1.8511

Table E 5: LR_{50} ratios between 3d plant and 2d glass tests for *A. rhopalosiphi*.

Test item	LR_{50} ratio [3d plant / 2d glass]	
	A. rhopalosiphi	
	Qualifier	Value
Dimethachlor (formulation)	>	35.28114664
Isopyrazam (formulation)	>	9.830097087
Spirotetramat OD 150 EU (formulation)	>	2.511423489
Imidacloprid SL 200 (formulation)		20.83333333
Emamectin benzoate (formulation)		1.068
Etridiazole (formulation)		5.702479339
Metaflumizone (formulation)		5.575862069
Formetanate 500 SG (Formulation)		4.428571429
Phosmet 50WP (formulation)		5.263589744
Methomyl 20 SL (formulation)		58.8
Cyanamide (formulation)		1.435185185
Prosulfocarb 800g/L EC (formulation)		73.70813397

REFERENCES

Candolfi MP, Barrett KL, Campbell P, Forster R, Grandy N, Huet M-C, Lewis G, Oomen PA, Schmuck R and Vogt H, 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, and SETAC-Europe, Brussels, Belgium.

DEFRA (Department for Environment, Food and Rural Affairs), 2001. Assessing the validity of current approaches for calculating off-crop exposure to pesticides. Defra Project code PS2001, 18 pp.

Koch H and Weisser P, 2001. Spray deposits of crop protection products on plants – the potential exposure of non-target arthropods. Chemosphere, 44, 307–312.

Koch H and Weisser P, 2004. Die Gesamtoberflaeche in Saumstrukturen als potentielle Retentionsflaeche fuer Driftpartikel, Retention Area Index (RAI). Nachrichtenblatt des Deutschen Pflanzenschutzdienstes, 56, 65–69.

Koch H, Weisser P and Landfried M, 2003. Effect of drift potential on drift exposure in terrestrial habitats. Nachrichtenblatt des Deutschen Pflanzenschutzdienstes, 55, 181–188.

Miller PCH, Lane AG, Walklate PJ and Richardson GM, 2000. The effect of plant structure on the drift of pesticides at field boundaries. Aspects of Applied Biology, 57, 75–81.

Appendix F. List of taxa that should be evaluated in representative agro-ecosystems in Europe (From de Jong et al., 2010)

	Minimum desired level of taxonomic precision	Arable (both cereals and leafy crops)	Orchard (including citrus)	Off-crop	Remark/examples
Insecta					
Heteroptera					
Sternorrhyncha	Superfamily	+/–	+	+	Generally target taxa. Aphidoidea, aleyrodoidea, coccoidea, psylloidea
Other	Family	+/–	+	+	Anthocoridae, miridae, lygaeidae, cicadellidae
Hymenoptera					
Apocrita	Superfamily	+	+	+	Ichneumonoidea, chalcidoidea, proctotrupoidea, vespoidea
	Family	+	+	+	Depending on abundance (e.g. Braconidae, ichneumonidae, chalcidoidea families, scelionidae. Formicidae)
	Lower level	0	0	0	Depending on abundance up to genus or species level (e.g. *Aphidius* sp., *aphelinus mali*)
Coleoptera					
	Family	+	+	+	Distinguish juveniles for families below
Carabidae	Species	+	–	+	For abundant taxa
Staphylinidae	Genus/species	+	+	+	For abundant taxa
Coccinellidae	Subfamily	+*	+	+	For abundant taxa
	Genus/species	+*	+	+	For abundant taxa
Lathridiidae	Juv./adults	–	+	+	At family level
Collembola	Suborder	+	+	+	Subsamples should be identified to a lower level (family/genus) to enable a characterisation of collembolan community composition
Dermaptera	Order	–	0	–	
Diptera	Suborder	+	+	+	
	Family	0	0	0	For abundant taxa
	Juv./adults	+	+	+	Syrphidae and others
Lepidoptera	Juv./adults	–	+	+	
Neuroptera	Family	–	+	–	Chrysopidae, (conyopterigidae), others
	Juv./adults	–	+	–	
Orthoptera	Order	–	–	+	
Psocoptera	Order	–	+	–	No experience at lower level of identification
Thysanoptera (adults)	Order	0	+	+	
Aranea					
Hunting spiders	Family	+	+	+	
Lycosidae	Genus/species	+	–	+	For abundant taxa
Thomisidae	Genus/species	–	+	+	For abundant taxa
Web spiders	Family	+	+	+	
Linyphiidae	Genus/species	+	–	+	For abundant taxa
Dictynidae	Genus	–	+	–	For abundant taxa
Araneidae	Genus	–	+	–	For abundant taxa (i.e. Araneus)
Acari					
Gamasida	Family	–	+	+	For abundant families (phytoseiidae) subsamples should be identified to species level to enable a characterisation of gamasid community composition

	Minimum desired level of taxonomic precision	Arable (both cereals and leafy crops)	Orchard (including citrus)	Off-crop	Remark/examples
Actinedida	Family	–	+	+	Subsamples
Oribatida	Suborder	+	+	+	

* For Coccinellidae the remark has to be made that species from this taxon can populate a certain area relatively quickly as a result of the presence of aphids. When aphids are not present and abundant, Coccinellidae will not appear; this does not render the test directly unreliable; however this phenomenon should be taken into account when evaluating the study.

+: Taxon should be present and identified at the level specified, else the test is not sufficiently comprehensive to be of general validity.

+/–: Taxon should be present in the south of Europe, but not necessarily in the north of Europe.

0: Test is less reliable (Ri 2) when sufficiently robust data at the indicated level of taxonomic precision are missing, but additional data are not required.

–: Specified taxon is generally not relevant for the specified cropping system(s).

When '+' taxa are lacking in the specified agro-ecosystem addition of appropriate data, for example from other (laboratory) studies is needed to make the test reliable, otherwise the test is considered unreliable. 'Off-crop' means non cropped lands in the vicinity of agricultural fields, e.g. meadows or woodlands.

Appendix G. Summary of (extended) laboratory test systems identified as potentially relevant

Organism	Ecology	Test design	Endpoints	Exposure	Guideline
IOBC ring-tested laboratory test guidelines (Table adapted from Brühl et al. 2013, Table 5.5–1)					
Aphidius rhopalosiphi (Hymenoptera: Braconidae)	Parasitoid	*Used for testing spray formulations.* Test units consisting of glass plates are treated with the test item and after glass plates are dried adult wasps are added to each test unit. Effects are assessed after 2, 24 and 48 hours	Mortality (adults)	2d contact (freshly dried residues on glass)	IOBC approved test guidelines published by Mead-Briggs et al. (2000)
		Surviving females are individually placed on aphid-infested cereal plants covered by cylinders. Females are removed after 24 hours. 10–12 days later the numbers of aphid mummies on the plants are recorded	Parasitation capacity	No exposure during reproductive phase	
Typhlodromus pyri (Acari: Phytoseiidae)	Predator	*Used for testing spray formulations.* Test units consisting of glass plates are treated with the test item and – after glass plates are dried protonymphs are added. Three days (optional) and seven days later mortality is recorded	Mortality (protonymphs)	7d contact (freshly dried residues on glass)	IOBC approved test guidelines published by Blümel et al. (2000a). Tests are performed using the Coffin cell (Bakker et al., 1992), the open glass method (Louis and Ufer, 1995) or the island method (Joisten, 2000).
		The reproduction of the surviving females (eggs and juveniles) is assessed three times between days 7 and 14	Reproduction	No exposure during reproductive phase	
Aleochara bilineata (Coleoptera: Staphylinidae)	Parasitoid on Diptera pupae	*Used for testing spray and solid formulations.* Test units are filled with moistened quartz sand (so called laboratory test) or standardised soil (so called extended laboratory test)	Reproduction, adult mortality and behaviour only optional	7d contact (residues on surface/in substrate)	IOBC approved test guidelines published by Grimm et al. (2000)
		Granule, powder or coated seed are incorporated into the substrate or applied on the surface. Spray formulations are sprayed on surface. Afterwards, adult beetles are added 7, 14 and 21 DAT* and the test substrate is mixed up with host pupae. 28 DAT, adult beetles are removed and the emergence is recorded		21d contact (mixed substrate during parasitation phase)	

Organism	Ecology	Test design	Endpoints	Exposure	Guideline
Chrysoperla carnea (Neuroptera: Chrysopidae)	Predator (larvae), adults feed on honeydew, pollen/nectar	*Used for testing spray formulations.* Larvae (first instar) exposed on glass plates treated with the test item. The surviving larvae remain on the glass plates until they have pupated. Hatching of the adults is detected. The fecundity of the females as well as the fertility of the eggs can be assessed by sampling all eggs laid within 24 hours twice a week	Initial mortality (larvae), reproductive performance of the emerging adults	Ca. 8d contact until pupation (freshly dried residues on glass)	IOBC approved test guidelines published by Vogt et al. (2000)
Coccinella septempunctata (Coleoptera: Coccinelidae)	Predator	*Used for testing spray formulations.* 3–5 day old larvae are placed individually on dried glass plates which have been treated with the test item. After the pupal stage, the surviving ecdysis beetles are removed and taken in non-treated breeding cages. During a period of two weeks, the eggs laid are collected and observed for fertility	Pre-imarginal mortality, reproductive performance of the ecdysed beetles	10–15d contact until end of ecdysis (freshly dried residues on glass)	IOBC approved test guidelines published by Schmuck et al. (2000)
Orius laevigatus (Heteroptera: Anthocoridae)	Predator	*Used for testing spray formulations.* The test units are treated with the test item. In each dried test unit 10 *O. laevigatus* larvae (2nd instar) are added for at least 9 days or until 80 % of the bugs are adult. To assess the fecundity of surviving females, they are placed individually on oviposition substrate and their egg production is noted for two consecutive 2-day periods	Mortality of juvenile bugs, egg production	At least 9d contact (freshly dried residues on glass). No exposure during reproductive phase	IOBC approved test guidelines published by Bakker et al. (2000)
Pardosa (Araneae: Lycosidae)	Predator	*Used for testing spray and solid formulations.* Test units are filled with moistened quartz sand (laboratory test) or standardised soil (extended laboratory test). Spray formulations: Field-collected spiders are introduced (1 individual per test unit). Afterwards, test units are treated. Solid formulations: Spiders are introduced after granule, powder or coated seed has been incorporated in the substrate of applied on the surface. Spiders are monitored for at least 14 days in which mortality and behaviour is recorded. Furthermore, food consumption is assessed	Mortality, behaviour, food uptake	At least 14d contact (overspray, fresh residues on substrate for at least 14d)	IOBC approved test guidelines published by Heimbach et al. (2000a)

Organism	Ecology	Test design	Endpoints	Exposure	Guideline
Poecilus cupreus (Coleoptera: Carabidae)	Predator	*Used for testing spray and solid formulations.* Test units are filled with moistened quartz sand (laboratory test) or standardised soil (extended laboratory test). Spray formulations: 6 individuals (3 males, 3 females) are placed in test units. Afterwards, test units are treated. Solid formulations: Beetles introduced after granule, powder or coated seed has been incorporated in the substrate of applied on the surface. Beetles are monitored for at least 14 days in which mortality and behaviour is recorded. Furthermore, food consumption is assessed.	Mortality, behavioural impacts	Contact (overspray, fresh residues on substrate for at least 14d).	IOBC approved test guidelines published by Heimbach et al. (2000b)
Trichogramma cacoeciae (Hymenoptera: Trichogrammatidae)	Parasitic wasp	*Used for testing spray formulations.* Female adults are placed in each test unit, consisting of a frame and two treated glass plates (fresh dried). 24 hours after exposure, surviving wasps are recorded. To get information about the parasitisation capacity, 24, 48 and 96 hours after treatment host eggs are introduced which are analysed at least 9 days after insertion.	Initial mortality (adults), parasitisation capacity	7d contact (freshly dried residues on glass)	IOBC approved test guidelines published by Hassan et al. (2000)
Extended laboratory test protocols from regulatory practice addressing issues not covered in IOBC ring-tested protocols (source – Draft Assessment Reports for PPP active substances)					
Syrphis corollae	Predator	*Used for testing spray formulations.* 2-day old larvae are placed on dried glass plates which have been treated with the test item. After pupation, reproduction of surviving females is assessed in an uncontaminated environment.	(Pre-) imaginal mortality, reproduction of developed females	Ca. 8d contact until end of pupation (freshly dried residues on glass)	Rieckmann (1989)
Further laboratory test methods that are recognised for being potentially helpful in addressing relevant traits, taxonomic groups, endpoints or exposure paths					
Philonthus cognatus (Coleoptera, Staphylinidae)	Predator	Adults: beetles are exposed for one week. Afterwards they are removed and placed to mate in reproduction chambers in non-contaminated medium. Between 6 and 10 weeks, the number of eggs and offspring hatched are counted	Adults: mortality & reproduction (number of eggs and offspring)	Adults: exposure duration 7 days; beetles observed for 6–10 weeks	Metge and Heimbach (1998)
		Larvae: first larval stage is exposed to the contaminant and. Survival, hatching weight, development time are assessed. Semi-field tests (larvae): first larval stage are exposed to the chemical and surface activity during the development to immature beetle is observed. The same parameters for the larvae test are measured	Larvae: mortality, development time)	Larvae: exposure and observation until adult	

Organism	Ecology	Test design	Endpoints	Exposure	Guideline
Lithobius mutabilis (Chilopoda: Lithobiidae)	Predator	Adults: individuals are kept in contaminated substrate and food is given. Parameters measured are survival, growth, locomotor activity and (in some cases) respiration rate. Exposure time varies between 4 and 12 weeks depending on the persistence of the chemicals	Mortality, growth rate, respiration rate, locomotor activity	Persistent chemicals: 10 weeks; degradable chemicals: 4 weeks	Laskowski et al. (1998)
Brachudesmus subterraneus (Diplopoda: Polydesmidae)	Saprophagous	Adults: individuals are kept in contaminated natural substrate and/or food. Parameters measured are survival, development, reproduction and feeding parameters. Exposure time is 10 weeks	Mortality, reproduction (number of nests & eggs per female)	10 weeks	Tajovsky (1998)
Porcellio scaber, Porcellionides pruinosus, (Isopoda, Porcellionidae)	Saprophagous	Adults: individuals are kept in contaminated natural substrate and/or food. Parameters measured are survival, reproduction, feeding parameters, behaviour. Exposure time varies according to the test parameters but can go up to 10 weeks	Larval mortality, adult mortality, percentage of parasite emergence	Different according to products and instars	No agreed standard test guideline available
Helix aspersa (Mollusca, Helicidae)	Saphophagous/phytophagous	Juvenile individuals are exposed to contaminated substrate and non-contaminated food is given. Exposure takes 28 days but the substrate is renewed every 7 days. Parameters measured are survival and growth	Adult mortality, juvenile mortality, development and growth	7–28 days	ISO 15956:2006
Pieris brassica larvae	Herbivorous	Larvae are fed with artificial substrate or cabbage treated with PPP	Larval mortality, LC_{50} or LD_{50}	7 days	No agreed standard test guideline available
Anticarsia gemmatalis, Spodoptera frugiperda larvae	Herbivorous	1-day-old second instar larvae are exposed in plastic cups or well plates to toxin mixed in an artificial diet for 48 hours. Subsequently larvae are transferred to test vessels containing uncontaminated artificial diet for 72 hours. Mortality at 48 hours and 120 hours is assessed (LC_{50}). This method was developed for testing the effects of *Bacillus thuringensis* strains	Larval mortality	48 hours	No agreed standard test guideline available

*DAT = days after treatment

REFERENCES

Bakker F, Grove A, Blumel S, Calis J, Oomen P, 1992. Side-effect tests for phytoseiids and their rearing methods. Bulletin OILB/SROP, 15, 61-81.

Bakker FM, Aldershof SA, Veire M vd, Candolfi MP, Izquierdo JI, Kleiner R, Neumann Ch, Nienstedt KM and Walker H, 2000. A laboratory test for evaluating the effects of plant protection products on the predatory bug, Orius laevigatus (Fieber) (Heteroptera: Anthocoridae. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium, 57-70.

Blümel S, Bakker FM, Baier B, Brown K, Candolfi MP, Gossmann A, Grimm C, Jäckel B, Nienstedt K, Schirra KJ, Ufer A and Waltersdorfer A, 2000a. Laboratory residual contact test with the predatory mite Typhlodromus pyri Scheuten (Acari: Phytoseiidae) for regulatory testing of plant protection products. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium, 128-150.

European and Mediterranean Plant Protection Organization (EPPO), 1989. Guideline 142: Guideline for the evaluation of side-effects of plant protection products: Encarsia formosa. OEPP/EPPO Bulletin 19, 355-372.

Grimm C, Reber B, Barth M, Candolfi MP, Drexler A, Maus C, Moreth L, Ufer A and Waltersdorfer A, 2000. A test for evaluating the chronic effects of plant protection products on the rove beetle Aleochara bilineata Gyll. (*Coleoptera: Staphylinidae*) under laboratory and extended laboratory conditions. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber.

Hassan SA, Halsall N, Gray AP, Kuehner C, Moll M, Bakker FM, RoembkeJ, Yousef A, Nasr F and Abdelgader H, 2000. A laboratory method to evaluate the side effects of plant protection products on Trichogramma cacoeciae Marchal (*Hymenoptera: Trichogrammatidae*). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium, 107-119.

Heimbach U, Wehling A, Barrett KL, Candolfi MP, Jäckel B, Kennedy PJ, Mead-Briggs M, Nienstedt KM, Römbke J, Schmitzer S, Schmuck R, Ufer A and Wilhelmy H, 2000a: A method for testing effects of plant protection products on spiders of the genus Pardosa (Araneae: Lycosidae) under laboratory conditions. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium. 71-86.

Heimbach U, Wehling A, Barrett KL, Candolfi MP, Jäckel B, Kennedy PJ, Mead-Briggs M, Nienstedt KM, Römbke J, Schmitzer S, Schmuck R, Ufer A and Wilhelmy H, 2000b. A method for testing effects of plant protection products on the carabid beetle Poecilus cupreus (Coleoptera: Carabidae) under laboratory and semi-field conditions. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium. 71-86.

Joisten M, 2000. The 'island method' - a new laboratory test method to determine side-effects of plant protection products on the predatory mite Typhlodromus pyri Scheuten (Acari, Phytoseiidae). Journal of Applied Entomology-Zeitschrift Fuer Angewandte Entomologie, 124, 267-268.

Laskowski R, Pyza E, Maryanski M and Niklinska M, 1998. Test on the Centipede Lithobius mutabilis L. Koch 1852 (*Chilopoda:Lithobiidae*). In Handbook of soil invertebrate toxicity. H Løkke and CAM Van Gestel (eds.). John Wiley & Sons. Chichester. p. 181-196

Louis F, Ufer A, 1995. Methodical improvements of standard laboratory tests for determining the sideeffects of agrochemicals on predatory mites (acari: Phytoseiidae). Anzeiger Fuer Schaedlingskunde Pflanzenschutz Umweltschutz 68:153-154.

Mead-Briggs M and Longley M, 1997. A standard 'extended laboratory' test to evaluate the effects of plant protection products on adults of the parasitoid, Aphidius rhopalosiphi (Hymenoptera, Braconidae). (unpubl.),

Mead-Briggs MA, Brown K, Candolfi MP, Coulson MJM, Miles M, Moll M, Nienstedt K, Schuld M, Ufer A and McIndoe E, 2000. A laboratory test for evaluating the effects of plant protection products on the parasitic wasp, Aphidius rhopalosiphi (DeStephani-Perez) (Hymenoptera: Braconidae). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium, 13-35.

Metge K and Heimbach U (1998) Tests on Staphylinid Philonthus cognatus Steph. 1832. In Handbook of soil invertebrate toxicity. H Løkke and CAM van Gestel (eds.). John Wiley & Sons. Chichester. p. 157-179.

Rieckmann W, 1989. Auswirkungen von Pflanzenschutzmitteln auf die Schwebfliege Syrphus corollae (fabr.) im Laboratorium, guideline part VI, 23-2.1.7 of the federal german biological agency for agriculture and forestry.

Schmuck R, Candolfi MP, Kleiner R, Jans D, Waltersdorfer A, Mead-Briggs M, Moll M and Wilhelmy H, 2000. A laboratory test system for assessing effects of plant protection products on the plant dwelling insect Coccinella septempunctata L. (*Coleoptera: Coccinellidae*). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck and H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium. 45-70.

Tajovsky K, 1998. Test on the Millipede Brachydesmus superus Latzel 1884. In Handbook of soil invertebrate toxicity. H Løkke and CAM van Gestel (eds.). John Wiley & Sons. Chichester. p. 197-205.

Vogt H, Bigler F, Brown K, Candolfi MP, Kemmeter F, Kühner Ch, Moll M, Travis A, Ufer A, Viñuela E, Waldburger M and Waltersdorfer A, 2000. Laboratory method to test effects of plant protection products on larvae of Chrysoperla carnea (Neuroptera: Chrysopidae). In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. MP Candolfi, S Blümel, R Forster, FM Bakker, C Grimm, SA Hassan, U Heimbach, M Mead-Briggs, B Reber, R Schmuck, H Vogt. IOBC, BART and EPPO Joint Initiative - IOBC/WPRS, Gent Belgium, 27-44.

Appendix H. Semi-field and field test systems

Potential (semi-)field and modelling approaches to the risk assessment of plant protection products (PPPs) towards non-target arthropods (adapted from Brown et al., 2009, available at: http://www.efsa.europa.eu/de/supporting/doc/16e.pdf)

Organism	Test design	Endpoints	Caveats	Guideline
Semi-field study				
Carabid beetle: *Poecilus cupreus*	Adult beetles confined in enclosures dug into crop situations. Enclosures are over-sprayed. Exposure scenario is realistic	Mortality assessed by recapture of beetles and sub-lethal effects assessed by consumption of Drosophila pupae as prey	*P. cupreus* is a burrowing beetle with a relatively thick cuticle. Those individuals that burrow beneath the soil surface for all or part of the exposure period of a study will be less exposed than those that remain on the soil surface	Heimbach et al., 1992
Invertebrates: Carabidae, Staphylinidae, Coccinellidae, Lycosidae, Chrysopidae, Anthocoridae	Four methods for use in cereals: barriered enclosures, 2 m^3 cages (*C. septempunctata*), sleeves and barriered large plots. In the first three methods, laboratory reared organisms were released into cages or enclosures shortly after treatment and their survivorship recorded after periods of time. Confinement and exposure of laboratory-derived insects in semi-field conditions increases realism of exposure while minimising the interspecies response variability	Mortality	A large rate of non-recovery of released organisms. Where sleeves were used organisms may avoid exposure by clinging to the untreated barriers.	Jepson and Mead-Briggs, 1992
Carabid beetle: *Pterostichus melanarius* *Lycosid hunting spider: Pardosa* sp.	Realistic exposure in 1 × 1 m enclosures especially for surface active predators where contact most relevant route of exposure	Mortality		Brown et al., 1990
Green lacewing: *C. carnea*	This semi-field approach incorporated laboratory-cultivated insect larvae released into orchard trees immediately prior to spraying. Larvae were re-captured with bait cards	Mortality	No information on how to extrapolate results to predict effects on other non-target invertebrates	Vogt et al., 1992

Organism	Test design	Endpoints	Caveats	Guideline
Field studies				
Non-target arthropods	General guidance on design, conduct and interpretation of non-target arthropod field studies. Advocates use of either arable or orchard as model crop system. Use realistic worst-case exposure and assessment of effects on phytophagous, detritivorous and predatory arthropods. Taxonomy to species level where possible			Candolfi et al., 2000b
Predatory mite field studies – vineyard and orchards				
Predatory mite: *T. pyri*	This method assesses the short- and long-term effects of products on phytoseiid mites in vineyards and orchards by sampling population density compared with that in a water-treated control at different time intervals after application. Naturally occurring populations are both directly and indirectly under realistic conditions. Ring tests in apple and vineyard showed that effects of greater than 50 % were statistically significant in 90 % of cases	Population development (mortality and reproduction)	Relatively expensive and variability of natural populations may reduce precision. Not clear how effects on mites reflect the response of non-target arthropods as a whole	Bluemel et al., 2000b
Predatory mite: *T. pyri*	This methodology describes the addition of overwintering sampling of predatory mites to existing methods (by dissection of leaf buds collected in February). Enhancing sampling to include surveys in an additional season generates useful data concerning effects at different life stages and the potential for recovery following exposure	Population development (mortality and reproduction)	Increased costs may be associated with sampling through leaf bud collection and dissection	Gyorffyne and Polgar, 1994
Mites: *Euseius finlandicus* and *T. pyri*	Assessment of effects on mites in orchards using similar methodology to Bluemel et al., 2000	Population development (mortality and reproduction)	Sensitivity of mite populations can vary because of previous exposure to pesticides. This could make extrapolation of results to other situations difficult	Sterk et al., 1994

Organism	Test design	Endpoints	Caveats	Guideline
Arable field studies				
Spring and autumn breeding carabid beetles, staphylinid beetles, spiders, aphid-specifics (parasitoids, coccinellids, neuropteran larvae, syrphidae and game-bird chick food insects)	Early reference that outlines some necessary components of field and semi-field studies in cereals, including at least four replicates of either small barriered or large unbarriered plots, and that data should be collected in two or more site-years. Large plot: a high degree of realism, with no need to erect barriers and no risk of over sampling; also provides data from a wide range of taxa, (especially polyphagous predators). Small barriered plot is more practical as only 1 ha required. Because of smaller size, selection of site with high and relatively uniform population is more achievable	Population development (mortality and reproduction)	Large plot: requires > 20 ha of cereals with homogeneous arthropod populations. This early guideline focused on predators, parasites and bird food insects. There is no mention of Collembola or mites and there was no investigation of off-crop effects.	Carter, 1993
Main taxa: Braconidae, Empidoidea, Carabidae, Staphylinidae, Linyphiidae	Experiment was carried out in two 4 ha fields, with one sprayed with a synthetic pyrethroid and one with a positive control substance, yearly for five years. Foliage and soil invertebrates were collected with D-vac and pitfall traps. The multiple year duration of the study increases realism and allows for observation of long-term effects on several invertebrate populations	Population development (mortality and reproduction)	Methodology incorporated only one plot per treatment which severely restricts statistical power of conclusions	Inglesfield, 1989
Predatory taxa plus specific assays of *Nebria brevicollis, Bembidion obtusum, Trechus quadristratus*	Movement of fauna between sprayed and unsprayed areas was estimated using a different method in each of the two years (traps on either side of barrier and then surface searches of fields and hedgerows). Enclosures were used to assess the mortality of key beneficial species.	Mortality and population development (mortality and reproduction)	Assessment of immigration and emigration. Significant effects were detectable in about 50 % of species tested when this method was utilised. The semi-field enclosures provided additional information for the key beneficial species	White et al., 1990

Organism	Test design	Endpoints	Caveats	Guideline
Invertebrates: Aphidae, Araneae, Carabidae, Chrysopidae, Coccinellidae, Entomophthorales, Staphylinidae, Syrphidae, Cicadina, Diptera, Heteroptera, Hymenoptera, Nematocera, Symphyta, Thysanoptera	Two 10 ha plots were established in existing crop fields (one control, one treatment), and multiple within plot invertebrate samples were collected over two years. Sampling over multiple years allowed researchers to track long-term effects and recovery of invertebrate populations	Population development (mortality and reproduction)	Replicates utilised during the course of this study were actually pseudo-replicates, and this design is therefore lacking in statistical rigor	Wick and Freier, 2000
Theridiidae, Linyphiidae, Tetragnathidae, Araneidae	Methodology was developed to determine the effects of BT maize and pesticide spraying on spider populations using 30 × 50 m sub-plots. Suction sampling was determined to be the most efficient and cost-effective methodology	Population development (mortality and reproduction)		Meissle and Andreas, 2005
Carabidae	Three-year field trial. The data from the first two years were used to develop models of the recovery process while data from the third year was used to validate the models. Pitfall traps were used to sample non-target epigeal invertebrates	Population development (mortality and reproduction) Model was able to pinpoint distance from field as crucial variable in determining population-level recovery	Need information on level of variation of recovery rate in families to assess effects on individual species	Thacker and Jepson, 1993
Linyphiidae				
Invertebrates: *Coccinella septempunctata*, *Propylea quatuordecimpunctata*, *Episyrphus balteatus*, *Chrysoperla carnea*	This paper presents an in-field method of examining susceptibility of foliage-dwelling invertebrate predators. Three replicates were utilised per treatment, and invertebrates were collected via beating and sweep nets	Population development (mortality and reproduction)	Methodology is limited in that it specifically assesses foliar dwelling predators, and may not be useful in assessing off-crop communities.	Jansen, 2000
Invertebrates: Carabidae, Staphylinidae, Linyphiidae, Collembola	This describes a large-scale field study in winter wheat using univariate analysis to identify changes at the family and species level for carabid, staphylinid beetles, linyphiid spiders and Collembola. Certain indicator species (identified through first order PRC analysis) may provide the most information on non-target arthropod effects in 1 ha plots	Population development (mortality and reproduction)	High cost	Brown and Miles, 2006
Syrphidae, Chrysopidae, Coccinellidae	Weed strips 1 to 3 m from the margin of a sprayed field were surveyed for resident insects	Population development (mortality and reproduction)	The numbers of beneficial non-target arthropods sampled from	Langhof et al., 2003

Organism	Test design	Endpoints	Caveats	Guideline
Bioassay: *Aphidius colemani*	following field spraying. Design of experimental methodology allowed for the calculation of median lethal drift rates for invertebrate taxa		plots was low and may decrease power of statistical assessment	
Invertebrates: non-target agroecosystem arthropods	Off-crop study in the margin of a wheat field. The 650 m long grass strip was divided into plots with spray drift and controls (no spray drift). Resident arthropods were sampled via biocoenometer surveys, pitfall traps, and grasshopper counting in quadrats following spray drift	Population development (mortality and reproduction)	Low density of arthropods reported in the grass margin; however, this may be as a result of the nature and number of subsamples taken per plot. Identification difficult because of wide range of arthropods	Freier et al., 2001
Field studies in fruit orchards				
Invertebrates: mites, psyllids	Large plots containing six rows of orchard trees (50 m length) were exposed via spray. Mite and psyllid populations were then sampled via beating methods	Population development (mortality and reproduction)	Untreated or water-treated controls are not included. No replication in the study. Replication of orchard studies is difficult to achieve. No sampling methods for arthropods on the soil surface or taxa that may predominantly live off-crop	Reboulet, 1994
Invertebrate: Hemiptera, predatory Heteroptera, Coleoptera, Neuroptera, Hymenoptera, Diptera, Araneae, Dermaptera, Lepidoptera, Orthoptera, Thysanoptera	This reference detailed the comparison between a small plot study (30 trees per plot) and an un-replicated large plot study (150 trees per plot) at the same orchard site, with respect to measuring invertebrate following exposure to PPPs	Population development (mortality and reproduction)	The effects observed in small plots were short-lived and transitory in nature. Effects seen in the large plot may be difficult to interpret with any statistical certainty	Brown, 1998

REFERENCES

Blümel S, Aldershof S, Bakker FM, Baier B, Boller E, Brown K, Bylemans D, Candolfi MP, Huber B, Linder C, Louis F, Müther J, Nienstedt KM, Oberwalder C, Reber B, Schirra KJ, Sterk G, Ufer A and Vogt H, 2000b. Guidance document to detect side effects of plant protection products on predatory mites (Acari: Phytoseiidae) under field conditions: vineyards and orchards. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative, IOBC/WPRS, Gent, Belgium, 145–158.

Brown K, 1998. Field studies with pesticides and non-target arthropods in orchardsEcotoxicology: pesticides and beneficial organisms. Ecotoxicology: pesticides and beneficial organisms, Dordrecht London Kluwer Academic 1998: 137-149.

Brown RA, McMullin LC, Jackson D, Ryan J and Coulson JM, 1990. Beneficial arthropod toxicity assessments with three insecticides in laboratory, semi-field and field studies Proceedings of the Brighton Crop Protection Conference, Pests and Diseases Volume 2, British Crop Protection Council Publications<Publisher Location> Thornton Heath.

Brown, KC and Miles M, 2006. ‘How much precision does a regulatory field study need?’ Pesticides and Beneficial Organisms IOBC/wprs Bulletin 29(10): 43-52.

Brown K, Tomlinson J, Duncan J and Hinchcliffe A, 2009. Critical comparison of available and potential higher tier testing approaches for the risk assessment of plant protection products, considering at least field and semi-field experimental designs, extrapolation from dose-response relationship, and increased dosages (aquatic and terrestrial). Literature reviews on ecotoxicology of chemicals with special focus on plant protection products. CFT/EFSA/PPR/2008/01. Exponent. UK. 192 p.

Candolfi M, Bigler F, Campbell P, Heimbach U, Schmuck R, Angeli G, Bakker F, Brown K, Carli G, Dinter A, Forti D, Forster R, Gathmann A, Hassan S, Mead-Briggs M, Melandri M, Neumann P, Pasqualini E, Powell W, Reboulet J-N, Romijn K, Sechser B, Thieme T, Ufer A, Vergnet C and Vogt H, 2000b. Principles for regulatory testing and interpretation of semi-field and field studies with non-target arthropods. Journal of Pest Science, 73, 141–147.

Carter N, 1993. Field trials to study the within-season effects of pesticides on beneficial arthropods in cereals in summer. Bulletin OEPP 23(4): 709-712.

Gyorffyne Molnar J and Polgar A, 1994. Effects of pesticides on the predatory mite Typhlodromus pyri Scheuten. A comparison of field and laboratory results <Original> [Effet des pesticides sur l'acarien predateur Typhlodromus pyri Scheuten. Comparaison des resultats au champ et au laboratoire] Working group Pesticides and beneficial organisms. Side-effects of pesticides on beneficial organisms: comparison of laboratory, semi-field and field results <Original> Groupe de travail Pesticides et organismes utiles. [Effets secondaires des pesticides sur les auxiliaires: comparaison des resultats au laboratoire, en semi-plein champ et en plein champ]. Bulletin OILB SROP (France) 17(10): 21-26.

Freier B, Kuehne S, Baier B, Schenke D, Kaul P and Heimbach U, 2001. Field study on effects of insecticide applications in wheat on the arthropod community of field boundaries. Mitt.Biol.Bundesanst.LandForstwirtsch. 383: 82-87.

Heimbach F 1992. Correlation between data from laboratory and field tests for investigating the toxicity of pesticides to earthworms. International Society of Soil Science, Avignon, France.

Inglesfield C, 1989. A long-term field study to investigate the effects of alpha-cypermethrin on predatory and parasitic arthropods. Mededelingen van de Faculteit Landbouwwetenschappen, Rijksuniversiteit Gent. Gent. 54: 895-904.

Jansen JP, 2000. A three-year field study on the short-term effects of insecticides used to control cereal aphids on plant-dwelling aphid predators in winter wheat.' Pest Management Science 56(6): 533-539.

Jepson P and Mead-Briggs M 1992. A discussion of methods used in semi-field studies to evaluate pesticide toxicity to beneficial invertebrates. Bulletin OILB/SROP 15(3): 4-17.

Langhof M, Gathmann A, Poehling HM, Vogt H, Heimbach U and Vinuela E, 2003. Field study on the effects of insecticide drift on non-target arthropods : residual toxicity, effect on populations and recolonisation processes. Bulletin OILB/SROP 26(5): 73-82.

Meissle M and Andreas L, 2005. Comparing methods to evaluate the effects of Bt maize and insecticide on spider assemblages. Agriculture, ecosystems & environment 107(4): 359-370.

Reboulet JN, 1994. Impact des produits phytosanitaires sur la faune auxiliare; methodologie d'experimentation en verger. Paris, Association de Coordination Technique Agricole.

Sterk G, Creemers P and Merckx K, 1994. Testing the side effects of pesticides on the predatory mite Typhlodromus pyri (Acari, Phytoseiidae) in field trials <Original> [Tests sur les effets secondaires des pesticides sur l'acarien predateur Typhlodromus pyri (Acari, Phytoseiidae), essais en champ] Working group Pesticides and beneficial organisms. Side-effects of pesticides on beneficial organisms: comparison of laboratory, semi-field and field results <Original> Groupe de travail Pesticides et organismes utiles. [Effets secondaires des pesticides sur les auxiliaires: comparaison des resultats au laboratoire, en semi-plein champ et en plein champ]. Bulletin OILB SROP (France) 17(10): 27-40.

Thacker JRM and Jepson PC, 1993. Pesticide Risk Assessment and Non-Target Invertebrates: Integrating Population Depletion, Population Recovery, and Experimental Design. Bull.Environ.Contam.Toxicol. 51(4): 523-531.

Vogt H, Rumpf S and Wetzel C, 1992. A field method for testing effects of pesticides on larvae of the green lacewing Chrysoperla carnea Steph. (Neuroptera, Chrysopidae) <Original> [Methode en champ pour tester les effets des pesticides sur la chrysope Chrysoperla carnea Steph. (Neuroptera, Chrysopidae)] Guidelines for testing the effects of pesticides on beneficial organisms: description of test methods. Meeting of the working group Pesticides and beneficial organisms, University of Southampton, UK, September 1991 <Original> [Directives pour l'evaluation des effets des pesticides sur les organismes utiles: description des methodes de test. Reunion du groupe de travail Pesticides et organismes utiles, Universite de Southampton, Royaume-Uni, septembre 1991]. Bulletiin OILB SROP (France) 15(3): 176-182.

White JS, Everett CJ and Brown RA, 1990. Assessing the effects of lambda-cyhalothrin on natural enemies using laboratory, small plot and field studies <Original> [Etude en laboratoire, en petite parcelle et en champ, des effets de la lambda-cyhalothrine sur les ennemis naturels] [International colloquium: Relations between phytosanitary treatments and animal reproduction, 25-26 April 1990, Paris] <Original> Colloque international: Relations entre les traitements phytosanitaires et la reproduction des animaux, 25-26 Avril 1990, Paris A long-term field study to investigate the effects of alpha- cypermethrin on predatory and parasitic arthropods. Relations entre les traitments phytosanitaires et la reproduction des animaux Published in Annales ANPP (France). Paris, ANPP Paris (France). 2: p 53-60.

Wick M and B Freier, 2000. Long-term effects of an insecticide application on non-target arthropods in winter wheat – a field study over 2 seasons. Anz. Schaedlingskd. Pflanz. Umweltschutz 73(3): 61-69.

Appendix I. Example of how to use Bayesian hierarchical statistical modelling together with elicited expert judgements for calibration of non-target arthropod risk assessment

The purpose of this appendix is to provide an illustration of how the principles established in section 5 of this opinion can be turned into a practical approach to calibration of an NTA risk assessment scheme.

As described in section 5 of this opinion, the underlying basis is a Bayesian hierarchical statistical model linking the various possible laboratory test and field study results to reference tier effects. The link is made by a 'dose–response'-like structure linking effects to application rates for each species in a possible test or field study and in the reference tier studies. Parameters of the dose–response will vary for different substances (PPPs), species and types of study. Because studies do not make an exact measurement of effects, we also need to account for the difference between measured effects and the 'true effects' described by the dose–response.

In this example, we link together only a few core elements of the wider NTA risk assessment: (i) the glass plate lethality test (GPLT) for *T. pyri*; (ii) the standard *T. pyri* field study (TPFS); and (iii) the in-field reference-tier multi-fauna field study.

Steps required in the example:

- Quantification of the uncertain relationship between GPLT LR_{50} for *T. pyri* and population effects for *T. pyri* at known application rates in a TPFS. For this, we have a relatively large amount of relevant data although it is not easy to interpret quantitatively. A detailed analysis is provided in section I.1.

- Quantification, for *T. pyri*, of the difference between the dose–response parameters for TPFS and the reference tier for a substance. For this, we have little or no direct data and may have to rely on expert judgement. Some hypothetical judgements, specified in section I.2, are used in the subsequent example illustrating computation of assessment factors.

- Quantification of the inter-species variation, relative to *T. pyri*, in dose–response parameters for the reference tier. For this, we have a small amount of direct data. Some thoughts about modelling are outlined in section I.2 and hypothetical expert judgements are specified for use in the subsequent example illustrating computation of assessment factors.

- Combining the three previous quantifications and looking at the implications for assessment factors. Within the Bayesian paradigm, this is both mathematically and computationally straightforward and follows as a consequence of the previous three quantifications.

One might consider that such complexity is not necessary. For example, it might be suggested that one could just write down assessment factors (AFs), also known as extrapolation/safety/uncertainty factors, for each step and then multiply them. There are three direct benefits to the Bayesian modelling approach: (i) greater transparency about the nature and size of each source of uncertainty; (ii) a coherent rational procedure for arriving at an overall AF; (iii) a rational procedure for reducing the overall AF when more or better data are available, for example when going up the tiers. The three benefits are in areas often seen to cause difficulties for the traditional approach to assessment factors.

I.1 *T. pyri*: glass plate LR_{50} → TPFS

In the Opinion, figure 20 shows the fundamental Bayesian network structure (Lauritzen, 1996) linking the various sets of dose–response parameters. We now focus on the parts of this network which relate to *T. pyri*, GPLT for *T. pyri*, TPFS and reference tier effect for *T. pyri*, and augment these with nodes corresponding to the GPLT and TPFS results. The resulting Bayesian network is shown in Figure I.1. Recall that, as in the Opinion, the notation $\theta^{\mathrm{GP},L}_{T.pyri}$ denotes the dose–response parameters for the relationship between lethality and application rate in the *T. pyri* GPLT for the substance being considered.

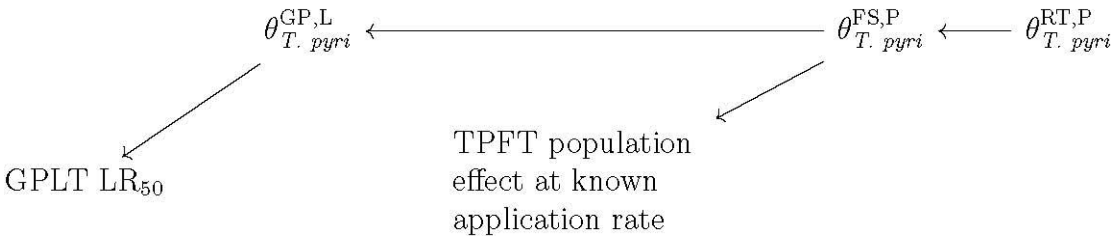


Figure I.1: Bayesian network underlying the *T. pyri* calibration considered in this appendix

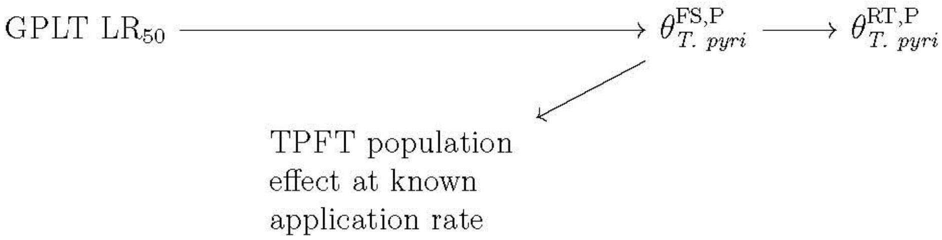


Figure I.2: Revised Bayesian network used for the *T. pyri* calibration considered in this appendix

However, the only data we have about $\theta^{\mathrm{GP},L}_{T.pyri}$ is a single measurement of the GPLT LR_{50} for each of several substances. For each substance, the only important role played by $\theta^{\mathrm{GP},L}_{T.pyri}$ is to link the GPLT result to the TPFS dose–response. Consequently, we can omit $\theta^{\mathrm{GP},L}_{T.pyri}$ from the network and draw an arrow directly connecting $\theta^{\mathrm{FS},P}_{T.pyri}$ to the GPLT LR_{50} measurement. Moreover, because our focus will be on observing the value of the GPLT LR_{50} for a new substance and making inference about $\theta^{\mathrm{FS},P}_{T.pyri}$ and $\theta^{\mathrm{RT},P}_{T.pyri}$, it is convenient for us to reverse the direction of those arrows as in Figure I.2. Note that generally it is not permissible to reverse the direction of arrows in a Bayesian network but that it is legitimate in this situation because of standard factorisation properties of probability distributions represented as Bayesian networks (Lauritzen, 1996). The importance of the arrow reversal is that it changes the modelling required and it does so in a way which makes it easier to bring the available data to bear. Having reversed the arrows, the problem is now to establish a probability distribution representing uncertainty about $\theta^{\mathrm{FS},P}_{T.pyri}$, given knowledge of the GPLT LR_{50}, and a probability distribution representing uncertainty and variability about the measured outcome of a TPFS at known application rate, given knowledge of $\theta^{\mathrm{FS},P}_{T.pyri}$.

For both of these distributions, we will use relevant data about the inter-chemical variation. We will build a simple statistical model which links the three quantities and use data to learn about the inter-chemical variation in the relationship. However, substances vary greatly in their fundamental toxicity. Consequently, the dose–response relationship would be expected to be quite different from one substance to another. In order to share information between substances, we need a relationship which we expect to show some stability. The obvious approach to this is to standardise the application rate used in a field study by dividing it by the GPLT LR_{50}. This is, of course, the HQ for the application rate used in the field study. We then express the dose–response as a model for field study effects depending on HQ (instead of application rate). This way of achieving comparability of results of field studies for different substances is exactly the method used by Campbell et al. (2000) and again by DEFRA (2007). We now hope to establish some form of general pattern to the dose response for multiple substances although we should remain open to the possibility that the dose–response will still differ between substances.

I.1.1 Relevant data

The data were obtained from field studies that were compiled from published Draft Assessment Reports for pesticide active substances. For one anonymised insecticide, dossier data were used. The measure of effect we use for a TPFS is the reported maximum reduction of predatory mite population size per season. Only valid and reliable studies were considered in the assessment. Predatory mite assessments in studies were all performed according to the guidelines Heimann-Detlefsen (1991) and Blümel et al. (2000a,b).

The data are not entirely straightforward. Many of the GPLT LR_{50} values are censored[12]: the study did not have a sufficiently high application rate to achieve 50 % mortality and so we only know a lower limit for the LR_{50}. The TPFSs vary in many ways, including: differing numbers of applications of product and of timings of applications; different timings for measurements of mite counts relative to applications; and different crops.

In what follows, we work with a single overall measure of the application rate obtained by applying the standard NTA multiple application factor (MAF) to the average rate used in the multiple applications. That proxy rate is then converted to a HQ by dividing it by the LR_{50}; where the LR_{50} is censored, the consequence is that we have only an upper limit for the HQ. Multiple effect measurements are also summarised: in some cases by the Henderson and Tilton (1955) calculation, in others by the Abbott (1925) calculation, and in one case the method is not recorded. For each field study, we have therefore a single measure of effect and a single, possibly censored, HQ.

Data analysis

Figure I.3 shows the data. Probably the most striking thing is how much variation in effect there is between TPFS effects for a single substance, even when the HQ is relatively similar in different studies. It is possible that this is partly because the proxy application rate is not a good summary of the application regime in a study. It is certainly quite difficult to see clear evidence of a positive association between HQ and effect within substances although there is some overall evidence of association.

Part of the difficulty in looking for association is that Figure I.3 is a bit cluttered. In Figure I.4, the original data are clustered: for a single substance, all data having HQ within a range of half an order of magnitude are combined to make a single point. Where multiple clusters result for a single substance, they are connected by black dashed lines. We can see that for the majority of substances, there is not enough variation in HQ to enable us to see an association with effect. For three substances there is a fairly obvious positive association, for one a hint of a negative association and for another no association.

This does not mean that there is no association but it means that empirical analyses are not enough. It will be necessary to rely on statistical modelling to make further progress.

[12] Statisticians use the term censored for data for which an exact measurement was not made and so it is known only that it would be greater than some value and/or less than some value.

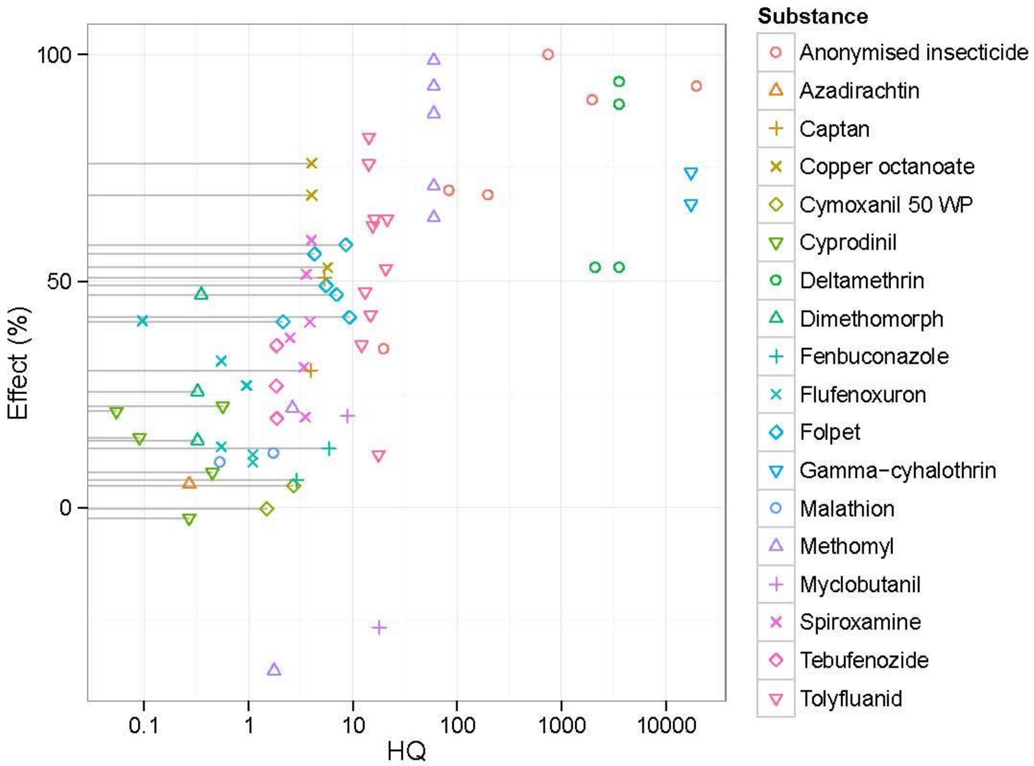


Figure I.3: Effect versus hazard quotient for *T. pyri* field studies. Grey horizontal lines denote data points where the LR_{50} was censored and so we have only an upper limit for the HQ

I.1.3 Statistical modelling

We lack the knowledge with which to build a sophisticated model. In particular, we have little detail with which to model inter-study variability (for the same substance) for *T. pyri* field studies[13]. This means that we lack a theoretical model for how the measured effect in a single field study relates to the real effect for the substance; the real effect would be the average effect across many (hypothetical) field studies for the same substance. To build such a model, we would need to separate out two components: (i) the difference between measured effect and real effect in a single study; and (ii) the variation in real effects between studies. Because both the Henderson-Tilton (1955) and Abbott (1925) methods work by taking the most extreme outcome, relative to control, over multiple measurement events, building a satisfactory statistical model, even of component (i), would require some model of the temporal population dynamics and it is fairly clear from the data underlying the Henderson-Tilton calculations that those dynamics vary considerably between studies[14].

The fundamental approach in what follows is to use some form of regression of effect on $\log_{10}(HQ)$. The censored LR_{50} values pose a problem but we will treat them as point values for now while recognising that they are an additional source of unquantified uncertainty.

[13] In principle there is some information from the ring-testing for *T. pyri* field studies but it is not obvious how to usefully exploit that information in order to build a suitable model for inter-study variation.

[14] However, it might still be possible (and possibly beneficial) to calculate some minimum level for the contribution made by (i) if the available data included not just the average mite count per leaf at each measurement event but also some measure of the variability and of the number of leaves involved.

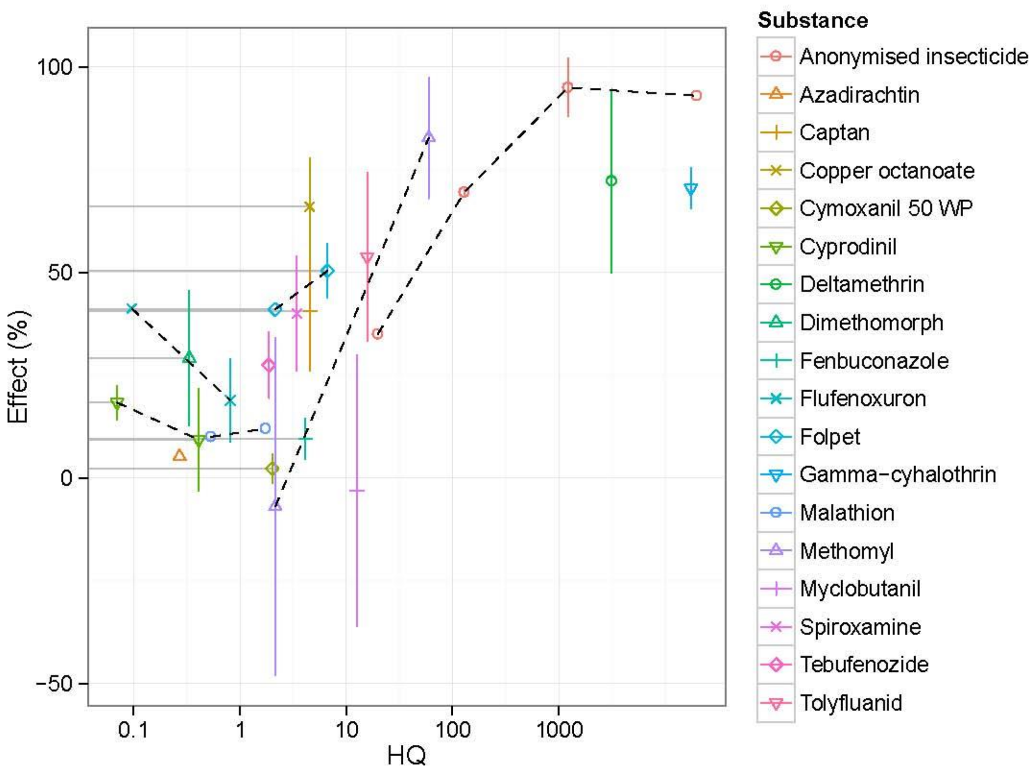


Figure I.4: Effect versus hazard quotient (HQ) for clustered data. Cluster HQ is geometric mean of individual HQ values; cluster effect is mean of individual effects. Vertical lines show plus/minus one standard deviation of effects within clusters. Black dashed lines connect clusters for the same substance

As indicated in this Opinion, common choices of dose–response family are the logit/logistic and probit families. For both, there exists a corresponding regression model and estimation procedure. However, the TPFS effects data we have do not lend themselves readily to such methods as both are designed for use with observation which are counts or percentages from binomial distributions with known sample sizes. The TPFS effects are recorded as percentages but there is no direct underlying binomial model since they derive from Abbott or Henderson-Tilton calculations. Moreover, we have one effect measurement at 100 % and others which are negative; these are difficult to fit easily into logistic or probit regression.

However, we clearly do want to retain the sigmoidal dose–response shape corresponding to the logit or probit families. A simple way to do so is to appropriately transform the response and then apply standard statistical methodology based around linear regression. The effects are being transformed to a scale where the dose–response is assumed to be linear.

DEFRA (2007) used a modified logit transformation of effects data so that the dose–response has a sigmoidal shape which is very close to a logistic dose–response, with small departures at extreme responses. Following that approach, we first adjust all negative measured effects to be 0 % and then pull both 0 % and 100 % slightly away from the boundary before computing the logit transformation. Specifically, we calculate

$$y = \log\left[\frac{\text{effect}^*}{1 - \text{effect}^*}\right]$$

using

$$\text{effect}^* = 0.5 + 0.98(\text{effect} - 0.5)$$

where effect in a field study is expressed as a number between 0 and 1. The choice of 0.98 for shrinking effects towards 50 % was made by examining scatter-plots of y versus HQ and considering their suitability for applying linear regression. On the one hand, one wants to use as large a value as possible as the shrinkage is artificial but, on the other hand, values larger than 0.98 led to plots suggesting potentially excessive influence for the most extreme effects values. Clearly, there is some uncertainty about the correct value to use and this should be taken into account when interpreting the results of the analysis

I.1.3.1 Simple linear regression

In simple linear regression, we assume a straight-–line relationship between effect and $\log_{10}$ HQ with independent normally distributed ‘errors’ in the y-direction: here ‘error’ actually means inter-study variation for the same substance. Mathematically, the model is

$$y_j = a + b \times \log_{10} \mathrm{HQ}_j + \epsilon_j \ (1)$$

where j indexes field studies, a and b are the intercept and slope of the line, respectively, and ϵ_j is the inter-study variation ‘error’. The distribution of ϵ_j is conventionally assumed to be normal with mean 0 and standard deviation σ_ϵ, which has to be learned from the data.

Figure I.5 shows the original data with the simple linear regression line and 90 % so-called confidence and prediction intervals overlaid. The confidence interval here represents uncertainty about the line that best predicts effect from $\log_{10}$ HQ, i.e. uncertainty about the values of a and b. The prediction interval represents uncertainty about the measured effect for a future field study with specified $\log_{10}$ HQ, i.e. it includes uncertainty resulting from the ϵ_j component in equation (1). Figure I.6 shows the same information but with the vertical scale transformed back to the original effects scale.

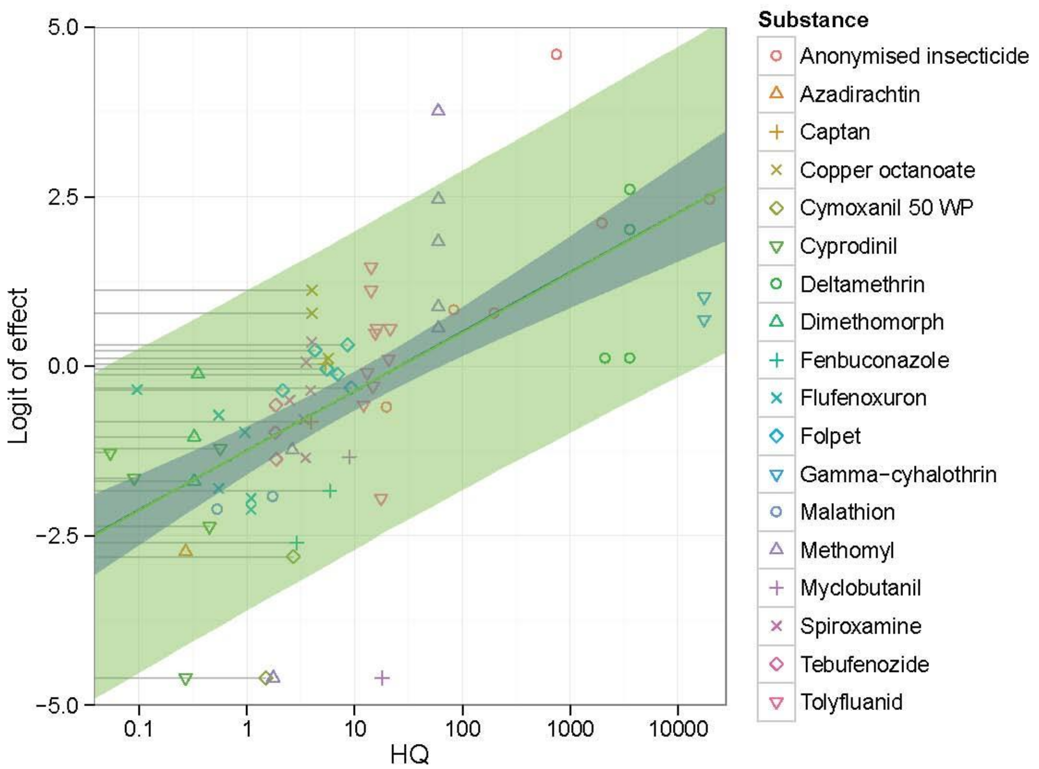


Figure I.5: Original data with overlays: (i) line estimated by simple linear regression; and (ii) 90 % confidence (inner band) and prediction (outer band) intervals for the dependence of logit of effect on HQ

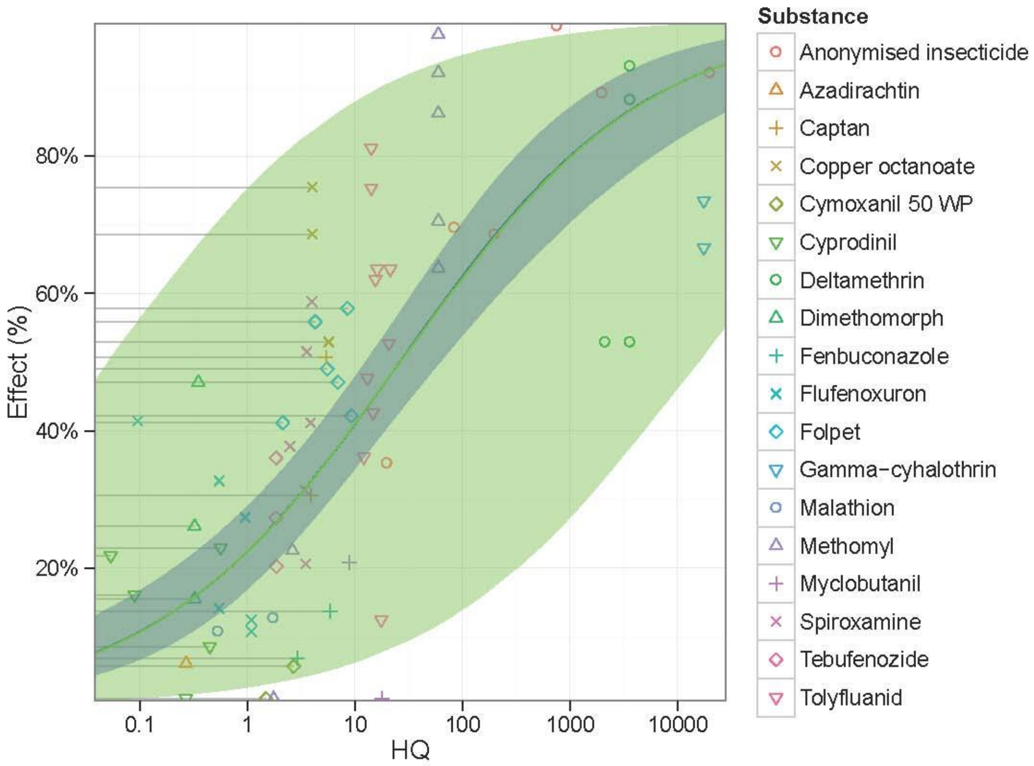


Figure I.6: Figure I.5 back-transformed to the original effects scale

The prediction intervals are wide. In particular, one would have to have HQ below 0.1 to have high confidence that the measured effect would be less than 50 %. However, given the earlier evidence of high variability in measured effect between field studies, it seems more reasonable to focus on the mean effect which would be observed if we performed many or very large field studies: in some sense, this is the true effect. DEFRA (2007) reached the same conclusion.

If the regression line were not subject to uncertainty, one might conclude that an HQ of less than 24 would give a true effect of less than 50 %. However, the uncertainty clearly should not be neglected. If the regression line were not subject to uncertainty, one might conclude that HQ<24 would result in less than less than 50% true effect. However, the uncertainty clearly should not be neglected. One might take the confidence interval (in the vertical direction) as the uncertainty about the true effect corresponding to a specified HQ. If so, one would conclude that HQ<13 would give high confidence of less than 50% true effect.

However, some caution is needed. Firstly, the confidence interval is based on the assumptions of the simple linear regression model: (a) underlying straight-line relationship; (b) independent 'errors'; (c) homoscedastic normally distributed errors. Figure I.5 looks perfectly consistent with assumptions (a) and (c) but it cannot show that they are correct. But the bigger issue is in relation to (b): we have multiple points from a single substance. If the relationship between true effect and $\log_{10}$ HQ differs between substances, the 'errors' in the simple linear regression are not independent.

The data suggest that we should at least consider this possibility. The pattern of residuals in Figure I.5 is not easy to read from the plot but does in fact suggest some dependence. There are seven substances for which there are multiple residuals, each of which has the same sign. This is empirical evidence suggesting that we should try to build a model which allows for differences between substances.

Moreover, it seems ecotoxicologically plausible that there will be variation between substances in the ratio of glass plate to field study LR_{50} values because of different properties of the substance, for example degradation sorption on the leaf material in the field[15]. There are other large sources of uncertainty in the *T. pyri* field studies, for example variability in dilution by vegetation (depending on crop or crop growth stage to which the product is applied) or the timing of application in relation to annual population dynamics of *T. pyri*, etc. Also, in computing the HQ, we used MAF for leaf material as issued in the ESCORT 2 guidance document and this factor takes no account of the size of interval between applications. Finally, the glass plate LR_{50} test result used to compute the HQ is itself subject to uncertainty in relation to the true glass plate LR_{50}. For all these reasons, it would be natural to expect the intercept of the dose–response to vary between substances.

I.1.3.2 Random effects model

The simplest statistical tool for dealing with this situation is the random effects linear regression model. This means is that instead of a single regression line for all substances, we have a different regression line for each substance. To keep things simple for now, we will make the slope the same for all substances and just allow the intercept to vary between substances. This is equivalent, on the original scale, to the sigmoidal dose–response moving to the left or right from one substance to another but being otherwise unchanged. Note that, in principle, the slope could also change for each substance but that the limited amount of per-substance evidence about association between effect and HQ makes inference from these data about such variability very difficult.

The precise version of the random effects model we will use is that inter-substance variability in the intercept follows a normal distribution. Taking the mean value of that distribution together with the common slope provides a central line relating effect and HQ. That central line is the average effect across all substances corresponding to each HQ value. When it comes to thinking about a new substance for which there is no field study, the intercept for that substance is a random-draw from the normal distribution of intercepts.

Mathematically

$$y_{ij} = a + \triangle a_i + b \times \log_{10} \mathrm{HQ}_{ij} + \epsilon_{ij}$$

where i indexes substances, j indexes field studies for the same substance, a and b are the coefficients for the central line representing the average outcome across substances, $\triangle a_i$ is the intercept adjustment (up or down) for substance i, and ϵ again represents inter-study variation 'error'. The random intercept adjustments $\triangle a_i$ are modelled as being drawn from a normal distribution, which has mean 0 and uncertain standard deviation σ_a which needs to be learned from the data.

Models were fitted using the lme4 (Bates et al., 2012) and MCMCglmm (Hadfield, 2010) packages for the statistical software package R (R Core Team, 2012). The former uses likelihood-based inference and the latter uses Bayesian inference implemented by a version of Markov Chain Monte Carlo (MCMC) (Gelman et al., 2013). Figures conveying uncertainty are based on the Bayesian output which is better suited to this task because uncertainties are quantified using probability distributions. However, a disadvantage of Bayesian inference is that a prior distribution is needed for parameters. For the coefficients a and b of the central regression line and the 'error' standard deviation σ_ϵ, there are uncontroversial defaults to use (uniform on the regression coefficients and the logarithm of σ_ϵ) but the situation is more difficult for σ_a, the standard deviation of the random intercepts. Gelman (2006) commends the uniform distribution on σ_a as a sensible simple solution. It should be recognised however that the choice of prior distribution is an additional source of uncertainty. With this choice of prior distributions, there is good agreement between the results of the Bayesian and non-Bayesian fitting procedures. It was verified that the two procedures gave similar central estimates for parameters.

[15] If this is only for leaf dwelling NTAs such as *T. pyri*, one could decrease this uncertainty by performing tests on leaf disks.

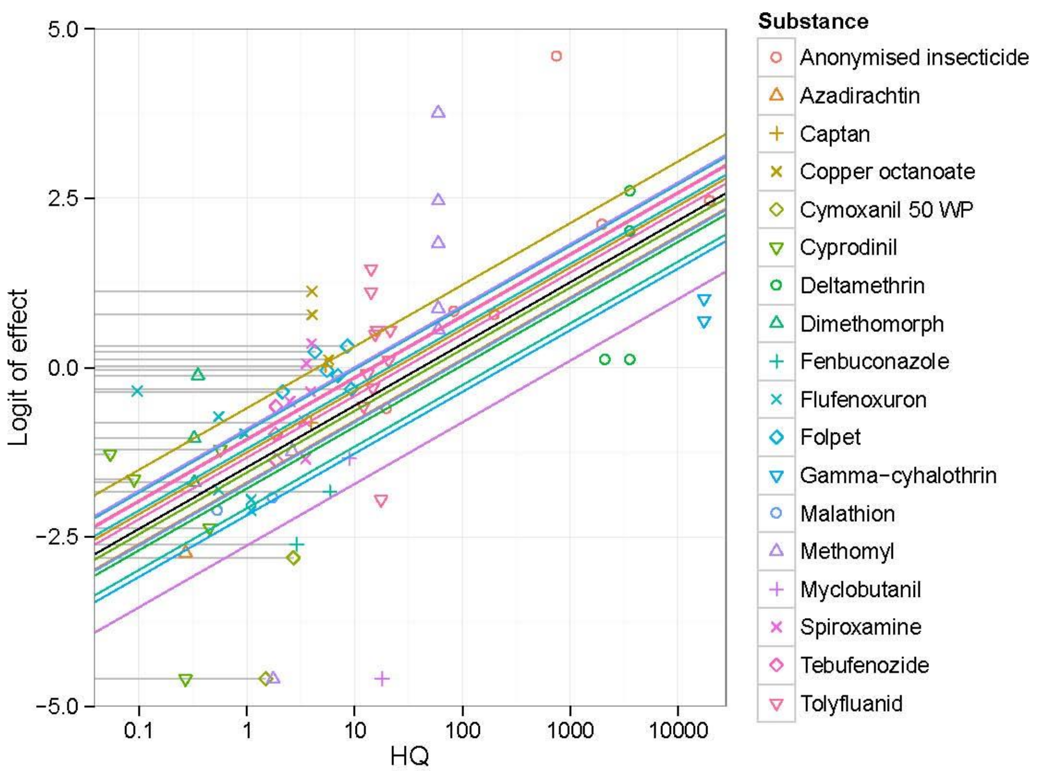


Figure I.7: Original data with overlaid estimated central regression line (black) and per-substance estimated regression lines

Figure I.7 shows the result of fitting the model. It is important to understand that we are not just doing regression separately for each substance. There are two aspects to this: each regression has the same slope; and we have a distribution modelling variability of intercepts and only limited data about most individual intercepts. The result is a phenomenon known as *shrinkage*: the lines for individual substances are often a bit closer to the central line than one might expect. This happens because the model is balancing the contributions of two forms of variability: 'error' variability and inter-substance variability of intercept. If we were to back-transform to the original effects scale, the result would be a collection of logistic curves instead of lines but the *horizontal* spacings would be exactly the same as for the lines in Figure I.7.

If we consider the random effects model to be preferable to the model without substance-specific effects, we can see immediately from Figure I.7 that the consequence is that we believe that there is variability between substances which we might not want to ignore. The model we are using is that there is a normal distribution for the inter-substance variation of the intercept. The standard deviation σ_a of that distribution describes the amount of variation and there is also uncertainty about the value of σ_a.

Now consider the situation where we know only the GPLT LR_{50} for a new substance. Given an application rate, or proxy application rate computed using the multiplication factor from the application regime, we can compute the HQ. There are two distinct sources of statistical uncertainty about the dose–response for the new substance. The first results from inter-chemical variability: where in the distribution of random intercepts does it lie? The second is uncertainty about the values of the parameters for the random effects regression model. In the Bayesian paradigm, for each value of the HQ we can combine these two sources of uncertainty into a single distribution representing uncertainty about the true effect for the new substance. The result is shown in Figure I.8. We see that not only is there uncertainty about the central line for all chemicals (inner dark-coloured band) but that there is additional uncertainty about the true effect at specified HQ for the new substance (intermediate olive-coloured band). For example, in order to have high (95 %) probability that the true effect for the new substance is less than 50 %, we would need an HQ of less than 0.9.

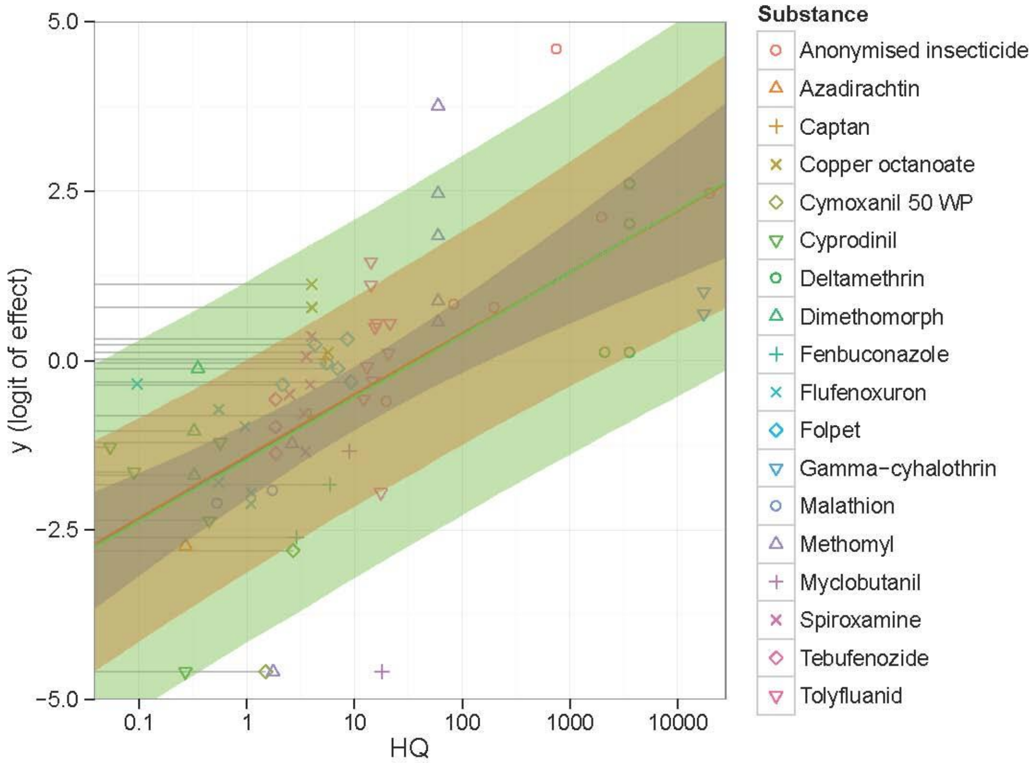


Figure I.8: Original data with overlays, based on the random effects model, reflecting: (i) uncertainty about the central regression line (inner band) for all chemicals; (ii) uncertainty about the true effect for a new substance at specified HQ (intermediate band); (iii) uncertainty about the measured effect in a field study for a new substance at specified HQ (outer band).

The conclusions from using the random effects model are somewhat different from those of simple linear regression. The random effects model is in principle a better description of the situation. Moreover, the random effects model includes the simple linear regression model as a special case (when the standard deviation of the substance intercepts is zero). The model with random effects would be chosen in preference to the simple linear regression model by standard statistical model comparison criteria such as DIC (Gelman et al., 2013).

I.1.4 Data and analysis from Campbell et al. (2000)

The data from Campbell et al. (2000) are also available in summary form in Table 1 in their paper. Figure I.10 shows their data overlaid on the data used here. Some of their LR_{50} values are censored, in both directions, and some effects are also censored. A few substances appear more than once.

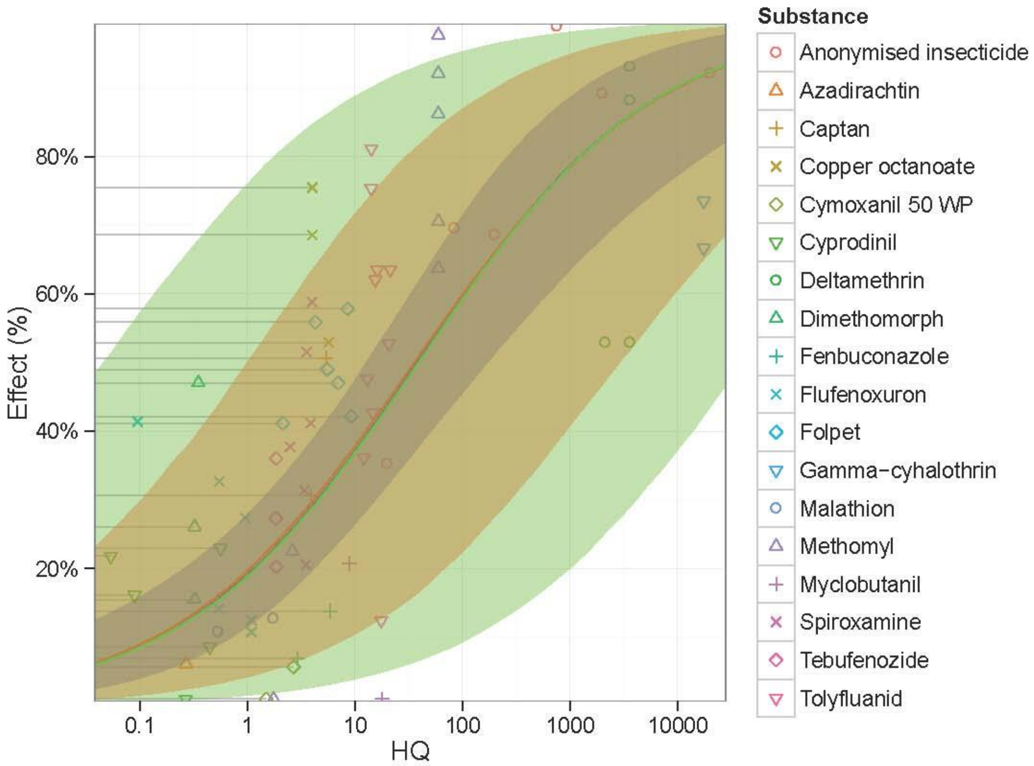


Figure I.9: The random intercepts regression for pseudo-logit effects back-transformed to the original effects scale. Overlays have the same meanings as in Figure I.8

Campbell et al. (2000) concluded[16] that an HQ value of less than 12 would lead to effects of less than 40 %. They obtained their conclusion by observing that there were no measured effects greater than 40 % for any field study in their data with an HQ value of less than 12: there are no black points in the top-left part of Figure I.10 where the is effect greater than 40 % and HQ less than 12. They did not apply any statistical reasoning in making their inference and made no assessment of statistical uncertainty about the choice of 12 as the proposed trigger value.

However, there is one point with HQ = 12.05 and 95 % effect and two points having 40 % effects with HQ values of 5.5 and 3.4. From the general pattern of variability, it is easy to see that there could easily have been field studies with HQ values of less than10 and more than 50 % effects; there just happened to be none in the dataset.

DEFRA (2007) set out to address the issue of uncertainty. They applied linear regression, taking the logarithm of HQ as the independent variable and the same kind of modified logistic as used earlier[17] as the response variable. Negative effects were mapped to 0 % and then the problem of making a logistic transformation for 100 % or 0 % effects was overcome by a numerical adjustment suggested in some statistical literature; there is no indication in their analysis that it caused significant difficulties. Some data points were removed: two on the basis of a personal communication from one of the authors of Campbell et al. (2000) and one (the study with HQ > 10 000 and 0 % effect) because it was considered to be an extreme outlier which would unduly affect the inference.

[16] During the ESCORT 2 workshop, data provided by German authorities suggested that the HQ proposals of 12 and 8 could underestimate effects on *T. pyri* and *A. rhopalosiphi* in the field and the HQ trigger was set to 2. This data was taken into account in the analysis presented here.
[17] It was noted that the choice of logistic transformation was somewhat arbitrary.

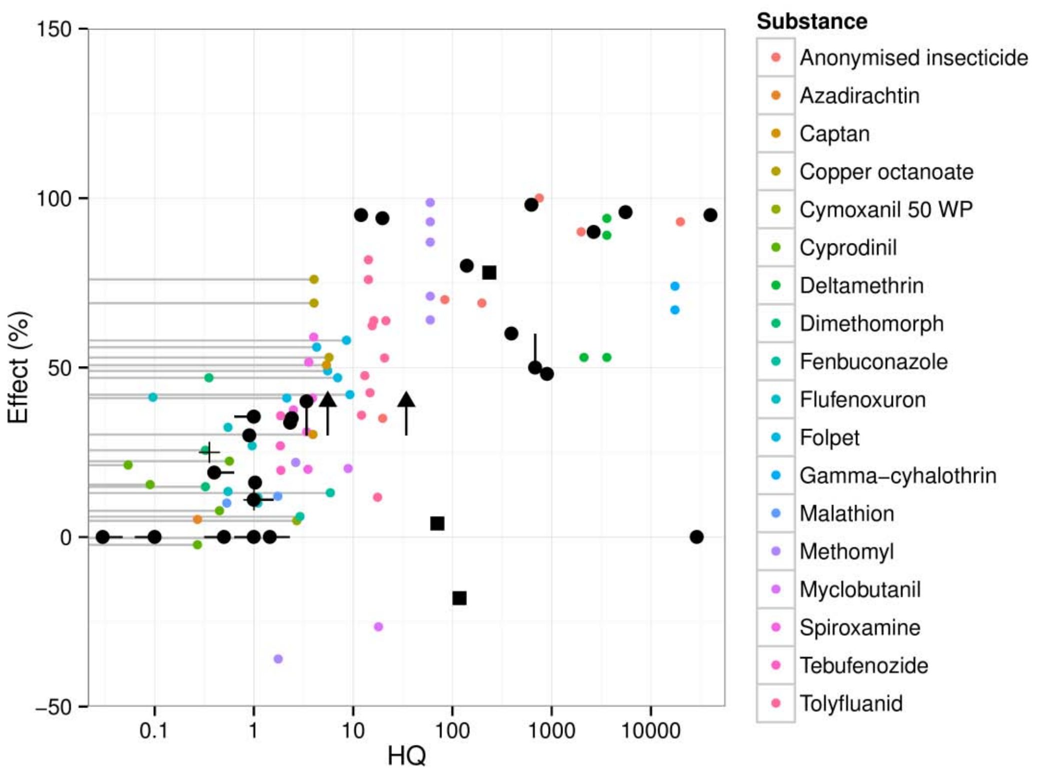


Figure I.10: Data from Figure I.3 with data from Campbell et al. (2000) overlaid as black shapes: 'circle', for substances having a single study, 'triangle', for study 1, 'square', for study 12, and '+' for substance 36. Horizontal black line stub indicates a censored LR_{50} and direction of possible value for HQ; vertical stub indicates a censored effect and direction of possible value.

DEFRA (2007) show that uncertainty about the measurement of field study effect for a specified HQ value, as expressed by the prediction interval, is very considerable using their model: for example, an interval from 0 % to 95 % for an HQ value of 2. They do not suggest that the prediction interval is appropriate for use for setting trigger values. They emphasise that using the confidence interval for the linear regression line as the basis for setting trigger values

'…would be sufficient for setting HQ trigger values if all pesticides followed the same HQ/field effects relationship. However, as explained above, it is unclear what proportion of the variation in the data is due to real differences between pesticides in the HQ/field effects relationship. If a large proportion of the variation was due to differences between pesticides, then an HQ trigger that was protective for one pesticide could be unprotective for another pesticide. The HQ/field effects relationship might vary between pesticides more widely than is indicated by the 95 % confidence intervals for the average relationship…'

They go on to say that a future re-analysis might be able to address the issue of differences between substances in the effect/HQ relationship if there was a sufficient number of substances included in multiple studies. They also note that, if the confidence interval from the regression is considered to be appropriate (i.e. that there are no differences between substances), then the extremely wide prediction interval means that a single field study for new substances would provide little, if any, information over and above the glass plate test outcome.

Neither Campbell et al. (2000) nor DEFRA (2007) addressed the issue of differences between substances in the effect/HQ relationship. The analysis in section I.1.3 suggests that such differences do exist and that they make some real contribution in the sense that the trigger value for 50 % effects would be set lower using the random effects model than using the simple linear regression model.

I.2 From the field study for *T. pyri* to the reference tier for other species

The remainder of this appendix tries to show that one might approach modelling of uncertainty for the steps from the *T. pyri* field study to reference tier outcomes for other species. It is not based on detailed data and is not intended to provide a complete solution at this time.

I.2.1 From the field study for *T. pyri* to the reference tier for *T. pyri*

For the step from the *T. pyri* field study to the reference tier outcome, consider the two dose responses involved. In section I.1.3, we made a simplifying assumption that the slope of the field study dose–response for *T. pyri* is the same for all substances. We now assume that the dose–response slope for *T. pyri* is the same in the reference tier as in the field studies. There is therefore only one parameter which changes: the intercept in the linear relationship between logit of effect and HQ.

In the absence of readily available data to use to learn about this step, we will show how to use expert judgement if it is available. The question to be asked of relevant experts is then the following: if one knew the true rate causing 50 % effects in *T. pyri* field studies, what would the effects be in the reference tier at the same rate? The answer would be expressed as a probability distribution. If preferred, the question could be phrased in terms of the ratio between application rates causing 50 % effects.

In what follows, purely for illustrative purposes, we will suppose that the experts expressed uncertainty about the ratio of rates causing 50 % effects using a log-normal distribution with a geometric mean equal to 1 and a geometric standard deviation equal to 1.5.

Note that if it were judged that there was no difference between true sensitivity in field studies and in the reference tier, this could be modelled simply by changing the geometric standard deviation to 1 so that there would be no uncertainty.

I.2.2 From *T. pyri* to other species in the reference tier

For the step from *T. pyri* to other species in the reference tier, we are dealing with inter-species variability, the domain of SSDs in some areas of ecotoxicology. We do not have the kind of data that would normally be required for the use of standard SSD methods but we can still use the same kind of conceptual model in order to help describe and address uncertainty.

The uncertainty we are concerned with is the difference in sensitivity in the reference tier between *T. pyri* and other species.

In order to keep the presentation simple, we will suppose that the dose–response slope is the same for each species in the reference tier. Therefore, once again, all that varies is the intercept or, equivalently, the application rate causing 50 % effects.

We have the results of two multi-fauna field studies of the type described in the reference tier readily available. However, they are for a single substance and would require a level of modelling which is beyond the scope of this document. Instead we show how expert judgement could be used.

The first question to be asked of experts is what distributional form would be used to model the inter-species variability. Here, we assume a log-normal distribution for the ratio of the reference tier ER_{50} for another species to the ER_{50} for *T. pyri*, i.e. that the logarithm (base 10) of the ratio is normally distributed. The two parameters of that distribution both have a real meaning. The standard deviation σ of that distribution quantifies the amount of inter-species variability in the ER_{50} for the substance: 90 % of species have ER_{50} values in a range of $2 \times 1.645\sigma$ orders of magnitude. Thus, if the standard deviation is 0.3, 90 % of species have ER_{50}s in a range of approximately one order of magnitude; if it is 0.6, the range becomes two orders of magnitude, etc. The mean μ of the distribution quantifies where in the distribution *T. pyri* lies. If μ is 0, then *T. pyri* is in the middle of the distribution, whereas if the mean is 1, *T. pyri* is 10 times more sensitive than species in the middle of the distribution, and if the mean is –1, *T. pyri* is 10 times less sensitive, etc.

Note that if one was sure that *T. pyri* really is a sensitive species in the field, this could be reflected in judgements about μ. One possible line of supporting argument might be the evidence that *T. pyri* is a sensitive species in the laboratory combined with scientific insight that species which are relatively sensitive in the laboratory should also be relatively sensitive in the field.

Expert knowledge elicitation (EFSA, 2014) would then be used to establish expert uncertainty about the parameters of the distribution. For illustrative purposes in what follows, it is supposed that experts expressed uncertainty using the normal/inverse-gamma (NIG) structure, which is a standard form of prior and posterior distribution used for Bayesian modelling of uncertainty about the parameters of a normal distribution (Gelman et al., 2013). There are various approaches to eliciting the parameters but all essentially involve asking for a central range of values for μ and a central range for σ (taking into account their interpretations) from which it is mathematically straightforward to deduce the corresponding particular inverse-gamma/normal distribution.

In the subsequent example of how to compute an assessment factor, we shall suppose that experts made the following specifications: a 50 % range for σ lies from 0.45 to 0.75, and a 50 % range for μ from −0.3 to 0.3. From these, we can deduce that the shape and rate parameters for the gamma distribution for $1/\sigma^2$ are 2 and 0.55, respectively, and that the mean and scale parameter for the conditional normal distribution for μ, given σ are 0 and 0.78 σ, respectively. Note that these hypothetical judgements expect that *T. pyri* lies in the middle of the distribution but allow some uncertainty about exactly where *T. pyri* lies.

Combining uncertainties and obtaining assessment factors

We now consider how to quantify the combined uncertainty about the reference tier outcome for other species given the GPLT LR_{50} for *T. pyri* and how to deduce the size of the assessment factor required.

In order to try to establish an assessment factor, two decisions need to be made first: what level of effect in the reference tier should be considered and what percentile of the inter-species distribution should be protected at that level of effect. For illustrative purposes in what follows, it will be supposed that the percentile of inter-species variability of interest is 5 % as is commonly used for SSDs in aquatic ecotoxicology (Aldenberg and Jaworska, 2000).

Following the procedure described in the opinion, the proposed application rate will be acceptable if it is less than or equal to the value obtained by dividing the GPLT LR_{50} by the chosen AF. An equivalent way of expressing this requirement is that the HQ computed using the proposed application rate must be less than or equal to 1/AF.

The question then is what is the lowest possible value for the AF so that we are sufficiently certain of acceptable effects in the reference tier. In order to establish this, we must first find a way to compute the probability of acceptable effects in the reference tier given a particular choice for AF.

Since a choice for AF translates into a value for HQ, and vice versa, the statistical model in section 1.3 provides a predictive distribution representing uncertainty about the true field study effect for *T. pyri*.

For the substance being considered:

$$y_{T.pyri}^{\mathrm{FT}}(\mathrm{HQ}) = a + \Delta a + b\ \log_{10}\mathrm{HQ}$$

$$y_{T.pyri}^{\mathrm{RT}}(\mathrm{HQ}) = y_{T.pyri}^{\mathrm{FT}}(\mathrm{HQ}) + b\,\varphi$$

$$y_p^{\mathrm{RT}}(\mathrm{HQ}) = y_{T.pyri}^{\mathrm{RT}}(\mathrm{HQ}) + b\,(\mu - z_p\sigma)$$

where $y_{T.pyri}^{\mathrm{FT}}(\mathrm{HQ})$ denotes the logit of true effect in the TPFS for the substance at the specified HQ, $y_{T.pyri}^{\mathrm{RT}}(\mathrm{HQ})$ is the corresponding reference tier logit of effect for *T. pyri*, and $y_{p}^{\mathrm{RT}}(\mathrm{HQ})$ is the logit of effect for the *p*th percentile species in the reference tier. Combining these three equations, we find that

$$y_p^{\mathrm{RT}}(\mathrm{HQ}) = a + \Delta a + b\,(\log_{10}\mathrm{HQ} + \varphi + \mu - z_p\sigma)\ (2)$$

where a and b are the uncertain mean and slope of the central regression line from section 1.3, Δa is the random intercept adjustment for this substance, sampled from the normal distribution of intercept adjustments with uncertain parameters, φ is the uncertain difference between field study and reference tier $\log_{10}$ LR_{50}s for *T. pyri* for this substance and μ and σ are the uncertain parameters of the normal distribution for reference tier inter-species variability.

We shall assume that the three sources of uncertainty involved (*T. pyri* GPLT LR_{50} to TPFS, TPFS to *T. pyri* reference tier, *T. pyri* to other species in the reference tier) are independent. From the output of the Bayesian analysis in section 1.3 using MCMCglmm (Hadfield, 2010), we have a random sample from the posterior distribution representing uncertainty about the parameters in the random effects model and we have independently specified uncertainty about φ using a normal distribution in section 2.1 and about μ and σ in section 2.2 using the NIG structure. We can express the latter two uncertainties by taking a Monte Carlo random sample from each. Consequently, all those samples can be used together to compute from equation (2), for any chosen value of HQ, a sample from the distribution representing uncertainty about the effect for the *p*th percentile species in the reference tier.

Figure I.11 shows, on the logit scale, the median estimate of the effect for the 5th percentile species in the reference tier, based on the hypothetical expert judgements detailed earlier, and a band representing 95 % and 5 % limits of uncertainty. Figure I.12 shows the same information on the original effects scale. We see that, for example, it would be necessary to have an HQ value of less than 0.025, i.e. an assessment factor of greater than 40, to give high confidence of less than 50 % effects for the 5th percentile species; this result is, of course, purely hypothetical since no real expert judgements were used.

It is worth noting that this kind of assessment factor calculation is generic. It would apply, in principle, to all substances for which the data (and expert judgements had they been available and used) were considered to be relevant.

Additional uncertainties

In addition to the uncertainties which have been quantified, a number of additional uncertainties have been introduced in the modelling process:

- Choice of dose–response shape.
- Consequences of assuming the same slope for each dose–response. In a technical sense, this could be addressed quite easily by allowing for variation of the slope. However, it would then be necessary to consider carefully where information about the variability would come from.
- Effect of the modification to the computation of the logit for extreme TPFS effects data.
- The choice of prior for σ_a (the standard deviation of inter-substance variability for the TPFS dose–response intercept).
- Use of censored LR_{50} values as though they were not censored.

Each of these additional uncertainties should be given some further consideration in the future.

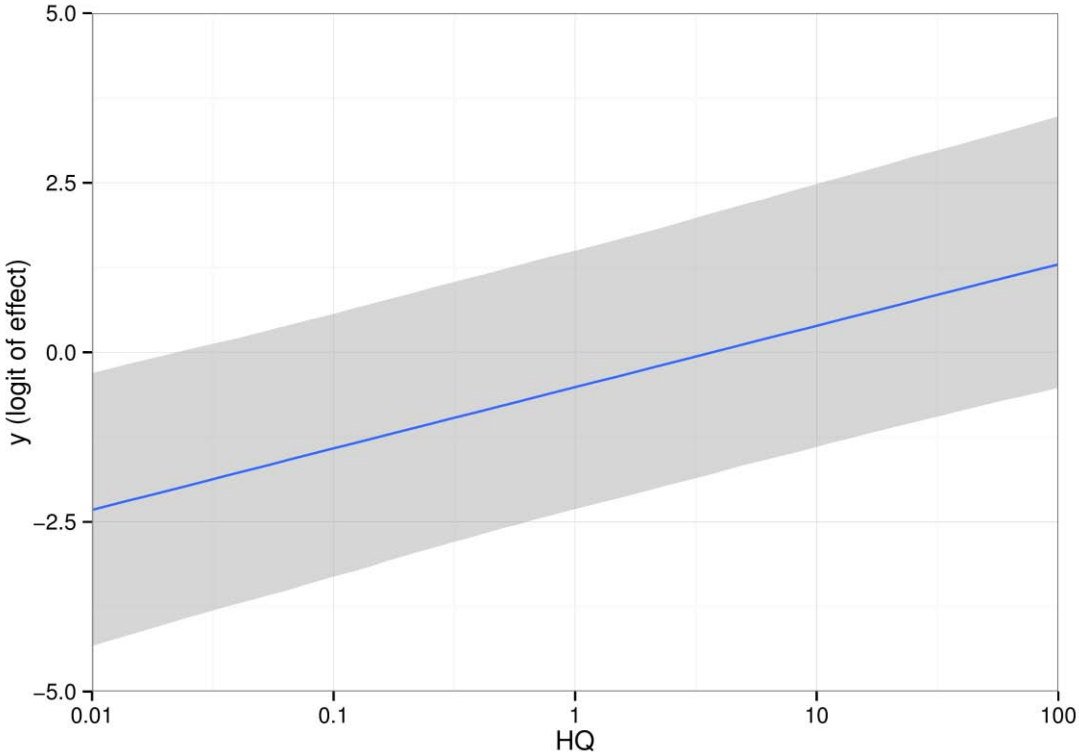


Figure I.11: Uncertainty about the logit of effect in the reference tier for the 5th percentile species depending on the HQ

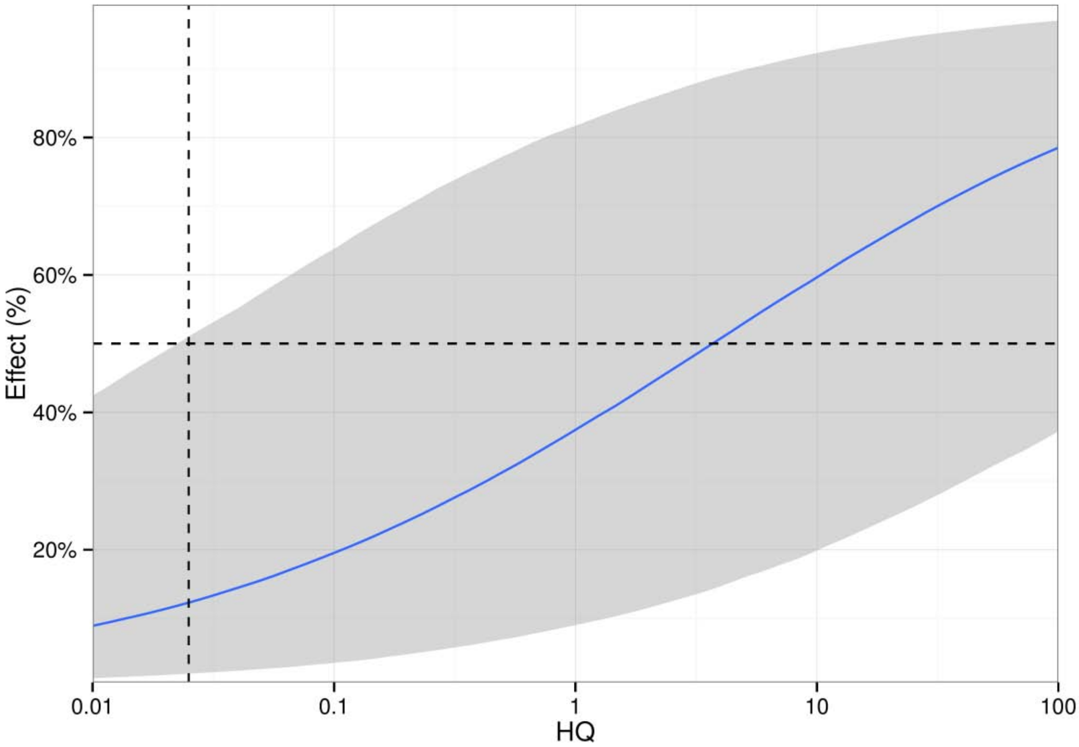


Figure I.12: Uncertainty, back-transformed to the effect scale, about the effect in the reference tier for the 5th percentile species depending on the HQ. The horizontal dashed line corresponding to 50 % effects and the vertical dashed line corresponds to HQ = 0.025

REFERENCES

Abbott WS, 1925. A method of computing the effectiveness of an insecticide. Journal of Economic Entomology, 18, 265–267.

Aldenberg T and Jaworska JS, 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotoxicology and Environmental Safety, 46, 1–18.

Bates D, Maechler M and Bolker B, 2012. lme4: Linear mixed-effects models using S4 classes. Available online: http://CRAN.R-project.org/package=lme4.

Blümel S, Bakker FM, Baier B, Brown K, Candolfi MP, Gossmann A, Grimm C, Jäckel B, Nienstedt K, Schirra KJ, Ufer A and Waltersdorfer A, 2000a. Laboratory residual contact test with the predatory mite Typhlodromus pyri Scheuten (Acari: Phytoseiidae) for regulatory testing of plant protection products. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative, IOBC/WPRS, Gent, Belgium, 128–150.

Blümel S, Aldershof S, Bakker FM, Baier B, Boller E, Brown K, Bylemans D, Candolfi MP, Huber B, Linder C, Louis F, Müther J, Nienstedt KM, Oberwalder C, Reber B, Schirra KJ, Sterk G, Ufer A and Vogt H, 2000b. Guidance document to detect side effects of plant protection products on predatory mites (Acari: Phytoseiidae) under field conditions: vineyards and orchards. In: Guidelines to evaluate side-effects of plant protection products to non-target arthropods. Eds Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs M, Reber B, Schmuck R and Vogt H. IOBC, BART and EPPO Joint Initiative, IOBC/WPRS, Gent, Belgium, 145–158.

Campbell PJ, Brown KC, Harrison EG, Bakker F, Barrett KL, Candolfi MP, Canez V, Dinter A, Lewis G, Mead-Briggs M, Miles M, Neumann P, Romijn K, Schmuck R, Shires S, Ufer A and Waltersdorfer A, 2000. A hazard quotient approach for assessing the risk to non-target arthropods from plant protection products under 91/414/EEC: hazard quotient trigger value proposal and validation. Journal of Pest Science, 73, 117–124.

DEFRA, 2007. Methods of addressing variability and uncertainty for improved pesticide risk assessments for non-target arthropods. Report for United Kingdom Department of Environment Food and Rural Affairs project PS2307.

EFSA, 2014. Guidance on expert knowledge elicitation in food and feed safety risk assessment. EFSA Journal 2014;12(6):3734, 278pp. doi:10.2903/j.efsa.2014.3734

Gelman A, 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1, 515–533.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB, 2013. Bayesian Data Analysis (3rd edition). Chapman and Hall/CRC, Boca Raton.

Hadfield JD, 2010. MCMC Methods for multi-response generalised linear mixed models: the MCMCglmm R package, Journal of Statistical Software, 33, 1–22. Available online: http://www.jstatsoft. org/v33/i02/

Heimann-Detlefsen D, 1991. Auswirkungen eines unterschiedlich intensiven Pflanzenschutz-und Düngemitteleinsatzes auf die Collembolenfauna des Ackerbodens. PhD thesis, Technical University of Braunschweig

Henderson CF and Tilton EW, 1955. Tests with acaricides against the brow wheat mite. Journal of Economic Entomology, 48, 157–161.

Lauritzen SL, 1996. Graphical Models. Oxford University Press, Oxford, UK.

R Core Team, 2012, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: http://www.R-project.org/

ABBREVIATIONS

BBCH	Code for the phenological development stages of plants (crops) http://pub.jki.bund.de/index.php/BBCH/issue/view/161
DRT	drift reducing technology
EFSA	European Food Safety Authority
FOCUS	Forum for the Co-ordination of Pesticide Fate Models and their Use (http://focus.jrc.ec.europa.eu/)
Effect classes	Effects on populations or functional groups are defined as large if the effect is > 65%, medium 65 % - 35%, small <35 % - >10%, negligible <= 10 % or a comparable non-detectable effect
GAP	good agricultural practice
LAI	leaf area index: the single sided surface area of the leaves per area soil surface
NTA	non-target arthropod
PPP	plant protection product