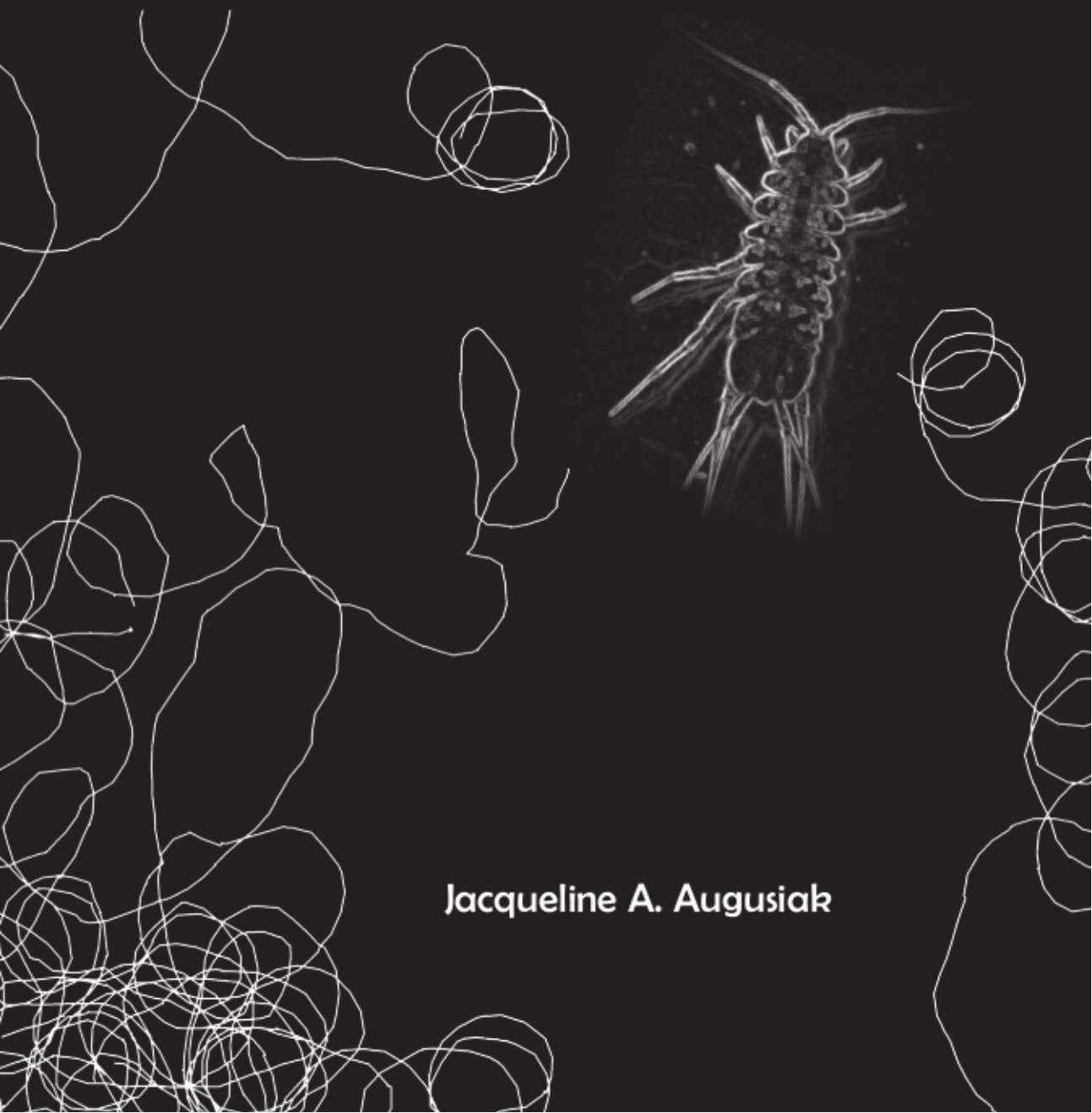
IMPROVING COMMUNICATION AND VALIDATION OF ECOLOGICAL MODELS

A case study on the dispersal of aquatic macroinvertebrates



Jacqueline A. Augusiak

Improving communication and validation of ecological models

A case study on the dispersal of aquatic macroinvertebrates

Jacqueline A. Augusiak

Thesis committee

Promotor
Prof. Dr Paul J. van den Brink
Personal chair, Aquatic Ecology and Water Quality Management
Wageningen University

Co-promotor
Prof. Dr Volker Grimm
Deputy-Chair, Ecosystem Analysis
Helmholtz-Centre for Environmental Research GmbH – UFZ, Germany

Other members
Prof. Dr Marc Naguib, Wageningen University
Dr Emiel E. van Loon, University of Amsterdam
Dr Udo Hommen, Fraunhofer Institute for Molecular Biology and Applied Ecology, Schmallenberg, Germany
Prof. Dr Ryszard Laskowski, Jagiellonian University, Krakow, Poland

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE).

Improving communication and validation of ecological models

A case study on the dispersal of aquatic macroinvertebrates

Jacqueline A. Augusiak

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 6 October 2016
at 11 a.m. in the Aula.

Jacqueline A. Augusiak

Improving communication and validation of ecological models: A case study on the dispersal of aquatic macroinvertebrates

204 pages.

PhD thesis, Wageningen University, Wageningen, NL (2016)
With references, with summary in English

ISBN 978-94-6257-937-8
DOI http://dx.doi.org/10.18174/390313

“We are at best guessing, hoping, elaborating, and our uncertainty is clearly notable in our many scientific, religious and philosophical interpretations.”

- Joseph Rain -

CONTENTS

1 Introduction

In an effort to make use of available resources and space, human activities continue to shape and change the Earth. Consequently, ecosystems are put under stress through exploitation and pollution so that the number of cases with irreversible environmental degradation keeps rising. Freshwater resource exploitation, deforestation, overfishing, waste disposal, and release of toxic chemicals are only a few examples that already left a lasting footprint on the planet. Ecosystem health, however, is crucial not only in regards to aesthetics or recreational purposes, but also for the maintenance of ecosystem services central to human well-being such as food or oxygen production, nutrient cycling, or erosion control (Millennium Ecosystem Assessment 2005). At the same time, it remains uncertain to which degree ecosystems are able to cope if human actions go unmanaged. This is not only due to the complexity of the various interacting stress factors, but also due to the systems' inherent dynamics, which usually are not fully understood either.

In order to protect ecosystems and their crucial provision to humankind and, at the same time, ensure the globally growing demands for food, water, and shelter, environmental management practices are implemented in more and more countries. Many countries now require an assessment of risks to the environment before granting the use of chemicals or other human activities. Environmental policy- and decision-makers, however, cannot always anticipate the full extent of their decisions and need a set of tools to grapple with the spatial dimensions and complex interactions occurring over time. As such, models are increasingly used to bridge knowledge gaps and to weigh different alternatives against each other.

Ecological models provide such a tool set that is increasingly used for policy advice. Biodiversity developments resulting from different management decisions (Pereira *et al.* 2010) or ecosystem responses to climate change (Cramer *et al.* 2001) can be explored with their help and taken into consideration for a final decision. By being able to capture species-habitat interactions at different levels of biological organization across large spatial and temporal scales, decision-makers can gain an overview of potential outcomes. Thus, ecological models already regularly support control responses to infectious diseases nowadays, from setting vaccination radii, for example against rabies (Källén *et al.* 1985), to weighing between simple exposure-avoidance or vaccination responses in the case of avian influenza (Stegeman *et al.* 2004). They can also provide an initial understanding of transmission routes to guide the development of appropriate action plans as in the case of the Zika virus (Bogoch *et al.* 2016). Furthermore, economic-ecological models are increasingly used to inform about the consequences of habitat destruction on economic gains. Jin et al. (2003), for example, simulated regional economic interactions in a marine food web and determined the effects of fishing activities and basic environmental quality on monetary gain due to fishing. In another example, Nautiyal et al. (2010) explored the effects of farming on Himalayan biodiversity. Forman and Alexander (1998), on the other hand, used a modelling approach to understand the impacts of road networks and different approaches on ecological flows across roads to define the most promising strategies to reduce road kill and reduce species loss. Neither of these questions could be sufficiently

addressed without accounting for the spatial or temporal components. More often than not, these scales are difficult or impossible to address in empirical studies.

As such, ecological effect models have been put forward in the field of environmental risk assessment of pesticides (Forbes *et al.* 2010). Pesticides are an important group of chemicals designed to optimize crop yields by targeting pests such as fungi, parasites, or weeds. They are intended to be biologically active and as such also pose a risk to non-target species in neighbouring terrestrial and aquatic habitats where these substances can end up through drainage, aerial drift or leaching through the soil. In the European Union, new pesticides, thus, need to undergo an extensive testing regime under Regulation (EC) No. 1107/2009, formerly Directive 91/414/EEC (EC 1999; 2009), to prevent unacceptable negative effects on the environment during and after use (Hommen *et al.* 2010). In the first tier, acute and chronic effects of pesticide exposure are assessed under standardized test conditions to determine worst-case estimates. If a pesticide fails during these preliminary tests, higher tier studies are performed that allow for the incorporation of more realism, but also more complexity, such as more complex exposure scenarios or additional test species to estimate the variation in sensitivity. Further higher tier studies include tests under semi-natural conditions in micro- or mesocosms where effects on the higher levels of biological organization, i.e. populations or community, are assessed. The results from these studies are subsequently translated into hazard quotients to derive an environmentally acceptable concentration that is considered "safe". This approach, however, mostly relies on empirical information rather than mechanistic understanding. Consequently, applying the respective findings to species that were not included in the prior test regime to larger spatial scales or different regions, or to effects of additional chemical exposure implies large uncertainties.

Ecological effect models, such as population models, can integrate the effects of pesticide exposure on individual survival and reproduction with the species-specific life history to estimate the resulting implications for the population. To this end, a suite of different modelling approaches that is able to translate effects of chemical exposure from the sub-organismal or individual level to higher levels of biological organization has been developed. Dynamic energy budget (DEB) theory, for example, takes into account a species' metabolism and resulting allocation of energy resources to individual growth, reproduction and maintenance activities such as food search and feeding as well as detoxification processes. Chemical exposure may shift energy allocations and subsequently affect reproduction or survival of an individual, which can impact the population as a whole (Jager *et al.* 2004). Toxicokinetic-toxicodynamic (TKTD) models, on the other hand, simulate the time course of processes leading to toxic effects on individuals. The toxicokinetic part describes the uptake and elimination rates of a chemical, which summarizes the concentration inside an organism given a certain external concentration of this chemical. The internal concentration is assumed to be the cause of damage and described in the toxicodynamic part (Ashauer *et al.* 2011). Depending on the chosen endpoint, damage may be defined as immobilization or mortality. Both approaches translate exposure to a given concentration to damage on individuals of a population.

Individual-based models (IBM) can further extrapolate such individual effects to the population level by adding habitat and life history traits to the model (Grimm 2008). Other types of models make use of more mathematical approaches like matrix or differential equations. Matrix models consider the different life stages that a species undergoes and the dynamics leading from one stage to the next. If the chemical affects one or several life stages, this will have rippling effects on the resulting population structure over time. Considering the entire demography of a species can thus yield a comprehensive estimation of population-level effects (Stark *et al.* 2004). Compared to the previous examples, models using differential equations generalize population dynamics and can thus be employed to extrapolate the effect level of a chemical at a given concentration to larger areas or time courses. Ultimately, the modelling approach and possible integration of other approaches can cover questions of larger spatial and temporal scale effects than would be possible with empirical testing alone.

Hence, models can add value to the environmental management process by expressing ecological risks in ways that are more closely related to the prescribed protection goals, which usually target whole populations, key species or biodiversity as a whole (Forbes *et al.* 2016). In this context, Hommen et al. (2010) identified five areas in the risk assessment procedure that would particularly benefit from including modelling studies:

1) Understanding the relevance of effects observed on individuals for the population level,
2) Extrapolating effects of a tested exposure pattern to other, untested, exposure patterns,
3) Extrapolating recovery processes, from individual to population level recovery, including recolonization,
4) Analysing and predicting possible indirect effects in communities, and
5) Estimating bioaccumulation and biomagnification within food chains or food webs.

Despite such promising potential and successful examples of model application, ecological effect models are still rarely used for the environmental risk assessment of pesticides (Schmolke *et al.* 2010b) while they, in contrast, are frequently employed to support ecological conservation strategies (Forbes et al. 2016, Starfield 1997). The complexity of model development and communication, as well as the uncertainty inherent to any model input and output, pose serious hurdles to the acceptance and increased use of ecological models in this context (Hunka *et al.* 2013a). Hunka et al. (2013) isolated a general lack of trust in modelling and the problem of uncertainty as the most pressing issues that need to be addressed before models can find their way to acceptance. Therefore, developing guidance for model use and evaluation, as well as model validation, rank high on the list of possible strategies to overcome those obstacles. Decision-makers need to be able to judge whether a model mimics the real world sufficiently well to answer the question at hand, and whether the model output is credible (Rykiel 1996). This, in combination with unfamiliar terminology and concepts used among ecological modellers, and the importance of their decision, puts decision-makers often enough into a position where they cannot fully trust a study that involves modelling.

To overcome the identified obstacles, good modelling practice and communication among the involved stakeholders from academia, industry and regulatory authorities are required (Schmolke et al. 2010b). After a series of exemplifying modelling studies and discussions, the LEMTOX workshop was organized to provide a platform for international modellers to discuss the strengths and pitfalls of ecological models in pesticide risk assessments (Forbes *et al.* 2009; Thorbek *et al.* 2009). The workshop was followed by the founding of the advisory group MEMoRisk (Mechanistic effect models for ecological risk assessment of chemicals) under the SETAC (Society of Environmental Toxicology and Chemistry) Europe umbrella (Preuss *et al.* 2009). In order to further increase overall performance, acceptance and credibility of ecological models, the European project CREAM (Chemical Risk Effects Assessment Models) was launched in 2009 (Grimm *et al.* 2009). The project brought together 13 full and nine associated partners from different European countries representing the three main sectors involved in chemical risk assessment (academia, large agrochemical companies and regulatory authorities). Moreover, in 2012 and 2013, the workshop series MODELINK was held in which representatives of the three stakeholder groups came together to discuss questions and experiences after actively working on a number of case studies (Hommen *et al.* 2016). The case studies comprised models of birds, terrestrial vertebrates as well as terrestrial and aquatic invertebrates and made use of different modelling approaches and temporal and spatial scales. Thus, the workshops did not only provide practical experience in regards to model applications, but also helped identifying which aspects and scenarios can be particularly helpful to the risk assessment procedure.

The present thesis is part of the CREAM project and is concerned with questions related to model credibility and validation.

1.1 Objectives

The aim of this thesis is to improve the understanding of different sides of the discourse concerning trust, credibility and acceptance of ecological effect models in policy-making processes, particularly in the field of environmental risk assessment of pesticides and to propose ways to overcome obstacles based on this improved understanding.

1.2 Outline

The thesis begins with an overview of the inconsistent and confusing terminology related to model evaluation, verification, and validation. In **Chapter 2**, I provide a review of publications from the field of ecological modelling that specifically targeted questions related to terminology and/or methodological approaches to ensure appropriate model quality and testing. In this study, we also draw on lessons learned from other scientific fields that employ modelling in decision-making processes.

Chapter 3 continues this work by presenting a revised version of the model documentation framework TRACE. This revision expands the applicability of TRACE as a tool for documenting modelling practice to a framework for planning, performing and documenting the entire modelling process. This way, model quality and credibility can be established and followed throughout all stages of model development, analysis, and application.

The next three chapters (**4**, **5** and **6**) focus on a case study for which I re-evaluate a previously developed population model for the aquatic arthropod *Asellus aquaticus*. This model employs a submodel for individual dispersal behaviour for which the parameters were only estimated. In **Chapter 4**, I present an experimental design to observe the movement behaviour of this species under different environmental conditions. **Chapter 5** discusses the results obtained from these experiments. In **Chapter 6**, I eventually revisit the submodel parameterization and discuss the implications of the extended data availability for the population model outcomes.

In **Chapter 7**, the results of this thesis are discussed and placed in a broader perspective to provide a prospective for the use of ecological effect modelling in environmental decision-making processes.

2 MERGING VALIDATION AND EVALUATION OF ECOLOGICAL MODELS TO 'EVALUDATION': A REVIEW OF TERMINOLOGY AND A PRACTICAL APPROACH

Jacqueline Augusiak, Paul J. Van den Brink, Volker Grimm

This chapter has been published in *Ecological Modelling*, 2014, **280**, 117-128.

Abstract

Confusion about model validation is one of the main challenges in using ecological models for decision support, such as the regulation of pesticides. Decision makers need to know whether a model is a sufficiently good representation of its real counterpart and what criteria can be used to answer this question. Unclear terminology is one of the main obstacles to a good understanding of what model validation is, how it works, and what it can deliver. Therefore, we performed a literature review and derived a standard set of terms. "Validation" was identified as a catch-all term, which is thus useless for any practical purpose. We introduce the term "evaludation", a fusion of "evaluation" and "validation", to describe the entire process of assessing a model's quality and reliability. Considering the iterative nature of model development, the modelling cycle, we identified six essential elements of evaludation: (i) "data evaluation" for scrutinising the quality of numerical and qualitative data used for model development and testing; (ii) "conceptual model evaluation" for examining the simplifying assumptions underlying a model's design; (iii) "implementation verification" for testing the model's implementation in equations and as a computer program; (iv) "model output verification" for comparing model output to data and patterns that guided model design and were possibly used for calibration; (v) "model analysis" for exploring the model's sensitivity to changes in parameters and process formulations to make sure that the mechanistic basis of main behaviours of the model has been well understood; and (vi) "model output corroboration" for comparing model output to new data and patterns that were not used for model development and parameterisation. Currently, most decision makers require "validating" a model by testing its predictions with new experiments or data. Despite being desirable, this is neither sufficient nor necessary for a model to be useful for decision support. We believe that the proposed set of terms and its relation to the modelling cycle can help to make quality assessments and reality checks of ecological models more comprehensive and transparent.

'I assert that whenever a dispute has raged for any length of time, especially in philosophy, there was, at the bottom of it, never a problem about mere words, but always a genuine problem about things.'

I. Kant (1786)

2.1 Introduction

Ecological models are increasingly used and needed for supporting environmental decision-making (Schmolke et al., 2010a). Often they are the only way to take into account the relevant spatial and temporal scales and the multitude of processes characteristic to ecological systems. Corresponding experiments can be impossible, and insights from descriptive studies do not necessarily provide enough mechanistic understanding to predict responses of ecological systems to new conditions.

Since models are simplified representations of real systems, a key challenge is, however, to show that the models are realistic enough to meet their intended purpose (Rykiel 1996). Before we can transfer inferences from model results to the real world, we have to demonstrate that the model reproduces observations for the right reasons, not just because it has been tweaked via calibration to do the right thing. If models are in fact used without being carefully checked for their validity, they might lead to erroneous decisions. Pilkey and Pilkey-Jarvis (2007) call inappropriate models “useless arithmetics” and find that "these types of applied models are frequently detached from reality - built on oversimplified and unrealistic assumptions about natural processes". Thus, scepticism with regard to using ecological models to support environmental decisions is a healthy attitude. It is up to the modellers to provide evidence and indicators that their model is realistic enough.

An example field of decision making, where scepticism regarding ecological models so far has prevented the use of models, is ecological risk assessment of chemicals, in particular pesticides (Forbes *et al.* 2009, 2010; Thorbek *et al.* 2009). Ecological risk assessments are required for pesticides to minimise potentially negative impacts on non-target flora and fauna and, thus, on ecosystems in general. Regulatory decisions on whether or not a certain pesticide can be used are, at least in the lower tiers of the risk assessment, based on highly standardised schemes. They focus on effects on individuals of a set of standard species, observed under standardised conditions in the laboratory.

Mechanistic effect models have long been identified as potentially useful tools to extrapolate the limited findings from standard tests to more realistic conditions such as fluctuating exposure profiles, higher levels of biological organisation, and larger temporal and spatial scales, thus making risk assessments ecologically more relevant (Forbes et al., 2010, 2009; Galic et al., 2010; Pastorok, 2002; Thorbek et al., 2010). Mechanistic effect

models comprise ecological and organism-level effect models. They are referred to as "mechanistic" to clearly separate them from descriptive, or statistical, models, and as "effect models" to separate them from physico-chemical models describing the fate and exposure of chemicals in the environment.

Despite the high potential of mechanistic effect models to improve the ecological realism of pesticide risk assessment, so far they have not often been used or accepted in regulatory risk assessments. A major obstacle is the doubt as to whether a given model represents the real world sufficiently well, which is reinforced by a lack of clear criteria for assessing a model's realism. Additionally, a comprehensive model assessment is often hampered by the ambiguous application of terminology within and between involved stakeholder groups. Academics, industry, as well as regulators each possess a different set of vocabulary, knowledge, and interests (Hunka et al., 2013; Jakeman et al., 2006), which interferes with both a more productive advancement and communication of methods, and with actually using models to support decision making.

Terminology regarding model assessment has in general proven to be a particular source of confusion (Oreskes et al., 1994a; Rykiel, 1996). To describe general tasks of quality assurance throughout a model's development and application, academics often use the term "validation" more or less intuitively, due to a lack of a clear and unambiguous definition. Yet, academics are at odds with each other as to what "validation" should mean in a modelling context, to which degree model validation would be generally feasible, and which methods or criteria should be applied to assess the compliance of a given model with its real counterpart.

This issue has been debated in the context of ecological modelling for the past 50 years and still no commonly accepted language and methodology could be agreed upon (see references in (Rykiel 1996)). This makes it very hard to clearly assess and communicate the credibility of models, which in turn makes it difficult, if not impossible, for decision makers, who are usually not trained in assessing whether a model is good enough, to let models influence their decisions. Other domains, e.g. hydrology, economics, meteorology, or environmental engineering, where mechanistic models are being used as well to support decision making, are facing similar problems (Ferson et al., 2008; Gass, 1983; Hodges and Dewar, 1992; Oriade and Dillon, 1997; Refsgaard et al., 2005).

In this article, we review and evaluate the literature concerning the terminology and methodology regarding model validation. We focus predominantly on literature related to ecological models but draw relevant lessons from other scientific fields with relations to regulatory frameworks to provide a pragmatic solution to the above-mentioned challenges. According to the most dominant trends that we could identify, we will propose a common vocabulary for the evaluation of applied ecological models. This can for example assist the risk assessment process by introducing a structured system of language. In particular, we will suggest the new, artificial term, "evaludation", which is a merger of "evaluation" and "validation".

Evaludation consists of several elements, or steps, that correspond to the different stages of iterative model development forming the “modelling cycle” (Grimm & Railsback 2005). They thus serve as the main structuring elements for the suggested terminological system. The modelling cycle consists of the following elements (see also section 3): formulation of the questions to be addressed; assembly of hypotheses that constitute our conceptual model of the system in question; choice of model structure, i.e. choice and representation of entities, state variables, and processes; implementation of the model via equations and/or a computer program; model analysis; and communication of model output.

Based on this approach, we will demonstrate that validation is not a binary criterion that is determined once a model’s development has been finished. Rather, overall model credibility arises gradually throughout the entire modelling cycle.

2.2 Terminology and concepts

Mechanistic modelling simplifies real-world processes to understand driving mechanisms well enough so that forecasts of a system’s response to certain conditions become feasible. This simplification implies the risk that not all relevant factors were captured or that relevant data are missing. Investigating these deficiencies in detail is not always feasible due to monetary, time, or other constraints. For this and other reasons, models inherently possess a level of uncertainty.

To reduce the likelihood of a flawed decision due to an uncertain, simplified representation, decision makers usually demand that a model should be validated. Typically, they ask for a comparison of model output with new empirical data to determine whether possible discrepancies render the model too unrealistic for use. Many scientists argue (correctly in our opinion) on the contrary that this approach to validation is too limited for at least three reasons. First, agreement between modelled and empirical data does not necessarily imply that a model is “correct”, but could also result from a combination of “wrong” input parameters and process representations (Oreskes & Belitz 2001). Second, this kind of direct validation often is impossible to achieve because such data do not exist, which is rather the rule than the exception in ecological and environmental systems. In fact, this is the reason why models are needed for these systems in the first place. Third, the genuine meaning of the word "validation" does not fully match with the uses of the term in ecological modelling and is accompanied by philosophical discourses about its legitimate usage.

It seems obvious that validation should not be mistaken with “truth”, although the term certainly implies a strong sense of legitimation (Oreskes et al., 1994b; Rykiel, 1996). Decision makers would appreciate having some form of quantifiable certification that increases confidence in a model’s appropriateness for application; or, as a risk assessor of pesticides once asked: “Isn’t there a kind of R-square to assess a model’s validity?” This desire is understandable but reflects a lack of knowledge and understanding of how modelling is usually done and should be used, i.e. the modelling cycle. If validation would

be defined to depend on only one or a few expressions of error, major flaws in the model structure could still mislead a decision.

However, decision makers cannot be blamed for lack of understanding of the above points. The roots of the controversy around validation reach much deeper and keep confusing modellers as well. One of the main reasons for disagreements concerning semantics and methodological approaches lies in the philosophical views on how science is performed and, in turn, what validation means in science in general.

Logical empiricism, or positivism, dominated scientific conduct between the middle of the 19th to middle of the 20th century. This school of thinking favoured inductive inferences building from singular observations and/or experiments to universal statements such as hypotheses or theories (Barlas & Carpenter 1990; Refsgaard & Henriksen 2004). The proposed hypotheses or theories are eventually to be tested in experiments that are designed to confirm or refute the general statement at hand. From a model validation perspective, such an approach would render the process of validation formal and algorithmic. Under such premises models would be assumed to be objective and absolute representations of the modelled system, such that they could only be either true or false. This perspective seems to be taken by many non-modellers.

Critics of this approach (Popper 1959; Kuhn 1962) argue that theories can be only falsified and never verified. Typically, they follow a more deductive approach towards science, where inferences are drawn from universal statements, such as theories or hypotheses, to more specified statements. Conclusions are derived logically from several statements, and predictions of empirical patterns must be formulated as deductive consequences from theories or hypotheses. If those conclusions and predictions can be shown to be true, the overarching hypothesis is deemed corroborated or confirmed (Popper 1959). The larger the wealth of confirming observations the more credible the respective hypothesis is deemed to be. However, no matter the number of confirmations, there is always a chance that an observation can be explained by more than one theory. Furthermore, a single falsifying incident is sufficient to reject the correctness of the scrutinised hypothesis; for example, seeing a single black swan falsifies the theory that all swans are white, which hitherto might have been "verified" by observing a million white swans (Taleb 2010).

From this rationalist, deductive perspective, validation becomes a less formal process since a valid model is assumed to be one of several probable representations of a real-world process. Barlas and Carpenter (1990) as well as Oreskes and Belitz (2001) and Oreskes et al. (1994a) argue that one such representation may be preferable over other alternatives, but that no model could claim absolute objectivity as each is also subject to the modeller's subjectivity, view and understanding of the world, and proneness to mistakes. Thus, models are neither true nor false but lie on a continuum of usefulness for which credibility can be built up only gradually (Barlas and Carpenter, 1990; Rykiel, 1996). The question is transferred from whether or not a model holds true to how likely it is to be sufficiently true in the light of accumulated, existing evidence and the model's purpose.

Ecological modellers have discussed model validation since the 1960s. The development of ideas and methodological concepts for validating ecological models underwent several turns since then. Levins stated in 1966 that validation of a model ought to be the generation of testable hypotheses rather than finding that a model is "true" but he left out any quantifiable measures of assessment. On the other hand, Goodall (1972) suggested that the degree of agreement between a model and its real counterpart would be an appropriate measure, which corresponds to today's most common understanding of validation. He furthermore suggested that model input data and the field data used for comparison should be statistically independent. This line was followed in 1977 by Overton. He viewed modelling as an iterative process of refinements and calibration until the output met specified performance criteria, that is, the model was capable of mimicking a predefined data set. He acknowledged that validation in the sense of absolute truth was not possible, as this approach does not necessarily allow identifying the most appropriate model from a set of candidate models.

Early ideas of evaluating a model according to its purpose were discussed by Holling (1966), May (1973) and Caswell (1976). Caswell distinguished between models used in an engineer-like fashion as predictive tools and models used as tools for scrutinising and testing scientific theory. He furthermore introduced the term "corroborate" for the latter class of models and "validate" for the first. He explained this choice by comparing the testing of scientific models with hypothesis testing in which a statement might be scientifically corroborated or refuted, whilst validation, as defined by Goodall (1972), would resemble a form of engineering performance testing. Caswell furthermore claimed that the two different uses would not have to be mutually exclusive. A model could well be predictively valid and be scientifically refuted at the same time. A famous example of such a model is the Ptolemaic model of the solar system, which makes precise predictions of the planets' visible trajectories, but is based on an incorrect view of the structure of the solar system. Understandably, such combinations should preferably be avoided if models are used to predict responses to changes in the environment.

Holling (1978) and Shugart (1984) both shared the view that models resemble complex hypotheses and that validation therefore is impossible to achieve and that only their falsification is possible. Holling went so far as to consider the request for validated models to be inappropriate. He argued that invalidation could be regarded as a tool to establish the limits of a model's credibility to establish a sufficient degree of belief in the model to justify its application. Shugart built on ideas of Goodall (1972) and Overton (1977) and defined model validation as the application of procedures to test a model's agreement with a set of data that is independent from that used for calibrating and parameterising the respective model. Complementary, he defined verification as a test of whether a model can be made correspond with a given data set.

Rykiel (1996) sought a technical and more pragmatic understanding of the term validation. He pointed out that ecological models usually aim to combine theory and practice and that

this duality leads to conflicts when model validation is sought to combine hypothesis testing and engineering practice, a conflict which remains unresolved until today. His pragmatism is in line with Beven's (2002) suggestion to extend the philosophical context in which environmental models are viewed. Beven suggests that one should explicitly account for underlying uncertainties and promotes Von Bertalanffy's idea of "equifinality", e.g. that more than one model can be reliably applied for a given situation. He considers it an option to compare different possible models (different structural models or parameter combinations) and their closeness to predefined performance criteria to gain a more complete understanding of the influence of alternative considerations. The range of plausible models can thus be limited over time as knowledge about the system grows. In contrast, Oreskes et al., (1994a) argue that equifinality would rather pose a source of doubt than help increasing trust in models. Nevertheless, both, Beven and Oreskes et al., share the view that absolute validation of environmental models is impossible to achieve, as environmental systems are open, which complicates strict deductive thinking.

Botkin (1993) and Oreskes et al. (1994a, 1994b) focused particularly on the semantics of validation and verification. Their concerns were that the usage of these terms would not agree with their original definitions, which, according to the authors' understandings, would follow the deductive school of thinking. Oreskes et al. (1994b) argued that the slight differences in meaning of various alternative terms for validation (namely corroboration, confirmation, verification) matter and that current usage of these terms would not follow a common school of thinking.

The term validation has not been used consistently in the literature. Different authors used different definitions depending on their view of the matter; others had similar meanings in mind but used different synonyms. The same holds for other terms commonly used in relation to evaluating the different stages in the modelling cycle. While (Popper 1959) used the term corroboration to describe the process of evaluating a model as a whole, Goodall (1972) named the same process testing. Nowadays, verification usually describes the process of checking a computer code for mistakes (e.g. Rykiel, 1996; Sargent, 2005; Van Waveren et al., 1999). On the other hand, Arthur et al. (1999) described the process of model evaluation with this term, while Borenstein (1998) used the same word for testing whether the correct model has been built, not if it had been built correctly. In contrast, Jakeman et al. (2006) understood verification as a step in which the accurate fit of model results is tested, a step that Jakeman et al. (2006) called substantiation, and a majority of publications validation (e.g. Beck et al., 1997; Gass, 1983; Rykiel, 1996; Van Waveren et al., 1999). These are just a few examples where different authors introduced differing connotations of particular terms. Table 2-1 gives an overview of the confusing usage of terms and synonyms that can be found in the literature. The term "validation" has been given virtually any possible meaning in this context (Table 2-1). A reason for this might be that this term seemingly prejudices expectations of the outcome toward the positive (i.e. the model is valid or the quality is assured), which is one of the major criticisms surrounding

the term. Yet, or maybe because of this positive reassurance, the term persistently remains and returns regularly in discussions.

To conclude, there is little agreement on terms and underlying notions in the literature, with the one exception that it has repeatedly been pointed out that the evaluation of a model should depend on its purpose (e.g. Hoover and Perry, 1989; Mankin et al., 1977; Mayer and Butler, 1993; Rykiel, 1996, 1984).

Table 2-1: Synonyms and definitions used in model testing and validation literature.

Definition	Term	Source
Entire process of forming the decision whether and when a model is suitable to meet its intended purpose by building confidence in model applications and increasing the understanding of model strengths and limitations.	Corroboration	Popper (1959), US - EPA (2009)
	Evaluation	Bart (1995), Borenstein (1998), Committee on Models in the Regulatory Decision Process (2007), Hodges and Dewar (1992), Jakeman et al. (2006), Loizou et al. (2008), Schmolke et al. (2010b)
	Testing	Goodall (1972)
	Validation	Bacsi and Zemankovics (1995), Barlas (1996), Borenstein, (1998), Gass (1983), Hodges (1991), Kirchner et al. (1996), Landry et al. (1983), Macal (2005), Sargent (2005)
	Verification	Arthur et al. (1999)
Tests to ensure that the 'right model' is being built.	Validation	Aumann (2007), Ormerod and Rosewell (2009)
	Verification	Borenstein (1998)
Assuring that the computer program and implementation of the conceptual model are correct.	Verification	Aumann (2007), Barlas and Carpenter (1990), Gass (1983), Hodges (1991), Loizou et al. (2008), Macal (2005), Oriade and Dillon (1997), Ormerod and Rosewell (2009), Refsgaard and Henriksen (2004), Rykiel (1996), Sargent (2005), Schmolke et al. (2010a), US - EPA (2009), Van Waveren et al. (1999)
Assessment of the implications of errors made in design and implementation for the model output and whether the output behaviour exhibits the required accuracy with regard to the model's intended purpose. The assessment is mainly built on comparing model output to data that were preferably not used for model development.	Validation	(Arthur et al. (1999), Beck et al. (1997), Ferson (1996), Gass (1983), Oriade and Dillon (1997), Ormerod and Rosewell (2009), Refsgaard and Henriksen (2004), Rykiel (1996), Van Waveren et al. (1999), Wang and Luttik (2012)
	Verification	Jakeman et al. (2006)
	Substantiation	Borenstein (1998)

2.3 Proposed terminology based on the modelling cycle

Many of the discussions listed above focus on general aspects of how validation should be defined, what it should comprise, or how it should be done. Most of them, however, do not consider structured approaches. Schmolke et al. (2010a) demonstrated that a structured documentation of the subsequent modelling steps already would support a more comprehensive assessment of a model. They proposed a generic structure for documenting modelling which is built on the structure of the modelling cycle. We propose a similarly structured approach towards model evaluation.

The central elements in model development are shown in Figure 2-1. Typically, basic or applied questions about an environmental system lead to a conceptualisation of the underlying processes. Once a conceptual model has been derived that seems to account for the most relevant processes to answer the question at hand, the conceptual model is translated into a computerised model. Proceeding from the conceptual to the computerised model works in two steps. First, the conceptual model has to be made quantitative and operational so that it can be run on computers (note that we here also refer to mathematically formulated models, which are numerically solved on computers, as computerised models). This step comprises the definition of entities and state variables for characterising the state of the model system; mathematical or algorithmic submodels that represent the processes included in the model; and a schedule of the model's processes. We call this the "written formulation" of a model. Second, the written formulation has to be translated into a program that can be run on computers, referred to as the "implementation of the model".

At all stages of the cycle, lack of knowledge and good quality data, and human imperfection, unavoidably induce uncertainty. The level of uncertainty can be reduced by applying a standardised evaluation scheme similar to quality assessment protocols (Refsgaard *et al.* 2005). For such a scheme to be practicable, the different elements in the modelling cycle should be examined separately. To distinguish between these, and to reduce currently prevailing misunderstandings between involved stakeholder groups, we follow the pragmatic recommendations from different scientific fields to split model evaluation into subparts (Figure 2-1; Barlas, 1996; Refsgaard and Henriksen, 2004; Rykiel, 1996; Sargent, 2005). The general logical order calls first for a test of the appropriateness of the chosen model structure before testing accuracy of model output.

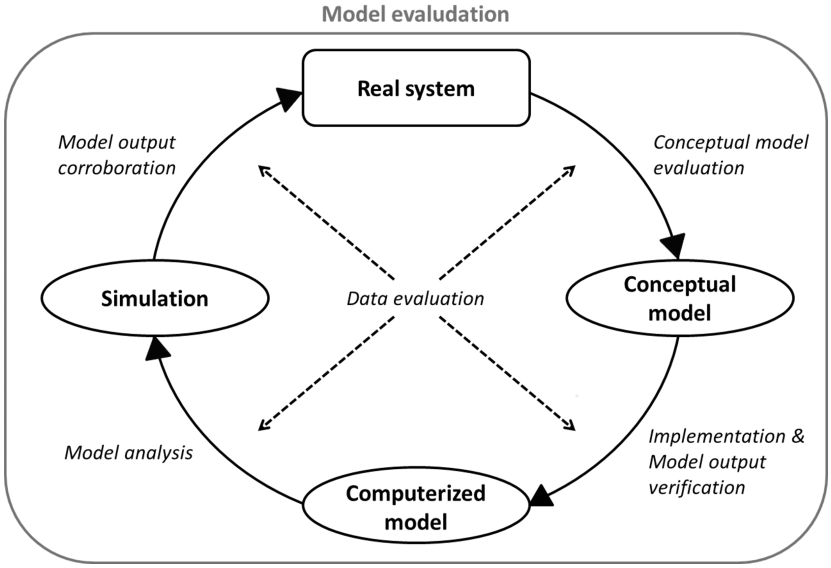


Figure 2-1: A simplified representation of the modelling cycle, consisting of the four main steps of model development and their corresponding elements of evaludation. The terms in italics comprise the terminology that we suggest to organise and communicate model evaludation. The four elements of model development were, in the context of model quality assurance, suggested by Refsgaard and Henriksen (2004) and Schlesinger (1979). Their definition of quality assurance corresponds to what we here refer to as 'evaludation' and what so far in ecological modelling usually has been implicitly referred to as evaluation and/or validation.

To combine the imprecise but important term "validation", and building on its implied meaning for assessing a model's quality, with the more neutral and complementing connotations of "evaluation", we introduce the new, artificial term "evaludation". We define "evaludation" as "the entire process of establishing model quality and credibility throughout all stages of model development, analysis, and application". We suggest this term for several reasons. Firstly, we aim to avoid using "validation" itself while keeping it still visible. Secondly, we aim to link the understandable request for validity assurance with the more neutral, multi-step process of evaluating the quality of not only the model output but also all other relevant aspects of modelling, which can affect model credibility and validity. Furthermore, a new term implies that it is more likely to be specified when it is used, which avoids misunderstandings and emphasises the multi-criteria character of model assessment.

Evaludation consists of six elements, which are necessary to communicate which uncertainties have to be faced at the different stages of the modelling cycle, which evaluation tools are applied for these elements, and which measures should or could be taken to reduce uncertainties. In the following, we will define the different elements that

constitute evaludation and discuss why they are important. Short discussions of possible approaches to tackle the different evaludation steps will be discussed in section 2.4.

2.3.1 Data evaluation

"Data evaluation" is defined as the critical assessment of the quality of numerical and qualitative data used to parameterise the model, both directly and inversely via calibration, and of the observed patterns that were used to design overall model structure. Thus, with data we here not only refer to numerical data, as in some data sheets or spread sheet tables, but also qualitative data, i.e. expert knowledge for which no hard numbers exist. Computer models can take into account such knowledge in the form of probabilistic if-then rules: if a certain state is given, various things may happen with certain probabilities. The term "data" also refers to patterns (Grimm & Railsback 2012) or, in economists' terminology, "stylised facts", which are general trends and signals in data, observations, and empirical knowledge.

To illustrate these different types of data, consider a census time series of a population of small mammals in a certain area over 30 years. The numerical data are the set of abundances, which are uncertain in themselves because abundance could not be determined directly. Expert knowledge could exist about individual behaviour, for example territoriality, and how it changes, in broad terms, in response to changes in population density. Another pattern could occur in a series of years with bad weather, during which variability of animal abundance differs from that in series of years with good weather. Such kinds of data are important for model development, although they are statistically uncertain.

The appropriateness, accuracy, and availability of data used throughout model development are a major source of uncertainty and often a reason for failed attempts to validate a model (Sargent 2005). Data are needed for the design of a conceptual model, to deduce relevant theories and derive mathematical and logical relationships that represent the modelled system sufficiently well to fulfil the model's stated purpose. Furthermore, data are needed to fully parameterise and calibrate a model. Finally, data, which were preferably not used for model development and calibration, are needed to test the model's underlying operational assumptions. Without such independent data, confidence in a model can be hard to establish (Rykiel 1996; Sargent 2005). However, it should be kept in mind that for ecological systems, independent numerical data often do not exist or even cannot exist. Instead, using additional qualitative data or general patterns that were not considered or known during the model development can and should be used ("pattern-oriented modelling": e.g. Grimm and Railsback, 2012; Grimm et al., 2005).

The quality of available numerical data can be corrupted by measurement errors (e.g. by quality of instruments and frequency of their calibration, data logging, etc.), flawed experimental design (e.g. choice of sampling site, small sampling sizes, etc.), and natural heterogeneity and stochastic variability inherent to environmental systems (Gass 1983;

Wang & Luttik 2012). Likewise, expert knowledge and the detection of patterns are notoriously prone to bias and therefore must be treated with particular caution. Another aspect to watch carefully concerns the extrapolation of data from one situation to another. This can include the usage of laboratory data to estimate effects in the field, as well as the usage of data from another climate zone or related, and not the actually studied species.

Data evaluation is needed to ascertain a high level of quality, a point laid out in several quality assurance and control protocols (Van Waveren *et al.* 1999; Refsgaard *et al.* 2005; US - EPA 2009). This is part of the reason why one cannot simply assume that data yield the best testing conditions for a model's structure or output, as data themselves do not always represent the real system sufficiently well. Additionally, experimental data are only gathered during a particular period or in a particular area and therefore represent only one of the many states of the ecosystem (Fagerstrom 1987; Topping *et al.* 2012). A model cannot be expected to provide more accuracy and clarity than what has been used to develop it in the first place.

2.3.2 Conceptual model evaluation

"Conceptual model evaluation" is defined as the critical assessment of the simplifying assumptions underlying a model's design. A conceptual model is our verbal or graphical model of the system of interest with regard to a certain question. As with any element in the modelling cycle, the conceptual model is very simple at first and subsequently develops gradually. Evaluating the conceptual model means to explicitly list, discuss, and justify its most important simplifying assumptions. Typically, assumptions include the choice of spatial and temporal scales; the choice of entities and processes to be represented in the model; considerations concerning stochasticity and heterogeneity; considerations of local versus global interactions; representation of environmental drivers; etc. Furthermore, conceptual model evaluation includes the assessment of whether the structure, underlying theories, concepts, assumptions, and causal relationships are reasonable to form a logically consistent model.

Conceptual model validity is mostly affected by a modeller's subjectivity, incomplete understanding and knowledge of underlying problem entities, and the quality of available data. Different modellers may make different decisions about the kind and form of processes to include in a model. Boesten (2000) found in a comparative study that despite equal starting conditions, i.e. having the same data sets and study objectives defined, different modellers obtained significantly different model results. He identified the expert judgment involved in establishing the process relationships as one of the major causes for this variation. Additionally, incomplete knowledge about the factors that control behavioural aspects of the modelled system, either due to a modeller's lack of awareness of relevant studies or a lack thereof, as well as limitations arising from simplifying assumptions, need to be considered and justified. As another guiding principle, Occam's

razor should be applied to ensure that the chosen model complexity does not introduce avoidable uncertainty (Beck et al., 1997; Clark, 2004; Jakeman et al., 2006).

2.3.3 Implementation verification

We define "implementation verification" as the critical assessment of (1) whether the computer code for implementing the model has been thoroughly tested for programming errors and (2) whether the implemented model performs as indicated by the model description. This element of evaludation is hence primarily concerned with checking the computer code for errors, bugs, and oversights. However, even an error-free program code might not actually implement the model as intended or described, which can be due to ambiguities in the model description or due to misinterpretations of ready-to-use procedures in the employed software platform (for an example of the latter, see Grimm and Railsback (2005, Chapter 5).

Although implementation verification mainly focuses on technical aspects of a modelling exercise, it is also essential for assessing whether a model is of sufficient realism and quality. Tests of independent model predictions (see below) might look promising but without a thorough evaluation of the implementation procedure, the risk could still be high that the model leads to wrong conclusions because the model might not work as we think it does.

2.3.4 Model output verification

"Model output verification" is defined as the critical assessment of (1) how well model output matches observations and (2) to what degree calibration and effects of environmental drivers were involved in obtaining good fits of model output and data.

Model development always includes a judgment of model output according to observed data and patterns and some criteria of similarity. After all, the purpose of models is to be "representations" of real systems, even if this representation has to be much simpler than reality. However, just considering "predicted vs. observed" figures is not sufficient either. Model users need to know how much calibration was involved to make the model fulfil verification quality criteria. The more parameters had to be fine-tuned via calibration, the higher the risk that successful verification was enforced by unrealistic parameter combinations, i.e. by a combination of factors which does not occur like this in the real system. Likewise, a good match of model output and observations might have been imposed by representing strong environmental drivers, for example weather, chemical disturbances, or predation risk so that model output rather reflects the dynamics of the drivers and not necessarily a realistic representation of the system's internal organisation.

Furthermore, when comparing model output to data it is critical to avoid comparing apples with oranges: environmental conditions, initial states of the model world, and data

sampling protocols (for example the timing of sampling a population) implemented in the model should match those underlying the data as close as possible (Zurell *et al.* 2012). For complex models, this can be a major task (Topping *et al.* 2012).

In general, the task of this element of evaludation is to demonstrate that the individuals and populations represented in the model respond to habitat features and environmental conditions in a way that is sufficiently similar to their real counterparts. What can be considered "sufficient" cannot be defined from the outset and also depends on other elements of evaludation, on the overall understanding and experience with managing the system in question, and on whether the model is supposed to deliver absolute or relative predictions. In the latter case, a qualitative agreement of model output and data might already be considered sufficient.

2.3.5 Model analysis

We define "model analysis" here as the assessment of (1) how sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood. Testing model sensitivity is essential since a good match of model output and data might also be the result of fine-tuning several parameters. The match might vanish as soon as one or more or the calibrated parameters are changed. Model evaluations, which do not include sensitivity analyses, are thus too limited.

Sensitivity analyses identify subsets of parameters that have strong effects on the model outputs. Since parameters represent the relative contribution of certain processes and feedbacks, we thereby learn which processes are most important for further considerations, which is an important first step to understanding which factors are most important in explaining model behaviour.

Understanding model behaviour is needed to avoid using a model as a black box. If we understand why and how a model produces certain outputs, we can, if the model is well evaludated, cautiously transfer this understanding to the real world, which would often be more important for supporting decisions than any kind of specific numeric model output. Evaludation thus implies that the modeller has tried several possibilities, and documented them, to understand and explain how model output emerged.

2.3.6 Model output corroboration

"Model output corroboration" is here defined as the comparison of model predictions with independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterised, and verified. The emphasis on new, independent data is important because with data known and used during model development, modelling will often end up with a model reproducing these data. This implies the unavoidable risk that the model has been "tweaked" to do the right thing for the wrong reasons.

Still, for models of complex systems, making a model match known observations can be difficult and it is a myth to believe that a model could reproduce or forecast any data with just enough model parameters. Thus, model output verification can already indicate whether the internal organisation of the real system has been captured sufficiently well, in particular when verification comprised not only one data set or pattern, but multiple ones.

However, even when using several patterns simultaneously to verify if a model is working correctly (Grimm & Railsback 2012), there is still a risk that the respective model might have been manipulated too much to produce the right behaviour for the wrong reasons. Multiple patterns are sometimes not independent from each other and thus do not necessarily reduce this risk (Latombe *et al.* 2011). In contrast, with new data and patterns, this risk is being eliminated. Not knowing these data or information and patterns makes artificial imposing of rules and tweaking impossible.

One form of new data consists of results from new and specifically designed experiments and field studies. This corresponds to one of the most common interpretations of validation: a model is considered "valid" if it made predictions that were confirmed by subsequent experiments. However, as mentioned earlier, such experiments or field studies are usually unfeasible for most ecological systems. Thus, for model output corroboration we usually have to resort to comparing model predictions with data and patterns that already exist, but have not been known or used by the modeller. The guide to finding such data and patterns can be the model itself. Does it predict any striking features, regularities, or patterns, which are robust and seem to emerge from the interaction of the key processes in the model? If so, can we find corresponding data or patterns in the literature, existing databases, or can we confirm them via expert judgement ("This is exactly what I have observed.")?

One example for this approach is a model of natural beech forests in Central Europe (Neuert *et al.* 2001; Rademacher *et al.* 2004). In this model, canopy trees were represented as individuals with certain ages and crown sizes. However, information concerning age and size were never used during model development and calibration. Rather, local stand structure was assigned to three so-called "developmental stages", which take into account the leaf cover in four different height classes. Temporal and spatial patterns regarding the developmental stages, which were know while the model was developed, were used for model output verification, and the model was published and used for its original purpose (Neuert *et al.* 2001; Rademacher *et al.* 2004), which was the estimation of the sizes of forest reserves that would be needed to enable natural spatio-temporal forest dynamics.

In a follow-up project, the age structure of model canopy trees was analysed as well as the spatial distribution of very large and old trees (Rademacher *et al.* 2001). The two patterns found in this analysis were that neighbouring canopy trees on average differed in age by 60 years and that 80% of all trees older than 300 years had a tree of similar size within a distance of less than 40 m. This pattern was confirmed after re-analysing the existing literature (Rademacher *et al.* 2001).

In the current literature, model output corroboration based on patterns identified in the model is the exception rather than the rule. This might partly be due to limited resources, but the main reason seems to be that the term "prediction" is often used in a very broad sense, which blurs the distinction between verification and corroboration of model output. The term "prediction" should be used only for new, independent - secondary - predictions that forecast something new, either results to be obtained in the future or patterns to be detected in the existing data and knowledge. Data used for verification should then be referred to as "model output", as no prediction is involved because the data and patterns "predicted" were already used in the model development.

A clearer distinction between model output verification and corroboration could actually lead to more systematic attempts of model output corroboration. And it should be noted that a model can still be considered realistic and fit enough to meet its purpose, even if corroboration was not possible due to lack of resources or data.

2.4 Evaludation: Planning and approaches

Considering the vast complexity of environmental issues that can be addressed with modelling approaches and the diverse set of modelling concepts, it is not possible to establish a detailed, fool-proof protocol for evaludating a model or declaring whether or not it can be deemed fit for application. Nevertheless, the systematisation of checking the different building blocks of a model throughout its lifecycle and evolution ensures reduced uncertainties, and, maybe more importantly, an easier communication of the capabilities and limitations of a model so that decision-makers feel more confident about using it. Some general concepts and considerations can help to add more structure to the task of model evaludation. Especially ideas derived from general quality assessment and control frameworks, as well as experiences from common practice, can help to establish a more consistent procedure. Some discussions on Good Modelling Practice and existing regulatory protocols in a number of fields have already succeeded in establishing a first, rough guidance (Rykiel 1996; Van Waveren *et al.* 1999; US - EPA 2009). Refsgaard et al. (2005) also provide a review of existing quality assessment guidelines and Matott et al. (2009) give an extensive overview of approaches to analysing model uncertainties.

In the following, we mainly list recommendations given in these contributions. It should be noted, that we focus on the overall scope and rationale of the methods discussed, not on technical details, which are described in the corresponding literature.

An important first question in evaludation is: Who is to carry out the evaludation? A common answer would be "the model user", but we claim that this perspective would be inefficient. As a matter of fact, most modellers perform all steps of model evaludation anyway because they are integral parts of model development, analysis, and testing, and modellers are usually no less interested in evaludating their models than decision makers

trying to use the models or their output. Thus, the correct answer is: both model developers and users.

Often, model developers might be biased and tend to overstate the structural realism of their models. It is thus advisable to either include potential model users in the model development process to establish model acceptance, as has been tried in the CREAM project (Grimm *et al.* 2009), or to follow what is called an "independent verification and validation" approach. The latter is derived from computer science and refers to an evaludation carried out by an external party, which was not previously involved in model development.

It is furthermore crucial to consider the timing of evaludation points. Evaludation measures can be taken either while the model is being developed, or after a model has been completely coded and parameterised. Common practice and experience favour evaludation to take place throughout the model development to reduce costs imposed by errors or misjudgements made early on. This corresponds also to our framework for model evaludation (Figure 2-1), which emphasises the iterative nature of model development: design and parameterisation of a model and its submodels are revised when the model did not pass certain performance criteria.

During the early stages of model development, sufficient time should also be invested in defining performance criteria and benchmarks. Benchmarks are metrics that allow an evaluation model output compared to empirical observations. Thus, they support defining meaningful points of reference for model output verification and corroboration (Jakeman et al., 2006; Kirchner et al., 1996). In many cases, goodness-of-fit parameters or confidence intervals are used as quantitative performance criteria to assess the statistical agreement of observed versus modelled data in form of a hypothesis test. A thorough understanding of the applied metrics is needed for this step to avoid potential misinterpretation due to a misunderstanding of a metrics' weaknesses (Bennett *et al.* 2013). A coefficient of performance, e.g., can be strongly influenced by low sample sizes or outliers, which in return could be a relevant feature of the investigated system. Another tool that is frequently used for qualitative benchmarking is the visual inspection of graphs that trace, for example, the behaviour of model entities.

Currently there seems to be trend in the modelling literature to require increasingly sophisticated statistical tests, in particular Bayesian methods. This trend is laudable, but should not lead to an underestimation of “face validation, which is defined by Klügl (2008) as: “all methods that rely on natural human intelligence” (p. 39). Examples listed by Klügl include: “structured walk-throughs, expert assessments of descriptions, animations or results". Klügl accordingly concludes: "face validity shows that processes and outcomes are reasonable and plausible within the frame of theoretic basis and implicit knowledge of system experts or stakeholder. Face validation may be applied from the early phases of the simulation study under the umbrella of conceptual validations. It is often also called

plausibility checking". This way of comparing model output to data is thus an integral tool for the evaludation steps "Conceptual model evaluation" and "Model output verification".

Finding the right benchmarks, or metrics, often is part of the problem to be solved in ecology and environmental systems. It can furthermore be necessary to adjust or extend the set of performance criteria. New knowledge or understanding gained during the modelling process can enforce changes not only to the conceptual or computerised model, but also to the way it is analysed. For defining suitable and representative benchmarks it is important to take natural stochasticity into account by using confidence intervals and by focussing on a set of benchmarks. The latter is the core idea of pattern-oriented modelling, i.e. to use multiple patterns for model output verification and corroboration, not just only one (Grimm *et al.* 2005; Grimm & Railsback 2012).

2.4.1 Data evaluation

At this step, a list of all parameters used in a model should be compiled with a description from which sources the parameter values were taken. Additionally, the parameter's units and where exactly (page number, Table number) in a publication they were found need to be provided. If multiple data sources exist for the same parameter it should be mentioned how much the corresponding values differed and whether the differences are caused by different environments, sampling protocols, or other reasons. If no hard data should exist for a given parameter, it should be noted on what grounds the parameter "guesstimation" was based, e.g. expert knowledge, data from similar species, theoretical considerations, etc.

Essentially, when assessing the quality of the data and patterns used, not only do the measurement protocols need to be evaluated but conclusions drawn from the data should be challenged as well. In some instances, wrong interpretations of data caused delays of model development (Holling 1978).

The main question in data evaluation is whether the available data are sufficient to support the choice of the model to be applied, and to ensure that the data are sufficiently characteristic of the system to be modelled to provide meaningful insights and comparisons to observations. It is therefore helpful to address these questions as early as possible in the modelling cycle and not postpone them until the end.

2.4.2 Conceptual model evaluation

There are hardly any specific testing strategies available to confirm conceptual model validity. Frequently, structural inconsistencies are only disclosed later, during model analysis. For example, for spatial processes like movement, visual model output of the implemented model can be decisive in spotting inconsistencies.

Especially for models with numerous entities or processes, the conceptual model becomes more difficult to evaluate. Under such circumstances, the option of evaluating several

alternative conceptual models should be considered (Beven 2006; Refsgaard *et al.* 2006; Troldborg *et al.* 2007). Later phases in model development may reveal major flaws in one or more of the alternatives, meaning that the underlying conceptual model has to be rejected. Alternative models may focus on alternative conceptual models with one or a few key processes, or behaviours, differing while keeping other parts of the model unchanged.

Consultation of expert knowledge in the form of a peer review process with or without a scoring system can be another helpful measure at this stage (Landry *et al.* 1983; Van der Sluijs *et al.* 2005) but requires the involvement of experts, potentially from various fields, which needs time and organisational preparation.

2.4.3 Implementation verification

Most of the testing at the stage of code verification involves techniques such as structured walkthroughs, correctness proofs, or an examination of program structure properties (Sargent 2005). Ferson (1996) suggests also using specifically designed software that detects common errors in computer code, such as dimensional and unit consistency, correlation matrices, constraints imposed by the biological domain (e.g., negative species abundance is not possible), or realisations of mathematical equations. Argent (2004) and Loizou et al. (2008) exemplify the automatic generation of model codes or equations. Scheller and Mladenoff (2006) and Scheller et al. (2010) demonstrate how current techniques from computer science can be used to manage and verify complex simulation models in ecology.

2.4.4 Model output verification

Several authors viewed it as crucial that performance criteria should be established early in the model development phase against which the model output can then be measured (Van Waveren *et al.* 1999; Refsgaard *et al.* 2005; Jakeman *et al.* 2006; Crout *et al.* 2008; US - EPA 2009). In any case the various criteria used for claiming that model is realistic enough should be communicated and justified. The choice of these criteria will be influenced by the overall quality of available data and the design of the conceptual model. Therefore, thorough performance of the previous steps of evaludation can reduce the effort required for model output verification.

2.4.5 Model analysis

Model analysis can be performed by a multitude of quantitative and qualitative methods. However, which particular approaches are acceptable may depend on the domain in which decisions are intended to be supported by the model, and on the model's purpose.

A method that is regularly performed to evaluate a model is a sensitivity analysis where the model's response to changes in model inputs is explored, i.e., computer program is

executed under different conditions to investigate how a model's response can be apportioned to changes in model inputs, i.e., parameter values and initial conditions (Saltelli *et al.* 2000). Sensitivity analysis is recommended as the principal evaluation tool for characterising the most and least important sources of uncertainty in environmental models. Local sensitivity analysis, where one parameter is varied a little at a time, is easy to perform but does not capture interaction between parameters and their processes and is restricted to linear effects. Global sensitivity analysis, where parameters are varied over their meaningful range and all possible parameter combinations are sampled, is usually only feasible for a small number of parameters due to run time limitations. In sensitivity experiments, one parameter is varied over its entire range, which can be combined with a second parameter in a contour plot. Sometimes, statistical models like ANOVA, GLMs, boosted regression trees or structured equation modelling can help to summarise the results of global sensitivity analyses.

In general, to understand a model, controlled simulation experiments are needed. The design of simulation experiments should follow the same principles as those of real experiments: keep all factors constant except one or two; explore simplified scenarios in which, for example, the environment is homogeneous and constant, or where some processes are de-activated; try different output metrics (also referred to as summary statistics, observations, or "currencies" (Grimm and Railsback, 2012)); etc. The overall approach is to try and understand simplified versions first and then gradually increase the number of possibly confounding factors. Sensitivity analyses and the testing of alternative model formulations are the most consistently applied methods for model analyses.

2.4.6 Model output corroboration

Similar approaches can be used as for model output verification. The only principal difference is that we here compare model output to new, independent data and patterns. Such data can sometimes be obtained from new experiments and field studies but more often will be taken from existing literature and expert knowledge. In the latter case, a first step is to identify patterns in the model, which can be considered independent, or secondary, predictions, as the model was not designed to reproduce these patterns.

2.5 Documentation

Documenting major elements of iterative model development and evaludation is crucial in communicating assumptions, justifications, and findings. Aber (1997) identified cryptic model descriptions as another source for mistrust in a model and Van Waveren et al. (1999) highlight in their Good Modelling Practice Handbook that all steps and actions taken ought to be described in a way understandable for the decision maker.

Schmolke et al. (2010a) proposed the TRACE (transparent and coherent ecological modelling) documentation scheme. The purpose of a TRACE document, which would

usually be provided as a supplement or appendix, is to provide additional evidence that the model has been carefully designed and thoroughly tested and analysed. The basic idea of TRACE was to introduce a common terminology and document structure, so that modellers and model users know exactly what to document and where to put and look for the different elements of evidence that a model is fit for its intended purpose. TRACE thus not only provides a common terminology and structure, but is also a checklist for model developers and users to make sure they addressed all important elements of model evaludation (see Grimm et al., 2014).

While a model is developed, the different steps and activities performed throughout the different stages of the modelling cycle should be documented in a modelling notebook, which corresponds to notebooks or journals kept in laboratories. If TRACE terminology is used for the entries in the notebook, it will be easy and efficient to extract the relevant information from the modelling notebook and assemble a TRACE document when the model is delivered.

TRACE has been tested in about 10 modelling projects (see Grimm et al., 2014). It turned out that TRACE, as originally described by Schmolke et al. (2010a) was not ready for being used. Grimm et al., (2014) present an update of TRACE and its rationale. The overall idea remains the same, but more specific guidelines for producing and reading TRACE documents were formulated. Most importantly, TRACE terminology and document structure was completely changed and now follows the terminology introduced here, including the six elements of evaludation. The focus of TRACE thus shifts from documentation, which is not necessarily linked to a specific purpose, to evaludation, which we here defined as "the entire process of establishing model quality and credibility throughout all stages of model development, analysis, and application". Consequently, a main purpose of TRACE documents is to report all elements of a model's evaludation. In addition, TRACE documents include a detailed problem formulation, a full model description, and a description and justification of the environmental scenarios explored with the model (for details, see Grimm et al., 2014).

2.6 Concluding remarks

Confusing terminology is one of the main obstacles to get a good understanding what model validation is, how it works, and what it can deliver. Attempts to clarify terminology were criticised by Hodges (2008): "There is a repeated call for ecological terminology to be standardised and for terms to be defined more concretely. These calls for the standardisation of definitions are based on faulty premises about the way language conveys meaning." (p. 35). We agree that terminological discussions can turn into hair-splitting exercises, but we hold, following the quote of Immanuel Kant that we chose as a motto for this article, that terminology discussions are also about genuine problems, not just words (see also Jax (2008)).

We therefore devised a standard set of terms related to validation that we derived from a literature review and from a consideration of the different elements of iterative model development (summary in Table 2-2). We believe that this set of terms and its relation to the modelling cycle can help to make model assessment more comprehensive and transparent. Our distinction of different evaludation steps offers a generic checklist, which makes it easier for modellers and model users to organize model evaluation and its communication.

We want to point out, however, that there can be reasons to perform one or more evaludation steps to only some limited degree. Reasons for this may include a lack of data, limited time and other resources, or ambiguities in the problem formulation. Our advice for such circumstances is to document and discuss possible limitations, their reasons, and how they could be overcome in the future. This enables model users to understand that limitations do not reflect oversights but limitations that the modeller could not overcome at a given time. It should also be noted that even despite partly performed evaludation steps, models can add important information to a decision-making process. Similarly, a fully performed evaludation does not guarantee that a model is good enough for application in a decision-making context. Thus, evaludation does not provide a yes/no criterion for whether a model can support decision-making. Rather, the different evaludation steps add to the "weight of evidence" (Weed 2005) that a model is fit for its purpose. For specific fields of application, it might be well possible to provide more respective guidance and requirements for the different evaludation steps, but such guidance can only be based on a joint activity of all stakeholders involved. We believe that the evaludation scheme that we presented here, and its documentation in TRACE documents, will facilitate such activities.

TRACE documents are a tool to put evaludation into practice and get it established, for both modellers and model users. Modellers will always benefit from keeping a modelling notebook, preferably on a daily basis. If they use TRACE terminology, which is based on the terminology introduced here, modellers can, at a later stage, easily assemble TRACE documents, which provide, in a structured and standardised way, various kinds of evidence that their model was well designed, thoroughly tested, and analysed. Modellers will thus directly profit from using and keeping a modelling notebook and from providing the kind of information that decision makers need to see to assess whether or not they can use model output as a basis for their decisions.

To conclude, we believe the suggested terminology and framework can ultimately contribute to establish an advanced culture of model development and evaluation, so that in the future better models are developed and actually used to support more environmental decisions in a productive and robust way.

Table 2-2: Summary of evaludation terminology.

Term	Definition
Evaludation	The entire process of assessing model quality and establishing model credibility throughout all stages of model development, analysis, and application.
Data evaluation	The assessment of the quality of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design overall model structure, whereby not only the measurement protocols need to be evaluated but conclusions drawn from the data should be challenged as well.
Conceptual model evaluation	The assessment of the simplifying assumptions underlying a model's design and forming its building blocks, including an assessment of whether the structure, essential theories, concepts, assumptions, and causal relationships are reasonable to form a logically consistent model.
Implementation verification	The assessment of (1) whether the computerized implementation the model is correct and free of programming errors and (2) whether the implemented model performs as indicated by the model description. The aim is to ensure that the modelling formalism is accurate.
Model output verification	The assessment of (1) how well model output matches observations and (2) to what degree calibration and effects of environmental drivers were involved in obtaining good fits of model output and data. The aim is to ensure that the individuals and populations represented in the model respond to habitat features and environmental conditions in a sufficiently similar way as their real counterparts.
Model analysis	The assessment of (1) how sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood. The aim is to understand the model and be able why which output is being produced to avoid drawing the wrong conclusions from model output.
Model output corroboration	The comparison of model predictions with independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterized, and verified. This step strengthens a model's credibility by proving that the model is capable of predicting/reproducing pattern and data that could not have influenced the model development.

3 TOWARDS BETTER MODELLING AND DECISION SUPPORT: DOCUMENTING MODEL DEVELOPMENT, TESTING, AND ANALYSIS USING TRACE

Volker Grimm, Jacqueline Augusiak, Andreas Focks, Béatrice M. Frank, Faten Gabsi, Alice S. A. Johnston, Chun Liu, Benjamin T. Martin, Mattia Meli, Viktoriia Radchuk, Pernille Thorbek, Steven F. Railsback

---

This chapter has been published in *Ecological Modelling*, 2014, **280**, 129-139.

Abstract

The potential of ecological models for supporting environmental decision-making is increasingly acknowledged. However, it often remains unclear whether a model is realistic and reliable enough. Good practice for developing and testing ecological models has not yet been established. Therefore, TRACE, a general framework for documenting a model's rationale, design, and testing was recently suggested. Originally TRACE was aimed at documenting good modelling practice. However, the word "documentation" does not convey TRACE's urgency. Therefore, we re-define TRACE as a tool for planning, performing, and documenting good modelling practice. TRACE documents should provide convincing evidence that a model was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose. TRACE documents link the science underlying a model to its application, thereby also linking modellers and model users, for example stakeholders, decision makers, and developers of policies. We report on first experiences in producing TRACE documents. We found that the original idea underlying TRACE was valid, but to make its use more coherent and efficient, an update of its structure and more specific guidance for its use are needed. The updated TRACE format follows the recently developed framework of model "evaludation": the entire process of establishing model quality and credibility throughout all stages of model development, analysis, and application. TRACE thus becomes a tool for planning, documenting, and assessing model evaludation, which includes understanding the rationale behind a model and its envisaged use. We introduce the new structure and revised terminology of TRACE and provide examples.

3.1 Introduction

Modelling is an iterative process. First model versions should deliberately be oversimplified to make the "modelling cycle" (Grimm & Railsback 2005; Figure 3-1a) start as soon as possible. Discrepancies between model output and observations then guide developing the next model versions. Thus, during iterative model development, many alternative submodels or even overall designs are tested, modified, improved, or discarded. As a result, models are usually a patchwork of elements that entered model development at different stages.

For example, in population models simple phenomenological submodels describing mortality due to senescence might have been introduced early on and never required intensive testing. Other submodels representing key behaviours, for instance home range dynamics (Liu *et al.* 2013), habitat selection (Railsback & Harvey 2002), or starvation (Martin *et al.* 2013), often require testing a large number of alternative versions, both in isolation and within the entire model.

When a model is finally published or made available to decision makers, most of the careful design and testing that went into the final model remains undocumented. However, without sufficient information about a model‘s design and testing it can be hard or even impossible to develop enough confidence to use it for supporting environmental decision making (Schmolke *et al.* 2010b). This situation is similar to a laboratory buying an expensive new analytical instrument: how do the lab‘s owner and clients know that the instrument works correctly and produces reliable results, and exactly how to calibrate and use it to produce credible data? They would require that documentation of the instrument‘s theoretical basis, its detailed design, and how it has been tested be available somewhere. This documentation might not be read routinely, but they are key components of quality assurance: lab instrument manufacturers know that customers expect full documentation and that flawed or incomplete documentation might make them cancel their purchase.

Unfortunately, in ecological and environmental modelling there is not yet a generally established culture of documenting the scope, design, and tests of our virtual laboratories, i.e., our models. Without such a culture, three bad things can happen: modellers develop models without employing basic mechanisms of quality assurance, leading to poor model designs; decision makers might not consider a model even though it is well-designed and tested; or, vice versa, they might use a model to support decisions although the model has major flaws (Pilkey & Pilkey-Jarvis 2007). Further, the modelling process itself becomes unnecessarily inefficient: analyses often must be repeated or revised because the original methods were not recorded; mistakes can be repeated; and unproductive approaches can be tried several times when the modeller does not document why they were unproductive the first time. Such a culture does however exist in other fields (e.g., in engineering and software development) that rely heavily on modelling and computation. There is in fact a vast literature on these topics (to get an idea of it, see the Wikipedia entries for topics such

as software testing, software documentation, and software specification; and Augusiak *et al.* 2014).

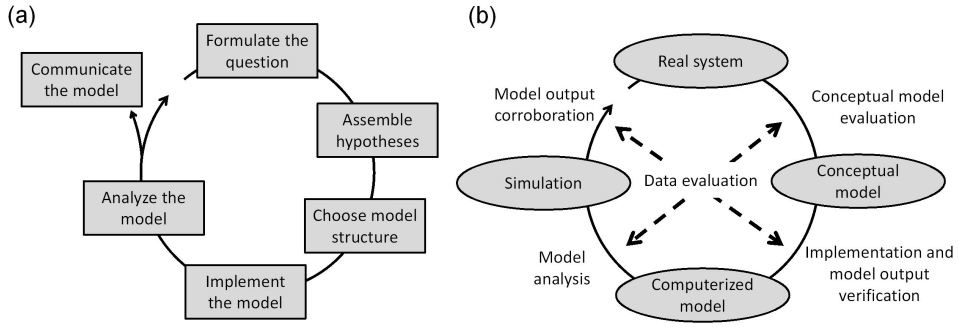


Figure 3-1: Correspondence of iterative model development (the "Modelling Cycle", Grimm & Railsback 2005) (a), and the elements of model evaludation as proposed by Augusiak *et al.* (2014) (b).

In ecological modelling, however, we do not yet have a culture of documenting model development, testing, and analysis, because clients of models usually do not know what kind of documentation they should require, hence model developers do not know what clients expect. Here, clients include other scientists trying to learn from a published model and decision makers trying to use a model or its output to make better decisions. To establish a culture of comprehensive modelling documentation we thus need to establish clear expectations: model clients need to have clear expectations and model developers need to be aware of clients‘ expectations.

Establishing such expectations has worked before, often via standardization. For example, the structure of scientific articles – Introduction, Methods, Results, Discussion – reflects expectations both of the readers and writers. Or, for individual- or agent-based models (IBMs), modellers increasingly are using a standard format, the ODD protocol (Overview, Design concepts, Details; Grimm *et al.* 2006, 2010), for describing the model, thereby increasingly making readers expect that IBMs are described in this format, with certain kinds of information at certain places in the model description.

Thus, to help establish a culture of comprehensive modelling documentation, Schmolke *et al.* (2010b) suggested a standard format and terminology referred to as "TRACE" (TRAnsparent and Comprehensive Ecological modelling documentation). The acronym also refers to the process of "tracing" model development and testing by "going backward over the evidence step by step" ("trace" in Merriam Webster online dictionary). Schmolke *et al.* (2010b) introduced the overall framework of TRACE but also made clear that subsequent work will be needed to establish TRACE and hence, the above-mentioned

culture of modelling documentation: "The TRACE documentation framework can only become established as a standard if it is applied and refined by numerous projects." (p. 484).

Here, we report a first such refinement, based on using and discussing TRACE in the EU-funded project CREAM (Grimm *et al.* 2009) and three further modelling projects. In CREAM, ecological and organism-level models were developed to assess the effect of chemicals, in particular pesticides, on populations and individual organisms. Ultimately, the hope is that such models are used to make regulatory risk assessment of chemicals more ecologically relevant (Forbes & Calow 2012). Nevertheless, although ecological risk assessment is a specific field of environmental decision making, the lessons learned about TRACE and how to actually use it are generic.

We first briefly summarize the original idea and structure of a TRACE document. Then we present the most important questions that came up in using TRACE, and from discussions at conferences, feedback from colleagues, and one publication (Wang & Luttik 2012). We provide a practical answer to each question and then present, as a result, a revised TRACE format and brief guidance for writing and reading TRACE documents.

The basic idea of TRACE remains the same but we completely revise the structure and terminology of TRACE documents to clarify the purposes of TRACE documents: to help model clients understand the model and assess the quality of the model and hence the reliability of the results; i.e., to provide comprehensive model evaluation and validation. For this, we adopted the terminology proposed by Augusiak *et al.* (2014), in particular merging the terms "evaluation" and "validation" into the new artificial term "evaludation", defined as: "the entire process of establishing model quality and credibility throughout all stages of model development, analysis, and application. TRACE thus becomes a tool for planning, documenting, and assessing model evaludation, which includes understanding the rationale behind a model and its envisaged use. TRACE is aimed at documenting model design and testing. Model application (i.e. the simulations carried out to answer specific environmental decision-making questions) will also need to be carefully documented. However, this is outside the scope of the TRACE documentation.

3.2 TRACE: the basic idea

The two basic tasks of using TRACE are: (i) keeping a modelling notebook in which you briefly document, preferably daily, what you did regarding model design, testing, and analysis, and what you learned from it; and (ii) using, in this modelling notebook, the standardized terminology used in TRACE documents.

The two basic ideas underlying TRACE are (i) TRACE and its terminology cover all elements of iterative model development, i.e., the modelling cycle; and (ii) by using a standardized terminology and document structure, readers and model users know exactly where in the TRACE document they can expect finding what kind of information. TRACE

also lets clients quickly check whether all important steps of model development were documented and how carefully the model was designed, parameterized, tested, and analysed.

To illustrate the potential benefits of TRACE documentation, imagine you developed an individual-based model of a small mammal, which includes home range behaviour, similar to that of (Liu *et al.* 2013). Due to the lack of appropriate data, you decided to use a phenomenological, not mechanistic, approach so that home ranges are more or less imposed rather than emerging from individuals making decisions in a heterogeneous landscape. You tested simple and complex ways of imposing home ranges and how they are related to habitat structure and then selected a rather simple approach, which takes into account vegetation cover but not food resources.

If this model were just factually described in a publication, reviewers, readers, and potential users might consider the design of the home range sub-model ad hoc, unrealistic, and not good enough to make the entire population model reliable. By contrast, if each time you worked on the home range model you kept notes in the modelling notebook labelled, e.g., "home range model/purpose", "home range model/data", or "home range model/sensitivity analysis/alternatives", you can easily extract relevant information from your notebook and compile it in a TRACE document. This document then shows that for the purpose of the overall model, a mechanistic home range model was not essential, that no data existed for relating home ranges to resource availability in time and space, and that alternative simpler models created artefacts and more complex models did not improve usefulness and realism of the entire model. Reviewers, readers, and users of your model would understand that model design was not ad hoc but that the chosen design reflects the overall model purpose, data limitations, and careful selection of submodel structure.

The structure of TRACE documents proposed by Schmolke *et al.* (2010b; Table 3-1) reflects all elements of model development, testing, and analysis. In their review of literature on good modelling practice, Schmolke *et al.* (2010b) found that virtually all authors agreed that quality assurance of models should address all elements of modelling, not only verification and validation (see also Augusiak *et al.,* 2014). For the terms used for the different elements of TRACE documents, Schmolke *et al.* (2010b) had to make choices, as terminologies vary considerably within and across disciplines.

Schmolke *et al.* (2010b) concluded from their review that most elements of good modelling practice have long been known but never got established. The main challenge thus is not so much to define good modelling practice but to get it established and widely used. Since producing TRACE documents requires additional effort, it is unlikely that TRACE will be used if it does not provide direct benefits to the modeller. Schmolke *et al.* (2010b) therefore suggested linking TRACE documentation to keeping modelling notebooks. Such notebooks have direct benefits to the modeller because they help organize and document the complex task of developing, testing and analysing a model. Extracting a TRACE document from a

notebook requires much less effort than producing it from scratch, after model development has ended.

Table 3-1: New structure, terminology, and contents of TRACE documents.

TRACE element		This TRACE element provides supporting information on:
1	**Problem formulation**	The decision-making context in which the model will be used; the types of model clients or stakeholders addressed; a precise specification of the question(s) that should be answered with the model, including a specification of necessary model outputs; and a statement of the domain of applicability of the model, including the extent of acceptable extrapolations.
2	**Model description**	The model. Provide a detailed written model description. For individual/agent-based and other simulation models, the ODD protocol is recommended as standard format. For complex submodels include concise explanations of the underlying rationale. Model users should learn what the model is, how it works, and what guided its design.
3	**Data evaluation**	The quality and sources of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design the overall model structure. This critical evaluation will allow model users to assess the scope and the uncertainty of the data and knowledge on which the model is based.
4	**Conceptual model evaluation**	The simplifying assumptions underlying a model‘s design, both with regard to empirical knowledge and general, basic principles. This critical evaluation allows model users to understand that model design was not ad hoc but based on carefully scrutinized considerations.
5	**Implementation verification**	(1) Whether the computer code implementing the model has been thoroughly tested for programming errors, (2) whether the implemented model performs as indicated by the model description, and (3) how the software has been designed and documented to provide necessary usability tools (interfaces, automation of experiments, etc.) and to facilitate future installation, modification, and maintenance.
6	**Model output verification**	(1) How well model output matches observations and (2) how much calibration and effects of environmental drivers were involved in obtaining good fits of model output and data.
7	**Model analysis**	(1) How sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood.
8	**Model output corroboration**	How model predictions compare to independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterized, and verified. By documenting model output corroboration, model users learn about evidence which, in addition to model output verification, indicates that the model is structurally realistic so that its predictions can be trusted to some degree.

3.3 How to actually use TRACE? Questions and answers

How well did the ideas underlying TRACE work? In contrast to the ODD protocol for describing individual-based and agent-based models (Grimm *et al.* 2006, 2010), which was immediately used by modellers in ecology and other fields and is becoming a widely accepted standard, no independent use of TRACE has yet been published, although it is being used for models for decision support (P. Thorbek, *pers. communication*). This suggests that TRACE as presented by Schmolke *et al.* (2010b) was not yet ready to use. The challenges in actually using TRACE became apparent when trying implementing it in 10 modelling projects. To summarize these challenges, we list and address the most frequently asked questions about TRACE.

How much detail should be in TRACE documents?

The modelling notebook should preferably be updated every day while working on a model. For complex models, which can take a long time to develop and test, this means that even if the entries in the notebook are concise, hundreds of pages of text could easily accumulate, in addition to sketches, figures, tables, links to program versions and data and model output files, etc. Of course, only a small proportion of all this should go into the TRACE documents.

The purpose of a TRACE document is to provide convincing evidence that a model was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose. Readers will first want to see an overview and only then decide whether and where to go into more detail. Thus, to allow for hierarchical reading and to keep TRACE documents concise and readable, it is critical to start the entire document and each of its sections with an executive summary.

For example, to document software testing the executive summary should describe the kinds of approaches and tools used to detect programming errors. This might include debug code that checks, while the program is running, that variables stay within meaningful ranges. The corresponding entry in the TRACE document might then read like: "The program includes 23 elements of debug code, which stop program execution and give an error message when a variable assumes values outside its meaningful range." Then a table might list all variables checked in this way plus their ranges. This information adds evidence that the software was thoroughly tested.

In general, summaries should always come first and details later; if details provide long and repetitive information, this information should be put into separate files, or moved to an appendix at the end of the TRACE document. For example, if a key submodel has been implemented independently in a spreadsheet, and its outputs were compared to output from the model‘s original implementation, the TRACE document would provide hyperlinks to the spreadsheet.

How are TRACE, the modelling notebook, and the model description, for example the ODD for individual-based models, related to each other?

The modelling notebook corresponds to lab journals or notebooks in laboratory research. It should be kept for its own, direct benefits, which include documentation of test procedures so tests can be replicated later, and supporting the careful planning, execution, and interpretation of simulation experiments. A notebook does not just document experiment design and factual results: "The act of writing in the notebook causes the scientist to stop and think about what is being done in the laboratory. It is in this way an essential part of "doing good science"." (Kanare 1985, p.1).

The link between the modelling notebook and TRACE documents is established by using TRACE terminology to label the entries in the notebook. Notebooks do not necessarily have the structure of TRACE documents but most often follow a chronological order, with each entry identified by its date and a label following TRACE terminology.

TRACE documents should include a full model description, which, if the model is individual-based or agent-based (or any other kind of simulation model), preferably should use the standard format ODD (Grimm *et al.* 2006, 2010); ODD can be used for any kind of simulation model, just by leaving out those "design concepts" that are specific to individual-based models (e.g., Meli *et al.* 2014).

Is there an overlap between model description (e.g., ODD) and the TRACE document?

There can be overlap between a TRACE document and one part of the ODD protocol, its description of the design rationale for model components and submodels. The ODD protocol includes design rationale because when it was developed TRACE did not yet exist, so that the justification of the model's biological background, structure, simplifying assumptions, and parameterization had to be in ODD. Now, if both an ODD model description and a TRACE document are provided, the model's underlying rationale would be described in both. This redundancy is unavoidable when an ODD must be complete by itself, for example in a journal publication. Moreover, due to space limitations, the explanations of the model's rationale will still be quite short in the ODD description, whereas in the TRACE document they can be more detailed and discuss alternative model designs that were tested and then discarded.

Who is going to read TRACE documents tens or hundreds of pages long?

TRACE documents are not designed to be read from cover to cover, but to provide additional information convincing users that all tasks of model development have been performed according to general good modelling practice and providing all elements of a

model‘s evaludation (for details on evaludation, see below). A TRACE document can be thought of as a reference manual where users can find particular details when necessary. Therefore, as explained above, a hierarchical structure in the entire document and each section and subsection is mandatory, with each unit starting with an executive summary.

Is TRACE a technical document, written by modellers for modellers (see Wang & Luttik 2012)?

Definitely not. TRACE is supposed to cover all aspects of model development, testing, and analysis, not only the technical ones. For example, an overview of the biological literature and reasoning that has been used to design the model and get parameter values is an integral and important part of each TRACE document.

There will certainly be elements that are more technical, for example documentation of software testing. However, TRACE‘s hierarchical structure requires that such elements are also first summarized in a non-technical way suitable for all users.

Do the “Parameterization”, “Calibration” and “Sensitivity analysis” elements overlap, making it difficult to decide where to put (or expect) what information?

These three elements certainly are related, but are also different enough to distinguish as separate elements within the modelling cycle. In the updated TRACE format (see below), we give them clearer definitions. We distinguish between "direct parameterization", obtaining parameter values directly from the literature or experts, and "inverse parameterization", obtaining parameter values inversely by calibrating the model to observations. Regarding sensitivity analysis, we now distinguish between "local sensitivity analysis", which is based on one parameter at a time, and "global sensitivity analysis", in which several or all parameters are varied over their whole ranges.

What about models developed before TRACE existed?

TRACE documents can of course be assembled even if no modelling notebook was kept. If no notes were made during model development, the corresponding analyses, reviews, and tests must then be performed and documented in retrospect. This effort can be substantial for complex models; it is also our main argument for keeping a modelling notebook. You should do these analyses anyway, so why not keep notes so that no analyses have to be repeated while putting the TRACE document together?

Many reviewers and readers never look at the Supplementary Material, so why should I produce the TRACE document if I don't get any credit for it, i.e. higher chances of getting the model published or used?

Just stating in an article or report that the Appendix or Supplement includes a TRACE document might not be sufficient to get credit for the work that went into producing it. We therefore suggest that, whenever a TRACE document has been produced, the main text or a printed appendix includes a "TRACE table" concisely summarizing the TRACE document (see example in Table 3-2). Providing such summaries will also help establishing the culture of model evaludation and its documentation: the more publications or reports submitted to decision makers include a TRACE table, the more often model clients will use it as a checklist for scrutinizing a model's evaludation. TRACE tables therefore could be critical to establishing the readers' expectation that we mentioned in the Introduction. Of course, once TRACE is more widely used, the credit for having provided a TRACE document will be immediate, as it will increase chances of getting published and used.

A similar development took place in ecological modelling over the last several decades: 20 years ago, few publications included a sensitivity analysis, whereas in 2009-2010, 24% of all articles published in Ecological Modelling included some sort of systematic sensitivity analysis (Thiele et al., *unpubl. manuscript*). Thus, nowadays, most reviewers expect a sensitivity analysis; modellers are aware of this expectation so they just include the analyses as a normal part of publication.

Do I need a full TRACE document for every model application?

Different applications of the same model can refer to the same documentation of model development, but model analysis, which includes the description and justification of the scenarios explored, needs to be updated. A similar situation often arises with using ODD for model description: what if only one or a few elements of a model were changed? A technical solution for both ODD and TRACE is to re-use the original elements and track changes by crossing out deleted text and emphasizing new text by colour or bold fonts.

3.4 TRACE: a first revision and short guide

Schmolke *et al.* (2010b) argue that TRACE documentation is critical in making a model fit for supporting environmental decision-making. However, just the word "documentation" does not convey this urgency. Therefore, we here suggest re-defining TRACE as a tool for planning, performing, documenting, and assessing a model‘s "evaludation" (Augusiak *et al.* 2014).

3.4.1 Evaludation

Augusiak *et al.* (2014) review the terminology and ideas around the terms "validation", "verification", and "evaluation", which all represent important elements of assessing whether a model is good enough for its intended purpose. The two main conclusions of their review are that (i) the term "validation" is a catch-all term that has been given so many different, partly contradicting, meanings that it cannot be used for any practical purpose; (ii) comparing model predictions to independent, new data is neither sufficient nor necessary to make a model useful for, e.g., decision support. Rather, all steps of iterative model development have to fulfil certain quality criteria: a model can reproduce existing data or make even correct new predictions, while still based on biased data, unreasonable assumptions, faulty software, and excessive parameter and submodel tweaking. Quality assurance of models should therefore include all elements of iterative modelling development (Figure 3-1a). Hence, Augusiak *et al.* (2014) suggest the new term "evaludation" for this kind of comprehensive quality assessment.

Evaludation consists of six elements, which largely correspond to the elements of the modelling cycle (Figure 3-1b). These elements are (i) "data evaluation", assessing the quality of numerical and qualitative data used for model development and testing; (ii) "conceptual model evaluation", scrutinizing the simplifying assumptions underlying a model‘s design; (iii) "implementation verification", checking the model‘s implementation in equations and software; (iv) "model output verification", comparing model output to the data and patterns that guided model design and calibration; (v) "model analysis", examining the model‘s sensitivity to changes in parameters and formulation to understand the model‘s main behaviours and describing and justifying simulation experiments; and (vi) "model output corroboration", comparing model output to data and patterns that were not used for model development and parameterization.

3.4.2 A new terminology for TRACE

Since both TRACE and evaludation relate to the iterative steps of the modelling cycle, their elements can be easily linked (Figure 3-1). Therefore we propose replacing the original TRACE terminology with the six elements of model evaludation, plus one element for problem formulation and one for model description (Table 3-1). By doing so, we also re-define the scope of TRACE from being a "standard format for documenting models and

their analyses" (Schmolke *et al.* 2010b) to being a tool for planning, performing, and documenting model evaludation. Accordingly, the "E" in the acronym TRACE changes from "Ecological modelling" to "Evaludation". TRACE thus now stands for "TRAnsparent and Comprehensive model Evaludation". The tasks documented in the original version of TRACE, including the documentation of scenarios tested with the model, remain largely the same, but have partly been renamed and re-grouped (Figure 3-2). One original element of TRACE is no longer included: "recommendations", because we believe that these are the main results of models for environmental decision making, so they should be presented in the main document.

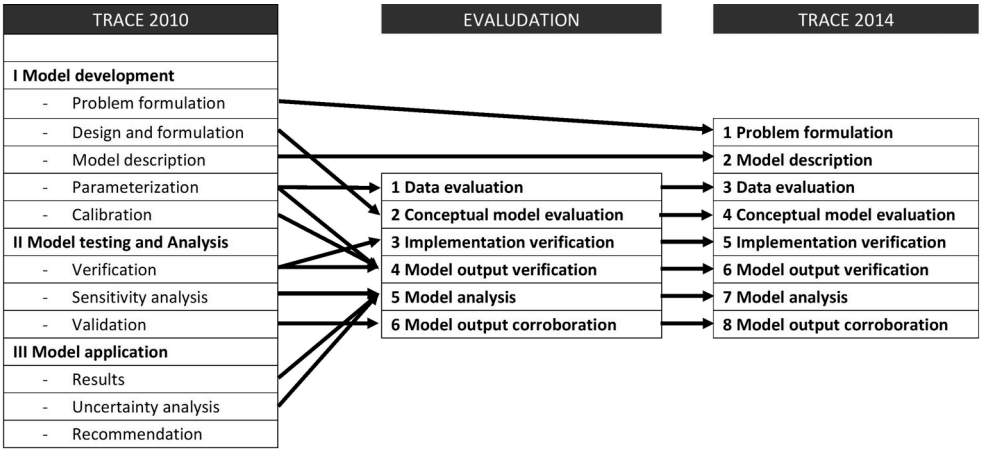


Figure 3-2: Structure and terminology of the original TRACE format (Schmolke *et al.*, 2010; TRACE 2010); model evaludation (Augusiak *et al.*, 2014; Chapter 2); and the new, updated TRACE format presented in this article (TRACE 2014).

3.4.3 An updated guide for using TRACE

A template for producing TRACE documents following the new structure and terminology defined in the previous section is provided in the Supplementary Material online (http://www.sciencedirect.com/science/article/pii/S0304380012005479). The questions and checklists at the end of each of the eight elements should be helpful for compiling coherent, comprehensive, and concise TRACE documents. Here we give only a short characterization of the eight elements of TRACE, and their subsections. For more detailed discussion of the six evaludation elements, see Augusiak *et al.* (2014).

Each element should start with an executive summary, which can be a short narrative, a bullet-point list, or a table of contents of this element. The summary should include references to corresponding page numbers and hyperlinks for convenient navigation in the electronic version of the TRACE document.

1. Problem formulation.

This element is largely unchanged from Schmolke *et al.* (2010b). It should describe: the decision-making context in which the model will be used; the type of model clients, or stakeholders, addressed; the precise question(s) that should be answered with the model and the necessary model outputs; and the domain of applicability of the model, including the extent of acceptable extrapolations. For regulatory models intended for comparison to field observations, what statistical measures of field data will the model be compared to? The assessment of the availability of knowledge and data included in this element by Schmolke *et al.* (2010b) is now covered by the third TRACE element, "Data evaluation".

2. Model description.

TRACE documents should include a complete model description that allows users to easily and fully understand a model and, in principle, replicate it independently. We recommend using the ODD protocol (see above, section 3.3). The ODD description should include a table with all model parameters, their meaning, units, reference values, range, and data source (if taken from publications, page numbers should be included). Parameters that were determined inversely via calibration should be clearly identified. Direct parameterization is documented in detail in the third TRACE element, "Data evaluation", and inverse parameterization is documented in the eighth element, "Model output verification".

Any verbal model description is likely to include ambiguities that prevent full replication. Therefore, the model‘s computer code should be provided as well, usually in a separate file. If it is not possible to provide the code, e.g., if it is proprietary, the executable program and all data and script files needed to run the model should be made available. Any other information needed to run the software (e.g., platform version, operating system limitations) should be provided. This material could be part of the Supplementary Material or made available on permanent repositories such as the CoMSES Computational Model Library maintained by the OpenABM consortium (http://www.openabm.org/models).

3. Data evaluation.

Augusiak *et al.* (2014) define "data evaluation" as the "critical assessment of the quality of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design the overall model structure". Data here refers both to quantitative data, taken from experiments, monitoring, or publications, and qualitative data, which often corresponds to expert knowledge. Data also include patterns in time, space, and organization, which are characteristic of the system to be represented ("pattern-oriented modelling"; Grimm *et al.* 2005; Grimm & Railsback 2012).

Concise text plus tables should summarize what data and knowledge went into the design and parameterization of the model, including references, data sources, and information about where and when those data were collected, and by whom. If possible, the reliability of the data used should be discussed, as data quality and ecological significance might be limited by measurement errors, inappropriate experimental design (e.g., number of replicates), and, in particular, the heterogeneity and variability inherent to environmental systems (Gass 1983; Wang & Luttik 2012). Likewise, expert knowledge and the detection of patterns are prone to bias and therefore must be treated with particular caution. The document should indicate which parameter values were used directly without calibration and which were determined inversely; the methods used for inverse parameterisation will be described in the TRACE element, "Model output verification".

The data description and evaluation allows model users to (1) see whether a model was mainly built on its authors‘ own data and knowledge, or on that of a certain expert or group of experts, or on a systematic evaluation of the literature, and (2) assess how uncertain the data are.

4. Conceptual model evaluation.

This element is defined by Augusiak *et al.* (2014) as "the critical assessment of the simplifying assumptions underlying a model‘s design". The design of any mathematical or simulation model is based on a conceptual model which reflects our preliminary understanding and perception of the system to be represented in the model. For example, we may focus on nutrients and energy, species composition, or individual organisms. In this TRACE element, the underlying conceptual model should be described and its choice explained and justified. The evaluation applies to the overall model structure and sometimes to submodels, for example of metabolism, competition among individuals, movement, or the physical environment.

In detail, this evaluation lists and explains the most important conceptual design decisions: spatial and temporal scales, selection of entities and processes, representation of stochasticity and heterogeneity, consideration of local versus global interactions, environmental drivers, etc. Moreover, conceptual models are often determined by certain theories, concepts, or, in particular, earlier models. Modellers should explain why they chose these elements and briefly contrast them, if applicable, to alternative conceptual designs that would have led to other model structures.

Explaining and justifying conceptual models allows model users to understand that model design was not ad hoc but based on carefully scrutinized considerations. It makes users also aware that each model is only one of many possible ways to represent a certain system with regard to a certain question. Blind trust in a model can thereby be prevented, but so can blind distrust: even crude simplifying assumptions can be trusted if they are justified well.

5. Implementation verification.

This term is defined by Augusiak *et al.* (2014) as "the critical assessment of (1) whether the computer code for implementing the model has been thoroughly tested for programming errors and (2) whether the implemented model performs as indicated by the model description". For instance, implementation verification might be conducted by peer-reviewing the code, i.e., other scientists thoroughly comparing it with the written formulation of the model, or by independently implementing submodels. This TRACE element provides evidence that the model software has been thoroughly tested and accurately implements the model description.

A second component of implementation verification is documenting how the model‘s software has been designed to make it usable for the model‘s purposes. In addition to accurately implementing a model, its software often must also provide the graphical interfaces necessary to understand and test the model‘s behaviour (e.g., to see the behaviour of individuals in individual-based models), automate simulation experiments, be designed and documented to make modifications and maintenance easy, and be operable by clients.

6. Model output verification.

Augusiak *et al.* (2014) define this element as "the critical assessment of (1) how well model output matches observations and (2) how much calibration and effects of environmental drivers were involved in obtaining good fits of model output and data". In developing any model, we try to make it reproduce some features or patterns of the real system before claiming that it is a good enough representation. In this TRACE element, we list the features we used plus the quantitative criteria for deciding whether a certain observation was matched by the model. Example features for population models include persistence, mean and standard deviation of population size, and metrics of size, age, or spatial distributions. The more observed features or patterns a model can reproduce simultaneously, the higher the chance that it has captured the internal organization of the real system sufficiently well ("pattern-oriented modelling"; Grimm *et al.* 2005; Grimm & Railsback 2012).

Output verification involves what often is referred to as "face validation" and more formal tests. Face validation can be defined as: "all methods that rely on natural human intelligence" (Klügl 2008, p. 39). Examples listed by Klügl (2008) include: "structured walk-throughs, expert assessments of descriptions, animations of results". Klügl (2008) accordingly concludes that face validity shows that a model‘s processes and outcomes are reasonable and plausible within its theoretical basis and the knowledge of system experts or stakeholders. It should be noted, however, that system experts and stakeholders may disagree on the type of data and knowledge they have. Therefore more formal tests are required that are based on multiple quantitative criteria for a model matching data (e.g., Railsback & Grimm 2012, Chapter 20.4.2).

Evaluation of output verification needs to consider such concerns as over-fitting and extrapolation. The higher the proportion of calibrated, guesstimated, or uncertain parameters (see TRACE element "Model analysis" below), the higher the risk that the model seems to work correctly (e.g., because it fits calibration data well) but for the wrong reasons, i.e., has not captured the mechanisms of the real system. Moreover, it is important to distinguish between system-level parameters and those related to lower level processes. A population model, for example, may be based on empirically determined demographic rates, but this restricts the scope of the model to environmental conditions under which those rates were determined. In contrast, if submodels, for example foraging or habitat selection, are parameterized for a wider range of environmental conditions, population-level phenomena are no longer imposed but emerge and the population model can be expected to predict responses to new conditions more reliably (Grimm & Railsback 2005; Railsback & Grimm 2012; Grimm & Martin 2013).

Finally, a good match of model output to data can sometimes simply reflect the overarching influence of environmental drivers. For example, if the egg-laying rate of a honeybee queen follows uni-modal seasonal dynamics, colony size will vary accordingly and thus look realistic, but this does not indicate that all other processes included in a honeybee colony model have been captured realistically enough (Becher *et al.* 2013). Thus, example model runs should be presented along with time series of important environmental drivers.

Model developers naturally often claim that their models are realistic enough for their purpose, but in this TRACE element they should summarize why they believe so, with supporting evidence. This information enables users to scrutinize the modeller‘s claim and to critically assess how well model output matches observations, the degree to which the match results from calibration and environmental drivers, and how much the model‘s reliability is limited by use of empirical parameters that reflect only a narrow range of conditions.

7. Model analysis.

This element is defined by Augusiak *et al.* (2014) as "the assessment of (1) how sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood". The purpose of the element is to prevent blind trust in the model output by asking "How did this output emerge?", and to challenge the model, which might look impressive, by asking "Does verification still look good if I change one or more parameters a bit?".

Thus, foremost here we document how we made sure that we understood a model‘s main mechanisms. For example, if recovery after disturbance is strongly affected by a certain parameter and, thus, the processes the parameter represents, we should be able to explain why this parameter was so important. We can learn much about a model by performing controlled simulation experiments: keeping most parameters constant and varying one or a

few over a wider range, and exploring the effect on one or more output variables. Simulation experiments should also include simplified model versions, in which the environment is made more homogenous and constant, system size is reduced, and certain processes are deactivated. Initial conditions and input data are other model components to which sensitivity should be analysed.

TRACE should not include details on all these experiments, but give an overview of what kind of experiments were performed and present results from experiments that significantly increased understanding but could not be included in the paper or report.

Local sensitivity analysis is important for developing a first understanding of a model by evaluating how sensitive output is to small changes in one parameter at a time. The analysis can produce conclusions about model uncertainty: if the parameters to which the model is most sensitive are the most uncertain ones, the entire model will be quite uncertain. Moreover, such parameters indicate which processes are most important for certain model outputs.

By varying more than one parameter at a time, local sensitivity analysis gradually becomes global analysis, which captures interactions among parameters by examining the entire parameter space, not only the local neighbourhood of a default parameter set. Run time, complexity, and stochasticity often limit global sensitivity analysis, but it should be performed for at least a subset of parameters. One way to summarize such sensitivity analyses is regression modelling, which quantifies the relative influence of parameters on model output. Uncertainty analysis can augment sensitivity analysis by demonstrating how uncertainty in model parameters translates into uncertainty in model output.

Parameters often represent entire processes that the modeller chose not to represent explicitly. Submodels represent processes that are represented explicitly in more detail; therefore, sensitivity analysis should also be applied to important submodels by contrasting alternative submodels. For example, a submodel describing movement might be based on complex decision making, but contrasting this submodel with simpler, or even more complex, alternatives can provide insights into how important or useful it was to choose this very model design. This sensitivity analysis of submodels corresponds to what Railsback & Grimm (2012) refer to as "pattern-oriented theory development": which submodel best causes the full model to reproduce a set of observed patterns?

Model users learn from this TRACE element how the model works, i.e., which processes and process interactions are important and explain major behaviours of the model system. Moreover, users learn how robust model results are to uncertainties in model parameters and submodel formulation.

8. Model output corroboration.

This term is defined by Augusiak *et al.* (2014) as the "comparison of model predictions with independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterized, and verified". Most scientists, in particular non-modellers, require this analysis, calling it "validation": for a model to be trusted it should make predictions that are subsequently confirmed in empirical experiments. Indeed, we consider this the "gold standard" for demonstrating that a model has captured the internal organization of a system sufficiently well. Corroboration is discussed in more depth by Augusiak *et al.* (2014).

Model output verification always includes "tweaking", i.e., we try to make a model reproduce certain observations by tuning parameters, environmental settings, and submodel formulation. Such adjustments are often necessary to compensate for processes not included in a model (due to insufficient information or to keep the model simple) but were important in the real system when the verification data were collected. Making a model simultaneously reproduce multiple observed patterns reduces the risk that the model is completely unrealistic, but does not eliminate this risk. Only when a model predicts phenomena that we even did not think about during model development and testing do we have the strongest indicator of its structural realism, because no tweaking could have been involved.

However, achieving this standard is rarely possible with ecological systems because the empirical experiments are infeasible: we often build models to address questions such as response to climate change exactly because empirical experiments are impossible. Instead, we can directly test independent predictions of submodels. At the system level, we can identify characteristic patterns in model output that are robust and seem characteristic. Then, we can consult the literature or experts to find out how accurate these independent predictions are.

Documenting model output corroboration provides model users evidence, in addition to model output verification, indicating the extent to which the model is structurally realistic so that its predictions can be trusted. The model‘s purpose should be a primary consideration in determining what model results need corroboration and how quantitatively and closely model results need to reproduce observations. If no corroboration was possible, the modeller should discuss here why, and why and to what degree the model still can be trusted. A classic example of model output verification that could be trusted was the structure of the DNA (a conceptual, not numerical, model), which Watson and Crick identified as a double helix because this structure was compatible with several patterns observed in DNA and its elements (Watson 1968). This verification was convincing enough, even without independent predictions, which were made and tested only later.

3.5 Examples

Three example TRACE documents are available online as Supplementary Information (http://www.sciencedirect.com/science/article/pii/S0304380012005479). They were not produced from scratch, but from existing TRACE documents produced according to the original TRACE format described in Schmolke et al. (2010b); those documents are supplementary material to Meli *et al.* (2013), Johnston *et al.* (2014) and Focks *et al.* (2014). Table 3-2 summarizes the three TRACE documents.

3.6 Discussion and Conclusion

Our experience producing TRACE documents following Schmolke *et al.* (2010b) led to a revised terminology, structure, and rationale for TRACE. The most important new feature is the link to the framework of model evaludation (Augusiak *et al.* 2014).

We hope that the new format is easier to use than the original and that the resulting documents are more efficient to use and understand by model clients, so that clients can better assess whether or not a model is realistic and robust enough to let it influence decisions affecting the real world. The three example TRACE documents we provide online follow the new format and terminology. However, they were compiled mostly from documents that followed the original TRACE format and terminology; TRACE documents following the new format and rationale from the start should be even more comprehensive and clear. Moreover, modelling notebooks following the new format should directly lead to more thorough model development, testing, and analysis because TRACE now provides a detailed checklist of all elements of modelling that have an influence on a model‘s credibility and usefulness.

In Table 3-2 we compiled the summaries of each of the eight TRACE elements. Similar tables might be a good way to summarize all the work that went into making a model fit for its purpose in the main text, or its printed appendix. However, Table 3-2 as well as TRACE in general do not provide criteria for when, for example, model output verification is good enough. TRACE by itself thus does not constitute good modelling practice. Nevertheless, the development of TRACE is the first step toward developing guidance and criteria for good modelling practice (Schmolke *et al.* 2010b). It might be possible to provide more detailed guidance for at least some TRACE elements, for example providing a checklist for implementation verification, or calculating an index that quantifies the proportion of calibrated versus uncalibrated parameters. To try this, though, a critical number of TRACE documents is needed; for the ODD protocol, an update and more specific guidance became possible after the protocol had been used about 50 times (Grimm *et al.* 2010).

Table 3-2: Overview and comparison of the three example TRACE documents provided in the Supplementary Material (online: http://www.sciencedirect.com/science/article/pii/S0304380012005479). The example documents' summaries are presented for each TRACE element.

TRACE element	Focks et al. (2014): Integrating chemical fate and population-level effect models for pesticides on the landscape scale: new options for risk assessment.	Johnston et al. (2014): An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides.	Meli et al. (2013): Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: an individual-based model of springtails.
1. Problem formulation	The MASTEP-regional model will be used by risk managers and scientists who are interested in effects of pesticides at the landscape scale. The model provides the possibility to assess the effects of multiple applications of a pesticide in space and time, as being typical for agricultural landscapes, on the population dynamics of aquatic species. The approach does not yet provide endpoints that are immediately operational for environmental risk assessment. Further research about what new indicators could look like is recommended. Extrapolations to other species and landscapes are intended and supported by the model.	The acquisition and expenditure of energy to life cycle processes depends on a combination of environment- and organism-specific conditions. In addition, exposure of individuals to chemical stress can alter a population's dynamics via physiological pathways. To investigate the sublethal effects of pesticides we develop and evaluate an energy budget agent-based model of the earthworm *Eisenia fetida*. *E. fetida* is used as a model species here due to its recommended use in lower tier toxicity tests, and therefore ample quantity of literature data available for model development at the individual level.	[1]Suitable habitat for soil organisms may be scarce, thus leading to locally high population densities, because soil, being more static than water or air, is heterogeneous: physical conditions often vary widely on a scale of a few centimetres. Moreover, toxic chemicals are likely to be unevenly distributed in the soil as well. The spatially explicit individual-based model presented in Meli et al. (2013) is developed to explore the consequences of these heterogeneities for the population dynamics of soil invertebrates, in particular the collembolan *Folsomia candida*. *F. candida* is a common arthropod that occurs in soils worldwide and is used as a standard test organism for estimating the effects of pesticides on non-target soil arthropods.
2. Model description	The MASTEP-regional model is described in detail. The model provides a framework that compiles the definition of a landscape structure, pesticide exposure time series and a population model (ODD format) into landscape-scale simulations. Concrete examples for the subparts of the MASTEP-regional approach are given. An overview about the modelling approach is given at first.	Here we present the complete model description. It is identical to the one given in Johnston et al. (2014), but nevertheless included here because it is relatively short and allows to keep all supplementary information in one document.	Here we present the complete model description following the ODD format.

3. Data evaluation	The model was not calibrated to experimental data. Overall, the available data for parameterization of the model parts was taken from peer-reviewed literature. For the parameterization of the pesticide fate model, several scientific publications were evaluated. The population model was parameterized based on a number of scientific publications that focused, however, primarily on size and fecundity related aspects. Previous applications of the population model indicate its reasonability. However, information on density dependence and dispersal parameters are scarce. The link between exposure and effects was parameterized based on an appropriate scientific publication.	Energy budget parameters for *E. fetida* have been directly derived from relevant literature data. As much of the data does not directly relate to energy equivalents, calculations were necessary to transform the literature data in to compatible units. The parameters represent energy acquisition and expenditure under optimal and constant environmental conditions. In suboptimal conditions, environmental variables (e.g., food availability) limit energy ingestion and subsequent allocation to life cycle processes. Methods used to parameterize the dose–response relationship between pesticide concentration and physiological parameters are also outlined below.	All life-cycle parameters of Folsomia candida used in the model, with the exceptions of those related to the energy expenditures, have been directly derived from empirical data published in the literature, as well as individual-level toxicity data for copper. Qualitative observed patterns were also used to design the overall model structure.
4. Conceptual model evaluation	The MASTEP-regional model builds on existing models whose model concepts make quite some simplifying assumptions. These simplifying assumptions are not discussed in this document. The concept for the landscape-scaled approach of the MASTEP-regional follows from embedding an already existing model into a spatially realistic landscape. Only a few simplifying assumptions had to be made and are discussed.	The conceptual model is represented in Figs. 1 and 2. The design concepts underlying model design are presented in Section 2, Model description. Further information regarding simplifying assumptions is presented in Section 3, Data evaluation.	The conceptual model is represented in Fig. 1. The design concepts underlying model design are presented in Section 2, Model description. Further information regarding simplifying assumptions is presented in Section 3, Data evaluation.

Table 3-2: continued

TRACE element	Focks et al. (2014): Integrating chemical fate and population-level effect models for pesticides on the landscape scale: new options for risk assessment.	Johnston et al. (2014): An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides.	Meli et al. (2013): Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: an individual-based model of springtails.
5. Implementation verification	In addition to standard verification tests such as code check being performed for compilation, two main approaches were followed to ensure a correct implementation of the MASTEP-regional upscaling approach. A species balance calculation is performed for each time step to ensure that individual processes are correctly linked to the upscaling framework. Specific test simulations using manipulated code ensured further integrity of the model code.	In order to ensure that the computer code implementing the model works according to its specification in the ODD model description, a series of tests has been performed. These tests included syntax checking of the code, visual testing through NetLogo interface, the use of print statements and spot tests with agent and patch monitors to check against calculations in Excel, stress tests with extreme parameters values and environmental variables, chemical exposure and concentrations and independent code reviews.	In order to ensure that the computer code implementing the model works according to its specification in the ODD model description, a series of tests has been performed. These tests included syntax checking of the code, visual testing through NetLogo interface, print statements, spot tests with agent monitors, stress tests with extreme parameters values, test procedures and test programs, and code reviews.
6. Model output verification	In this study, no calibration of model parameters was executed in the sense of optimizing parameters to a given data set. Information on how well model simulations match observations are presented in Model output corroboration.	In this section it is described how many and which parameters were inversely determined via calibration. As the energy budget parameters in Table 1 were all directly calculated from literature data sources, information on these parameters are confined to Section 3, data evaluation. Here, details on the modelling of the toxicity submodels are presented. To inversely determine the most plausible toxicity submodel (by altering physiological parameters according to the dose–response relationships in Section 3, Data evaluation), we set up the model as in the corresponding empirical study and evaluated the model output against several patterns observed in the respective laboratory populations (following “Pattern-Oriented Modelling (POM)” and “Akaike Information Criterion (AIC)”.	In this section it is described how many and which parameters were inversely determined via calibration. To inversely determine the values of these parameters we made the model reproduce several patterns observed in laboratory populations at different scales and levels of biological organization (“pattern-oriented modelling”; Grimm et al., 2005).

7. Model analysis	A comprehensive sensitivity analysis of the MASTEP-regional model is due to the relatively high computation times not possible. However, the sensitivity of the model outcomes was evaluated in a set of simulations covering a wide range of pesticide toxicity and persistence. The simulation results indicate a reasonable and meaningful response of the model.	The sensitivity of the model to the values of its parameters is presented in Table 9. The model was run with the parameter values of Table 1 (N = 100) and again with parameter values increased one at a time by 10% (N = 100) and changes in model outputs (adult biomass, juvenile biomass and cocoon production per adult) are shown in Table 9. Also shown in Table 9 are the sensitivity of the model to the baseline values of the environmental variables varied individually; these were soil temperature: 25◦C; soil moisture: 60%; and food density: 20 g per patch.	A sensitivity analysis was performed to explore the behaviour of the model in response to variations in the values of parameters that were not directly determined from the literature. Two different model outputs, final population size and average weekly population growth rate, have been used in this sensitivity analysis. Results are shown in Table 5.
8. Model output corroboration	Given the spatial dimension and resolution of the landscape-scale simulations, data that can be used to corroborate model results is hard to find. We used data from field monitoring campaigns in the Netherlands to corroborate at least the undisturbed population dynamics as simulated with a local MASTEP population model.	A number of patterns on the individual life cycle processes and population dynamics of E. fetida have been identified as reproducing well the available literature data. The studies used to evaluate model output use variable laboratory conditions (e.g., temperature, food density). The energy budget model is parameterized with data relating to optimal environmental conditions, and so good model fits to variable conditions show our model to realistically represent E. fetida physiology. Good model fits to sublethal effects of the pesticides copper oxychloride and chlorpyrifos further show the methods for identifying how chemicals achieve their effects. At the population level, good fits to population density, biomass and stage structure show the potential of the model to extrapolate to more natural conditions. Simulation details of all results are available in Johnston et al. (2014).	Three patterns have been identified from the literature, which have been numbered 3–5 to distinguish them from the patterns used for calibration (1–2).

TRACE is not intended to establish, in the end, good modelling practice that corresponds to Good Laboratory Practice (GLP). GLP is a formalized means for ensuring a defined quality of chemical tests by standardising every single step of analysis. Ecological models, however, are completely different from chemical analyses (Wikipedia Contributers 2013); they are scientific tools, and as such not amenable to something like GLP. Standardising the documentation of model development, testing and analysis does not mean standardising models; likewise, standardising the structure of scientific articles or of the description of individual- or agent-based models using the ODD format does not impose any restrictions on scientific creativity.

Nevertheless, the purpose of TRACE is to establish a culture of model development, testing, and analysis more likely to produce models that are useful, in particular for supporting environmental decision-making. TRACE thus is intended to establish expectations of what modellers should clearly communicate when presenting their model, for example a clear model description, sensitivity analyses, and a detailed description of the empirical information that went into the model‘s design and testing. "Culture" here means that you just do all these things as well as you can because you know that peers and model clients are expecting you to; there is no point any more in complaining about "additional effort" for these things.

In empirical sciences, results cannot get published until methods - including quality control - are fully described. Laboratory experiments require evaluation of instrument error, reagent reliability, etc.; field experiments require evaluation of observation error; and data analysis typically requires comparison of alternative models and evaluation of error and uncertainty. Similarly, the culture of good modelling practice mentioned above already exists in many fields. As ecological modelling matures as a scientific (and regulatory) approach we must expect the same kind of scrutiny of methods as clients become more sophisticated and more demanding of careful practice. In fact, standards for publishing models and accepting their results have increased by several leaps already since the beginning of computer modelling. Further increases in the sophistication with which clients scrutinize models must be expected as models are used for increasingly high-consequence assessments such as predicting effects of climate change and pesticides. As a guide to how to model (beyond its role in documentation), TRACE should be especially valuable for ecologists and other scientists who are self-taught or otherwise lack training in modelling skills such as software testing and model analysis.

We hope that the new TRACE format presented in this article will be widely used, so that it can further be developed and refined. To facilitate TRACE‘s refinement, the template provided in the Supplementary Material should be used unchanged. Furthermore, in parallel to TRACE, establishing a culture of keeping modelling notebooks (Grimm et al., *unpubl. manuscript*) that use TRACE terminology will also improve the culture of ecological modelling.

4 STUDYING THE MOVEMENT BEHAVIOUR OF BENTHIC MACROINVERTEBRATES WITH AUTOMATED VIDEO TRACKING

Jacqueline Augusiak, Paul J. Van den Brink

This chapter has been published in *Ecology and Evolution*, 2015, **8 (5)**, 1563-1575.

Abstract

Quantifying and understanding movement is critical for a wide range of questions in basic and applied ecology. Movement ecology is also fostered by technological advances that allow automated tracking for a wide range of animal species. However, for aquatic macroinvertebrates such detailed methods do not yet exist.

We developed a video tracking method for two different species of benthic macroinvertebrates, the crawling isopod *Asellus aquaticus* and the swimming fresh water amphipod *Gammarus pulex*. We tested the effects of different light sources and marking techniques on their movement behaviour to establish the possibilities and limitations of the experimental protocol and to ensure that the basic handling of test specimens would not bias conclusions drawn from movement path analyses. To demonstrate the versatility of our method, we studied the influence of varying population densities on different movement parameters related to resting behaviour, directionality, and step lengths.

We found that our method allows studying species with different modes of dispersal and under different conditions. For example, we found that Gammarids spend more time moving at higher population densities, while Asellids rest more under similar conditions. At the same time, in response to higher densities, Gammarids mostly decreased average step lengths, whereas Asellids did not. Gammarids, however, were also more sensitive to general handling and marking than Asellids.

Our protocol for marking and video tracking can be easily adopted for other species of aquatic macroinvertebrates or testing conditions, for example presence or absence of food sources, shelter, or predator cues. Nevertheless, limitations with regard to the marking protocol, material, and a species' physical build need to be considered and tested before a wider application, particularly for swimming species.

Data obtained with this approach can deepen the understanding of population dynamics on larger spatial scales and of the effects of different management strategies on a species' dispersal potential.

4.1 Introduction

Movement ecology has experienced increasing attention over the years with technological advancements yielding ever more precise location devices to gain a better understanding of what influences the movement and distribution of animals (Schick *et al.* 2008; Nathan *et al.* 2008). So far, studies of movement behaviour focussed mostly on larger animals living in environments where their movement can be followed rather easily. Examples range from observations of migrating birds, to wandering whales, to mice and other rodents (e.g. Edwards *et al.* 2007; Gurarie *et al.* 2009; Humphries *et al.* 2012). With improving technology, the number of studies on smaller species has increased, whereby terrestrial examples like collembolans and ants are frequently chosen as study objects (Amorim *et al.* 2008; Robinson *et al.* 2008). Aquatic invertebrates and their population distributions, however, are mostly studied in time and labour intensive field surveys where a defined area is chosen and the occurring species quantified (Céréghino *et al.* 2001; Malmqvist 2002). Mark and recapture studies (e.g. Davy-Bowker 2002) are used as a variation of this method. Despite improving insights into dispersal times and patterns, they can over- or underestimate realized dispersal by overlooking patch specific effects on individual behaviour (Ovaskainen 2004; Van Dyck & Baguette 2005). Hawkes (2009) reviewed studies that aimed to link dispersal and population processes to investigate different ways in which they can be combined to yield an understanding of spatial population distributions. He found that the resulting metapopulation models were sensitive to small differences in the dispersal estimates. Consequently, he proposes that in order to estimate dispersal more realistically, individual variability of behaviour should be accounted for.

Long-distance dispersal can be estimated from the small-scale behaviour of a species (Turchin 1998). Small-scale behaviour studies, e.g. in the lab via video tracking, make it possible to investigate mechanistic drivers of movement behaviour. This facilitates the estimation of dispersal distances under various conditions with reduced efforts compared to field surveys. Currently, the behaviour of small organisms is typically recorded via cameras installed above an arena and the obtained paths are analysed with computer software (Martin 2004). Often, the observed individuals are marked. However, choices concerning marking protocols depend strongly on the research question as well as detection requirements of the applied tracking software and the animals' capability to cope with a marker and the marking procedure (Hagler & Jackson 2001).

Compared to terrestrial species, additional technical challenges need to be overcome for studying aquatic macroinvertebrates. Such problems include refraction and light reflection interferences at the air/water boundary, positioning of the light source, and suitable marking techniques. Probably due to these technical challenges, so far only a few behavioural studies have been conducted for aquatic macroinvertebrates (e.g., Englund & Hambäck 2004). Holyoak *et al.* (2008) also found in a review that most reported studies on invertebrate movement were performed at the population-level without quantifying

individual variation of behaviour. This limits the understanding of factors that control respective behaviour.

Learning more about the movement of benthic macroinvertebrates is urgently needed. As consumers at the intermediate trophic level, macroinvertebrates fulfil an important role in the nutrient cycling of aquatic ecosystems (Wallace & Webster 1996). Chemical or physical disturbances due to human activities such as agricultural or engineering practices can lead to local population declines (Vaughn 2010). The immigration of unaffected, or temporary emigration of affected individuals, can support the recovery of disturbed populations (Brederveld *et al.* 2011; Galic *et al.* 2013).

We developed an experimental method to overcome the technical challenges described above to enable the study of movement behaviour of aquatic macroinvertebrates. We tested our method with two species with different modes of dispersal, the crawling isopod *Asellus aquaticus* and the swimming freshwater amphipod *Gammarus pulex* (Figure 4-1). The developed method allows studying individuals of small aquatic macroinvertebrates under various test conditions, which is demonstrated in this paper by varying the population densities in the test setups.

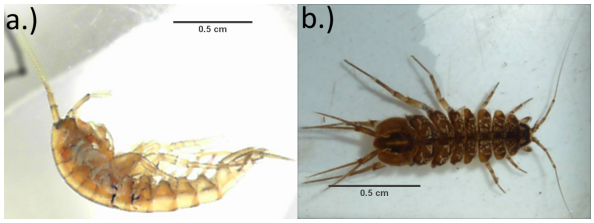


Figure 4-1: Specimens of (a) adult *Gammarus pulex* and (b) adult *Asellus aquaticus* used in the experiments.

4.2 Materials & Methods

4.2.1 Test organisms

Gammarus pulex is an amphipod species that disperses over short distances by swimming, whereby *Asellus aquaticus* is an isopod that moves along the benthos by crawling. Both species are widely spread throughout freshwater habitats in Europe. Despite their different dispersal modes, the predominant dispersal plane is 2-dimensional for both species.

Adult *A. aquaticus* and *G. pulex* were collected during springtime in 2011 and 2012, respectively, from a non-contaminated pond (Duno pond, Doorwerth, The Netherlands) using sweeping nets. To obtain a narrow body size range, specimens of *A. aquaticus* that were larger than approximately 0.5 cm and *G. pulex* larger than approximately 1 cm were transferred to the laboratory and kept in separate, aerated 30L tanks in a climate controlled room at 20ºC and a 10:14 light-dark cycle. Prior to the experiments, the organisms were

acclimated to copper-free water in a sequential diluting process of the original pond water with copper-free water during one week. Dried poplar leaves were supplied as food source ad libitum.

4.2.2 Experimental setup

The movement observations were performed in a climate-controlled room at 20ºC. The test setup consisted of a digital single-lens reflex camera (EOS 1100D, Canon) mounted above an aquarium of approximately 1 m², which was filled with a 0.5 cm layer of quartz sand and 10 cm of copper-free tap water. The camera was directly connected to a computer. Four of such aquarium-camera combinations were installed and used in parallel.

Before the observations, individuals for the experiments were randomly chosen from the stock (mean size *A. aquaticus*: 6.4 mm ± 0.66; mean size *G. pulex*: 13.1 mm ± 1.76) and marked (see below). After 1 hour recovering from the tagging procedure, they were introduced to the aquarium. After another 30 minutes for acclimation, animal movements were recorded for 1 hour and the tracks analysed. All experimental trials were replicated twenty times with different individuals. For those setups designed to investigate the influence of population densities, only one individual of the group was marked and observed, while the unmarked ones served as "background" population. When the recording was finished, the marked individual was exchanged for another marked one. The background population was exchanged after four hours to prevent potential starvation induced behavioural changes, such as, in the case of *Gammarus*, cannibalistic tendencies.

Water temperature, pH and dissolved oxygen were measured twice a day. All experiments were carried out during daytime in a dark room. The average water temperature was 20 ± 0.8 ºC, average pH was 7.6 ± 0.3 (pH323, WTW Germany) and average dissolved oxygen levels varied around 8.6 ± 0.3 mg/L (Oxi330 with CellOx 325 sensor, WTW Germany).

4.2.3 Tagging procedure and marker choice

For the tagging procedure, individual animals were removed from the water, placed in a petri dish, and their backs carefully dried with a lint-free tissue. Rectangular pieces of a fluorescent material (approx. 2 x 2 mm) were then fixed with a small amount of cyanoacrylate (Pattex, Gold Gel) to the back of the selected individuals and the animals put back into fresh water. The time limit for animals to be out of the water was set to 2 minutes to avoid over-stressing the marked individuals.

The employed marking material had to fulfil requirements related to size, weight, and toxicity to ensure that it would not influence the animals mechanically or by chemical release. A strong fluorescence under UV light and easiness to handle during preparation and marking were especially important. We found in preliminary experiments (see

Appendix 1) that regular printing paper was most suitable for *Asellus*, while neon coloured rubber-like plastic met the requirements best for *Gammarus* (UV Gear, Mark SG Enterprises, Surrey, UK; www.uvgear.co.uk).

4.2.4 Movement behaviour studies

a.) Tagging induced effects

To estimate potential influences of the tagging procedure and marker choice on movement behaviour, we recorded marked and unmarked organisms under white light conditions. We used full-spectrum light tubes (JBL, Solar Tropic T8) as light sources, which in combination with the quartz sand substrate enabled the observation of either marked or unmarked specimens. The tubes were adjusted in positions that allowed approximately even illumination of the arenas with as little light reflection on the water surface as possible. In our case, the best positions for the light tubes were slightly to the left and right of the aquaria (Figure 4-2b) at a height halfway between water surface and camera, yielding an average light intensity of 2.0 ± 0.7 µmol s^{-1} m^{-2} (LI-250A Light Meter, LI-COR Biosciences). Due to limitations with extracting movement paths of multiple individuals from the movies, only single animals were introduced to the tanks and recorded. Both treatments, tagged and untagged, were alternated randomly.

b.) Light induced effects

Gammarids and Asellids are generally more active under dark than under light conditions (Wallace *et al.* 1975; Andrikovics 1981). We tested different lighting conditions to investigate light mediated differences in movement behaviour. For tests under dark conditions, i.e. excluding the visible wavelength spectrum, the animals were tagged with a fluorescent marker (see above) and their movement recorded whilst UV-A light tubes were used for illumination instead of the above mentioned full-spectrum tubes. Figure 4-2c illustrates the observation of a marked *Asellus* under such conditions. Single specimens were introduced into the aquaria and the recorded movement data compared to the previously acquired data of the movement of marked specimens under full-spectrum light conditions.

c.) UV light and population density effects

We used UV light and fluorescent markers to differentiate single individuals from a background population of unmarked specimens. This made it possible to investigate the effects of population density on the behaviour of individual Asellids and Gammarids by introducing 0, 50, 100, and 200 unmarked animals in the aquaria along with a single marked individual. These setups were performed with 20 replicates.

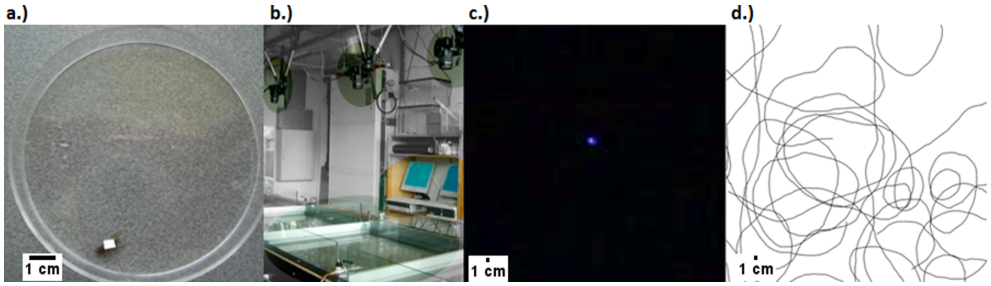


Figure 4-2: Marked *Asellus* specimen (a), the experimental setup (b), the resulting observation under UV light illumination (c) and extracted path representation (d).

4.2.5 Data analysis

The open source software ImageJ (Abramoff *et al.* 2004) was used to process and extract animal tracks from the recorded movies. Tracks within a 10 cm margin of the arena's walls were left out to exclude bias due to fence behaviour (Cant *et al.* 2005). One image per second was processed and resulted in a series of (x, y)-coordinates of an individual at time *t*. The obtained tracks were analysed using R software (R Core Team 2013) and the R package "adehabitat" (Calenge 2006).

Step length, turning angle, and overall activity are key parameters in the analysis of movement paths. Therefore, we analysed the obtained trajectories by the distance between subsequent time points (step length); by the angle between successive moves measured as deviation from straight locomotion in degrees (±180˚); and by the time spend resting (see Figure 4-3a for a schematic representation of path components).

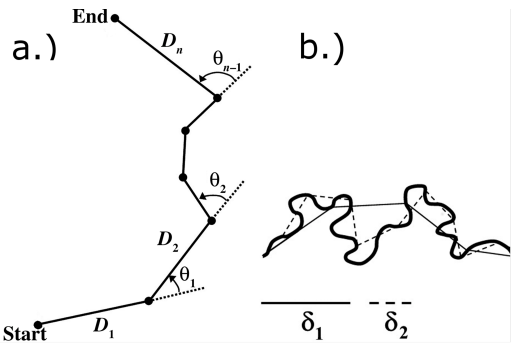


Figure 4-3: a.) Illustration of the components of a movement path. Solid lines represent the distance D_i travelled per time interval (step length). The dashed lines indicate the turning angle (θ) as the deviation from straight-line locomotion measured in degrees (±180˚). b.) Schematic of the divider method. Two steps of the analysis are shown, using two different divider lengths δ (adapted from Seuront *et al.* 2004).

The resting times were calculated from the data as the fraction of time points when the observed individuals did not move. The smallest detectable steps were in a range of ± 0.5 mm in *x*- and *y*-direction. We determined this value by placing paper chips used to mark

Asellus specimen into the aquaria, recorded them for 10 minutes, and processed the movies like the movies with animal observations. Due to slight movements of the water phase, slight vibrations of the installed cameras or inconsistencies in camera sensor performance, the estimated so-called *centres of gravity* of the recorded paper chips could vary by some pixels in either direction and thus lead to an error of up to 0.5 mm in the position determination. Considering that both species breathe and perform other small movements when resting, we assumed that for the determination of the resting times a larger error margin needs to be applied. We thus extended our analysis by manually choosing recording excerpts from times that we knew the animals to not move and found an error margin of up to 1 mm. Consequently, we defined steps larger than 1 mm as relocation and steps smaller than 1 mm as resting moments.

These metrics are scale-dependent and vary depending on the physical or temporal scale at which they are measured. We used fractal analysis to analyse path tortuosity scale-independently (Seuront *et al.* 2004a). The fractal dimension D of a trajectory ranges between $D = 1$ (straight line) to $D = 2$ (Brownian motion, eventually filling a 2 dimensional-plane). We used the Fractal Mean estimator in the Fractal software made available by Nams (1996) to calculate the fractal dimension for each path. If multiple paths were obtained for one individual, a mean value was estimated. The software makes use of the divider method (Mandelbrot 1967) and calculates the trajectory length (L) over a range of divider sizes (δ; see Figure 4-3b for a schematic illustration) such that

$$L(\delta) = \mathrm{k}\delta^{1-D}$$

where k is constant, and D the fractal dimension of the trajectory.

The fractal dimension can be calculated from a subsequent regression of $\log(L)$ as a function of $\log(\delta)$. We used 200 divider sizes (δ) ranging from approximately half of a species' body size (*Asellus*: 0.25 cm; *Gammarus*: 0.5 cm) to the observation scale of 100 cm.

Since the Fractal Mean estimator excludes paths with less than 5 locations from the analysis to enable a robust regression result, we limited the remaining metric calculations for movement length, turning angle, and resting time to the same range to keep the data as comparable as possible. To test whether the resting time or fractal dimension ($\log(D\text{-}1)$ transformed) varied among testing conditions, we used Welch's t-test, or in case of comparing more than two treatments, ANOVA. Because the step length data were not normally distributed, significance of differences between treatments was assessed with the Mann-Whitney test. The turning angles were analysed by taking the circular nature of the data into account (Batschelet 1981; Cain 1989), i.e. 180° referring to the same direction as -180°. We used a method proposed by Abuzaid *et al.*(2011) to represent the obtained data in form of a boxplot. For the analysis of experimental effects, data were pooled from the relocation data from all replicates for each treatment. Since the distributions of the turning angles exhibited varying concentration parameters κ (defines how evenly distributed the

data are, similar to the standard deviation of the Normal distribution), we used the non-parametric Watson-Wheeler test to compare treatments.

4.3 Results

Since we decided to exclude the outer 10 cm range of the aquaria from the data analysis, we did not obtain tracking information for all time points. In Table 4-1, we list the number of data points analysed for each testing regime along with the number of paths and their average duration. In the case of *G. pulex*, we furthermore experienced a loss of information due to the marking material. The fluorescence of the plastic markers was not as strong as the paper's. At certain angles of the swimming *Gammarus* towards the camera, the fluorescent surface was not recordable for the camera and thus also not detectable by the image processing software.

4.3.1 Animal activity and resting behaviour

a.) Effects of experimental conditions

The marking had little influence on the average resting time of *A. aquaticus*, although the variability in resting time increased when the animals were marked (compare light-unmarked with light-marked in Table 4-1 and Figure 4-4b). Under UV light conditions, this variability decreased and the overall distribution of resting times approached that of unmarked Asellids. Furthermore, the mean resting time dropped by almost 10% under UV light conditions compared to full-spectrum light with marked test specimens (Table 4-1 and Figure 4-4b). Due to the relatively high variability of average resting times, this difference was not statistically significant (Table 4-2).

The resting behaviour of *G. pulex*, in contrast, was significantly affected by the marking procedure (Figure 4-5b, Table 4-2). The mean resting time increased drastically (Figure 4-5b). We also found in further analysis that the number of stops per distance increased strongly (Appendix 2).

b.) Effects of population density

Population density did not affect the resting behaviour of *A. aquaticus* significantly, which was the case for *G. pulex* (Table 4-2). Increasing the population density of *Asellus* from one to fifty individuals per aquarium yielded the strongest change of mean resting time for that species. Further increases of *Asellus* population size returned resting times between the two testing regimes with one and fifty individuals. While the presence of unmarked individuals led to a small increase in resting time for *A. aquaticus*, the opposite occurred for *Gammarus* at densities of 50 and 100 individuals. For both species, the mentioned trends were reversed at a density of 200 individuals per m^2 (Figure 4-4b and Figure 4-5b).

Furthermore, increasing population sizes caused a small increase in variation of resting times for *Asellus*, while the opposite occurred for *Gammarus* (Table 4-1).

Table 4-1: Average values and standard deviations for movement parameters estimated for the different experimental regimes with *Asellus aquaticus* and *Gammarus pulex*.

	1 Individual		**50 Individuals**		**100 Individuals**		**200 Individuals**		**Light - marked**		**Light - unmarked**	
	A. aquaticus	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*
Available data points	27895 (39%)	1911 (3%)	23563 (33%)	2329 (3%)	21073 (29%)	3449 (5%)	29511 (41%)	3846 (5%)	13035 (18%)	12021 (17%)	12162 (17%)	13817 (19%)
Number of available paths	328	65	375	134	321	161	408	104	172	256	157	793
Path length	84.8 (117.8)	26.7 (80.1)	62.0 (94.3)	15.0 (19.5)	64.9 (105.7)	19.4 (28.2)	72.0 (85.8)	36.0 (60.2)	74.1 (143.1)	46.3 (98.9)	77.3 (109.0)	16.3 (19.9)
Animal activity and resting behaviour												
Resting time	30.2 % (12.4)	39.5 % (33.7)	40.2 % (13.8)	26.0 % (28.1)	36.9 % (14.2)	20.2 % (22.9)	38.5 % (15.6)	45.7 % (18.3)	40.1 % (21.4)	47.9 % (28.6)	41.4 % (11.6)	18.2 % (18.0)
Step length pattern												
Step length	0.72 cm (0.26)	1.31 cm (1.47)	0.54 cm (0.25)	2.14 cm (2.27)	0.59 cm (0.26)	2.83cm (2.25)	0.57 cm (0.26)	0.67 cm (0.79)	0.65 cm (0.42)	1.30 cm (0.92)	0.61 cm (0.20)	4.13 cm (1.56)
Turning behaviour												
Turning angle	0.74° (7.28	34.29° (88.79)	-0.72° (7.95)	12.28° (67.38)	-0.07° (7.77)	6.01° (16.12)	-7.63° (35.12)	-19.64° (115.96)	0.97° (13.0)	1.93° (13.89)	0.07° (6.26)	-1.8° (±6.68)
Fractal dimension *D*	1.17 (0.13)	1.20 (0.17)	1.10 (0.13)	1.11 (0.10)	1.11 (0.12)	1.09 (0.09)	1.11 (0.10)	1.29 (0.21)	1.10 (0.11)	1.13 (0.09)	1.10 (0.08)	1.05 (0.04)

Table 4-2: Summary statistics of the statistical tests to estimate the significance of the effects of experimental conditions on movement parameters from observations of *Asellus aquaticus* and *Gammarus pulex*.

	Resting times [a,b]		**Step length** [c,d]		**Turning angle** [e]			**Fractal dimension** [*] [a,b]	
	t	p	W	p	W	p	df	t	p
Marking									
A. aquaticus	-0.23	0.82	166	0.86	2.56	0.28	2	0.05	0.96
G. pulex	3.96	<0.01	29	<0.01	18.21	<0.01	2	3.57	<0.01
Light									
A. aquaticus	-1.69	0.11	220	0.60	3.06	0.20	2	1.81	0.08
G. pulex	-0.62	0.55	72	0.71	3.72	0.16	2	1.20	0.26

Density												
	df	F	p	df	X^2	p	W	p	df	df	F	p
A. aquaticus	41.47	2.21	0.11	3	5.47	0.14	4.98	0.55	6	41.31	1.20	0.32
G. pulex	19.09	3.66	0.03	3	10.88	0.01	17.99	0.01	6	21.96	4.69	0.01

[a] Welch's t-test for 2-sample comparison
[b] ANOVA for multi-sample comparison
[c] Wilcoxon's rank sum test for 2-sample comparison
[d] Kruskal-Wallis test for multi-sample comparison
[e] Watson-Wheeler test for 2- and multi-sample comparison
[*] Fractal dimension was log(*D*-1) transformed prior to statistical testing.

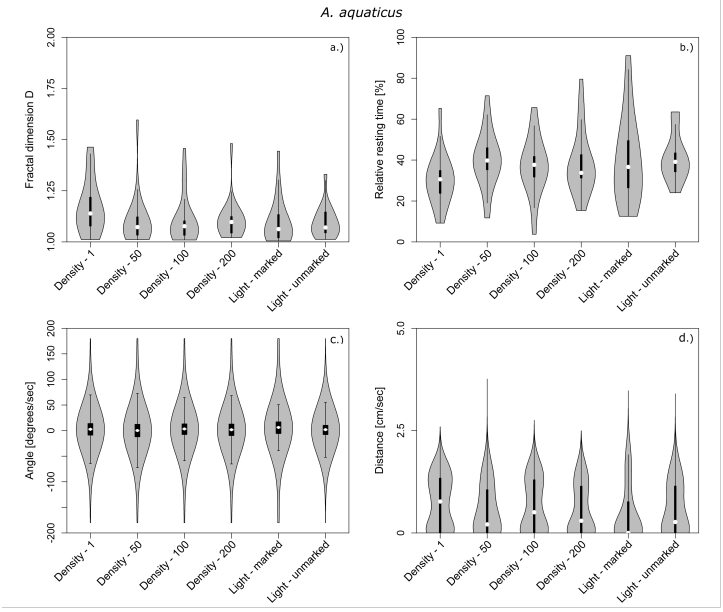


Figure 4-4: Box and Whisker plots combined with violin plots showing the effects of the different treatments on a.) the fractal dimension *D*; b.) resting times; c.) turning angles; and d.) step lengths of *Asellus aquaticus*. Violin plots are a combination of box- and kernel density plots and display the probability distribution of parameters at different values (Hintze & Nelson 1998).

4.3.2 Step length patterns

a.) Effects of experimental conditions

The marking procedure affected the step lengths of *Asellus* only slightly and was statistically not significant (Table 4-2). The average step length of *Asellus* remains about the same with the marker applied but increases when the light regime is changed from full spectrum light to UV (Table 4-1). The distribution of step lengths follows an exponential pattern under the full spectrum light conditions whereas it changes to a Lévy walk pattern where a series of small steps is interchanged with a few larger steps under dark conditions. The violin plots in Figure 4d and 5d depict the distribution of data points around the boxplot representation. A Lévy walk pattern would typically be characterized by a violin with two “bulbs”, whereby the lower one would be bigger due to the presence of more short steps than large steps. An exponential distribution exhibits a broad "base bulb" with a lengthy neck.

Step lengths of *G. pulex,* are significantly reduced (more than 50%) by the marking procedure (Table 4-1, Table 4-2). The distribution of step lengths changed from a Lévy pattern to a more exponential one when a marker was applied (Figure 4-5d). We did not observe any significant changes of average step lengths when comparing light full spectrum to UV exposure although the different light sources lead to increased step lengths and a stronger Lévy pattern in the UV setup (Figure 4-5d).

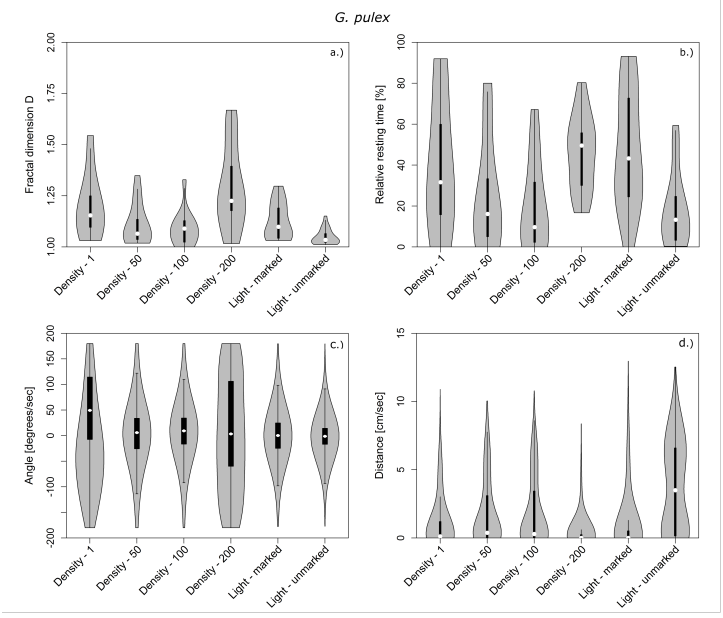


Figure 4-5: Box and Whisker plots combined with violin plots showing the effects of the different treatments on a.) the fractal dimension *D*; b.) resting times; c.) turning angles; and d.) step lengths of *Gammarus pulex*.

b.) Effects of population density

Changes in population density did not significantly affect the observed step lengths for *Asellus* (Table 4-1, Table 4-2). The average step length was highest when the *Asellus* were alone in the arena, but remained virtually unchanged at higher densities. The form of the exhibited Lévy pattern in step length distributions also remained similar at higher densities of Asellids (Figure 4-4d).

Step lengths of *Gammarus*, on the other hand, were significantly affected by population density (Table 4-2). The average step lengths and their standard deviation increased up to a density of 100 Gammarids/m^2, and decrease again at the highest density (Table 4-1), where the resting time was also clearly higher than at the two intermediate population densities.

4.3.3 Turning behaviour

a.) Effects of experimental conditions

Asellus hardly changed their turning behaviour when marked (Figure 4-4c). The increase in turning angle variability due to marking and using full spectrum light reduces the dominance of angles around 0º (forwards) not significantly (Table 4-1, Table 4-2). At the same time, the path tortuosity, as represented in the fractal dimension, remains almost unchanged, and is only slightly wider distributed after marking. A change of the light conditions from full spectrum to UV light reverses the change of turning angle variability and leads to a distribution similar to that of unmarked conspecifics under full light spectrum conditions. The path tortuosity, however, became slightly more variable (Figure 4-4a).

The marking had a significant effect on the turning angle of *Gammarus* (Table 4-2). Although the average direction remained approximately the same, the variability of angles exhibited by marked individuals was greater than of unmarked ones and the path tortuosity increased significantly as displayed in Figure 4-5a. Changing of the light regime from full spectrum to UV light also induced a strong change of the average turning angle as well as the turning angle distribution (Table 4-1), but due to the variability of this parameter in both treatments, no statistical significance of light conditions on turning angles could be detected (Table 4-2).

b.) Effects of population density

Population density hardly affected the turning angle distribution of *A. aquaticus* (Figure 4-4c, Table 4-1, Table 4-2). Density also had no statistically significant influence on the fractal dimension. The higher the density, though, the narrower the distribution of *D* (Table 4-1).

As with the previous metrics, the overall directionality of Gammarids was significantly affected by population density (Table 4-2). The single Gammarids performed sharper turns with an average direction that would not yield less straight-line relocations. This is also observed in the fractal dimension, which has a higher distribution and average value compared to the two intermediate population densities. At the highest density level, the turning angle distribution becomes almost uniform (Figure 4-5c, Table 4-1).

4.4 Discussion

We developed a method for automated video tracking of individual, aquatic macroinvertebrates, which allows collecting detailed information about their behaviour under different conditions such as varying population densities, sediment composition, light regimes, or presence/absence of other factors like food, shelter or stress. The presented tagging and light regime methods can also be adapted to accommodate different species with different modes of dispersal. Furthermore, the spatial and temporal scales as well as the data analysis remain flexible, which can be beneficial and important, depending on the relevant scales of either aspect for the study (Skelsey *et al.* 2012). The application of UV lamps and fluorescent markers proved to be a cost efficient solution to observing aquatic macroinvertebrates while avoiding light reflections on the water surface that can interfere with the image analysis. Additionally, the differences in coloration of the study objects and the substrate, i.e. sediment, are usually smaller than between the species and quartz sand that we used. In this respect, fluorescing markers can be a useful means to overcome object detection difficulties during the image processing, especially when relatively big arenas (compared to the body size of the species) are used for the experiments and only a few pixels are available to represent the animal. However, several factors require careful consideration before the method can be adopted in a meaningful way for new species.

The marking procedure affected both species, but *Gammarus* much more strongly than *Asellus*. However, while *Gammarus* showed effects in all analysis parameters, all of them also statistically significant, *Asellus* exhibited a slightly increased variability in turning angles and path tortuosity. The crawling mode of dispersal and a lower centre of gravity make Asellids more stable on even grounds and thus less prone to an increase of the water resistance due to attached markers. Any device attached to an aquatic animal will exhibit a drag, which affects the animal's movement mechanics depending on the size and weight differences between device and animal. A recent study by Jones *et al.* (2013), illustrated that marking devices mounted on marine turtles exhibit a drag that influences energy expenditures and behaviour of the turtles. In order to be visible to the camera, we had to size and position the markers on the test specimens in a way that made the markers extend slightly wing like. This may alter the hydrodynamics and thus affect the movement of *Gammarus*, especially the directionality. It was also more difficult to mark *Gammarus* individuals because they were more agile when removed from the water phase than *Asellus*

and exhibited unpredictable, erratic turns. This increased the risk of *Gammarus* to be stressed more, leading to a stronger impact on the overall movement behaviour despite an acclimation period prior to the experiments. The mean resting time and mean number of stops made per covered distance increased along with the variability of both parameters (Figure 4-4b and Figure 4-5b, Appendix 2). This is most likely not only due to the physiological stress response by the (more sensitive) Gammarids but also due to the mechanical, physical impairment that the chosen material, or the way it was fixed, may have had on the swimming. Nevertheless, previous studies as the one by Freilich (1989) applied similar marking methods successfully to other macroinvertebrate species in the lab and in the field although the study organisms, stonefly larvae, were larger (approx. 2-5 cm) and more robust than Gammarids. Also, the rubber pieces could not be designed smaller as they were not as brightly fluorescent under UV light as the paper markers and otherwise not yield sufficient visibility. Another material choice, preferably of white colour and inedible material, could overcome these problems and allow for the study of smaller or swimming species. Aiken & Roughley (1985), for example, successfully used small pieces of a plastic waterproof tape that they applied to aquatic beetles. Most other techniques of marking that can be used for terrestrial invertebrates, such as powder coating or dyes, cannot be applied for aquatic invertebrates as the materials would either wash off or require dry surface tissues for fixation, which the water bound organisms may not survive. Mutilation techniques may also alter the hydrodynamics and thus affect the movement behaviour already on a mechanical level. Feeding coloured or fluorescent compounds, as is often done with microorganisms or smaller and short-lived species, could exhibit stronger potentials for intoxications of the marked organism (Hagler & Jackson 2001). Here, a possible intoxication could occur due to the use of cyanoacrylate. During the polymerization process of the glue, the surrounding water can induce a hydrolysis reaction leading to the release of small amounts of formaldehyde and alkyl cyanoacetate. A previous study, in which we tested the safety and toxicity of the chosen marking regime, however, did not indicate any severe effects on the animal's survival or behavioural endpoints (results shown in the Appendix 1).

Comparing the behavioural changes of both species due to marking, we would suggest that the presented marking technique would need to be refined for species that swim and/or are small, and where maintaining hydrodynamic stability thus is a bigger concern than for species that live close to the benthic areas or have a flatter body design like *Asellus*.

Another factor to consider in regard to the experimental setup is the application of UV lamps. Some species of aquatic invertebrates react to this wavelength spectrum and may use it as a reference to guide diurnal or mating behaviour pattern (Frank & Widder 1994). We could not find any relevant information on the photosensitivity for our particular test species and whether their retinae allow the detection of UV light. However, considering the studies of Goldsmith & Fernandez (1968) and Aarseth & Schram (1999) on spectral sensitivities of crustaceans and comparing the behavioural responses from both species when changing the light regime, we conclude that neither *Asellus* nor *Gammarus* will be

affected by the UV range. Goldsmith & Fernandez (1968) investigated the light receptors in the eyes of different species of freshwater crustaceans and a terrestrial isopod but found only scarce occurences of UV sensitivity for the crustaceans. Aarseth & Schram (1999) compared the vertical migration profiles of two copepod species under exposure to visible wavelengths (VIS) and a coombination of VIS and UV wavelength. They found, that one species gathered deeper in the water phase when UV light was used. The other reacted only to the VIS-UV combination when they were kept in a shallow beaker closely to the light source. We did find a reduction in resting times for both species as the most notable behavioural change when using UV instead of the full spectrum lights. Allema *et al.* (2012) found a similar response in terrestrial, nocturnal beetles when comparing full spectrum to red light conditions. They furthermore concluded that near infrared (NIR) light would be the most suitable to study the behaviour of nocturnal organisms but that the more practical and more readily available red light lamps would still allow a representative observation of the animals as total darkness would be rarely found in ecological environments. We draw the same conclusion for the species studied here and the application of UV lights. Goldsmith & Fernandez (1968) attributed a similar conclusion for crustaceans in general to an absence of UV wavelengths in most of the relevant aquatic habitats. The light source could be changed to NIR or red light for species that respond more strongly to UV. However, a lot of contrast would be lost between observed object and the background. Given the dimensions of our setup, either a better camera needs to be used under such circumstances or the camera would need to be lowered to increase the number of pixels representing the object of interest, which in return would mean that only a smaller part of the arena could be monitored.

We tested further limits of the developed protocol by studying the movement behaviour of *Asellus* and *Gammarus* in different population densities. For *Asellus*, we generally found the most striking differences in behaviour between the lowest densities of 1 individual/m^2 and 50 individuals/m^2 (Table 4-1). Parameter values determined at higher population densities fell into ranges that were in-between these two densities. Resting times changed the strongest. The exhibited increase in activity when alone compared to the higher densities suggests a search for conspecifics as protection mechanism against predation. A similar phenomenon was reported for the movement speed of mussels by Van de Koppel *et al.* (2008). They explained their findings by suggesting that an initial slowing at increasing densities was initiated by small-scale cluster formations as protection against predators. At higher densities, they found movement speeds to increase again, which was hypothesized to release intraspecific competition. Additional work by De Jager *et al.* (2013), furthermore suggests that changes in movement behaviour at increasing population densities can be explained by conspecific encounter rates. We find similar effects of density on both our species with an increased number of stops made per meter, reduced average step lengths and more variable turning angles at the highest population density compared to the intermediate ones.

G. pulex showed a different behavioural pattern at the different population densities regarding the resting time, with the biggest overall differences occurring between the intermediate densities and the 200 individuals/m^2 experiments (Table 4-1) and appears most active in the intermediate density ranges. This duality in inactivity, resting similarly much when alone or at higher densities, could be influenced by the marking. The presence of conspecifics could trigger a searching or escaping mode of behaviour despite the negative influence of the markers on the hydrodynamics. Once the population density, and thus the encounter rate, becomes too high it might energetically be more advantageous for the marked individual to stay inactive rather than searching for food or trying to escape conspecifics.

Nevertheless, we rarely found statistical significance when comparing testing regimes, with the strongest indication of marking affecting the behaviour of *G. pulex*. The high variability of individual behaviour is a reason for this, which is amplified by the observation of 20 individuals per setup. Despite the rare statistical significances, trends in the data could possibly be magnified with appropriate methods in a modelling exercise to determine if the small local changes in behaviour yield a significant effect on a larger scale. Considering, that the scale-dependent parameters exhibited patterns that are similar to the scale-independent fractal dimension indicates that our observations are representative and might not change much if a different temporal or spatial scale was applied for the analysis. In general, the data analysis, the estimation of summary statistics such as a net squared displacement, and adjusting of the experimental environment can be designed and performed according to the respective research question. The basic experimental setup could furthermore be applied in semi-natural environments in outdoor systems if the water phase is clear enough. However, the UV fluorescence would not hold up under such circumstances, as the effect of the lamps would be deterred under natural light conditions.

To extrapolate the experimental findings to more complex scenarios or spatial scales than could be captured with a camera, modelling can be used to translate these findings from the small-scale behaviour to large-scale dispersal. Models can thus help to understand how localized factors relate to dispersal events and pattern as well as the resulting distribution of populations and their connections. The experiments might only reflect a small aspect of an overall behaviour on a population level in a larger, heterogeneous environment but can provide first insights into the behavioural drivers for species, which so far were not studied because of technical limitations or could be used as building blocks in mixed modelling approaches. Holdo & Roach (2013), for instance, demonstrated that Monte Carlo simulation could serve as a tool to extrapolate from small sample sizes to the population and to account for potentially different behavioural modes to capture population dispersal more realistically.

5 THE INFLUENCE OF INSECTICIDE EXPOSURE AND ENVIRONMENTAL STIMULI ON THE MOVEMENT BEHAVIOUR AND DISPERSAL OF A FRESHWATER ISOPOD

Jacqueline Augusiak, Paul J. Van den Brink

This chapter has been published in *Ecotoxicology*, 2016, **25 (7)**, 1338-1352.

Abstract

Behaviour links physiological function with ecological processes and can be very sensitive towards environmental stimuli and chemical exposure. As such, behavioural indicators of toxicity are well suited for assessing impacts of pesticides at sublethal concentrations found in the environment. Recent developments in video-tracking technologies offer the possibility of quantifying behavioural patterns, particularly locomotion, which in general has not been studied and understood very well for aquatic macroinvertebrates to date.

In this study, we aim to determine the potential effects of exposure to two neurotoxic pesticides with different modes of action at different concentrations (chlorpyrifos and imidacloprid) on the locomotion behaviour of the water louse *Asellus aquaticus*. We compare the effects of the different exposure regimes on the behaviour of *Asellus* with the effects that the presence of food and shelter exhibit to estimate the ecological relevance of behavioural changes.

We found that sublethal pesticide exposure reduced dispersal distances compared to controls, whereby exposure to chlorpyrifos affected not only animal activity but also step lengths while imidacloprid only slightly affected step lengths. The presence of natural cues such as food or shelter induced only minor changes in behaviour, which hardly translated to changes in dispersal potential.

These findings illustrate that behaviour can serve as a sensitive endpoint in toxicity assessments. However, under natural conditions, depending on the exposure concentration, the actual impacts might be outweighed by environmental conditions that an organism is subjected to. It is, therefore, of importance that the assessment of toxicity on behaviour is done under relevant environmental conditions.

5.1 Introduction

Arthropod populations form an integral part of freshwater ecosystems and are, as such, often exposed to chemical and physical disturbances such as nutrients, pollutants, habitat destruction and flow alterations (Dudgeon *et al.* 2006). In agro-ecosystems, pesticides used for plant protection in particular can enter surface waters through spray drift, run off, and draining, and affect non-target animal populations. Hence, environmental risk assessments are required for pesticides to minimize undesired side effects. Standard tests comprise a battery of mortality, immobilization and reproduction studies on single species in the lower tiers of the assessment process. In the higher tiers, micro- and mesocosms may be employed to evaluate ecological community responses to different exposure concentrations (Brock *et al.* 2006).

To improve the determination of ecologically relevant risk levels, behavioural endpoints are increasingly investigated in ecotoxicological studies (Rodrigues *et al.* 2016). They have been shown to be relevant and useful in acute and chronic environmental risk assessments because they link physiological functions with ecological processes. Behavioural endpoints are also very sensitive towards environmental stimuli and chemical exposure (Dell'Omo 2002), and several studies assessing the environmental risks of pesticides reported behavioural effects at concentrations significantly below those causing mortality (for examples see Böttger et al. 2013; Agatz et al. 2014). Locomotor behaviour is particularly vital to animal life as it facilitates feeding, predator avoidance, reproduction, or migration, and thus may link the effects of individual stress to the population level (Bayley *et al.* 1997). This type of behaviour can be studied easily via video tracking (Augusiak & Van den Brink 2015; Rodrigues *et al.* 2016).

In aquatic environments, relocating macroinvertebrates are likely to encounter contaminated stretches with residue concentrations of pesticides. Depending on the mode of action and concentration of the encountered pesticide, travelling animals may be affected and their movement behaviour may be likely to change under such conditions. Especially neurotoxic substances might adversely affect orientation and activity. The observed alterations in activity, furthermore, correlated with the measured contamination gradient. Baatrup & Bayley (1993) showed that cypermethrin exposure disrupted the general movement pattern and activity of the Wolf Spider *Pardosa amentata*. However, studies on the behavioural effect of toxicants on aquatic crustaceans, so far mainly focused on feeding responses (Böttger *et al.* 2013; Agatz *et al.* 2014), induction of drift (Beketov & Liess 2008), breathing activity, and immobilization (for example Rubach et al. 2011). Fewer studies attempted quantification of more complex behaviour such as precopulatory mate guarding (Blockwell *et al.* 1998) or predator-prey interactions (Brooks *et al.* 2009) after sublethal pesticide exposure. To estimate the impact of chemical exposure on arthropod populations in an ecologically more meaningful way, ecological effect models are increasingly often applied to integrate different habitat, species, and exposure related information to assess population recovery timeframes (Galic *et al.* 2013; Focks *et al.*

2014a). Accounting for immigrating and emigrating individuals is essential to improve the mechanistic understanding derived from such modelling studies (Focks *et al.* 2014a; Hommen *et al.* 2016).

With the present study, we present a method to test the effects of chemical exposure on macroinvertebrate movement, and to improve the understanding of the potential effects of exposure to neurotoxic pesticides, in this case chlorpyrifos and imidacloprid, on the water louse *Asellus aquaticus*. To establish a broader knowledge of the background levels and variance of the movement responses we included observations of non-exposed specimens under environmentally relevant scenarios such as the presence or absence of food and shelter items.

Imidacloprid is a selective and systemic insecticide belonging to the group of neonicotinoids that agonistically affect nicotinic acetylcholine receptors (nAChRs) of insects (Matsuda *et al.* 2001). Chlorpyrifos, on the other hand, is an organophosphate insecticide that inhibits acetylcholine esterase, which is essential to nerve function in insects, humans, and other animals (Pope 2010), thus acting as a broad-spectrum agent (Song *et al.* 1997). Exposure to either substance, however, can eventually cause paralysis and death. We aimed to test whether the differences in mode of action would lead to different effects on the locomotion behaviour and whether the responses are concentration-dependent.

A. aquaticus is widely distributed throughout Europe, and is relatively sensitive to insecticides (Wogram & Liess 2001). As consumers at an intermediate trophic level, they also fulfil an important role in the nutrient cycling of aquatic ecosystems (Wallace & Webster 1996). Their population recovery processes are limited since the species has a fully aquatic life-cycle with virtually no possibility to reoccupy exposed patches by air. Recovery, hence, depends mostly on the intrinsic reproduction potential and dispersal of individuals within a water body from uncontaminated patches towards exposed ones. This species also appeared to be easily studied using automated video tracking (Augusiak & Van den Brink 2015).

5.2 Materials & Methods

5.2.1 Test species

Adult *A. aquaticus* were collected from a non-contaminated pond (Duno pond, Doorwerth, The Netherlands) with sweeping nets, and organisms larger than approximately 5 mm were transferred to the laboratory. The specimens were kept in a 30 L aquarium in a climate-controlled room at 18ºC and a 10:14 light:dark cycle. Prior to the experiments, the organisms were acclimatised to copper-free water over one week by a sequential diluting process of the original pond water with copper-free water. Dried poplar leaves were provided as food source ad libitum and aeration was constantly supplied. Individuals for

the experiments were chosen randomly from this stock (mean body length ± standard deviation: 6.4 mm ± 0.66).

5.2.2 Experimental Setup

The movement observations were performed in a climate-controlled room at 20ºC. The test setup consisted of a camera mounted above an aquarium of 1m², which was filled with a 0.5 cm layer of quartz sand and 10 cm of copper free tap water. Before the observations, individual specimens were marked with rectangular paper snippets of approximately 2 x 2 mm, left for 1 hour to recover from the marking procedure, and introduced into the aquarium. Small droplets of cyanoacrylate (Pattex, Gold Gel) were used to fix the marker to the backs of the *Asellus*. After introduction into the aquarium and 30 minutes acclimation time, animal movements were recorded for 1 hour and the tracks statistically evaluated to determine movement related parameters. We used a digital single-lens reflex camera (EOS 1100D, Canon) for the recordings, which was connected to a computer. Four of such aquarium-camera combinations were installed in parallel within a water bath that maintained constant temperatures. See Augusiak & Van den Brink (2015) for further details about the used methodology.

Water temperature, pH and dissolved oxygen were measured twice every day to ascertain stable conditions throughout the experimental period. All experiments were carried out at a water temperature of 20 ± 0.8ºC, an average pH of 7.6 ± 0.3 (measured with electrode pH323, WTW Germany) and an average dissolved oxygen level of 8.6 ± 0.3 mg/L (measured with oximeter Oxi330 equipped with sensor CellOx 325, WTW Germany).

a. Test chemicals - application, sampling, and analysis

Exposure concentrations were derived from toxicity tests performed prior to the behavioural study (see Appendix 3 for details). Solutions of chlorpyrifos were prepared by spiking copper-free water with an aqueous stock solution of chlorpyrifos (480 g/L) to reach exposure concentrations of 0, 0.6 and 1.5 µg/L (48 h-EC50 = 3.2 µg/L, 48 h-EC10 = 2.7 µg/L, Appendix 3).

Water samples from the controls and exposure vessels were taken at the start and after 48 hours of exposure to confirm concentrations. In the beginning, 200 mL samples were taken from the spiked batch volume; at the end, 200 mL per exposure vessel were sampled. Chlorpyrifos was measured by liquid-liquid extraction with 20 mL n-hexane followed by gas chromatography coupled with electron capture detection (GC-ECD). The specifications for the sample analysis via GC-ECD were in accordance with the study by Rubach *et al.* (2011).

Dosing solutions of imidacloprid were prepared by mixing a soluble formulation containing 200 g imidacloprid/L into copper-free water, yielding an 80 ppm stock solution, which was

used to spike the exposure solutions of 0, 37.5 and 75 µg/L (48 h-EC50 = 603 µg/L, 48 h-EC10 = 225 µg/L, Appendix 3). Water samples from the controls and exposure vessels were taken at the start and after 48 hours of exposure to confirm concentrations. For this, samples of approximately 3 mL were transferred into 4 mL glass vials that contained 1 mL acetonitrile. After mixing, the vials were stored at -20ºC prior to analysis. Specifications for the water sample analysis via liquid chromatography–tandem mass spectrometry (LC-MS/MS) were analogous to the study by Roessink *et al.*(2013).

b. Test conditions

To study the effects of sublethal pesticide exposure on the dispersal behaviour, specimens were exposed to the respective pesticide concentration for 48 hours prior to the marking and video observation procedure. After 48 hours, the animals were removed from the exposure vessels and transferred into clean, copper-free tap water. Water quality parameters were measured in the beginning and the end of the exposure phase and water samples taken for chemical analysis at the same time. During the chlorpyrifos exposure, the water temperature was 20.1 ± 1.6ºC, the average pH was 6.8 ± 0.8 (measured with electrode pH323, WTW Germany) and the average dissolved oxygen level was 7.9 ± 0.2 mg/L (measured with oximeter Oxi330 equipped with sensor CellOx 325, WTW Germany). During the imidacloprid exposure the water temperature was 20.0 ± 1.4ºC, the average pH 7.8 ± 0.2 and the average dissolved oxygen level was 7.5 ± 1.2 mg/L. Control groups were kept under similar conditions, except that no pesticide was added.

To test the effect of potential food items being present, we cut leaves found in the animals' native environment into 5 x 5 cm rectangular pieces and hung four such fragments at evenly distributed spots into the water in the arenas. We used simple threads to fix the leaves and adjusted the vertical position in the water phase so that the leaf material was just immersed. Shelter experiments, on the other hand, were conducted with 5 x 10 cm big rectangles of stainless steel mesh wire structures that were placed at six evenly distributed spots in each arena. Control groups were handled similarly, except that no items were added to the arena. All experiments were conducted with two population densities, one and fifty individuals per arena, respectively, and were replicated twenty times each (Augusiak & Van den Brink 2015).

c. Data analysis

We used the open source software ImageJ (Abramoff *et al.* 2004) to extract animal tracks from the recorded movies. Tracks within a 10 cm margin of the arena's walls were dismissed to exclude potential bias due to edge behaviour (Creed & Miller 1990). The obtained time series of (x, y)-coordinates of the animals' positions were analysed using the R software (R Core Team 2013) and the package "adehabitatLT" (Calenge 2006).

We defined relocations of less than 1 mm as resting moments (Augusiak & Van den Brink 2015), and calculated resting time per individual as the percentage time that the respective individual spent not moving. During periods of activity, behaviour was further characterized by step lengths and turning angles. Step length is defined as the distance covered per time interval, whereas angles between successive moves were measured as deviation from straight locomotion in degrees (±180˚) (see Figure 5-1a for a schematic representation of the path components). Since these metrics depend on the physical or temporal scale at which they are measured, we chose to further calculate the fractal dimension of each individual's path. The fractal dimension is a measure of a path's tortuosity and quantifies an object's ability to cover the Euclidian space through which it navigates scale-independently (Seuront *et al.* 2004c). The parameter values range between $D = 1$ (straight line) to $D = 2$ (Brownian motion). We used the Fractal Mean Estimator contained in the Fractal software made available by Nams (1996) to calculate the fractal dimension for each path. If multiple paths were obtained for one individual, a mean value was estimated. The software makes use of the divider method (Mandelbrot 1967) and calculates the trajectory length (L) over a range of divider sizes (δ; see Figure 5-1b for a schematic illustration) such that

$$L(\delta) = k\delta^{1-D}$$

where k is constant, and D the fractal dimension of the trajectory. The fractal dimension can be calculated from a subsequent regression of $\log(L)$ as a function of $\log(\delta)$. We used 200 divider sizes (δ) ranging from approximately half of a species' body size (*Asellus*: 0.25 cm) to the observation scale of 100 cm. Movement tracks shorter than 5 relocation points were excluded from the estimation of fractal dimension values to facilitate a robust regression. For consistency among compared parameters, we limited the remaining data analysis to the same range.

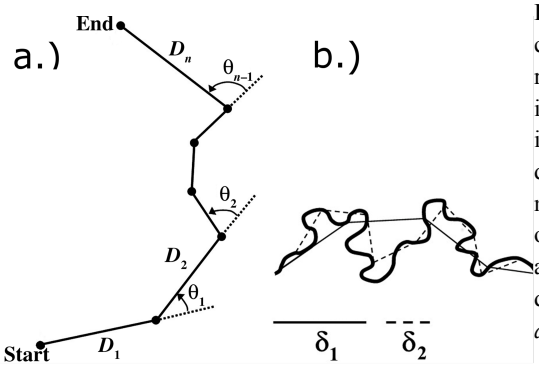


Figure 5-1: a.) Illustration of the components of a movement path. Solid lines represent the distance D_i travelled per time interval (step length). The dashed lines indicate the turning angle (θ) as the deviation from straight-line locomotion measured in degrees (±180˚). b.) Schematic of the divider method. Two steps of the analysis are shown, using two different divider lengths δ (adapted from Seuront *et al.* 2004).

The assumption of normality was violated for all variables, except a transformed version of the fractal dimension (log(*D*-1) transformed), restricting us to mostly non-parametric tests to assess differences between experimental conditions. Wilcoxon's rank sum tests were applied to test for pairwise differences of resting times and step lengths between treatments, Kruskal Wallis tests were used for comparing more than two treatments. To determine differences between fractal dimension values, we used the Welch's t-test, or in case of comparing more than two treatments, ANOVA. Standard methods of circular statistics were used to analyse the turning angles. Since the angular distributions exhibited varying concentration parameters κ, we used the non-parametric Watson-Wheeler test to compare treatment effects (Batschelet 1981). Significances were assessed at a 95% confidence level.

The paths recorded under different experimental conditions were further analysed for deviances with a correlated random walk (CRW) model following the steps laid out in Turchin (1998). This type of model is suitable for evaluating paths in homogeneous environments and can be used to estimate the population dispersal rate within the respective substrate (Turchin 1998). For an analysis of movement paths according to the CRW model framework, a series of statistical approaches needs to be applied to test whether model assumptions are met.

The primary assumption in CRW models is that the organisms exhibit some degree of directional persistence, i.e. the stronger the directional persistence, the faster the population is assumed to spread. This can be checked visually via the frequency distribution of observed turning angles. CRW models furthermore assume that step lengths and turning angles within a path are not serially correlated (Turchin 1998). Such correlations can influence the model output and need to be interpreted accordingly (Turchin 1998; Westerberg *et al.* 2008; Dray *et al.* 2010). Auto-correlation for step-length and turning angles was estimated according to the procedures defined by Dray et al. (2010). The correlation between the magnitude of turning angles and step length was estimated using Spearman's correlation.

For verifying the applicability of the CRW formulation, net-squared displacements (R^2_n) were calculated and comparisons made between estimated (theoretical) and observed (actual) values. Observed net-squared displacements were calculated as the squared distance between each location in an individual's track and the individual's original location. Directional information thereby is removed by using the square of the distances. According to the CRW framework, R^2_n can be estimated and extrapolated as follows:

$$R_n^2 = nL_2 + 2L_1^2 \frac{c}{1-c}\left(n - \frac{1-c^n}{1-c}\right)$$

where L_1 is the mean move length (cm), L_2 is the mean squared move length (cm^2), n is the number of consecutive moves, and c is the mean cosine of turning angles (Kareiva & Shigesada 1983; Turchin 1998). The 95% confidence interval for the estimated R^2_n was constructed following a procedure described by Turchin (1998).

5.3 Results

Due to excluding short tracks and tracks within the outer 10 cm margin of the aquaria from the data analysis, we did not obtain tracking information for all time points. The number of data points analysed for each test regime along with the number of paths and their average duration are summarised in Table 5-1. Furthermore, Table 5-1 lists the intended and measured concentrations of the two studied pesticides. The achieved chlorpyrifos concentrations were approximately 40% below the intended levels at the start of the exposure phase. During the course of the exposure the concentrations dropped due to evaporation, chemical degradation, and sorption processes. However, the concentration difference remained at a factor of approximately 2 between the higher and the lower concentration treatments, indicating that observed changes in behaviour were still comparable among the different exposures. Achieved imidacloprid concentrations, on the other hand, were slightly above the intended levels, with concentrations decreasing less strongly as in the case of chlorpyrifos.

Observed movement and dispersal

In Figure 5-2 the relationship between the observed net-squared displacements (R^2_n) of *A. aquaticus* under different testing conditions and the number of consecutive steps they have made is represented with dashed lines. Net-squared displacement describes the ability of an organism to disperse, i.e. the smaller its value the closer an individual is to its original location. An individual's R^2_n over time is influenced by the combination of step lengths and turning angles it uses. The more active an animal is and the longer and more directed its subsequent steps are, the faster it will move away from its original location.

Pesticide exposure. Observed net-squared displacements were reduced by pesticide exposure compared to the respective controls (Figure 5-2a-e). Higher exposure concentrations thereby caused stronger decreases in R^2_n for both substances, except for the application of the higher chlorpyrifos dosage in the higher density setup. That treatment also changed the observed pattern of single individuals dispersing farther than their counterparts in a group (Figure 5-2b). Compared to the controls, chlorpyrifos exposure increased resting times and decreased step lengths more than imidacloprid exposure did. The standard deviations of either parameter also increased but were, irrespective of the substance, concentration, or population density, overall in a more similar range than the mean values (Table 5-1). The control group exhibited slightly bigger average turning angles with lower variability than the exposed groups did, which however hardly affected the fractal dimension of the analysed paths. Resting times were affected significantly for all single-specimen observations, while step lengths were affected significantly or marginally significantly for both single- and 50-specimens observations (Table 5-2). Chlorpyrifos exposure had an overall statistically more significant effect on those parameters than imidacloprid exposure had. Turning angles and fractal dimension were statistically less affected by either exposure (Table 5-2).

Table 5-1: Basic path information and mean values of movement parameters estimated for the different experimental regimes with *Asellus aquaticus*.

	Chlorpyrifos low (0.6 µg/L)		**Chlorpyrifos high (1.5 µg/L)**		**Imidacloprid low (37.5 µg/L)**		**Imidacloprid high (75 µg/L)**		**Control (starved)**		**Control (fed)**		**Food**		**Shelter**	
Density	1	50	1	50	1	50	1	50	1	50	1	50	1	50	1	50
Available data points	29760	33098	28132	31668	11295	27450	21432	23127	19484	23212	27807	23263	26189	25569	11291	12119
(Percentage of total recording time)	41%	46%	39%	44%	17%	38%	30%	32%	27%	32%	39%	32%	36%	36%	16%	17%
Number of available paths	256	384	244	421	336	330	379	394	336	448	328	375	314	289	176	186
Average path duration (sec)	114.4	85.2	113.8	74.2	35.4	81.9	55.1	57.4	56.7	50.8	83.8	60.9	82.4	87.2	62.4	63.1
	(138.7)	(139.2)	(156.7)	(96.2)	(56.8)	(143.0)	(117.4)	(105.3)	(99.1)	(85.9)	(117.8)	(94.2)	(132.7)	(136.8)	(97.0)	(92.2)
Average measured concentrations (t_{0h}, t_{48h}, µg/L)	0.40, 0.28 (0.03, 0.06)		0.83, 0.75 (0.05, 0.21)		42.09, 40.67 (3.80, 3.94)		80.82, 77.61 (2.80, 3.38)		-		-		-		-	
Resting time	51.5%	53.8%	56.7%	44.4%	28.4%	37.9%	36.3%	35.4%	29.5%	31.2%	30.2%	40.2%	35.7%	45.4%	44.2%	44.2%
	(26.7)	(29.4)	(29.9)	(21.4)	(16.7)	(24.5)	(27.4)	(22.6)	(10.3)	(15.9)	(12.4)	(13.7)	(13.5)	(19.1)	(19.4)	(11.0)
Step length (cm/sec)	0.79	0.71	0.53	0.75	0.82	0.81	0.74	0.92	1.12	1.13	0.99	0.86	0.94	0.80	0.86	0.69
	(0.37)	(0.36)	(0.31)	(0.28)	(0.30)	(0.33)	(0.42)	(0.36)	(0.25)	(0.29)	(0.25)	(0.25)	(0.25)	(0.29)	(0.32)	(0.28)
Turning angle	1.55°	-0.91°	0.93°	1.19°	1.14°	3.74°	-1.57°	-0.15°	2.92°	-2.73°	2.56°	0.09°	-4.35°	-6.70°	-1.48°	0.15°
	(28.10)	(35.41)	(44.87)	(32.85)	(37.00)	(37.19)	(45.18)	(43.20)	(25.09)	(26.31)	(27.77)	(36.71)	(28.25)	(34.21)	(28.33)	(38.57)
Fractal D	1.14	1.12	1.30	1.11	1.24	1.29	1.17	1.10	1.16	1.19	1.23	1.25	1.18	1.17	1.10	1.11
	(0.18)	(0.08)	(0.26)	(0.09)	(0.18)	(0.28)	(0.13)	(0.13)	(0.17)	(0.19)	(0.26)	(0.31)	(0.19)	(0.19)	(0.11)	(0.06)

() - The respective standard deviations are given in brackets.

Table 5-2: Summary statistics of the statistical tests estimating the significance of the effects of experimental conditions on the movement behaviour of *Asellus aquaticus*. Parametric tests were applied for evaluating effects on resting times and a transformed version of the fractal dimension, while non-parametric tests were chosen in the case of step lengths and turning angles. For additional insights into effect sizes, the correlations of step lengths and turning angles were estimated for each treatment.

		Resting times [a,b]		**Step lengths [c,d]**		**Turning angle [e]**			**Fractal D [a,b,*]**		**Spearman's rank correlation between turning angle and step length**	
		t	p	U	p	W	p	df	t	p	r	p
Pesticides												
Chlorpyrifos low	*1*	-3.26	<0.01	238	0.02	2.24	0.33	2	0.22	0.83	-0.29	<0.01
	50	-0.08	0.94	246	<0.01	2.23	0.33	2	-2.20	0.03	-0.38	<0.01
Chlorpyrifos high	*1*	-3.74	<0.01	312	<0.01	5.96	0.05	2	-1.73	0.09	-0.49	<0.01
	50	-1.05	0.31	233	<0.01	4.37	0.11	2	-0.54	0.59	-0.40	<0.01
Imidacloprid low	*1*	-3.10	<0.01	330	<0.01	6.70	0.04	2	1.01	0.32	-0.41	<0.01
	50	-1.16	0.26	298	<0.01	0.37	0.83	2	-1.55	0.13	-0.42	<0.01
Imidacloprid high	*1*	-2.25	0.03	340	<0.01	3.83	0.15	2	1.36	0.18	-0.51	<0.01
	50	-0.75	0.46	247	0.05	3.89	0.14	2	1.97	0.06	-0.36	<0.01
Controls												
Control (starved)	*1*	-2.43	0.02	226	0.19	4.78	0.09	2	1.93	0.06	-0.25	<0.01
	50	-2.12	0.04	311	<0.01	3.89	0.14	2	-0.71	0.48	-0.23	<0.01
Control (fed)	*1*										-0.25	<0.01
	50										-0.39	<0.01
Environmental factors												
Food	*1*	-1.19	0.32	235	0.35	3.73	0.15	2	0.65	0.52	-0.22	<0.01
	50	-0.84	0.41	233	0.06	0.91	0.63	2	1.72	0.10	-0.21	<0.01
Shelter	*1*	-0.87	0.39	217	0.46	5.25	0.07	2	1.05	0.30	-0.34	<0.01
	50	-0.35	0.73	221	0.24	4.15	0.13	2	-0.90	0.38	-0.43	<0.01
Pesticide concentrations												
		df	F	p	df	X^2	p	df	W	p	df	F	p
Chlorpyrifos	*1*	28.4	10.75	<0.01	2	18.69	<0.01	4	7.42	0.12	35.75	2.15	0.13
	50	35.9	5.71	<0.01	2	17.94	<0.01	4	12.92	0.01	36.97	0.94	0.40
Imidacloprid	*1*	28.8	0.55	0.59	2	9.71	<0.01	4	4.57	0.33	36.35	2.73	0.08
	50	33.5	0.75	0.48	2	9.23	0.01	4	3.90	0.42	37.16	6.67	<0.01

[a] Welch's t-test for 2-sample comparison
[b] Welch's ANOVA for multi-sample comparison
[c] Mann-Whitney U test for 2-sample comparison
[d] Kruskal-Wallis test for multi-sample comparison
[e] Watson-Wheeler test for 2- and multi-sample comparison
* Fractal dimension was log(D-1) transformed prior to statistical testing

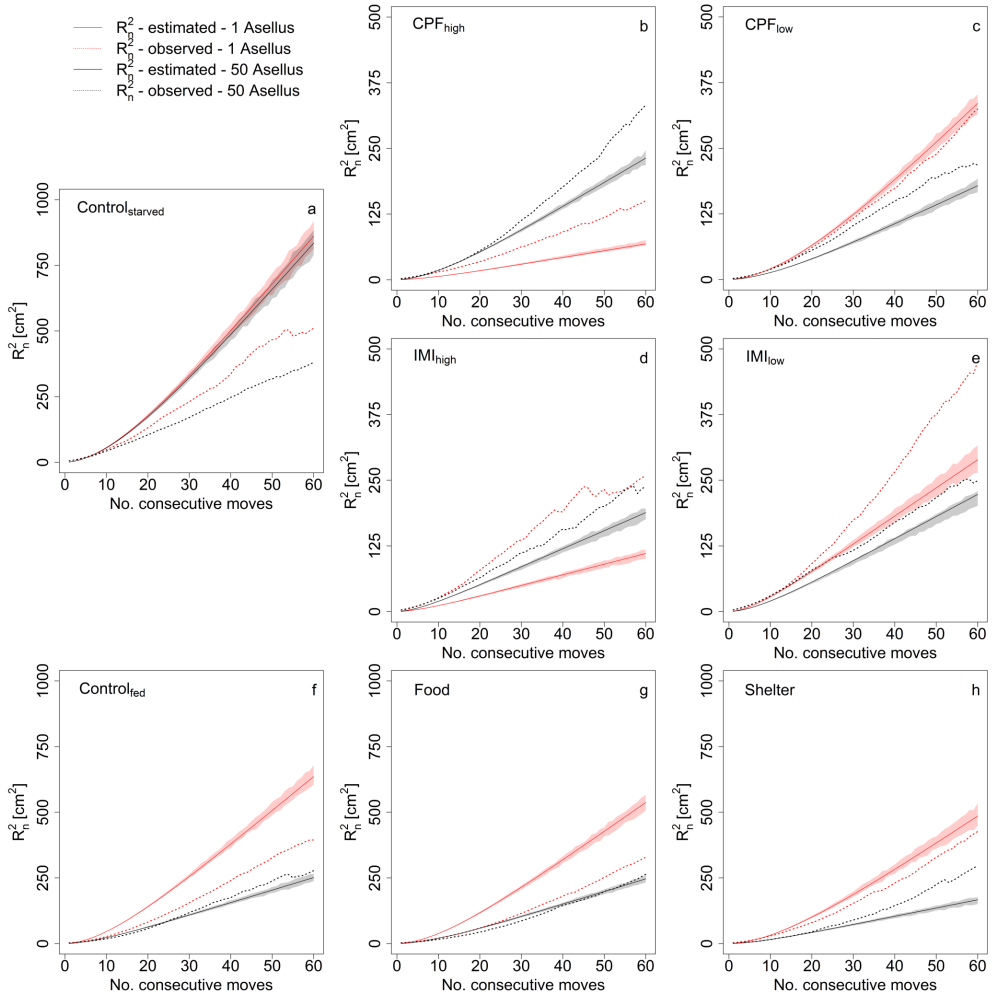


Figure 5-2: Relationship between the mean net-squared displacement (R_n^2; cm^2) and the number of consecutive moves made by *Asellus aquaticus* under different experimental conditions. •••••: observed mean net-squared displacement obtained by averaging over 20 observed individuals; ——: estimated net-squared displacement obtained by applying the observed average move distances and turning angles; solid: 95% confidence interval of the estimated net-squared displacement; red stands for the single-*Asellus* studies and black for the 50-*Asellus* studies.

Environmental stimuli. Observed R^2_n were more similar to each other in the food, shelter, and their respective control tests (Figure 5-2 f-h) than was the case for the pesticide tests. The presence of food items slightly decreased R^2_n in the single individual setup, whereas the presence of shelter items did not cause any observable changes. The biggest effect on observed R^2_n in these three setups was caused by population density. Higher population densities led to decreased R^2_n (Figure 5-2f-h). Resting times increased compared to the

controls when shelter or food items were introduced to the arena (Table 5-2). In the presence of shelter, resting times were equal among the different population densities. When food items were present, the single- and 50-individual specimen maintained the approximate 10% difference that we also found in the control groups. Average step lengths remained virtually the same in the presence of food items, and were slightly lower, although not significant, when shelter items were available. Amongst the different treatments, the observed individuals increased resting times and decreased average step lengths when they were with conspecifics compared to the respective single-specimen setups, probably due to the increased "traffic". Average turning angles increased in the presence of food items, while the presence of shelter items left this parameter unaffected. The fractal dimension decreased slightly more when shelter items were available than when food items were present (Table 5-1). The variability of these parameters was less affected by either treatment than observed in the pesticide exposure experiments, and no statistical indication of treatment effects could be detected. These changes indicate that the observed *Asellus* started searching for food when food items were present, while the presence of shelter provided structures for resting.

Food availability before the experiments had the overall biggest influence on the observed movement behaviour. The pesticide control groups did not receive food for 48 hours prior to the experiment. The control groups for testing the influence of external factors, on the other hand, had access to food until shortly before the recording. The lack of food caused an increase in observed net-squared displacement (Figure 5-2 a and f), which can be explained by a statistically significant reduced resting time and increased step lengths (Table 5-1). While the turning angle range hardly changed, the fractal dimension decreased slightly, indicating that the observed animals changed to overall more linear movements. Additionally, the differences in resting times and step lengths found between the single- and 50-specimen setups disappeared when the individuals were starved (Table 5-1).

Correlation and autocorrelation.

Most observed individuals in the various treatments displayed directional persistence forwards (Table 5-2), meeting the central assumption made under the CRW framework. Turning angles were also significant positively auto-correlated at lag 1 in most cases, and remained significant for several lags (see Online Resource 2 for detailed results, http://link.springer.com/10.1007/s10646-016-1686-y), representing a tendency to make sequential turns in the same direction. Furthermore, auto-correlations in step lengths were significant positive at lag 1 for almost all individuals, and remained significant for a number of lags (Online Resource 2, http://link.springer.com/10.1007/s10646-016-1686-y), which suggests that most individuals maintained similar walking speeds for a number of steps. In all treatments, step lengths and turning angles were significant negatively correlated (Table 5-2), i.e. larger changes in direction were performed only when the individuals slowed down, and average angles decreased with increasing walking speed.

Dispersal estimates

Figure 5-2, furthermore, compares the observed and estimated net-squared displacements (R^2_n) of *A. aquaticus* under different testing conditions. The CRW model overpredicts observed R^2_n in cases where the observed path is more tortuous than assumed by the model. In cases of underestimation, the observed path is straighter or the animal activity lower than expected.

Generally, we found that estimated R^2_n exceeded the observed values for the non-pesticide, single-specimen observations, while observed R^2_n were mostly underestimated after pesticide exposure. Exceptions are the lower chlorpyrifos and the starved control treatments. At the higher population density this pattern changes and all observed R^2_n exceed the estimated values except for the starved control group (Figure 5-2a-e). In the latter case, the model fits the observed pattern better for the non-pesticide treatments during the initial steps compared to the pesticide treatments. However, the CRW models do not provide a good overall fit to the observed displacements (Figure 5-2). The closest fits were found for the higher population density when the observed individuals were fed, and when food items were present (Figure 5-2g).

5.4 Discussion

This study aimed to improve insights into the small-scale movement behaviour of *A. aquaticus* and to evaluate its potential as endpoint in ecotoxicological studies with aquatic macroinvertebrates. The employed video-tracking method (Augusiak & Van den Brink 2015) allowed the detection of already small changes in the exhibited behaviour, although the high inter-individual variability of the analysed parameters made it difficult to detect statistical significant treatment effects. Our results indicate that the locomotory behaviour and dispersal potential of *A. aquaticus* were negatively affected by exposure to sublethal concentrations of chlorpyrifos and imidacloprid, while the presence of food or shelter items reduced the dispersal rate less significantly. In most cases, an increased population density lowered dispersal rates further. The observed effects on the small-scale behaviour also affected the displacement extrapolations.

The pesticides were chosen because of their relatively low elimination rates, making it likely that exposed individuals still experience pesticide related effects when placed in clean water that then can be observed. Rubach et al. (2010) report a 95% depuration time of 16.2 days for chlorpyrifos in *A. aquaticus* and of 7.5 days for adult *Gammarus pulex*, a freshwater shrimp species. In the case of imidacloprid, Ashauer et al. (2010) determined a 95% depuration period of 11.2 days for *G. pulex*. We assumed a continued causation of damage on the nervous system of *A. aquaticus* during the experimental time frame also in the case of imidacloprid. First estimations based on acute toxicity data of imidacloprid exposure, yielded a 95% depuration period of about 4.4 days for *Asellus* (A. Focks, *personal communication*).

The fact that *G. pulex* exhibits significantly higher sensitivities to both chemicals with regard to mobility and survival indicates that surviving individuals could possess a more efficient elimination pathway compared to *Asellus*, allowing the conclusion that the internal concentrations in our study should be stable over the period of time of observation. To test whether changes in locomotion are still observable at sublethal levels, we aimed to apply about 50% and 25%, respectively, of the observed 48h-EC10 of 2.7 µg/L in the case of chlorpyrifos (Rubach et al., 2011: 48h-EC10 = 3.3 µg/L). Due to a wider range of reported ECx values, we opted for a slightly higher safety factor for imidacloprid and chose to continue with about 30% and 15%, respectively, of the observed 48h-EC10 value of 225 µg/L (geometric mean of studies reported by Roessink et al. (2013) and Van den Brink et al. (2015): 48h-EC10 = 54 µg/L). The applied concentrations are also likely to occur in the environment. Concentrations of up to 10.8 µg/L of chlorpyrifos were detected in freshwater habitats throughout the past decade (Marino & Ronco 2005; Ensminger *et al.* 2013), while imidacloprid has been found at concentrations of up to 320 µg/L (Van Dijk *et al.* 2013; Ensminger *et al.* 2013).

In natural environments, the dispersal and local recruitment of aquatic macroinvertebrates is strongly driven by the availability of food, shelter, and population density (Holyoak *et al.* 2008). Food items may release chemicals during the degradation process, which then can be sensed by an organism equipped with the respective sensing systems (Collin & Marshall 2003). This can subsequently cause an alteration in the organism's searching behaviour, for example a switch from long, straight moves to a Brownian pattern for local searching together with a change of activity (Collin & Marshall 2003). Similarly, a lack of food may drive animals away from their current location to search for new resources. Shelter, on the other hand, can impact overall movement by providing protection from high temperatures, light, or predators (Obermüller *et al.* 2007). However, there is a lack of understanding to which degree the presence of food or shelter items can influence the movement and searching behaviour of aquatic invertebrates, or how it may additionally be driven by population density, either by compensating for interspecies competition or improving mating chances (Smith *et al.* 2008; Delgado *et al.* 2013).

Understanding the innate nature of movement behaviour, and to which degree different factors influence it, can help extrapolating small-scale observations to gain an impression on the ecological consequences of chemical or physical disturbances (Getz & Saltz 2008). In Table 5-3, we summarize a number of studies aiming to highlight the influences of chemical exposure or naturally occurring drivers, such as predator cues, on the movement behaviour of aquatic macro invertebrates. We found that most published studies on aquatic invertebrates either focused on environmental cues or chemical exposure, while none related the extent of behavioural changes under sublethal exposure conditions to the innate behavioural range to draw conclusions about potential ecological impact. Observational studies that do investigate such relationships usually use food consumption rates or preferences as endpoint instead of movement (for examples see De Lange et al. 2006; Agatz et al. 2014). The study by (Rodrigues *et al.* 2016) forms a rare exception, where the

effects of sublethal exposure of freshwater planarians to chlorantraniliprole are investigated through observing changes in feeding behaviour and locomotion.

The strong reductions in observed dispersal distances after pesticide exposure were mostly caused by decreased step lengths and increased resting times, which agrees with previous reports of hypoactivity caused by both substances (Rice *et al.* 1997; Suchail *et al.* 2001). Step lengths were significantly reduced by all pesticide treatments, while resting time was more affected by exposure to chlorpyrifos than to imidacloprid. The turning behaviour, i.e. directionality, was not significantly different from that observed in the controls after pesticide exposure, although the variability was higher after exposure (Table 5-2). These effects are in accordance with the modes of action of the used insecticides. Both substances disturb neural signal regulation to a degree that neurological activity of nerves remains lastingly stimulated, which eventually leads to muscle spasms and paralysis. Chlorpyrifos does so by inactivating the enzyme that hydrolyses acetylcholine, and imidacloprid by activating nACh receptor. The more pronounced effects we found in the case of chlorpyrifos exposure, i.e. the increase in resting time coupled with a decrease in average step length, might be associated with the irreversibility of the enzyme activation, while the nAChR stimulation through imidacloprid is reversible. The reduced step lengths and changes in resting behaviour indicate that muscle malfunction may have set in already at the time of observation. The increased variability of turning angles can be explained by either muscular impairment or additional neurological effects affecting the individuals' ability to navigate. Based on a study by Azevedo-Pereira et al. (2011) we would speculate to find effects of exposure to chlorpyrifos and imidacloprid to converge further after an extended exposure duration or at increased concentrations. In their study, Azevedo-Pereira et al. (2011) measured AChE activity along with behavioural endpoints after exposure of *Chironomus riparius* larvae to imidacloprid and found that AChE activity also decreased with increasing concentration after 96 hours of exposure onward. The chain of physiological effects of AChE inhibition in *Asellus*, respectively, would lead to a decrease in overall activity as would be the case after exposure to chlorpyrifos, which directly inhibits AChE activity.

Dose-response or population density related effects were less conclusive in our study. While at the higher concentrations, the higher population densities appear to incite higher activity and slightly larger step lengths, compared to their single-individual equivalents, no such pattern could be identified for the lower concentration treatments. This aspect, together with the high individual variability in behaviour only demonstrates that more research is needed fully understand the sublethal impacts of pesticide exposure on ecologically relevant functions. Eventually, reduced locomotion is likely to interfere with foraging activities as observed by Agatz et al. (2014) in the case of Gammarids. Decreased energy available from feeding and increased energy expenditure for internal repair mechanisms, in turn, may lead to reduced growth and mating (Martin *et al.* 2012).

In our study, the impact on organisms exposed to imidacloprid may be less drastic compared to chlorpyrifos due to the higher safety factor that we assumed. However, the significance of pesticide exposure becomes clearer, when seen in comparison to the non-pesticide treatments. The presence of food slightly lowered the dispersal potential by affecting orientation moments and variation of turning angles, indicating that the animals were indeed adjusting their searching efficiency. Shelter items on the other hand caused a comparable reduction in dispersal. However, mechanistically it resulted from an effect on activity by reducing step lengths and increasing resting times. The presence of conspecifics affected reorientation less as could probably be expected than that it increased resting times in most cases, respectively reducing overall dispersal. The differences between the fed and starved control groups, however, indicate that the feeding state could potentially change this and reduce the need of shelter availability.

To improve the risk level estimation of chemical exposure on aquatic arthropod populations in an ecologically more meaningful way, ecological effect models can be applied that integrate different habitat, species, and exposure related information to assess population recovery timeframes (Galic *et al.* 2013; Focks *et al.* 2014a). Accounting for immigrating and emigrating individuals can help to further the mechanistic understanding derived from such modelling studies (Van den Brink *et al.* 2013; Hommen *et al.* 2016). The simplified dispersal estimation via the correlated random walk framework as part of this study failed to capture the underlying correlations between turning angles and step lengths, as well as the autocorrelation structures of either of these two parameters. Westerberg et al. (2008) studied the effects of population density and food availability on collembola described a similar phenomenon. The mechanistic links of the *Asellus* decision making remain to be elaborated for a better model parameterization. Aggregating the step length data may be one of those approaches to eliminate the CRW assumption of non-autocorrelated steps. The high variability of individual behaviour expressions is another factor that complicates simple modelling approaches, although it is an often observed factor in observational studies (Seuront *et al.* 2004b; Nørum *et al.* 2010). Hawkes (2009) consequently propose to account explicitly for this variability when designing models of habitat use and dispersal, respectively, an approach that is ignored by the application of simple average values in our study. Integrating findings such as ours into a more complex model can facilitate a better understanding of the complex interactions of chemical exposure and resource availability and their impacts on population recovery times, allowing also for the study of long-term impacts of exposure events.

Table 5-3: Literature survey of studies investigating the influence of chemicals and/or environmental conditions on aquatic macroinvertebrate locomotion in the laboratory.

Observational method	Species	Experimental dimension	Variable	Movement related metrics	Reference
Camera	*Asellus aquaticus, Gammarus pulex*	Aquaria (100 L)	Population density	Speed, turning angles, fractal dimension	Augusiak and Van den Brink 2015
	Acilius sulcatus	Aquaria (100 L)	Kairomones	Distance	Åbjörnsson et al. 1997
	Balanus amphitrite	Petri dishes	Various antifouling biocides, Heavy metals, Neurotoxic pesticides	Swimming speed	Faimali et al. 2006
	Brachionus calyciflorus	Glass chamber	Copper, Pentachlorophenol (PCP), Lindane	Speed, sinuosity	Charoy and Janssen 1999
			Food presence, nutritive state		Charoy 1995
			Copper, Pentachlorophenol (PCP), Lindane, 3,4-dichloroaniline		Charoy et al. 1995
		Well-plates	Dimethoate	Speed, sinuosity, turning angles	Guo et al. 2012
	Brachionus calyciflorus, Asplanchna brightwelli	Well-plates	Dimethoate	Speed	Chen et al. 2014
	Brachionus plicatilis, Artemia sp.	Petri dishes, well-plates	Zinc pyrithione, Macrotrol® mt-200, Eserine	Speed	Garaventa et al. 2010
	Daphnia pulex	Exposure cells (20 mL)	Isopropanol, Ethanol, Caffeine, Imidacloprid, Sertraline, Copper sulfate, Fipronil, Carbofuran, Esfenvalerate, Cypermethrin, Abamectin, Trichlorfon	Speed, turning angles, activity	Chevalier et al. 2015
		Beaker (200 mL)	Carbaryl, Kairomones	Speed, turning angles, diel movement	Dodson et al. 1995
		Well-plates	Chlorpyrifos, Nicotine, Physostigmine	Distance, turning angles	Zein et al. 2014
	Eurytemora affinis	Beaker (200 mL)	Nonylphenols	Speed	Cailleaud et al. 2011

Table 5-3 (continued)

Observational method	Species	Experimental dimension	Variable	Movement related metrics	Reference
Camera	*Gammarus pulex*	Petri dishes, stream mesocosms	Lambda-cyhalothrin	Speed, activity, drift	Nørum et al. 2010
		Petri dishes	Cypermethrin	Speed, activity	Nørum et al. 2011
	Litopenaeus vannamei	Aquaria (7 L)	Methamidophos	Activity, qual. Observations	García-de la Parra et al. 2006
	Oncaea venusta	Small plastic tanks	Inherent individual variability	Speed, distance	Seuront et al. 2004
	Rana temporaria tadpoles	Small plastic tanks	Endosulfan	Speed, activity	Denoël et al. 2013
Multispecies Freshwater Biomonitor	*Chironomus larvae*	Beaker (ca 200 mL)	Imidacloprid	Ventilation, activity	Azevedo-Pereira et al. 2011
	Daphnia magna		Dipterex, Malathion, Parathion, Dimethyl sulfoxide	Motility	Ren et al. 2007
			Dichlorvos, Malathion, Parathion, Methyl parathion		Ren et al. 2008
	Gammarus pulex		Pharmaceuticals	Ventilation, activity	De Lange et al. 2006a De Lange et al. 2009
			Time of day	Ventilation, activity	Peeters et al. 2009
	Echinogammarus meridionalis, Hydropsyche pellucidula, Choroterpes picteti		Acidic mine drainage	Ventilation, activity	Macedo-Sousa et al. 2008
Visual inspection	*Asellus aquaticus, Dendrocoelum lacteum*	Crystallization dishes (500 mL)	Tebuconazole, Lambda-cyhalothrin	Activity, predator-prey interaction	Bundschuh et al. 2012
	Asellus aquaticus, Gammarus pulex	Aquaria (1.5 L)	Polycyclic aromatic hydrocarbons	Avoidance	De Lange et al. 2006
	Brachionus calyciflorus	Glass chamber	Copper, Pentachlorophenol (PCP), Lindane, 3,4-dichloroaniline	Distance walked	Janssen et al. 1994
	Chaoborus flavicans larvae	Aquaria (12 L)	Kairomones	Height in water column	Dawidowicz et al. 1990
	Rana catesbeiana tadpoles, Rana septentrionalis tadpoles	Aquaria (15 L)	Kairomones	Mobility	Ferland-Raymond et al. 2010

6 EFFECTS OF EXTERNALLY INDUCED CHANGES OF DISPERSAL ON AQUATIC INVERTEBRATE RECOVERY AFTER PESTICIDE EXPOSURE

Jacqueline Augusiak, Andreas Focks, Hans Baveco, Paul J. Van den Brink

Abstract

Ecological effect models are nowadays often valued as tools for estimating the environmental risks associated with the use of chemicals or profound landscape management actions. They are also promoted as ways to improve the environmental risk assessment of pesticides in the EU, where their use is underrepresented to date. To meet the protection goal of prospective risk assessments of pesticides, which is set at the population level, ecological can bridge the gap between individual-level experiments and the population-level on larger spatial and temporal scales.

As such, considering dispersal in such extrapolations is necessary. In this study we revisit the MASTEP model, an individual-based model, used to simulate the population response of aquatic arthropods after an exposure event. We refined the dispersal related submodel based on empirical findings that were obtained from experiments specifically designed for this purpose. We compare the recovery time estimations from previous versions with the updated adaptation. Moreover, we were able to integrate an increased level of environmental complexity that could not be addressed before due to a lack of data. We subsequently implemented density dependence of movement rates, heterogeneously spread food availability and assumptions concerning restrained movement rates for life stages when younger individuals cannot cover as much distance as adults may.

Compared to former versions of the population models, recovery times did not change significantly for the investigated species, the water louse *Asellus aquaticus*, when "adult only" movement was assumed for all simulated individuals. This indicates that the previous assumptions yielded robust estimations. Accounting for life stage dependent movement restraints prolonged recovery when exposure was assumed to occur shortly before a reproduction cycle.

Based on these findings, we conclude that an increase of assumptions regarding environmental complexity needs to be judged carefully on a case-by-case basis. Increased realism in models can introduce an unwarranted increase in model complexity and uncertainty, which is not always supporting an improved credibility level of a model. However, such judgments can be only done in the light of available data, which are not always available or impossible to gather.

6.1 Introduction

Ecological effect models are nowadays often valued as tools for estimating the environmental risks associated with the use of chemicals or profound landscape management actions. As such, these types of models have received an increased interest over the past decade in environmental risk assessments of pesticides (Grimm *et al.* 2009; Hommen *et al.* 2016). Pesticide risk assessments performed in Europe, for example, follow a tiered approach where simple toxicity and worst-case scenarios are assessed in the lower tiers. If a pesticide fails at this level, higher tier assessments can be performed that address the points of concern in ecologically more realistic contexts (Brock *et al.* 2006). At this stage, ecological effect models can help to extrapolate the effects observed in the lower tiers to larger spatial and/or temporal scales (Forbes *et al.* 2011). However, ecological models also regularly face scepticism due to their inherent simplification of real-world processes and assumptions made in this regard. Different degrees in model complexity and uncertainty about appropriate scenario selection do not make it easier for authorities to decide whether to accept a model and the information it provides to help with the decision-making (Wang & Luttik 2012; Dohmen *et al.* 2015).

Nevertheless, the protection goal of prospective risk assessments of pesticides is set at the population level rather than the individual level. The necessary extrapolation from lab based experiments or studies performed under semi-natural conditions to ecologically more relevant spatio-temporal scales via models can introduce uncertainty if the implemented processes are based on insufficient understanding and knowledge of the system to be modelled. Conroy et al. (2010) respectively identified potentially incorrect choices of model structures and parameter estimates as sources of unreliable model output. Due to the number of aspects such as, among others, population dynamics, dispersal and density-dependence that need to be integrated in spatially explicit population models, errors can accumulate in a non-additive manner.

An example of a model integrating such a set of information is MASTEP (Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides) developed by Van den Brink et al. (2007). MASTEP is an individual-based model that simulates populations of the water louse *Asellus aquaticus* in aquatic landscapes taking into account mortality processes, life history traits, dispersal and density dependence of population regulation. The model's primary aim is to estimate the effects of pesticide exposure on populations of aquatic macroinvertebrates and their recovery therefrom. In the past, the model received criticism regarding the uncertainties attached to the implementation of density-dependence of singular life history and mortality parameters as well as the way with which dispersal was estimated and integrated (P. van den Brink, A. Focks, N. Galic*, personal communication*).

In this study, we address concerns related to the dispersal submodel incorporated in MASTEP to exemplify measures that can be taken to improve model credibility and

acceptance. Van den Brink et al. (2007) estimated daily movement rates with the help of a separate random walk model for which parameters were estimated from experimental results. Englund and Hambäck (2004) describe the step length frequency distribution of individuals of *A. aquaticus* moving within a 30 cm wide channel. The step lengths where measured only along one dimension where the majority of steps fall into a range between 0 and 4 cm. This can lead to an underestimation of step lengths when considering a two dimensional pane. Van den Brink et al. (2007) used a spatially unrestricting random walk model to estimate a correction factor. No estimations were made on the daily step frequency in the study by Englund and Hambäck (2004) and Van den Brink et al. (2007) subsequently presumed the daily activity of individuals to be limited to one third of the day. Simulated individuals were assumed to be resting, or at least not dispersing, for the remaining time of the day. Based on these assumptions, a time distribution was estimated to describe how long an individual may remain within a 1 m^2 grid which was ultimately used in the population model. Galic et al. (2012), who adapted MASTEP, used a slightly different approach to overcome the different time scales used in the population model (days) and at which movement occurs (minutes or seconds). They still simulated dispersal separately but instead of residence times, they estimated how far individuals could disperse within 24 hours based on the above-mentioned assumptions. Those estimations were summarized in a probability distribution, which in turn was used in the population model to describe the likely displacement of individuals at each time step. Moreover, Galic et al. (2012) implemented the model in a programming environment that explicitly facilitates the interaction of individual entities with each other and with their environment. However, they did not explore the effects of behaviour changing due to local environmental factors that may have attracting or repelling influences on the movement behaviour.

This "double-estimation" and the lack of more informative data aroused doubts whether the obtained dispersal rates were representative or realistic. Since dispersal plays a central role in MASTEP due to the *Asellus* elsewise very low recovery potential, this was one of the central discussion points for the model's acceptance in regulatory contexts. To this end, we performed experiments to study the movement behaviour of *Asellus aquaticus* in the laboratory. The experimental design was inspired by the model dimensions and scales applied in MASTEP and took possible behavioural changes due to population density, sublethal pesticide exposure, and environmental cues into account (Augusiak & Van den Brink 2015, 2016). Parameters for the movement behaviour could be extracted from the observed paths, which were then used to obtain daily dispersal estimates for *Asellus*.

In this study, we revisit a MASTEP version developed by Galic et al. (2012) aiming 1) to compare the recovery times obtained by Galic et al. (2012) with those obtained based on the movement patterns described by Augusiak and Van den Brink et al. (2015) and 2) to study the influence of animal density, being fed or exposed to the insecticides chlorpyrifos or imidacloprid prior the movement observation and the presence of food and shelter during the observation and exposure as studied by (Augusiak & Van den Brink 2016) on the recovery times after different mortality events. We further tested how much model

complexity needs to be implemented to capture the population dynamics after a stress event such as pesticide exposure. To achieve this, we implement a degree of landscape heterogeneity in the MASTEP version used by Galic et al. (2012) and compare whether the previously chosen approach of assuming homogeneous conditions within the aquatic habitat yields recovery times different from our version and whether recovery patterns vary between both approaches.

6.2 Materials & Methods

6.2.1 Model species

The water louse *Asellus aquaticus* is a common species in European freshwater habitats. The species' population dynamics are mostly determined by climate and reproduction can range from univoltine in Northern Europe to bivoltine in central, or constant in southern Europe (Tadini *et al.* 1988). Our work focuses on Northwestern to Central European populations, which usually exhibit a bivoltine reproduction pattern (Økland 1978). *A. aquaticus* follows a purely aquatic life cycle and the spring generation produced by the overwintering population normally shows lower abundances than the summer generation. Asellids are detritivorous crustaceans with a relatively low dispersal potential (Moon 1968), and together with its limited reproductive capacity this makes it a species with a low potential for population recovery after a stress event. Rico and Van den Brink (2015) also found the species to be relatively sensitive to pesticide exposure, which makes it a suitable model organism for conservative risk assessments of pesticides.

We used an adapted version of the MASTEP model (Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides) developed by Galic et al. (2012), which they used to explore the influences of exposure timing throughout a year on population recovery, while Focks et al. (2014) applied it to investigate the impact of heterogeneous pesticide exposure on an upscaled landscape level on the same endpoint. Both versions relied on assuming homogeneous landscape features and behavioural patterns that were derived from experiments by Englund and Hambäck (2004) who determined step lengths one-dimensionally in a channel that was 30 cm wide and 1 m long. Prior to the modelling exercise, we thus performed experiments in microcosms of 1 by 1 m to observe the movement behaviour of asellids under different environmental conditions such as food or shelter items being present or not, sublethal pesticide exposure, or varying population densities (Augusiak & Van den Brink 2015, 2016). From the recorded animal paths, basic movement parameters were extracted and used to parameterize a simple random walk model, which we used to determine the dispersal potential, i.e. distributions of daily dispersal distances, under the respective conditions. These dispersal distributions are used as input in MASTEP, enabling the incorporation of landscape heterogeneity by assigning a certain distribution to a patch representing given environmental circumstances.

We adapted the IBM used by Galic et al. (2012). The model was mostly retained as described, only the submodel concerning dispersal was adjusted to rely on drawing parameter values from experimental observations by (Augusiak & Van den Brink 2016) instead of probability distributions. In the following, we provide a brief overview of the modelled life history of *A. aquaticus*, following the ODD protocol for describing individual- and agent-based models is provided (Grimm et al. 2006, 2010), while a full description of the model assumptions and structure are in Appendix 4 together with more information on the species life-history and data obtained from literature.

6.2.2 Model purpose

The model's purpose is to assess the recovery of a population of *A. aquaticus* after pesticide exposure, by accounting for the influence of environmental cues on dispersal potential and the timing of stress events throughout the year.

6.2.3 Entities, state variables, and scales

Entities in the model are individual female asellids, and square patches forming the landscape. Individuals are distinguished between juvenile and adult life stages and their state variables are individuals' age [days], size [length in mm], hatching date [day in the calendar year], reproductive state, lifespan [days] and location [continuous X and Y coordinates] in the habitat.

The simulated landscape comprises a ditch consisting of a string of 100 patches representing aquatic habitat (Figure 6-1) on a square grid with periodic boundaries, i.e. the grid forms a torus avoiding edge effects. The state variables of patches contain the local mortality of *Asellus* induced by chemical stress and density dependence. Each patch represents 1 m^2. Each time step represents one day of a calendar year consisting of 365 days. Simulations start on day 0 (Jan 1) und end after six years or when the modelled population goes extinct. The first two modelled years are not considered for further analysis to avoid transitional effects during the model initialisation phase. The model is implemented in the NetLogo platform v. 5.2 (Wilensky 1999; downloadable from http://ccl.northwestern.edu/netlogo).

Figure 6-1: A picture representing the ditch habitat in the model. The ditch consists of a string of 100 aquatic (blue) patches. The light blue patch is the locally observed patch (1 m^2); it does not differ from other aquatic patches in any other way.

6.2.4 Process overview and scheduling

Processes in the model are mortality, movement, growth, and reproduction, which are scheduled for individuals in a randomized sequence. Every time step each individual ages and its mortality probability is calculated, the surviving individuals change position in the modelled habitat, followed by juvenile and adult growth. Finally, individual's mortality probability increases if exposed to stress. All individual state variables are updated immediately (asynchronous updating). Survival, growth and movement functions are constant over the whole year.

Mortality consists of background, density-dependent and chemical stress imposed mortality. Background mortality was implemented by assigning individuals different lifespans (in days) at birth; lifespans were exponentially distributed with a mean of 90 days (Table 6-1). This resulted in 1% of individuals from the initial modelled population surviving more than 400 days, while it has been found that *Asellus* can survive up to 600 days according to (Vitagliano *et al.* 1991), thus rendering our estimate conservative. Density-dependent mortality assumes a negative effect of local densities on individual survival; whereas exposure to stress in certain times of year increases individual mortality probability. Since short-term, spatially correlated exposure to pesticides is most common in habitats adjacent to agricultural fields, we imposed the chemical stress for only one day, i.e. patch regeneration was almost instantaneous.

Individuals grow deterministically following a von Bertalanffy growth function (Von Bertalanffy 1957). The maximum growth rate is affected by local density, i.e. the density in each patch. This is based on observations and measurements by Hynes & Williams (1965) who experimentally showed that water lice populations produce more offspring when housed in larger jars, even though the amount of food in their experiments was the same.

In order to mimic the observed bivoltine *reproductive pattern* in North-western Europe we introduced two periods in a year when individuals are able to reproduce. The reproductive periods start in the beginning of May and mid-July, and last four and six weeks (Table 6-1), respectively. In these reproductive periods each individual female releases its offspring once, and the maximum realized fecundity is set to 100 juveniles (Tolba & Holdich 1981). Individual fecundity is positively correlated with the size (length) of the female at the time of release (Chambers 1977), resulting in bigger females releasing more juveniles.

Movement is modelled through a distribution of distances each individual can cover in one day that was obtained via a separate movement model. In the separate movement model, simulations of a random walk were carried out for a large number of individuals. Parameters for the basic parameters characterising this type of movement representation were comprised of the combination of step lengths and turning angles as well as general movement activity. We extracted these parameters from paths that we observed in experimental setups designed to test the influences of different environmental factors on the movement behaviour of *Asellus* (Augusiak & Van den Brink 2015, 2016). The simulations

of daily movement resulted in normal distributions of distances each individual can move to from its initial position in the modelled habitat. We obtained different distributions per experimental setup, which were used as input for the MASTEP population model (see Appendix 4 and Appendix 5 for details). For comparisons with the original model design by (Galic *et al.* 2012), we furthermore made use of the parameters used by them.

Table 6-1: List of model parameters.

Submodel	Parameter	Distribution	Value	Unit	Reference
Habitat	System carrying capacity, K	Constant	10 000	ind	Constant K per patch (growth related)
Mortality	Lifespan	Exponential	Mean 90	day	adapted from Vitagliano (1991)
	μ1	Constant	0.001		Based on Van den Brink et al. (2007)
Reproduction	Onset of 1st reproductive cycle	Constant	120 (April/May)	day	mimicking western European conditions
	Onset of 2nd reproductive cycle	Constant	200 (July)	day	
	Maximum clutch size	Constant	100	ind	Adapted from Tolba and Holdich (1981)
	Length of 1st reproduction period	Uniform	1 to 28	day	Adapted from Chambers (1977)
	Length of 2nd reproduction period	Uniform	1 to 45	day	Adapted from Chambers (1977)
Growth	Maximum size	Constant	12	mm	Økland (1978), Marcus *et al.* (1978), Arakelova (2001)
	Minimum size	Normal	Mean 1, SD 0.2	mm	Adcock (1979)
	Kappa, k	Constant	0.02	/day	determined by calibration
	Age at maturity	Constant	45	day	(Williams 1962)
	Density-dependent factor, *y*	Constant	1		Estimation by Galic *et al.* (2012)
Dispersal	benchmark	Normal	Mean 0.004, SD 4.444	m	Estimation by Galic *et al.* (2012)
	CPF low - 1	Normal	Mean -0.08, SD 2.07	m	Own simulation (more in Dispersal simulations)
	CPF low - 50	Normal	Mean 0.01, SD 1.71	m	Own simulation (more in Dispersal simulations)
	CPF high - 1	Normal	Mean -0.06, SD 1.29	m	Own simulation (more in Dispersal simulations)
	CPF high - 50	Normal	Mean -0.06, SD 2.07	m	Own simulation (more in Dispersal simulations)
	Imi low - 1	Normal	Mean -0.21, SD 3.14	m	Own simulation (more in Dispersal simulations)
	Imi low - 50	Normal	Mean -0.12, SD 2.41	m	Own simulation (more in Dispersal simulations)
	Imi high - 1	Normal	Mean -0.01, SD 1.85	m	Own simulation (more in Dispersal simulations)
	Imi high - 50	Normal	Mean 0.09, SD 2.38	m	Own simulation (more in Dispersal simulations)
	Control (starved) - 1	Normal	Mean -0.09, SD 4.09	m	Own simulation (more in Dispersal simulations)
	Control (starved) - 50	Normal	Mean -0.02, SD 3.86	m	Own simulation (more in Dispersal simulations)
	Control (fed) - 1	Normal	Mean -0.06, SD 3.41	m	Own simulation (more in Dispersal simulations)
	Control (fed) - 50	Normal	Mean -0.04, SD 2.39	m	Own simulation (more in Dispersal simulations)
	Food - 1	Normal	Mean -0.02, SD 2.99	m	Own simulation (more in Dispersal simulations)
	Food - 50	Normal	Mean 0.12, SD 2.15	m	Own simulation (more in Dispersal simulations)
	Shelter - 1	Normal	Mean -0.11, SD 2.82	m	Own simulation (more in Dispersal simulations)
	Shelter - 50	Normal	Mean 0.08, SD 1.93	m	Own simulation (more in Dispersal simulations)

6.2.5 Control scenario

Populations in the control scenario were not exposed to any stress; they were, otherwise, identical to treatment scenarios.

6.2.6 Exposure timing scenario

We followed the study of the effect of timing of stress on the recovery of *Asellus* by Galic *et al.* (2012) and chose the following time points for a hypothetical pesticide exposure: 1) just before (Julian day 110) and 2) after the first generation (Julian day 160), and 3) just before the second generation peak (Julian day 210) and 4) after the reproductive season (Julian day 260). The chosen schedule is consistent with insecticide applications in NW Europe, where the first applications of insecticides typically start between March and April, and while applications in autumn do not occur as frequently, some crops e.g. bulbs, fruits and vegetables are still treated in October (CBS, 2008). In the model, the toxic effect lasts for one day, and increases the mortality probability of each individual in exposed patches. Stress related mortality probability was set either to 0.5 or to 1 (details below).

6.2.7 Spatial exposure scenarios

We simulated three exposure scenarios: a) all 100 aquatic patches were exposed to a mortality probability of 0.5 per individual for one day (named hereafter "homogeneous exposure scenario"); b) only 50 connected aquatic patches exposed to a mortality probability of 1.0 per individual for one day (hereafter "heterogeneous exposure scenario"); c.) only 50 random aquatic patches exposed to a mortality probability of 1.0 per individual for one day (hereafter "patchy exposure scenario"). All of these scenarios resulted in, on average, 50% of the population dying due to inflicted stress, so the differences in recovery speed were solely caused by the distribution of the stress event itself.

6.2.8 Test scenarios - Habitat heterogeneity and population density effects

To understand how the estimation of recovery times would change due to implementing the information from our movement observations (Augusiak & Van den Brink 2015, 2016), we subsequently increased the complexity of the individual dispersal behaviour in four steps. The different scenarios are listed in Table 6-2. As a first step, we simulated the above mentioned scenarios by simply exchanging the daily dispersal distribution from Galic *et al.* (2012) with those obtained for the different experimental test regimes, which we call *benchmark scenario*.

As a following measure we aggregated the experiments performed at *different population densities*, by counting the number of conspecifics in a model patch before the next movement takes place. If the number was above a given threshold, the following dispersal distance was drawn from the dispersal distribution obtained with the higher density

experiment. Respectively, if the number of surrounding *Asellus* was below the threshold, a distance was drawn from the dispersal distribution representing the single-individual experiment.

Table 6-2: Overview of model experiments.

Scenario	Dataset used	Landscape	Exposure day	Pesticide toxicity	Exposure [% of landscape]
Step 1: *benchmark scenario*	benchmark	Ditch - homogeneous conditions	110	LC50	100
	CPF low - 1		160		
	CPF low - 50		210	LC100	50 (random patches)
	CPF high - 1		260		50 (connected patches)
	CPF high - 50				
	Imi low - 1				
	Imi low - 50				
	Imi high - 1				
	Imi high - 50				
	Control (starved) - 1				
	Control (starved) - 50				
	Control (fed) - 1				
	Control (fed) - 50				
	Food - 1				
	Food - 50				
	Shelter - 1				
	Shelter - 50				
Step 2: *aggregated population density scenario* (if > 50 individuals in patch, draw dispersal distance from higher population dataset)	CPF low	Ditch - homogeneous conditions	110	LC50	100
	CPF high		160		
	Imi low		210	LC100	50 (random patches)
	Imi high		260		50 (connected patches)
	Control (starved)				
	Control (fed)				
	Food				
	Shelter				
Step 3: *patch heterogeneity scenario* (if > 50 individuals in patch, draw dispersal distance from higher population dataset)	Control (fed), on empty patch for < 24 h	Ditch - 50% with food, 50% empty	110	LC50	100
	Control(starved), on patch for > 24 h		160		
	Food, on patch with food		210	LC100	50 (random patches)
			260		50 (connected patches)
Step 4: *life stage scenario*	juveniles can make steps, which are 10% of the adult step lengths	Respectively applied to Steps 1 to 3.			

Next, we increased model complexity by introducing *patch heterogeneity*. We assumed 50 of the aquatic patches to form empty habitat without food or shelter, which in nature may

be hard to find considering that even dredged ditches have sediments with organic particles, which may serve as food. For our model, however, we still chose to let the modelled *Asellus* draw their next movement distance from the distribution obtained from the control group that was fed until start of the experiments but else, no other options or factors besides population density were tested. We found during our experiments with sublethal pesticide exposure, where exposed individuals did not receive any food for 48 hours prior to the camera observations, that the respective control group was more active than the "fed" control group mentioned before. In the MASTEP model, the individuals would switch from the fed control dispersal distribution, to the starved control distribution if they did not enter a patch with food within 24 hours. The remaining 50 patches of the simulated ditch were assumed to contain food, and individuals on these patches would switch to drawing their next movement from a distribution that we obtained, when we had hung pieces of leaves into the water. The leaves came from the same habitat where we had collected the *Asellus* and were a known food source. By hanging them into the water phase, being out of reach for the observed individuals, we tested if *Asellus* would be able to "smell" the leaves and if they would change their movement accordingly.

Since Galic *et al.* (2012) found that reduced dispersal distances could impact the recovery times and considering that juveniles with a lower body size compared to adults may not be able to cover steps as big as adults, we tested a respective assumption in a third step, in the *life stage scenario*. In the random walk model, we allowed individuals to move following the same rules as in all other random walk models described in Appendix 5. However, step lengths were assumed to be only 10% of what the observed adults displayed in the experiments. In the population model, juvenile and adult life stages thus drew their following daily distance from the respective distributions, including the density assumption described for the previous step. This assumption was applied to all three above-mentioned scenarios.

6.2.9 Scenario analysis

The outputs from the treatment scenarios were compared to the control in the corresponding landscape. For the analysis of recovery times, we used 20 replicate simulations for each of the treatments including the control. We compared daily abundances in controls and treated populations and considered a treated population recovered if its abundance reached 95% of the abundance of the control population. Both in real aquatic systems and this model there is stochastic variation of abundances at the local scale, which may mask recovery or lack thereof, for instance, as water louse move from patch to patch that may lead to a very short-term recovery of that patch as the abundance would decrease again as soon as the water louse moved on. Consequently, we defined recovery as reaching the 95% or higher abundances than those of control populations for five days out of a period of ten days (Galic *et al.* 2012). Once this condition was met, the day of recovery was considered to be in the middle of this period.

Daily population abundances of each of the 20 treated replicates were compared with each of the 20 control replicates, yielding 400 estimations of recovery times per treatment. All resulting distributions of recovery times are presented in violin plots (Hintze & Nelson 1998) combined with boxplot representations. Violin plots are a combination of boxplots and kernel density plots, showing the probability density of data at different values. We compare median values in different scenarios, as medians are a more robust statistic for central tendency than means in non-normally distributed datasets.

6.3 Results & Discussion

The simulated populations followed a bivoltine reproduction pattern with distinct spring and summer generations as was expected based on the work by Galic *et al.* (2012). The median times population recovery in the different scenarios are presented in Table 6-3 and Figure 6-2 to 6-4.

6.3.1 Benchmark scenario

The results from the simulations using the parameters for the dispersal distribution by Galic *et al.* (2012) yielded recovery times comparable to their reported results in the homogeneous and heterogeneous exposure scenarios. For the homogeneous exposure scenario, our results varied by 4 to 8 days from Galic *et al.* In the case of the heterogeneous exposure scenario, our results deviated by no more than 4 days, except for the exposure in August, where median recovery required 46 days longer than reported by Galic *et al.* However, this is a time close to the second reproduction peak and the probabilistic assumptions concerning the reproduction, growth and mortality submodels may exert bigger effects at this stage.

Comparing "our" benchmark scenario with the scenarios where we made use of other dispersal distributions (Table 6-2: Step 1), we also find similar recovery times. Furthermore, after exchanging the original parameter set for the dispersal distributions determined with the help of our experimental findings (Augusiak & Van den Brink 2016), we found similar trends in recovery times as reported by Galic *et al.* (2012). The closer a reproductive period was, the shorter it took the populations to recovery. When pesticide exposure was simulated to occur in September, most populations required one or two reproduction periods before recovery was observed (Table 6-3). Galic *et al.* (2012) tested the influence of varying dispersal distribution parameters in a sensitivity analysis. During their tests, they found that reduced dispersal caused increased recovery times. Even though all mean 24-hour distances were higher than the one used by Galic *et al.* (2012), respective standard deviations were mostly clearly below their standard deviation of 4.44 m (Table 6-2).

Table 6-3: Median days to recovery in all exposure scenarios (total population) when assuming different dispersal distributions obtained from experiments performed under different conditions (Table 6-1). Applicable distributions were dependent on the population density in a patch (adult movement only considered as well as different dispersal distributions for adults and juveniles).

		adult movement only				life stage dependent movement			
	Spatial exposure scenario	**Exposure timing**				**Exposure timing**			
		April	**June**	**August**	**September**	**April**	**June**	**August**	**September**
benchmark	homogeneous	118	86	120	307				
	heterogeneous	131	112	218	311				
	patchy	119	83	152	316				
CPF low - 1	homogeneous	115	80	135	321	134	85	164	318
	heterogeneous	126	197	296	322	282	300	378	338
	patchy	118	81	141	313	131	171	201	306
CPF low - 50	homogeneous	118	85	136	312	136	196	192	319
	heterogeneous	162	215	292	327	311	325	384	345
	patchy	119	84	136	316	175	162	229	246
CPF high - 1	homogeneous	116	77	140	320	125	130	162	247
	heterogeneous	165	224	340	330	368	333	296	341
	patchy	122	79	103	308	189	164	273	333
CPF high - 50	homogeneous	123	94	98	286	141	101	174	251
	heterogeneous	128	159	277	310	224	273	283	310
	patchy	127	86	106	247	156	143	218	251
Imi low - 1	homogeneous	121	78	81	309	135	162	203	330
	heterogeneous	122	89	242	309	300	288	299	322
	patchy	116	85	93	307	182	143	238	322
Imi low - 50	homogeneous	130	106	131	304	151	153	195	331
	heterogeneous	133	125	278	309	377	332	381	355
	patchy	123	103	123	292	140	193	234	337

Table 6-3 (continued)

		adult movement only				life stage dependent movement			
	Spatial exposure scenario	**Exposure timing**				**Exposure timing**			
		April	**June**	**August**	**September**	**April**	**June**	**August**	**September**
Imi high - 1	homogeneous	116	97	128	321	152	108	165	337
	heterogeneous	157	170	338	322	251	334	383	347
	patchy	124	87	134	308	128	151	236	336
Imi high - 50	homogeneous	122	83	90	239	127	133	215	325
	heterogeneous	135	134	250	319	119	226	368	334
	patchy	122	86	105	312	138	176	202	327
Control (starved) - 1	homogeneous	125	108	93	316	131	110	156	309
	heterogeneous	123	98	211	307	148	187	277	331
	patchy	119	98	132	307	139	146	181	319
Control (starved) - 50	homogeneous	120	78	79	265	132	145	133	313
	heterogeneous	124	112	174	326	160	179	282	333
	patchy	122	78	108	309	150	150	227	328
Control (fed) - 1	homogeneous	115	88	143	310	140	129	134	309
	heterogeneous	124	121	278	321	231	310	283	341
	patchy	113	95	125	316	134	158	179	309
Control (fed) - 50	homogeneous	128	105	103	317	172	113	176	326
	heterogeneous	130	131	277	320	213	327	322	333
	patchy	121	115	132	255	169	163	262	338
Control (fed) - 100	homogeneous	118	82	81	252	129	92	165	308
	heterogeneous	120	99	242	318	155	274	280	329
	patchy	121	82	109	299	134	125	180	323
Control (fed) - 200	homogeneous	114	85	88	257	132	113	168	307
	heterogeneous	122	127	277	315	207	237	292	319
	patchy	110	98	123	253	134	127	175	313

Table 6-3 (continued)

		adult movement only				life stage dependent movement			
	Spatial exposure scenario	Exposure timing				Exposure timing			
		April	June	August	September	April	June	August	September
Food - 1	homogeneous	116	86	126	314	125	120	170	308
	heterogeneous	123	91	247	323	209	219	281	338
	patchy	121	85	100	317	127	140	140	259
Food - 50	homogeneous	121	86	117	308	134	91	159	322
	heterogeneous	135	155	278	325	214	327	284	330
	patchy	120	108	137	256	135	172	216	318
Shelter - 1	homogeneous	126	85	143	233	145	142	131	313
	heterogeneous	134	162	277	316	234	286	360	331
	patchy	126	109	135	303	187	145	191	327
Shelter - 50	homogeneous	119	93	101	312	162	127	180	248
	heterogeneous	131	191	282	318	302	285	357	323
	patchy	120	88	118	318	134	138	223	323
Light - marked	homogeneous	123	107	138	310	127	113	136	304
	heterogeneous	127	140	279	329	180	194	363	330
	patchy	118	86	134	312	134	152	156	277
Light - unmarked	homogeneous	127	123	134	314	131	143	190	313
	heterogeneous	126	168	292	325	173	265	278	320
	patchy	122	102	146	307	143	143	193	309

When we assumed juveniles to be capable of shorter steps compared to adults, median recovery times increased in the heterogeneous and patchy exposure scenarios, while recoveries in the homogeneous exposure scenario remained very similar to the previous assumption where all individuals were assumed to move according to the same rules (Table 6-3). In the latter case, the presence of surviving individuals in the entire ditch generally serves as a basis for recovery by reproduction, which was also an effect observed by Galic *et al.* (2012).

We also did not observe differences in recovery for any exposure scenario when exposure was imposed in September. The duration until the following reproduction period in April allows for a redistribution of surviving individuals while they also reach adulthood at the same time, thus providing similar conditions for reproduction and dispersal at that time. However, in cases where exposure was supposed to occur earlier in the year, the slower redistribution potential of juveniles exerted a clear effect on recovery times especially in the heterogeneous exposure scenario. The random distribution of exposed patches in the patchy exposure scenario benefitted from immigrants relocating from the unstressed patches in between. Surviving individuals thus remained more distributed throughout the ditch, in that sense resembling the homogeneous exposure scenario to some degree with its better foundation for recovery. In contrast, recovery due to immigration was more limited in the heterogeneous where an entire stretch of 50 patches was exposed.

Both the heterogeneous and patchy exposure scenario can occur in real settings depending on the nature of how a pesticide may enter the water body and how persistent it is. In cases of spray-drift, only small sections of a ditch may receive a dose while leaching or runoff can affect a wider stretch of a water body.

6.3.2 Exposure timing scenario

As observed by Galic *et al.* (2012) time to population recovery was shorter the closer the next reproductive period was. This was also true for our simulations, where the shortest recovery times were estimated to occur after exposures in April (Julian day 110) and August (Julian day 210), just 10 days before the onset of the first and second reproduction cycle, respectively. The longest recovery times were found when pesticide exposure was simulated to occur in September after the reproductive season ended. In the latter case, we also found a bimodal distribution of recovery times in most instances, indicating that some populations recovered within the same season, while other populations needed the following reproductive season to recover, both on local and total population scale (Figures 6-2 and 6-3; Appendix 3).

6.3.3 Spatial exposure scenarios

As mentioned above, in both the heterogeneous and patchy exposure scenarios, population recovery took longer than in the homogeneous exposure scenario, where survivors remained in all patches and could form a basis for recovery through reproduction. Respectively, recovery times were generally lower in the patchy exposure scenario compared to the heterogeneous exposure scenario (Table 6-3 and Figures 6-2 and 6-3). These patterns were found not only in the benchmark scenario but remained after increasing the model complexity during all steps listed in Table 6-2.

April
June
Homogeneous exposure
Heterogeneous exposure
Patchy exposure
Days to 95% recovery
CPF_low
CPF_high
IMI_low
IMI_high
Control_weed
Control
Food
Shelter

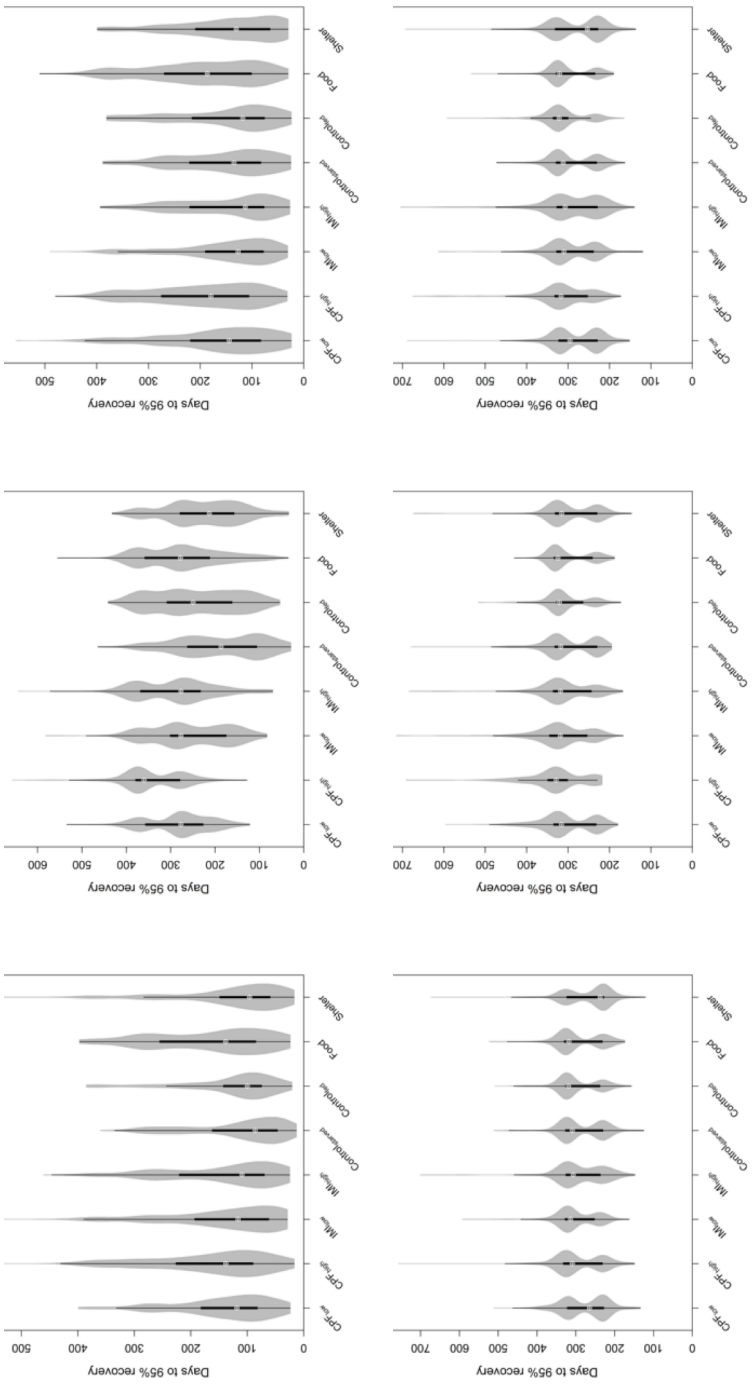


Figure 6-2: Time to population to recovery (in days) when assuming different dispersal distributions obtained from experiments performed under different conditions (Table 6-2, Step 2). Applicable distributions were dependent on the population density in a patch (adult movement only considered).

April
June
Homogeneous exposure
Heterogeneous exposure
Patchy exposure
Days to 95% recovery
CPF_low
CPF_high
IMI_low
IMI_high
Control_reveal
Control
Food
Shelter

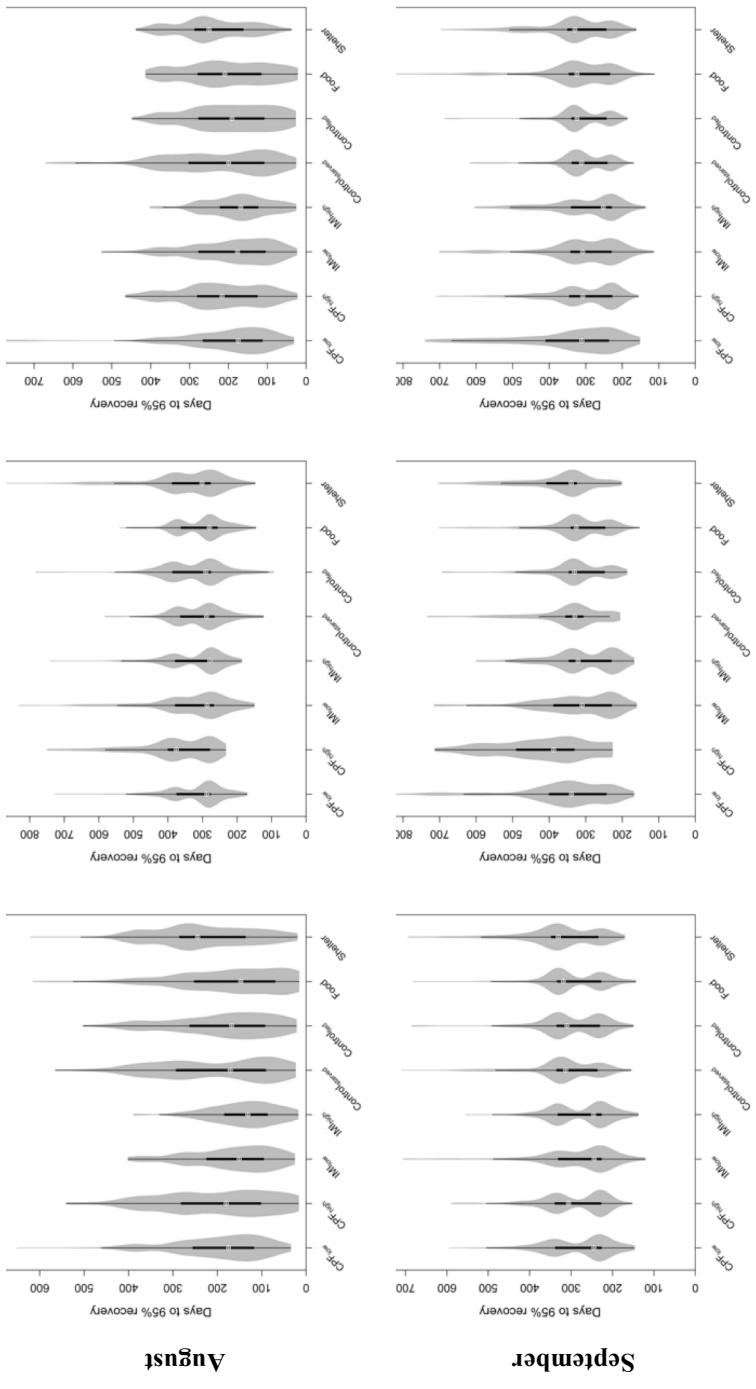


Figure 6-3: Time to population to recovery (in days) when assuming different dispersal distributions obtained from experiments performed under different conditions (Table 6-2, Steps 2 & 4). Applicable distributions were dependent on the population density in a patch (different movement for juveniles and adults considered).

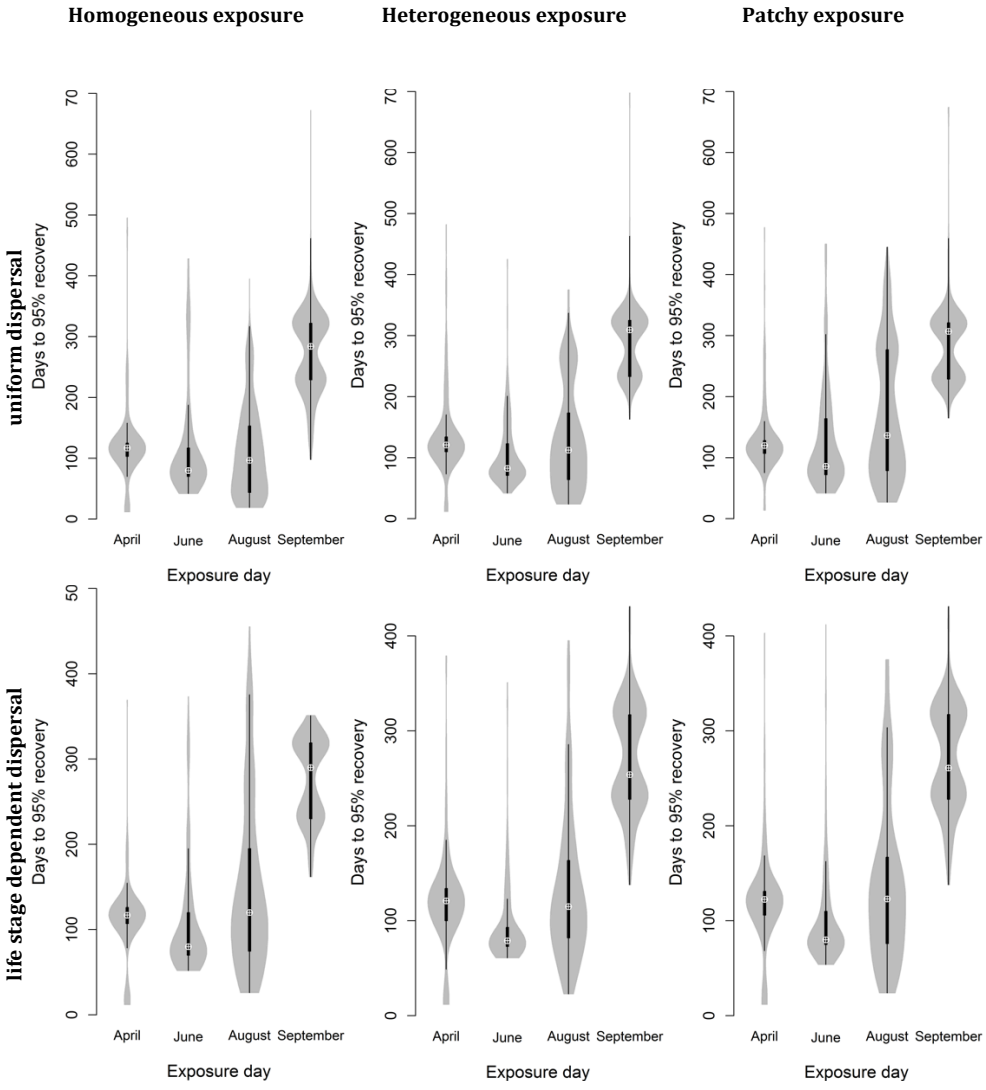


Figure 6-4: Time to population to recovery (in days) when assuming different dispersal distributions obtained from experiments performed under different conditions (Table 6-2, Step 3 and Step 3 & 4). Applicable distributions were dependent on the population density in a patch, whether food was available, and whether life stage dependent dispersal rates were included or not.

We also analysed recovery times for a single focal patch (results not shown). This patch was always exposed and in the middle of the exposed area in case of the heterogeneous and always exposed in the patchy exposure scenario. In our study, the corresponding *local* recoveries are shorter than the local recoveries reported by Galic *et al.* (2012), which is based on the fact that we focussed on a single 1 m^2 patch, while Galic *et al.* used a 10 m^2 stretch to define a local population. Recovery times of the local populations in our model were comparable to the homogeneous exposure scenario. In the heterogeneous exposure scenario, immigrating individuals first need to reach the exposed patch, which caused recoveries to take much longer. This need of dispersal to facilitate recovery also explains the more frequent occurrence of bi- and tri-modal recovery time distributions for the total population in the ditch in the heterogeneous scenario. Such distributions indicate that two or three reproduction periods were needed before recovery.

6.3.4 Environmental complexity

Population density. Using different dispersal distributions depending on the number of surrounding individuals in a patch, did not affect the previously described results in a marked way. Most dispersal distribution parameters based on the empirical data were quite similar and varied most strongly in terms of standard deviation, which in some instances changed the form of the recovery time distribution but not the corresponding median (Figure 6-2).

Life stage dependent dispersal. In contrast to the low impact of density dependent dispersal distributions, assuming a impaired movement pattern for juveniles based on their smaller body size, however, did cause an increase of population recovery times after pesticide applications in June and August (Table 6-3). Recovery from pesticide-introduced mortality was little affected when the exposure was imposed before the onset of the first reproductive cycle in May or after the reproductive season ends. In the first case, hardly any juveniles are present in the simulated ditch and all individuals in the system move according to the same rules at the time of exposure. In the latter case (exposure in September), differences between the exposure scenarios are masked by the time until the next reproductive season starts. During this time, individuals have time to redistribute in the patch as discussed earlier.

Landscape heterogeneity. Including different dispersal behaviours based on the patch quality, i.e. food present or not, did not change the overall patterns to population recovery neither for the "adult only" dispersal scenarios nor the life stage dependent dispersal scenarios. For this slowly dispersing species, intrinsic recovery potential rooted in the species' bivoltinism appears to take precedence unless longer stretches need to be repopulated again as observed earlier.

6.4 General discussion & Conclusions

The estimation of population recovery times can lend crucial information in the higher tiers of environmental risk assessments of pesticides, when the recovery principle is followed (Hommen *et al.* 2010). Recovery can be difficult to determine in experimental studies. Even under semi-natural conditions in mesocosms, the required spatial scales may not be sufficient or the study duration may be too short to allow focal populations. Particularly species with low reproductive potential may require long periods of time to recover from stress events (Niemi *et al.* 1990; Whiles & Wallace 1992). This is, where ecological effect models can support an improved understanding of a system's response to stress such as pesticide exposure (Forbes *et al.* 2011). Due to their inherent uncertainties, however, decision- and policy-makers are regularly hesitant to accept added information from modelling studies (Hunka *et al.* 2013a).

In our study, we addressed concerns regarding a population model that was developed to estimate population recovery times for the water louse *Asellus aquaticus*. After using the experimental data to parameterize the dispersal submodel in MASTEP, we performed modelling experiments following a study by Galic *et al.* (2012). Most conclusions drawn by Galic *et al.* still hold true. While the overall patterns strongly agreed between theirs and our studies, some recovery time estimates may require reconsideration given the additional information that we were able to include in the model based on the prior experiments. This was expected, as Galic *et al.* (2012) found recovery times to be relatively robust to varying dispersal distributions, unless dispersal was strongly decreased. This shows that recoveries of populations for this particularly species, which is a slow disperser, are mostly driven by reproduction and, to a lesser degree, landscape connectivity and passive movement, i.e. drift.

These findings might not hold for stronger dispersers who can cover larger distances within a day. Under such circumstances, landscape heterogeneity might form a more relevant aspect with stronger implications on recovery time estimates. Rico and Van den Brink (2015), for example, used the trait of dispersal strength to evaluate a species' vulnerability towards insecticides. The further a species can disperse, the more likely it is to contribute to extrinsic recovery through recolonization of exposed areas or by escape. We did not investigate heterogeneity configuration though. In a study about invertebrate dispersal and recolonization processes in marine environments, Boström et al. (2010) observed that the centres of vegetated areas took longer until they were repopulated compared to seagrass meadows or seagrass-sand interfaces. For *Asellus*, high degrees of vegetation may also affect real dispersal more significantly than what we observed in our study here. Our movement data from experiments with food sources present did affect the searching behaviour of observed individuals (Augusiak & Van den Brink 2016). However, we did not provide any additional structures in the arena at that time, thus allowing the animals to roam freely. At the same time, we found that resting times increased when abiotic structures were introduced in a similar setup. Vegetation in a ditch may not only serve as a

food source but shelter as well. An increase in resting time and relative surface area may consequently reduce real dispersal distances more strongly.

The applied two-step process of aggregating the experimental information from a fine time scale to a coarser 24-hour resolution may work in this kind of modelling approach. However, deriving daily dispersal from the experimental location data is a crucial element and may still lead to erroneous conclusions. Westerberg et al. (2008) showed that the correlation between step lengths and turning angles, as well as the autocorrelation among those parameters themselves, can yield wrong dispersal estimates if not taken into account appropriately. We found that step lengths and turning angles were strongly correlated in our study, and respectively chose to use fixed pairs from the experimental data sets. Randomly drawing either parameter from the experimental values would not have made a big difference in the dispersal distribution estimation (Appendix 5), but it would have introduced an additional layer of uncertainty.

Across all scenarios, however, the resulting recovery patterns were quite consistent and it seems unlikely that more detailed elaboration of the dispersal submodel would yield vastly different results. Considering that natural variability and factors related to pesticide exposure, or other stressors, can vary substantially in the real-world, the results of this modelling study can still serve as indications of whether recovery can be expected to occur within an "acceptable" time frame and which scenarios or pesticide applications may require mitigation measures to be taken. An increase of assumptions regarding environmental complexity, however, needs to be judged carefully on a case-by-case basis. Increased realism in models can introduce an unwarranted increase in model complexity and uncertainty, which is not always supporting an improved credibility level of a model. However, such judgments can be only done in the light of available data, which are not always available or impossible to gather.

7 SYNTHESIS & GENERAL DISCUSSION

Ecosystems worldwide face ever-increasing pressure due to rising demands from a growing human population, climate change, and pollution. In the quest to optimize land use while preserving biodiversity, landscape managers and policymakers need tools that help them understand and judge the consequences of their decisions and actions (Stillman *et al.* 2015). In recent years, policy responses began to rely increasingly on the support of modelling approaches (MacGillivray & Richards 2015). Responses to spreading infectious diseases, infrastructure planning or food production already take advantage of the more comprehensive overview that models can provide concerning the outcomes of alternative management decisions. By being able to incorporate complex interactions across large spatial or temporal scales, models undeniably are helpful and indispensable tools to prevent further damage to the global ecosystem, or at least to keep the consequences at an acceptable level.

When making a decision of far-reaching consequences, one generally wants it to be the right one and one rooted in thoroughly researched facts. The perception of what is considered sound and "beyond doubt" is eventually not only a question of the source of information but also a psychological one. This is not only true in regular day-to-day life but also in matters of public concern. Many fields facing the introduction of model-based support for decision-making consequently react sceptical towards new and unfamiliar tools. In the end, decision-makers are still held accountable for their choices, which keeps them hesitant despite recognising the promising possibilities that models offer. Any tool that is complex and unfamiliar may face comparable apprehension and caution before being accepted. As such, it is no big surprise that ecological effect models are rarely accepted and used for environmental risk assessments of pesticides.

In an effort to gain a better understanding of the opinions and concerns about models aiding the environmental risk assessment procedure for pesticides, Hunka *et al.* (2013) interviewed representatives of the three main stakeholder groups involved. They found that the prospective of ecological models is generally appreciated and their introduction welcomed. Interviewees agreed that successful use of models could not only support a more comprehensive understanding of a substance's ecological effects but also reduce costs related to testing and monitoring procedures while benefiting animal welfare by reduced testing needs. On the other hand, Hunka *et al.* were also able to identify a lack of trust in model performance as a main obstacle in the way to overall model acceptance together with a lack suitable models.

Within the CREAM project, the latter part was particularly successfully addressed and relevant models for different animal groups were developed from terrestrial and aquatic invertebrates (Zimmer *et al.* 2012; Meli *et al.* 2013; Martin *et al.* 2013; Focks *et al.* 2014b; Hamda *et al.* 2014; Kulkarni *et al.* 2014; Nyman *et al.* 2014), to fish (Stadnicka *et al.* 2012; Ibrahim *et al.* 2014), small mammals (Liu *et al.* 2013) and birds (Kulakowska *et al.* 2014). These studies also explored different modelling techniques ranging from Dynamic Energy Budget (DEB) theory, toxicokinetic-toxicodynamic (TKTD) concepts and matrix models to

individual-based models (IBM). Moreover, some of these projects investigated how different approaches can be integrated to extrapolate between different levels of biological organization, species, or spatial scales to specifically target requirements for chemical risk assessment. By doing so, questions relevant to the risk assessment procedure can be explored to gain a more profound insight into, for instance, how effects on the sub-organismal or the individual level may affect the entire population, how lethal or sublethal effects on the population may relate to ecosystem effects, whether recovery can be expected to occur within the same season, or whether the frequency of pesticide application needs to be limited to a certain time of the year to prevent unacceptable effects on non-target species.

However, such examples are not enough to entirely overcome the hesitation with which those models are regarded. The uncertainties attached to model input and output, the often unclear descriptions of a model's design and limitations, and how it should be used and interpreted result in a "black box" image that decision-makers cannot ignore. Such issues leave them with the impression that risk assessments may become more cumbersome and complex rather than easier or more comprehensive (Hunka *et al.* 2013). In order to express their uncertainty about a given model's performance and relevance, regulators thus frequently ask for a model to be validated. This implies that *some* empirical proof that the model performs as claimed should be delivered so that one would not need to fear that a decision informed by the model in question would be flawed.

This thesis aimed to explore factors in the way of model acceptance and are related to the request for model validation. To this end, the several layers that are linked to this issue and the request for validation in itself were addressed. Most of the identified issues during this research are to a great extent rooted in communication and reliability issues that can be overcome with a thorough approach.

The importance of language and clear communication

Lev-Ari & Keysar (2010) showed that the accent spoken by non-native speakers could have strong implications on the lives of such individuals. They asked participants to judge the truthfulness of trivia statements ushered in native or accented speech and found that native speakers were considered more frequently to speak the truth compared to non-native speakers. The authors concluded that the accent reduced the "processing fluency" of the ushered statement but that instead of acknowledging their difficulties of understanding the ushered statement, participants subconsciously perceived the speaker as less credible. Hunka *et al.* (2013b) used a discourse analysis approach when interviewing representatives of the major stakeholder groups involved in pesticide risk assessment, i.e. representatives from academia, industry and regulators. In their analysis, Hunka *et al.* evaluated the language used by each stakeholder group. While all groups were found to use a similar, highly technical language specific to environmental risk assessment terms, all groups also perceived communications among the groups as inefficient. This perception of inefficiency

may be further enforced by parts of the terminology used within circles of ecological modellers that might be foreign or unfamiliar to regulators. Mathematical concepts, complex statistical analyses, or technical descriptions of a computer code can sound like a foreign language to an untrained ear. Drawing on the conclusions by Lev-Ari & Keysar, this may explain partly why regulators find it difficult to follow particularly technical discussions about a model and its intricacies. The time and concentration needed to study the theory around those intricacies add to the time pressure apart from the time pressure that they often experience in their daily work life and the resulting lack of time and concentration to study the theory around those details. This aspect is not eased by scientists or industry who often criticise the choice of terminology used by regulators when they attempt to explain themselves and their expectations and needs. A mutual understanding of the terminology used by both parties is needed before regulators can feel more ready to accept ecological effect models to inform a final decision about a substance's safe usage.

When regulators ask for a model to be validated, they do not only ask for a mere comparison of a model's output against real-world data. What they innately assume to be asking for is some form of evidence that the model performs as expected, that it produces the right results for the right reasons. As such the term "validation" is used as a synonym for "disambiguation". It is also a strong word that carries reassurance and something "valid" is generally perceived as something "certain" and "safe". However, in conversations with academics, definitions do play an important role in reducing confusion (Grimm & Wissel 1997). In academia, and more so in fields concerned with modelling, computers and technologies, "validation" is defined differently. In **Chapter 2** of this thesis, I reviewed the different meanings that this word can stand for and found that its definition strongly depends on the context it is used in. In some fields, mostly the technical branches, "validation" is defined as a comparison of real-world data against model output. "Verification" is a term often used at the same time to describe the testing and review that the respective computer code would undergo. To make things more complicated, some authors chose to switch the meaning of both words depending on whether a scientist chooses to follow the school of deductive or inductive reasoning, where "validation" is defined as "absolute truth". While the inductive approach distinguishes quite clearly between "right" and "wrong" conclusions, deductionists do not believe in an absolutely "right" theory. Confirming observations may strengthen and "corroborate" their hypotheses but they are aware that one false observation may break their logic. Some publications make use of additional synonyms to describe the two above-mentioned processes of model output testing and code checking.

Neither context is able to fully capture what regulators seem to be asking for. Some dimensions were left unaddressed when sticking to the one-dimensional definitions given earlier. To overcome this, I compiled a framework of questions and issues that should be considered during model development. The sequence of steps taken to develop and refine a model, the modelling cycle as illustrated in Figure 7-1, as well as comparable discussions in other scientific fields concerned with model quality served as blueprint.

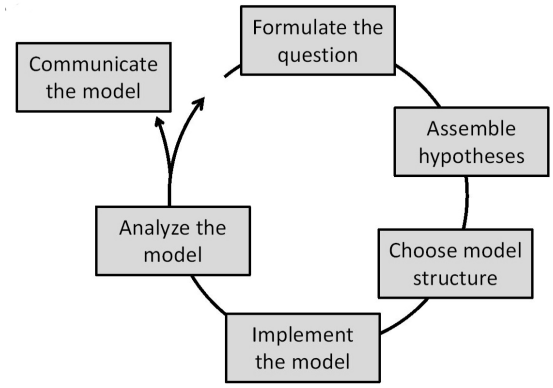


Figure 7-1: The iterative steps of model development summarized in the modelling cycle.

According to this framework and the nature of model development, the following aspects need to be addressed subsequently:

1. “Data evaluation” for scrutinising the quality of numerical and qualitative data used for model development and testing;
2. “Conceptual model evaluation” for examining the simplifying assumptions underlying a model’s design;
3. “Implementation verification” for testing the model’s implementation in equations and as a computer programme;
4. “Model output verification” for comparing model output to data and patterns that guided model design and were possibly used for calibration;
5. “Model analysis” for exploring the model’s sensitivity to changes in parameters and process formulations to make sure that the mechanistic basis of main behaviours of the model has been well understood; and
6. “Model output corroboration” for comparing model output to new data and patterns that were not used for model development and parameterisation.

This framework is called "evaludation", which represents a merger of "evaluation" and "validation", thus combining the emotional strength and affinity of one word with the more comprehensive and neutral outlook of the other. Adding additional terms to the already existing set of interchangeably used ones, especially a newly derived word, may seem counterproductive at first. The unfamiliarity with this term can make a conversation halt for a moment, requiring an explanation before moving on. This is an indented effect, as it distracts from the one-dimensional definition of the term "validation", that seems so manifested in some minds, to create room for expressing the bigger picture behind. Chaffin (1997) describes a series of tests done to understand how new words are learned in adults.

He found that unfamiliar words may shift one's attentional focus from easily understanding a conveyed message to focussing on understanding the unfamiliar word. This explains why people might react irritated when coming across "evaludation". On the other hand, Li (1988) found, as part of his study on how new words are being memorized, that words and their contextual meanings were more easily memorized the higher they scored in terms of word inference. This study raises hopes that the term and its associated framework will help overcoming a communication barrier by opening a discussion on common ground for the involved parties.

However, learning a new term and having a list of questions to consider while developing a model is not enough to facilitate a more effective and clearer communication of a model and its capabilities. After interviewing the different stakeholder groups about their expectations and requirements for introducing ecological effects model to the pesticide risk assessment procedure, Hunka *et al.* (2013a) learned that regulators expected models to be easy to understand and use as well as be transparent in regards to their design and implemented mechanisms. Transparent and easy to follow documentation can help addressing these elements and also help understanding more complex models. In that sense, Schmolke *et al.* (2010) proposed the model documentation scheme TRACE (TRansparent And Comprehensive Ecological modelling) after finding during a literature survey that the elements of good modelling practice have long been identified but not properly put into action for ecological models. This documentation framework offers a structured way to describe the steps and measures taken to develop a model. TRACE is as flexibly applicable to the different modelling techniques as the evaludation framework. However, after some time of testing, ecological modellers expressed their dissatisfaction with the documentation structure. Based on this feedback and the conclusions from **Chapter 2**, **Chapter 3** introduces a revised version of TRACE that ensures an improved explanation of all model development and testing phases.

The combination of evaludation and the adjusted TRACE documentation comes close to the definition of "validation" used by the WHO (2006), where "validation" is considered to summarize the "action of proving and documenting that any process, procedure or method actually and consistently leads to the expected results". The WHO's guidelines on good manufacturing practice follow a comparable rationality by first requiring all relevant technical subparts to be examined before it is tested whether a specific process consistently meets predefined quality specifications. This correlates with first testing whether a model is built on reliable data and theories, correctly implemented (i.e. the "technical subparts"), and whether it produces the expected output for the right reasons. The WHO guidelines also include a definition of "revalidation" as the "repeated validation of an approved process (or a part thereof) to ensure continued compliance with established requirements". This agrees strongly with the iterative nature of the modelling cycle. In one aspect, the guideline goes one step further by requiring a "validation master plan" in which the overall validation approaches, quality standards, and timeline are defined. Considering that such a master plan might be easier to define for manufacturing processes than for the development of

ecological models, this might lead to disproportionally high work pressure and may thus not be advisable to this degree for model development. However, defining a set of particular quality standards before beginning with the actual model implementation may help to ascertain a modeller's impartiality during model output verification. Moreover, the structure of evaludation and TRACE already promote some form of planning for the model development.

Credible data and the relevance of model re-testing

In order for a model to be credible and of use in regulatory settings, uncertainties of the model output need to be minimized and the model thus must be built on relevant data, sound theory and be tested thoroughly aside from being well and clearly communicated. In most instances when regulators have to decide whether or not they should accept a model, no feedback regarding the particular points of concern would be returned to the model developer (Hunka *et al.* 2013a), which makes it difficult to address such issues. One of the few exceptions is the MASTEP model (Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides) developed by Van den Brink et al. (2007). This model can be used to simulate the population dynamics of the water louse *Asellus aquaticus* before and after pesticide exposure and has been used in previous regulatory submissions. Galic et al. (2012) and Focks et al. (2014) developed adapted versions of MASTEP to explore the influences of timing of stress throughout a year and the impact of heterogeneous pesticide exposure on population recovery. All versions relied on the assumption of homogeneous landscape features and behavioural patterns, which was mostly owed to a lack of better information concerning behavioural responses of aquatic invertebrates to environmental cues and the impact of those cues on dispersal.

Criticism was directed at the model's implementation of the behavioural responses of the studied species, amongst others. In **Chapter 4** and **5**, I used an experimental setup to collect information about *Asellus* movement behaviour under different experimental conditions. These conditions covered effects exerted by natural factors ranging from different lighting conditions to presence/absence of food and shelter items. Furthermore, effects of sublethal pesticide exposure were tested since pesticides are not always found in the environment within the lethal concentration range, but are still present at concentrations that can impact the overall behaviour. Agatz *et al.* (2014), for example, found that exposure to sublethal concentrations of imidacloprid could severely affect the feeding behaviour of the freshwater shrimp *Gammarus pulex*, a macroinvertebrate that inhabits similar habitats as *Asellus* and occupies a similar ecological niche. The observed decrease in feeding rate can consequently lead to reduced fecundity or increased starvation, both processes with direct impact on population dynamics.

In **Chapter 6**, I compared the recovery time estimations from the previous version by Galic et al. (2012) with a version where I made use of the information from the experiments. I was also able to integrate an increased level of environmental complexity that could not be

addressed before due to a lack of data. In a series of subsequent steps, I implemented density dependence of movement rates, heterogeneously spread food availability and assumptions concerning restrained movement rates for life stages during which younger individuals cannot cover as much distance as adults may. Compared to former versions of the population model, recovery times did not change significantly, when "adult only" movement was assumed for all simulated individuals. This indicates that the previous assumptions yielded robust recovery estimations. Accounting for life stage dependent movement restraints, though, prolonged recovery when exposure was assumed to occur shortly before a reproduction cycle.

I chose to parameterize the dispersal with observational information rather than a mechanistic understanding. Linking the behavioural responses with, for example, a dynamic energy budget model would probably support scaling a juvenile's movement rate and needs for food supply more appropriately than an approach based on the relative body size difference to adults. The latter part is similar to assuming that a child, that is only half of its mother's size, would be capable of making steps that are only 10% the length of the steps that the mother can make. In the modelling study, I arbitrarily chose a step length of 10% of the adults' step length to get an idea of whether such an assumption would impact recovery time estimations and would thus require further study to quantify juvenile movement rates. In cases of slowly dispersing species, such as *Asellus*, this might need to be considered. A further advantage of linking the behavioural observations with a more mechanistic modelling approach is that it would allow for the inclusion and explanation of the observations I collected when studying the movement of individuals that were exposed to sublethal concentrations of neurotoxic pesticides.

As shown in **Chapter 6**, recoveries of extinct stretches required longer times until recovery. In my experiments, I made a start at collecting more information on the movement behaviour but did not include a setup where I studied the effects of vegetation on the behaviour. Vegetation can serve as both food source and shelter and, as such, may not only alter the searching behaviour, but also, due to an increased relative surface area, retain individuals in a given patch much longer. This could have significant impacts on dispersal and consequently on the recovery time estimation. In order to overcome this, experiments under semi-natural conditions would need to be performed; either in the lab with camera observations or outdoors. Both approaches hold additional challenges to overcome. Plain camera observations in the lab, for example, could become more difficult to analyse due to the lower contrast between background and animal coloration and the fact that individuals can "hide" under the provided substrate.

The lessons learned from this work help to underscore the robustness of the chosen dispersal modelling approach. Further investigations for stronger dispersers may require an alternative approach. Moreover, the step-wise inclusion of increasing amounts of complexity such as density or patch heterogeneity, as simple as they might have been in my work, can help to overcome the effect of overwhelming for someone who is not familiar

with this particular model or modelling in general. Topping *et al.* (2015) argue that a model with a high degree of realism and complexity would provide more predictive power than a simpler one like MASTEP. However, to a regulator who already carries doubts about model credibility, such complex models may be too much to take in. A process of building up confidence can help to overcome a sense of being stunned, and this is where the improved TRACE documentation can prove useful.

Conclusions and outlook

New technologies bring new methods and possibilities to old problems, while other problems remain the same. Ever increasing calculation capacities of computers have transformed the field of ecological modelling to a degree that such models now can serve as strong tools to support assessments of risks posed by humankind to the environment. While some management arenas have already accepted them, pesticide risk assessment still lags behind and this mostly due to reasons that lie outside a modelling exercise, such as the lack of trust or transparency as was shown by the work of Hunka *et al.* (2013a). It seems that once models are established within a decision-making context, they are used with comfort. In the realm of pesticide risk assessment, concentrations of plant protection products in groundwater and surface water in the EU are routinely estimated with the available FOCUS (FOrum for the Co-ordination of pesticide fate models and their USe) scenarios. The development of the FOCUS scenarios also required a lot of time. The working group, which was specifically appointed for this purpose, needed over a decade to establish the model scenarios as they are currently being used (Boesten *et al.* 1995).

Granted, chemical fate models may be somewhat easier to trust since they are based on first principles hat exhibit less variability as biological systems do. Nevertheless, the "black box" image of computer coded tools needed to be surpassed and new standards and methods for quality assessments needed to be established before FOCUS could be fully established. It should also be noted, that fate models "only" have to consider the fate of a chemical. Ecological systems are more complex and effects on one level of biological organization or one place within a sensitive ecosystem can exert rippling effects all along the food chain with possibly devastating consequences, including the accumulation of a dangerous substance all the way onto our plates. Such instances, where empirical studies fail to elucidate long-lasting, wide-ranging effects, can lead to a release of chemicals into the environment that later may need to be re-addressed. This is the case, for instance, with the current glyphosate debate in Europe where regulators have to address public concerns about the possible carcinogenic effects of this substance during the process of renewing the approval license (Clausing 2015).

Considering the high expectation pressure that regulators face in this field, it seems only natural that they look for easier ways to answer complex questions. They furthermore expect models to be easy to use and re-create, yet sufficiently complex to tackle multiple questions such as changes in species sensitivity when exposed to multiple toxicants at once,

extrapolating from one landscape type to another or generally to larger spatial and temporal scales, or even translating findings for one species to another. One of the biggest hopes expressed by regulators is that they would gain from a tool that would yield a form of binary output indicating whether the ecological effects on a species could be considered acceptable or not Hunka *et al.* (2013a). For lack of better knowledge, they bring such overly high expectations to the table in regards to what a model should be able to accomplish. However, models, correctly designed, can answer questions the best, when the question in itself is clearly defined. Considering the current state of protection goal definitions, this is not necessarily the case in pesticide risk assessment where no "unacceptable effects" should result from exposure (Brown *et al.* 2016).

This thesis addressed ways to overcome the notorious underrepresentation of models in ERA. In order to do so, I used the modelling cycle and approached the issues or model validation and credibility from two angles. The evaludation framework and TRACE scheme can help to improve planning and keeping track of a modelling study. A study performed by Courbaud *et al.* (2015) tested a spatially explicit, individual-based model of forest tree dynamics, and found that the evaludation approach was helpful in identifying weaknesses of model understanding that then could be addressed and improved. I explored further points of consideration of model credibility in form of a case study by entering a cycle of model revision. The simplified modelling approach can provide a strong basis for more elaborate techniques, which however may require more communication efforts.

Eventually, models need to be used to build up experience and to free oneself from the "black box" impression that new technologies can bring about. As such, studies like the one by Courbaud *et al.* (2015) that try a new approach and report on their experience are valuable in improving frameworks and building a sense of trust. Respectively, initiatives like the MODELINK workshops (Hommen *et al.* 2016) are indispensable for building a platform to exchange and provide experience with using particular models. In the MODELINK workshops, representatives from all major stakeholder groups involved in the risk assessment of pesticides came together to practically test and work with selected models to break the vicious cycle of models not being used because they are not being trusted, while they are not trusted because of not being used either.

APPENDICES

APPENDIX 1: TOXICOLOGICAL EFFECTS OF MARKERS

Materials & Methods

To test if the selected materials or the marking procedure would induce effects on mobility and mortality, tests were performed with both species (*Asellus* and *Gammarus*) and both materials (regular printing paper and UV reactive balloons obtained from UV Gear, Mark SG Enterprises, Surrey, UK; www.uvgear.co.uk). To this end, 10 individuals of either species were marked with one of the materials and kept in a 1L borosilicate beaker filled with copper-free water for 48 hours. Stainless steel hook-shaped gauze pieces were provided as structural elements and aeration was provided throughout the test duration. Unmarked individuals were treated similarly and kept under the same conditions as a control group. All treatments were set up with three replicates. Investigated parameters were immobilisation and mortality after 4, 24, and 48 hours. Criteria for defining these parameters were taken from Rubach et al. (2011). Individuals with affected movement compared to controls after agitation with forceps were categorised as immobile. Immobile individuals that did not react visibly within 30 seconds after repeated agitation were counted as dead.

Results

We found that neither marker material influenced the mobility or mortality of *A. aquaticus* in the 48-hour toxicity assessment essays. *G. pulex*, however, was more sensitive to the marking. While the plastic based markers exhibited relatively mild initial effects on mortality and mobility, the paper markers affected both endpoints more strongly during the first day (Table S1.1). We noticed that individual gammarids marked with paper were quickly surrounded by conspecifics, an effect that was not apparent for asellids. A follow-up experiment revealed (results not shown) that the gammarids perceived the paper marks as food sources and were similarly strongly attracted to them as to leaf disks. To avoid unnecessary effects of the marker material on the behaviour, we decided to use the plastic markers despite their lower fluorescence strength.

Table S1.1: Marker toxicity averaged over three replicates per treatment.

		Immobility			**Mortality**		
		4 h	24 h	48 h	4 h	24 h	48 h
G. pulex	**Paper**	27%	37%	37%	23%	37%	37%
	Plastic	10%	20%	40%	7%	20%	33%
	Control	0%	10%	33%	0%	10%	33%
A. aquaticus	**Paper**	0%	0%	0%	0%	0%	0%
	Plastic	0%	0%	0%	0%	10%	10%
	Control	0%	0%	0%	0%	0%	10%

APPENDIX 2: ADDITIONAL DATA ANALYSIS

Materials & Methods

From the obtained data files, the following additional parameters were extracted and analysed:

1. Number of stops per walked meter (gross distance) to better understand animal activity and resting behaviour;
2. A delta-statistic to determine whether an oriented random walk or unoriented correlated random walk model would describe the movement data best (Marsh & Jones 1988; Westerberg *et al.* 2008).

The delta statistic (Δ) was calculated for all consecutive path sequences of length n and equals the difference of the mean resultant vector lengths of the global compass directions of the subsequent steps (θ_i) and the turning angles between the subsequent steps (ω_j):

$$\Delta = \left(\frac{(\sum \cos\theta_i)^2 + (\sum \sin\theta_i)^2}{n^2}\right) - \left(\frac{(\sum \cos\omega_j)^2 + (\sum \sin\omega_j)^2}{(n-1)^2}\right).$$

A Δ –value greater than zero indicates that the distribution of compass directions is concentrated around a certain value compared, which means that the move pattern is directional. A negative Δ–value, on the other hand, indicates that an animal moves around without specific orientation.

To determine whether the treatments influenced the stop frequency and degree of orientation (delta-statistic) we performed Welch's t-tests, or, in case of more than two treatments, an ANOVA.

Results

Number of stops per walked distance

The marking had little influence on the average number of stops of A. aquaticus, although the variability increased when the animals were marked (compare light-unmarked with light-marked in Table S2.1 and Figure S2.1a). Under UV light conditions, this variability decreased and the overall distribution of number of stops per meter approached those of unmarked Asellids (Figure S2.1a, Table S2.1). Due to the overall high variability of the average number of stops, however, this difference was not statistically significant (Table S2.2).

The resting behaviour of G. pulex, in contrast, was significantly affected by the marking procedure (Figure S2.1b, Table S2.2). Both, the mean resting time and the mean number of stops made per "walked" meter increased, as did the variability of both parameters (Figure S2.1b).

We did not find statistically significant effects of population density in general on the resting behaviour of *A. aquaticus* or *G. pulex* (Table S2.2). The number of stops per walked meter did not show clear trends for either species. Population density affected this parameter significantly only in the case of Gammarus (Table S2.1, Table S2.2). However, it seems that Asellus stops more frequently at higher population densities, which fits the simultaneously increased resting times. Gammarus, on the other hand, appears to stop less frequently in accordance with the reduced average resting time (Table S2.1).

Delta statistic

The delta statistic is negative in all cases for both species (Table S2.1) indicating that an unoriented walk model could apply to represent the collected data.

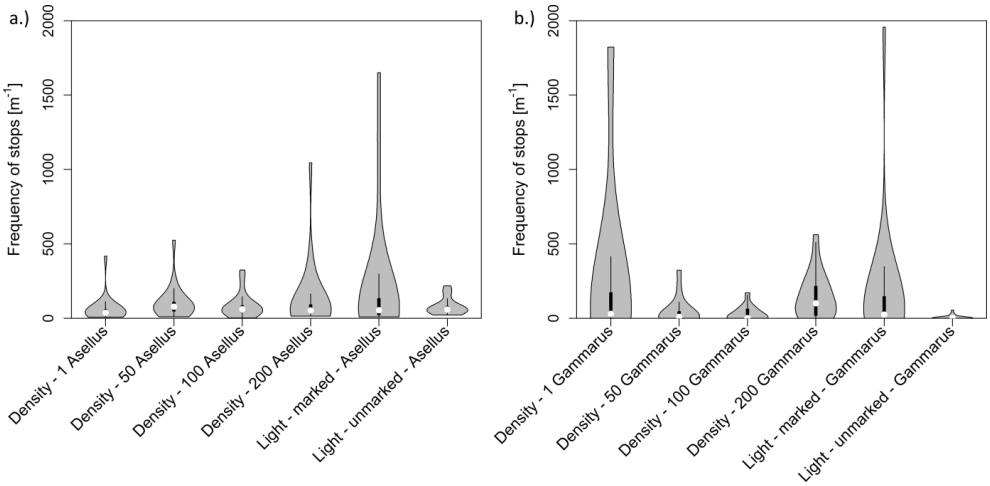


Figure S2.1: Violin and boxplots indicating the effects of different testing regimes on the distribution of the number of stops made by individual *A. aquaticus* (a) and *G. pulex* (b) per meter.

Table S2.1: Average values and standard deviations for movement parameters estimated for the different experimental regimes with *Asellus aquaticus* and *Gammarus pulex*.

	1 Individual		**50 Individuals**		**100 Individuals**		**200 Individuals**		**Light - marked**		**Light - unmarked**	
	A. aquaticus	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*	*A. aquaticus*	*G. pulex*
Available data points	27895 (39%)	1911 (3%)	23563 (33%)	2329 (3%)	21073 (29%)	3449 (5%)	29511 (41%)	3846 (5%)	13035 (18%)	12021 (17%)	12162 (17%)	13817 (19%)
Number of available paths	328	65	375	134	321	161	408	104	172	256	157	793
Path length	84.8 (117.8)	26.7 (80.1)	62.0 (94.3)	15.0 (19.5)	64.9 (105.7)	19.4 (28.2)	72.0 (85.8)	36.0 (60.2)	74.1 (143.1)	46.3 (98.9)	77.3 (109.0)	16.3 (19.9)
Animal activity and resting behaviour												
Resting time	30.2 % (12.4)	39.5 % (33.7)	40.2 % (13.8)	26.0 % (28.1)	36.9 % (14.2)	20.2 % (22.9)	38.5 % (15.6)	45.7 % (18.3)	40.1 % (21.4)	47.9 % (28.6)	41.4 % (11.6)	18.2 % (18.0)
Stops per meter	63.2 (87.9)	327.2 (633.1)	108.3 (113.5)	82.0 (134.8)	88.5 (85.5)	34.4 (54.8)	134.5 (233.9)	175.4 (167.5)	227.8 (425.8)	192.3 (439.4)	81.6 (57.2)	8.4 (14.2)
Step length pattern												
Step length	0.72 cm (0.26)	1.31 cm (1.47)	0.54 cm (0.25)	2.14 cm (2.27)	0.59 cm (0.26)	2.83cm (2.25)	0.57 cm (0.26)	0.67 cm (0.79)	0.65 cm (0.42)	1.30 cm (0.92)	0.61 cm (0.20)	4.13 cm (1.56)
Turning behaviour												
Turning angle	0.74° (7.28	34.29° (88.79)	-0.72° (7.95)	12.28° (67.38)	-0.07° (7.77)	6.01° (16.12)	-7.63° (35.12)	-19.64° (115.96)	0.97° (13.0)	1.93° (13.89)	0.07° (6.26)	-1.8° (±6.68)
Delta	-0.16 (0.14)	-0.13 (0.22)	-0.03 (0.07)	-0.07 (0.16)	-0.06 (0.15)	-0.08 (0.11)	-0.09 (0.09)	-0.03 (0.06)	-0.07 (0.09)	-0.07 (0.06)	-0.12 (0.08)	-0.11 (0.06)
Fractal dimension *D*	1.17 (0.13)	1.20 (0.17)	1.10 (0.13)	1.11 (0.10)	1.11 (0.12)	1.09 (0.09)	1.11 (0.10)	1.29 (0.21)	1.10 (0.11)	1.13 (0.09)	1.10 (0.08)	1.05 (0.04)

Table S2.2: Summary statistics of the statistical tests to estimate the significance of the effects of experimental conditions on movement parameters from observations of *Asellus aquaticus* and *Gammarus pulex*.

	Resting times [a,b]		**Number of stops** [a,b]		**Step lengths** [c,d]		**Path tortuosity** [a,b]		**Delta index** [a,b]		**Turning angle** [e]		
	t	p	t	p	W	p	t	p	t	p	W	p	df
Marking													
A. aquaticus	-0.23	0.82	1.52	0.14	166	0.86	0.59	0.56	-0.31	0.76	2.56	0.28	2
G. pulex	3.96	<0.01	1.87	0.08	29	<0.01	-3.30	<0.01	1.29	0.20	18.21	<0.01	2
Light													
A. aquaticus	-1.69	0.11	-1.68	0.11	220	0.60	-1.92	0.06	-0.54	0.59	3.06	0.2	2
G. pulex	-0.62	0.55	0.55	0.59	72	0.71	-1.18	0.26	-1.04	0.33	3.72	0.16	2
Density													
	df	F	p	df	F	p	df	X^2	p	df	F	p	df	F	p	W	p	df
A. aquaticus	41.47	2.21	0.11	40.61	0.97	0.42	3	5.47	0.14	41.18	1.73	0.18	39.85	3.37	0.03	4.98	0.55	6
G. pulex	19.09	3.66	0.03	19.09	3.99	0.03	3	10.88	0.01	22.06	5.48	0.01	19.07	0.21	0.89	17.99	0.01	6

[a] Welch's t-test for 2-sample comparison
[b] ANOVA for multi-sample comparison
[c] Wilcoxon's rank sum test for 2-sample comparison
[d] Kruskal-Wallis test for multi-sample comparison
[e] Watson-Wheeler test for 2- and multi-sample comparison

APPENDIX 3: TOXICITY TESTS

To study the effects of the pesticides on mortality and immobilization, separate toxicity tests were conducted for each pesticide. Ideally, the concentrations chosen for the movement studies should be below the observed EC10 levels and lead to observable behavioural changes but not complete paralysis or death.

Additionally, the movement study was designed to make use of individuals that were exposed for 48 hours, then removed from the exposure solution, rinsed with copper free water and placed in cosms that contained clean copper free water. Thus, the studied substances should not be readily detoxified by the test species. Toxico-kinetic data were used as measure for depuration time (imidacloprid: Ashauer et al. (2010) and chlorpyrifos: Rubach et al. (2010)).

Materials & Methods

Exposure solutions of chlorpyrifos and imidacloprid were prepared as described in the main manuscript. For chlorpyrifos the selected exposure concentrations were 0, 0.6, 1.5, and 3 µg/L, for imidacloprid 0, 100, 200, 400, 800, 1600 µg/L. Three replicates with 10 animals each were prepared per concentration level and stainless steel hook-shaped gauze pieces were inserted into the beakers to provide a physical substrate for the animals. In the case of chlorpyrifos, 2.5L Weck beakers were used and filled with 1.5 L of exposure solution, in case of imidacloprid 1.5 L Weck beakers filled with 0.75 L exposure solution were used. Test animals ranged from 0.5 to 0.9 cm in body length and no artificial aeration was supplied to minimize the evaporation of the chemicals. Beakers were closed with lids throughout the test. The light/dark regime was adjusted to 16/8 hours. Water temperature, pH and dissolved oxygen were monitored in the beginning, and after 24 and 48 hours to confirm stable conditions throughout the experimental period. All experiments were carried out at a water temperature of 19.6 ± 1.0 ºC, an average pH of 8.0 ± 0.1 (measured with electrode pH323, WTW Germany) and an average dissolved oxygen level of 8.6 ± 0.7 mg/L (measured with oximeter Oxi330 equipped with sensor CellOx 325, WTW Germany), all corresponding to the levels occurring in the movement experiments.

To determine the chlorpyrifos concentrations, water samples from all test vessels were taken and analysed for chemical concentrations as described in the main document at t = 0 and after 24 and 48 hours. The test lasted 48 hours and animals were scored after 24 and 48 hours with regard to mobility and mortality. Criteria for defining these parameters were taken from Rubach et al. (2011).

The imidacloprid toxicity study lasted for 96 hours since no reported EC10 values for *A. aquaticus* could be found in the literature at the time of the study and values determined for other species covered a broad range. The numbers of immobile and dead animals were counted after 24, 48, 72, and 96 hours in each replicate. Water samples were taken from

each beaker at the start of the experiment and after 96 hours, and processed as described in the main manuscript. However, samples with concentrations above 400 µg/L were diluted by factor 10 prior to analysis.

Results

Exposure

The measured exposure concentrations of chlorpyrifos and imidacloprid are given in Table S3.1. The chlorpyrifos concentrations in the exposure solutions were well within the range of the intended concentrations and decreased slightly over time as expected. Due to the high volatility of the substance and associated contamination routes, one of the controls received cross-contamination. This was tolerated because the control immobility and mortality were below 10%. Similarly, the nominal concentrations of imidacloprid were well achieved in the exposure solutions, which, remained more stable over time than was observed with chlorpyrifos. The increased variability of measured concentrations in the samples of the highest three concentration levels may be due to the dilution step.

Effects on mortality and mobility

For both pesticides, mortality and immobility increased with increasing concentrations, and over time (Table S3.1). This was more pronounced for imidacloprid than for chlorpyrifos in our study but was also expected for the chosen concentration ranges. Van Wijngaarden et al. (1996) report an 48h-EC10 for chlorpyrifos of 2 µg/L, and an 48h-EC50 of 4.3 µg/L, respectively, for *A. aquaticus* in laboratory based experiments. Rubach et al. (2011) found an 48h-EC50 of 6.16 µg/L (48h-EC10 = 3.3 µg/L) for *A. aquaticus* under similar conditions. Our results (see Table S3.2) fall into a comparable range.

Reported ECx and LCx values for imidacloprid exposure of *A. aquaticus* range wider than those for chlorpyrifos. Lukančič et al. (2010) determined a 48h-LC50 of 8.5 mg/L, and a 24h-EC50 of 0.8 mg/L for *A. aquaticus*. Roessink et al. (2013) found a 96h-LC50 of 316 µg/L (96h-LC10: 61.6 µg/L) and a 96h-EC50 of 119 µg/L (96h-EC10: 24.7 µg/L). Van den Brink et al. (2015) compared the work of Roessink et al. (2013) on a summer generation with an overwintering generation of Asellus and reported an additional 96h-EC50 of 78 µg/L. Similar to chlorpyrifos, the results obtained for imidacloprid in this study (see Table S3.2) lie in between those reported levels.

Conclusions

Our findings correlate well with other studies, indicating that the population we worked with was similarly sensitive and, thus, representative. Furthermore, based on the findings and the motivation for conducting the toxicity experiments, we chose to continue the study of effects of sublethal exposure on the movement behaviour with the concentrations listed in Table S3.3.

Because we intended to investigate if concentrations below the immobilization level would already lead to observable changes in behaviour, we decided to continue working with about 50% and 25%, respectively, of the observed EC10 in the case of chlorpyrifos. We had more experience working with this substance in combination with the chosen model species. Respectively, we opted for a slightly higher safety factor for imidacloprid and chose to continue with about 30% and 15%, respectively, of the observed EC10 value. Both decisions were also driven by the fact that the resulting exposure concentrations are likely to occur in the field (Muschal & Warne 2003; Marino & Ronco 2005; Van Dijk *et al.* 2013; Ensminger *et al.* 2013; Agatz *et al.* 2014; Papadakis *et al.* 2015).

Table S3.1: Test concentrations of chlorpyrifos and imidacloprid over, including controls, and results of the acute toxicity studies given as percentage of affected individuals.

	Immobility				**Mortality**				**Measured concentration [µg/L] (SD)**			
	24 h	48 h	72 h	96 h	24 h	48 h	72 h	96 h	0 h	24 h	48 h	96 h
Chlorpyrifos												
Control	0%	3%	-	-	0%	3%	-	-	0 (0.00)	0 (0.00)	0.02 (0.01)	-
0.6 µg/L	3%	7%	-	-	3%	7%	-	-	0.62 (0.00)	0.54 (0.04)	0.45 (0.01)	-
1.5 µg/L	0%	3%	-	-	0%	3%	-	-	1.26 (0.00)	1.14 (0.02)	1.09 (0.02)	-
3.0 µg/L	17%	40%	-	-	10%	23%	-	-	3.04 (0.00)	2.5 (0.07)	2.24 (0.01)	-
Imidacloprid												
Control	0%	0%	0%	0%	0%	0%	0%	0%	0 (0.00)	-	-	0.3 (0.05)
100 µg/L	0%	0%	3%	3%	0%	0%	0%	0%	97.47 (0.53)	-	-	97.64 (1.44)
200 µg/L	3%	10%	10%	17%	3%	3%	3%	3%	164.99 (11.63)	-	-	162.38 (5.85)
400 µg/L	20%	30%	37%	37%	7%	10%	20%	23%	381.25 (42.59)	-	-	429.54 (2.56)
800 µg/L	60%	67%	67%	83%	33%	47%	57%	67%	778.8 (62.27)	-	-	829.16 (23.10)
1600 µg/L	83%	90%	90%	97%	23%	40%	53%	70%	1651.75 (194.90)	-	-	1544.5 (7.30)

Table S3.2: Results of the acute toxicity study summarised as concentrations causing 10% or 50% lethality or immobility (LC10 and LC50, or EC10 and EC50, respectively; plus 95% confidence intervals) after 48 hours.

	Mortality				**Immobilization**			
	LC50 (µg/L)	(95% CI)	LC10 (µg/L)	(95% CI)	EC50 (µg/L)	(95% CI)	EC10 (µg/L)	(95% CI)
Chlorpyrifos	3.4	NC	2.8	NC	3.2	NC	2.7	NC
Imidacloprid	1517	(989-2327)	332	(191-578)	603	(487-747)	225	(155-326)

NC: could not be calculated

Table S3.3: Concentrations of chlorpyrifos and imidacloprid selected for studying the effects of acute sublethal exposure on the movement behaviour of *A. aquaticus.*

Chlorpyrifos	
Control	0 µg/L
Low	0.6 µg/L
High	1.5 µg/L
Imidacloprid	
Control	0 µg/L
Low	37.5 µg/L
High	75.0 µg/L

APPENDIX 4: POPULATION MODEL DESCRIPTION

The model description follows the ODD protocol for describing individual- and agent-based models (Grimm *et al.* 2006, 2010) and is largely adapted from Galic et al. (2012).

Model purpose

The model's purpose is to assess the recovery of a population of *A. aquaticus* after pesticide exposure, by accounting for the influence of environmental cues on dispersal potential and the timing of stress events throughout the year.

Entities, state variables, and scales

Entities in the model are individual female asellids, and square patches forming the landscape. Individuals are distinguished between juvenile and adult life stages. The simulated landscape comprises a ditch consisting of a string of 100 patches representing aquatic habitat (Figure S4.1) on a square grid with periodic boundaries, i.e. the grid forms a torus avoiding edge effects. Each patch represents 1 m^2. State variables of the different entities are listed in Table S4.1.

Each time step represents one day of a calendar year consisting of 365 days. Simulations start on day 0 (Jan 1) und end after six years or when the modelled population goes extinct. The first two modelled years are not considered for further analysis to avoid transitional effects during the model initialisation phase. The model is implemented in the NetLogo platform v. 5.2 (Wilensky, 1999; downloadable from http://ccl.northwestern.edu/netlogo).

Figure S4.1: A representation of the ditch in the model consisting of a string of 100 aquatic patches.

Table S4.1: Entities and their respective individual state variables.

Entity	State variable	Unit
Individuals	Age	[days]
	Size	[body length in cm]
	Hatching date	[day in calendar year]
	Lifespan	[days]
	Location	[x and y coordinates, continuous]
	Reproductive state	[binomial, depending on whether individual reproduced or not]
Patches	Carrying capacity	
	Local mortality	[%, induced by stress and/or population density]

Process overview and scheduling

Every time step (day), the following processes, or submodels, are scheduled for all individuals in a randomized sequence for each process; state variables are updated immediately (asynchronous updating):

Aging (Increase age of individuals by one time step)

If **Mortality** of juveniles and adults (background and density-dependent)

Delete from the population

Else **Move** – same for all

Grow – both juveniles and adults grow until they reach maximum size

Mature – when they reach 45 days, juveniles become adults

Reproduce - once per each individual adult, 2 generations in 1 year

Pesticide mortality

Update plots and outputs

Design concepts

Basic principles. The model relies locally on an observed phenomenological representation of density-dependent effects. Another principle explored is the effect of movement on recovery, and of metapopulation and rescue effects, i.e. of individuals immigrating from unstressed areas.

Emergence. Population dynamics and in particular the response of the population following pesticide-induced mortality arise from individual behavior (reproduction and movement) and local, within-patch density-dependent effects.

Interaction. Individuals interact indirectly via local density-dependent effects on growth, reproduction, and mortality. In individuals that have not reached their maximal size, daily growth increment is hampered by increasing density in their local environment (single patch), and the size at reproduction time determines the final clutch size. Probability of dying increases with increasing density of individuals in a patch. Where included, the patch characteristics, i.e. whether or not it provides food sources, determines the behaviour of individuals.

Stochasticity. Values of most parameters are drawn from probability distributions obtained from literature data to represent natural variability observed in water louse populations. All parameter values and distributions are shown in Table S4.1.

Observation. For model testing and analysis, the dynamics of a local (10 patches) and total population (all 100 aquatic patches), population death rate and size distribution of the

whole population are observed. Furthermore, time to recovery, i.e. the number of days needed for exposed population to reach 95% of the reference population, which had not been exposed to pesticides is calculated.

Initialization

The initial population of adult individuals is set to 300, 3 individuals per patch, each with a given size, drawn from a normal distribution (mean 3 mm, SD 0.2, based on Chambers (1977)). Individuals are randomly distributed within their aquatic habitat and have their variables defined at the start of the simulation (Table S4.2).

The following pseudo-code gives an overview of the initialization process

initialize all patches:

set pesticide mortality level…

initialize *Asellus* individuals within the patch

set size

set age

set location …

Table S4.2: Initial parameter set.

Parameter	Distribution	Value	Unit	Reference
Size (length)	normal	mean 3, SD 0.2	mm	Okland (1978), Arakelova (2001) Marcus et al. (1978)
Lifespan	exponential	mean 90	days	adapted from Vitagliano (1991)
First reproduction day	uniform	1 to 28	days	adapted from Chambers (1977)
Second reproduction day	uniform	1 to 45	days	adapted from Chambers (1977)

Pesticide exposure

Patches in the habitat can be exposed to pesticides or not; those that are exposed simulate a 24h LCx (where x can range from 1 to 100%) on different dates, resulting in 1 time (1 time step) effects on the population. Fate of pesticides in the environment is not considered in this version, i.e. mortality is only caused on the day of application.

Input data

The model does not include external input, i.e. there are no external drivers of system behavior such as temperature, rainfall etc.

Submodels

The model is comprised of four submodels: mortality, dispersal, growth and reproduction.

Mortality:

Individuals are subject to three different causes of mortality, i.e. natural background, density dependent and pesticide induced, the probabilities of each not being additive but sequential.

The *natural lifespan* of *A. aquaticus* in NW Europe is related to age, reproductive status and predation and can range up to 600 days (Vitagliano et al. 1991). We furthermore assume that juveniles experience a higher mortality rate than adults (Van den Brink et al. 2007). Under this assumption, we assign each individual a lifespan from an exponential distribution with a mean of 90 days, resulting in less than 1% of individuals surviving longer than 400 days. Once the number of time steps in the simulation reaches an individual's lifespan, it will die and be erased from the population.

Density-dependent mortality is an indirect way of modelling resource competition, which we otherwise do not explicitly include in this model. Even though underlying mechanisms of density dependence are not known for this species, it is known that populations of *A. aquaticus* are regulated to a certain degree by their densities (see e.g. Iversen & Thorup (1988), Adcock (1979), Van den Brink et al. (2007)); we follow the mortality based on local densities, μ_{dd}, from Van den Brink et al. (2007):

$$\mu_{dd} = \mu_1 \cdot N \qquad \text{eq. 1}$$

where μ_1 is a parameter governing the steepness of the density dependence (m^2/ind*d) and N represents the local density (ind/m^2). Density-dependent mortality is patch based, so all individuals in one patch have the same probability of dying due to overcrowding. Eq. 1 is the simplest assumption on effects of density, where each individual exerts a certain effect on the survival of each of its conspecifics within the patch. Galic et al. (2012) performed a sensitivity analysis for this parameter by testing a logistic function as alternative and found unrealistic effects on local abundances. Particularly at the higher end of the spectrum, after the release of juveniles, a whole patch could go extinct because of the highly increased mortality probability. Based on these findings and lack for better information, the linear relationship between abundance and mortality probability seemed the most appropriate assumption.

Pesticide-induced mortality is implemented as a patch specific mortality probability (i.e. all individuals in a given patch have the same probability of dying), which ranges from 0 to 1. The (hypothetical) pesticide is assumed to be toxic only on the day of application (see Table S4.3 for more information on application days).

Table S4.3: Simulated model experiments.

Landscape	Exposure day	Pesticide toxicity	Exposure [% of landscape]
Ditch	110	LC50	100
	160		
	210	LC100	50 (random patches)
	260		50 (connected patches)

Dispersal:

Movement and dispersal are crucial for recolonization and recovery dynamics (Niemi *et al.* 1990; Holomuzki & Biggs 2000; Albanese *et al.* 2009). We estimated daily movement distances in the following way. We assumed that *Asellus* follows a random walk. Parameters for the three basic parameters characterising this type of movement representation are comprised of the combination of step lengths and turning angles as well as general movement activity. We extracted these parameters from paths that we observed in experimental setups designed to test the influences of varying environmental factors on the movement behaviour of *Asellus* (Augusiak & Van den Brink 2015, 2016). The experimental treatments and observed summary statistics per treatment are listed in Table S4.4. Pairs of step lengths and turning angles were drawn from the pooled set of observations for a given experimental setup. Movement activity was correspondingly included by drawing pairs of zeroes for step length and turning angle. Once a data pair was drawn from the observed data, each individual rotates according to the chosen angle and moves the distance of the chosen step length.

Because small, individual movement occurs on a very fine time scale (in seconds), while the basic time step in the population model is 1 day, we chose to simulate dispersal separately and include only the distances covered after 24 hours in the population model. To this end, we simulated dispersal of 100,000 individuals in a ditch, for 24 hours, all starting from the same position in one patch, counted the number of individuals in each of the patches after the simulation and, from that, estimated the probability distribution of distances each individual is expected to cover in one time step of the population model. See Table S4.4 for the obtained dispersal distribution parameters.

In the population model, all individuals of the initial population and their subsequent offspring are positioned randomly within a patch. Each time step, an individual is assigned a random number from the previously simulated dispersal distribution. The individuals then change their x coordinates accordingly by moving to the left or to the right from their initial position. In that way, individuals move from patch to patch, but keep the same location within the patch. Since conditions within each patch are considered to be uniform, the location within each patch is not important.

Growth:

Individuals grow following the von Bertalanffy growth equation, leading to a logistic growth curve, which is observed in most isomorphs under constant food conditions:

$$l(a) = l_{\max} \cdot (1 - e^{-k \cdot a}) \qquad \text{eq. 2}$$

where l_{max} is the maximal length an individual can reach, κ is the daily growth rate and a is the age (days). The maximum size of individuals varies depending on the region, between 11 and 12 mm in the publication by Økland (1978), Chambers (1977) reports 9 mm, Arakelova (2001) 10 mm and Marcus et al. (1978) up to 12 mm.

The newly hatched individuals are assigned an initial size, normally distributed with a mean of 1 mm and SD of 0.2 mm (Adcock 1979). Following the growth function (with a growth rate as in Table S4.5), an individual starting with 1 mm length needs 145 days to reach 95% of the maximum size, i.e. 11.4 mm.

Daily growth increment is then the derivative of eq. 2:

$$\frac{dl}{da} = \kappa \cdot l \cdot (\frac{l_{\max}}{l} - 1) \qquad \text{eq. 3}$$

We assume that growth is density dependent, and decreases exponentially with high densities in a patch (Figure S4.2). The density dependent growth factor (*DDG)*, is expressed as

$$DDG(density) = e^{-y/\kappa} * density \qquad \text{eq. 4}$$

where y is the scaling factor of the function, and κ is the local (patch specific) carrying capacity.

Hynes and Williams (1965) experimentally showed that waterlouse populations produce more offspring when housed in larger jars; even though the amount of food in their experiments was the same, the lower productivity indicates certain effects of competition resulting in lowered production, possibly interference. We approximate the effect of less preferred habitat or scarce resources by including the effects of density on daily size increments, i.e. individual growth rate, in such a way that when the density in a patch reaches the assigned carrying capacity, each individual size increment decreases by 60% (green line in Figure S4.3).

Figure S4.3 shows (mean individual) growth trajectories at different constant densities (density dependent scaling parameter is set to 1, 50 individuals in the population), where the carrying capacity of the system varies. At κ =100, population abundance of 50 individuals is at the middle of its capacity, at κ =50 a population of 50 individuals is just at its capacity, and, finally, at κ =10 the population is five times over its capacity.

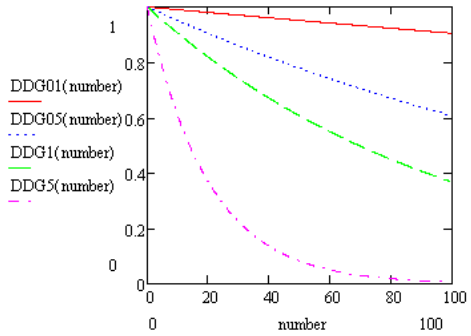


Figure S4.2: The strength of individual growth dependence on local densities is expressed with a scaling parameter. The carrying capacity in this figure is 100 individuals, and 4 different values are depicted, namely 0.1, 0.5, 1 (the default value) and 5.

With no density dependent effects, individuals reach their maximal size by day 200, whereas it takes about 300 days longer when the density is half of the carrying capacity (κ =100), and much longer in case the population is at its κ (κ =50). If the density exceeds the carrying capacity by 5 times (κ = 10, Figure S4.3) individuals almost stop growing.

In the default parameter set, patch carrying capacity is set to 100 while there are between 10 and 60 individuals per patch on average. Individuals change their position at each time step and the effect of densities on individual growth rate is calculated with new densities each new time step.

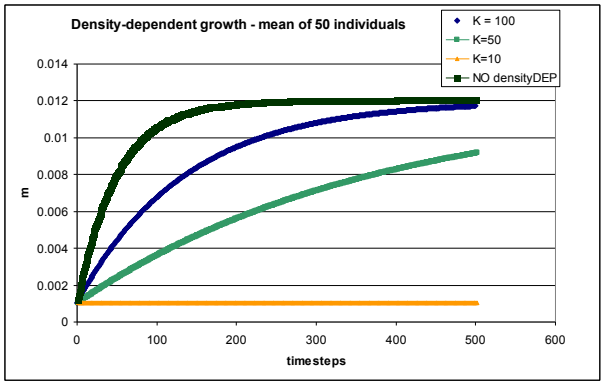


Figure S4.3: Mean values of individual growth trajectories of populations under different densities. *m* is the size in meters, and "timesteps" are in days. There are 50 individuals in this population (that only go through the growth procedure), their growth is modeled under no density-dependent effects, and under set κ of 10, 50 and 100.

Reproduction:

Voltinism, i.e. number of generations in a year, of aquatic arthropods is typically governed by environmental conditions, mainly through water temperatures. As we do not include external temperature data in the model, we introduced two periods in a year when individuals are able to reproduce. The reproductive periods start in the beginning of May (Julian day 210) and mid July (Julian day 200), and last 28 and 45 days (Table S4.5),

respectively. In these reproductive periods, each individual female releases its offspring once. The timing of offspring release is assigned via a random number from a uniform distribution, corresponding to the length of each reproductive period that represents the day after the onset of each reproductive period at which a female will release her offspring. For instance, after winter, an adult female is assigned a number 12 (from a uniform distribution between 1 and 28) and respectively will release her young at day 132 (onset 120 + 12).

Individuals reproduce only once in their lifetime as this is most commonly observed pattern (Chambers 1977). The number of offspring (clutch size) is size-dependent and is positively correlated with size (Chambers 1977) (Figure S4.4):

$$N_{offspring} = \left(\frac{current\_size}{\max\_size}\right) \cdot \max\_clutch\_size \qquad \text{eq. 5}$$

Each adult, thus, gives birth to a given number of juveniles and dies immediately after. Females from the winter generation are bigger and will have more offspring per female, but are fewer to start with; summer females are smaller by the time they start reproducing (after around 80 days and less) and will be maximum 9 mm long (according to Økland (1978), summer females are up to 7 mm in length). Different authors report different clutch sizes in water lice females with a size of 12 mm, ranging from maximum of ca. 100 (Arakelova 2001), 125 (Graça *et al.* 1993), 250 (Ridley & Thompson 1979) to more than 300 (Tolba and Holdich 1981). Here, we fixed the maximum clutch size to 100 individual females (we do not model male individuals).

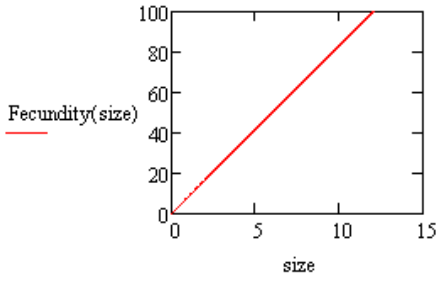


Figure S4.4: Relationship between size and number of offspring each female releases in 1 reproductive cycle.

Table S4.4: Basic path information, mean values and standard deviations (SD) of movement parameters observed for the different experimental regimes with *Asellus aquaticus* and parameters for the estimated 24h dispersal distributions (Normal distributions) considering the movement behaviour exhibited under the different experimental conditions. See (Augusiak & Van den Brink 2016) for an explanation of the different experimental setups.

	Chlorpyrifos low (0.6 µg/L)		**Chlorpyrifos high (1.5 µg/L)**		**Imidacloprid low (37.5 µg/L)**		**Imidacloprid high (75 µg/L)**		**Control (starved)**		**Control (fed)**		**Food**		**Shelter**	
Density	*1*	*50*	*1*	*50*	*1*	*50*	*1*	*50*	*1*	*50*	*1*	*50*	*1*	*50*	*1*	*50*
Available data points (Percentage of total recording time)	29760 41%	33098 46%	28132 39%	31668 44%	11295 17%	27450 38%	21432 30%	23127 32%	19484 27%	23212 32%	27807 39%	23263 32%	26189 36%	25569 36%	11291 16%	12119 17%
Number of available paths	256	384	244	421	336	330	379	394	336	448	328	375	314	289	176	186
Average path duration (sec)	114.4 (138.7)	85.2 (139.2)	113.8 (156.7)	74.2 (96.2)	35.4 (56.8)	81.9 (143.0)	55.1 (117.4)	57.4 (105.3)	56.7 (99.1)	50.8 (85.9)	83.8 (117.8)	60.9 (94.2)	82.4 (132.7)	87.2 (136.8)	62.4 (97.0)	63.1 (92.2)
Average measured concentrations (t_{0h}, t_{48h}; µg/L)	0.40, 0.28 (0.03, 0.06)		0.83, 0.75 (0.05, 0.21)		42.09, 40.67 (3.80, 3.94)		80.82, 77.61 (2.80, 3.38)		-		-		-		-	
Resting time	51.5% (26.7)	53.8% (29.4)	56.7% (29.9)	44.4% (21.4)	28.4% (16.7)	37.9% (24.5)	36.3% (27.4)	35.4% (22.6)	29.5% (10.3)	31.2% (15.9)	30.2% (12.4)	40.2% (13.7)	35.7% (13.5)	45.4% (19.1)	44.2% (19.4)	44.2% (11.0)
Step length (cm/sec)	0.79 (0.37)	0.71 (0.36)	0.53 (0.31)	0.75 (0.28)	0.82 (0.30)	0.81 (0.33)	0.74 (0.42)	0.92 (0.36)	1.12 (0.25)	1.13 (0.29)	0.99 (0.25)	0.86 (0.25)	0.94 (0.25)	0.80 (0.29)	0.86 (0.32)	0.69 (0.28)
Turning angle	1.55˚ (28.10)	-0.91˚ (35.41)	0.93˚ (44.87)	1.19˚ (32.85)	1.14˚ (37.00)	3.74˚ (37.19)	-1.57˚ (45.18)	-0.15˚ (43.20)	2.92˚ (25.09)	-2.73˚ (26.31)	2.56˚ (27.77)	0.09˚ (36.71)	-4.35˚ (28.25)	-6.70˚ (34.21)	-1.48˚ (28.33)	0.15˚ (38.57)
24-hour dispersal (m)	$2.8*10^{-3}$ (2.09)	$6.2*10^{-3}$ (1.74)	$5.8*10^{-3}$ (1.27)	$-3.8*10^{-3}$ (2.07)	$1.7*10^{-2}$ (3.01)	$6.4*10^{-3}$ (2.34)	$4.9*10^{-5}$ (1.78)	$2.3*10^{-3}$ (2.34)	$2.0*10^{-3}$ (4.02)	$5.8*10^{-3}$ (3.82)	$-7.8*10^{-3}$ (3.45)	$-7.3*10^{-3}$ (2.39)	$1.6*10^{-3}$ (3.07)	$7.7*10^{-3}$ (2.17)	$6.3*10^{-3}$ (2.74)	$9.0*10^{-3}$ (1.98)

() - The standard deviations are given in brackets.

Table S4.5: List of model parameters.

Submodel	Parameter	Distribution	Value	Unit	Reference
Habitat	System carrying capacity, K	Constant	10 000	ind	Constant K per patch (growth related)
Mortality	Lifespan	Exponential	Mean 90	day	adapted from Vitagliano (1991)
	μ1	Constant	0.001		Based on Van den Brink et al. (2007)
Reproduction	Onset of 1st reproductive cycle	Constant	120 (April/May)	day	mimicking western European conditions
	Onset of 2nd reproductive cycle	Constant	200 (July)	day	
	Maximum clutch size	Constant	100	ind	Adapted from Tolba and Holdich (1981)
	Length of 1st reproduction period	Uniform	1 to 28	day	Adapted from Chambers (1977)
	Length of 2nd reproduction period	Uniform	1 to 45	day	Adapted from Chambers (1977)
Growth	Maximum size	Constant	12	mm	Okland (1978), Arakelova (2001), Marcus et al. (1978)
	Minimum size	Normal	Mean 1, SD 0.2	mm	Adcock (1979)
	Kappa, k	Constant	0.02	/day	determined by calibration
	Age at maturity	Constant	45	day	Williams (1960)
	Density-dependent factor, *y*	Constant	1		Estimation by Galic et al (2012)
Dispersal	benchmark	Normal	Mean 0.004, SD 4.444	m	Own simulation (more in Dispersal simulations)
	CPF low - 1	Normal	Mean -0.08, SD 2.07	m	Own simulation (more in Dispersal simulations)
	CPF low - 50	Normal	Mean 0.01, SD 1.71	m	Own simulation (more in Dispersal simulations)
	CPF high - 1	Normal	Mean -0.06, SD 1.29	m	Own simulation (more in Dispersal simulations)
	CPF high - 50	Normal	Mean -0.06, SD 2.07	m	Own simulation (more in Dispersal simulations)
	Imi low - 1	Normal	Mean -0.21, SD 3.14	m	Own simulation (more in Dispersal simulations)
	Imi low - 50	Normal	Mean -0.12, SD 2.41	m	Own simulation (more in Dispersal simulations)
	Imi high - 1	Normal	Mean -0.01, SD 1.85	m	Own simulation (more in Dispersal simulations)
	Imi high - 50	Normal	Mean 0.09, SD 2.38	m	Own simulation (more in Dispersal simulations)
	Control (starved) - 1	Normal	Mean -0.09, SD 4.09	m	Own simulation (more in Dispersal simulations)
	Control (starved) - 50	Normal	Mean -0.02, SD 3.86	m	Own simulation (more in Dispersal simulations)
	Control (fed) - 1	Normal	Mean -0.06, SD 3.41	m	Own simulation (more in Dispersal simulations)
	Control (fed) - 50	Normal	Mean -0.04, SD 2.39	m	Own simulation (more in Dispersal simulations)
	Food - 1	Normal	Mean -0.02, SD 2.99	m	Own simulation (more in Dispersal simulations)
	Food - 50	Normal	Mean 0.12, SD 2.15	m	Own simulation (more in Dispersal simulations)
	Shelter - 1	Normal	Mean -0.11, SD 2.82	m	Own simulation (more in Dispersal simulations)
	Shelter - 50	Normal	Mean 0.08, SD 1.93	m	Own simulation (more in Dispersal simulations)

APPENDIX 5: RANDOM WALK MODEL - TESTING

We tested the robustness of the random walk model to changes in parameters and assumptions. Not much is known of the movement behavior of *Asellus* and our experiments are among a few to observe individual paths of this species under various conditions. Additionally, essential principles governing the movement of animals are generally not well understood yet in the field of movement and resource ecology. Simulating animal paths and dispersal thus remains a challenge to date.

To obtain daily dispersal distributions for the population model based on the experimental observations, we used a separate random walk model. Parameters for the three basic parameters characterizing this type of movement representation were comprised of the combination of step lengths, turning angles and general movement activity. We extracted these parameters from paths observed in experimental setups that were designed to test the influences of varying environmental factors on the behavior of *Asellus* (Augusiak & Van den Brink 2015, 2016).

Following Galic et al. (2012), we simulated dispersal in a ditch with a homogeneous environment and consisting of a string of 100 patches, each patch representing $1m^2$. Individual *Asellus* were modeled regardless of age or gender assumptions. The time step in this model was 1 second, the same time resolution as used for extracting the relocation information from the recorded paths in the experiments. Corresponding to the experimental setups, a data set containing all recorded pairs of step lengths and turning angles for this particular setup was used to derive the information for the next modeled step. Each time step, an individual would randomly draw a step length-turning angle pair from this data set, reorient itself according to the angle and move a distance according to the step length. Resting moments were incorporated as pairs of zeroes. Individuals reaching the land-water border would only reorient themselves but not move if the next step would cause them to leave the water. We simulated a time frame of 24 hours with all individuals starting from the same position but with random orientation. At the end of the simulation, the horizontal coordinates of all modeled individuals were collected and used to estimate the probability distribution of distances each individual is expected to cover in one time step of MASTEP. For practical reasons, all simulations used for comparisons were performed by simulating 1000 individuals, instead of the 100,000 that were used to obtain the dispersal distributions used in MASTEP.

Testing

Landscape scenario. To test whether the assumption concerning the individual behavior at the land-water border had a significant impact on the obtained 24-hour dispersal distributions, we alternatively used a homogeneous landscape of 50 x 50 patches, each patch representing $1m^2$. Individuals in this landscape could move without any spatial

restrictions. We also used the unrestricted walk model as benchmark to compare the following options.

Correlation scenario. One factor introducing error in models concerns the representation of correlation among the different parameters that characterize relocations, and their inherent correlation and autocorrelation structures (Westerberg *et al.* 2008). Since turning angles and step lengths, i.e. velocity, are often found to be correlated, we estimated the Pearson correlation coefficient for these two parameters for each experimental setup after pooling all observed step lengths and turning angles (see Figure S5.1). Sharp turning angles were regularly correlated with reduced step lengths while larger step lengths were correlated with straighter angles. To investigate the relevance of the correlations between step lengths and turning angles, the model was adjusted to draw step lengths and turning angles independently from each other. The only limiting assumption here was that the individuals would redraw their next step information in cases when a turning angle of zero was associated with a resting moment in the experimental data set while the drawn step length was not.

Physical limitation scenario. MASTEP makes use of juvenile and adult lifestages of *Asellus* and accounts for differences between these stages in regards to reproduction but not in terms of movement ability. However, due to their smaller size, juveniles may not be capable to cover steps as big as those from adults. To account for the possible physical limitation, we modeled a population comprised of only juveniles that were allowed to move only 10% of the step length that an adult would be able to cover.

Analysis

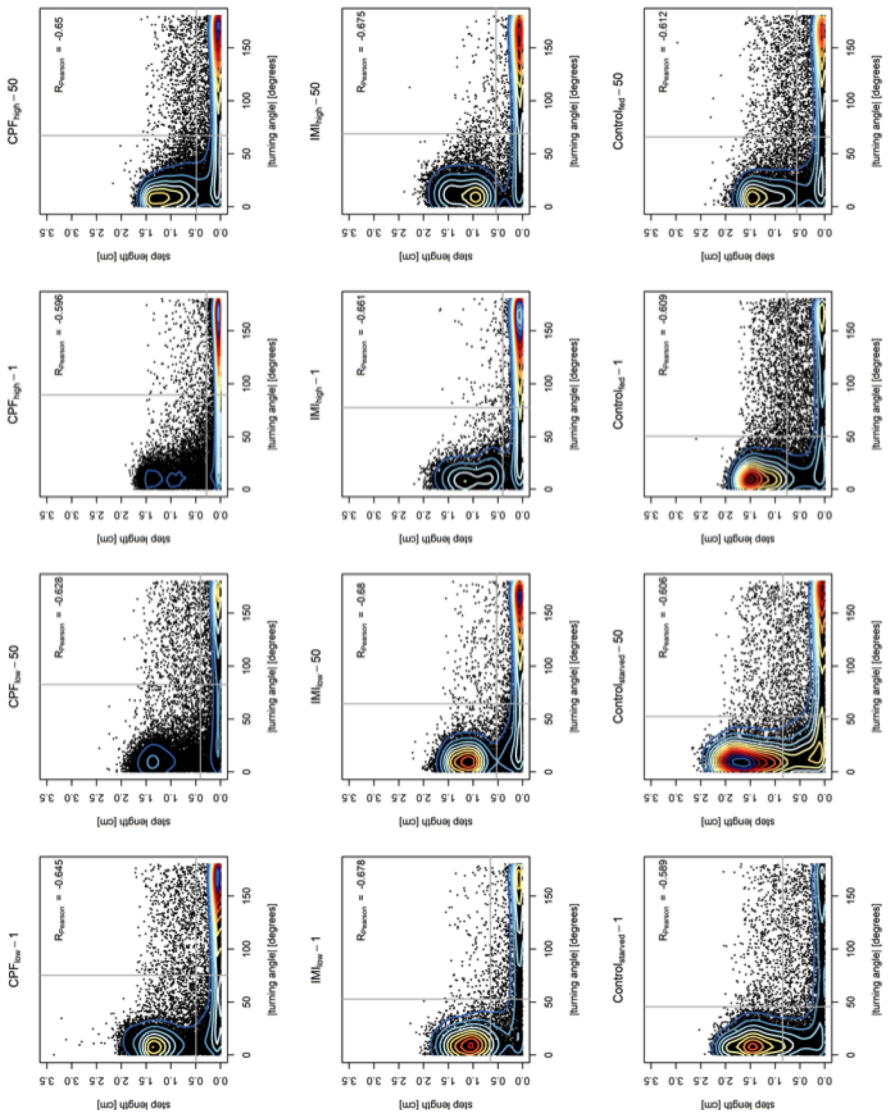
All obtained location distributions were normally distributed (Kolmogorov-Smirnoff test) and mean values and standard deviations were calculated (Table S5.1). We used the unrestricted landscape scenario with fixed turning angle-step length pairs to statistically compare all other scenarios with via a Student's t-test (significance level $\alpha = 0.05$). Grey boxes in Table S5.1 mark where differences were found to be significant.

Results

The individual simulated dispersal distances after 24 hours are summarized in Figure S5.2. Locations found to the left of the starting point were designated with a negative sign while locations to the right of the starting point received a positive sign. We used violin plots in combination with boxplots to display relevant statistical parameters. Violin plots consist of kernel density plots and show the probability density of data at different values.

The calculated mean and median values were all in a range close to zero, while the biggest differences were found in the standard deviations, which could vary depending on how directed an individual moved or how much it rested. All estimated dispersal distributions fell into a range tested by Galic et al. (2012) when they performed a sensitivity analysis of the influence of changes in the applied dispersal distribution on calculated recovery times

in MASTEP. During this investigation, they found that the obtained recovery times were quite robust to changes of the dispersal distribution unless the distance decreased and limited movement affected the immigration of surviving individuals. Galic et al. (2012) assumed a 24-hour dispersal distribution of 0.004 meters and a standard deviation of 4.44 meters. Our mean values were comparable to that used by Galic et al. (2012). However, in most cases, we calculated lower standard deviations, which can depending on the assumptions made and used in the population model, lead to higher recovery times, at least on the local population level. Particularly the possibility of reduced step lengths for juveniles is likely to have a strong influence on recovery times and this aspect was thus included in the main manuscript, although we did not have experimental data to confirm lower movement rates for the juvenile life stage of *Asellus*.


CPF_low – 1
CPF_low – 50
CPF_high – 1
CPF_high – 50
IMI_low – 1
IMI_low – 50
IMI_high – 1
IMI_high – 50
Control_starved – 1
Control_starved – 50
Control_fed – 1
Control_fed – 50
step length [cm]
|turning angle| [degrees]

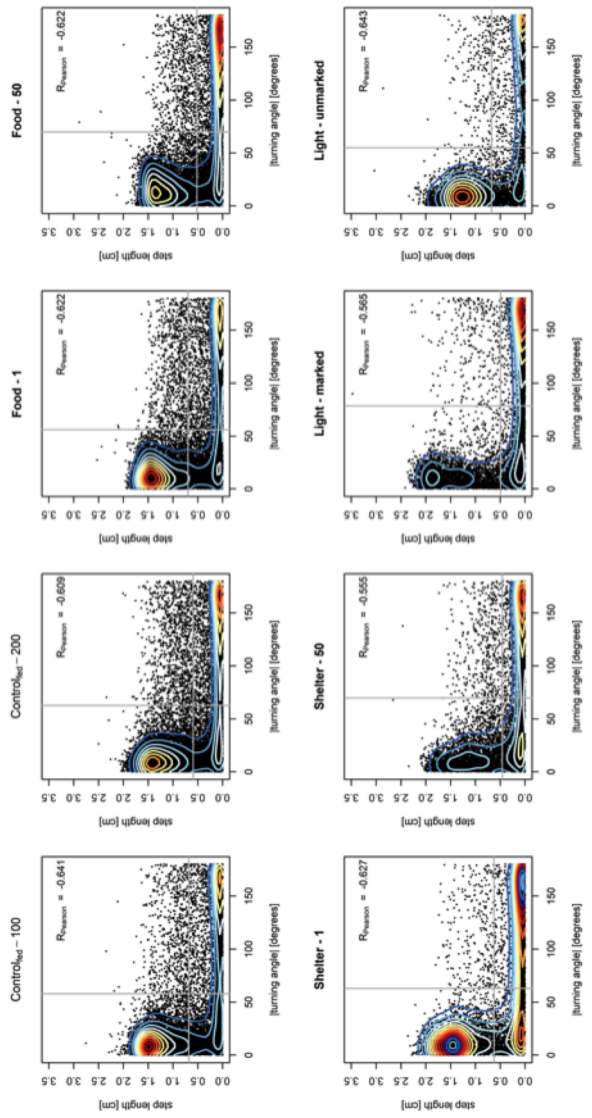
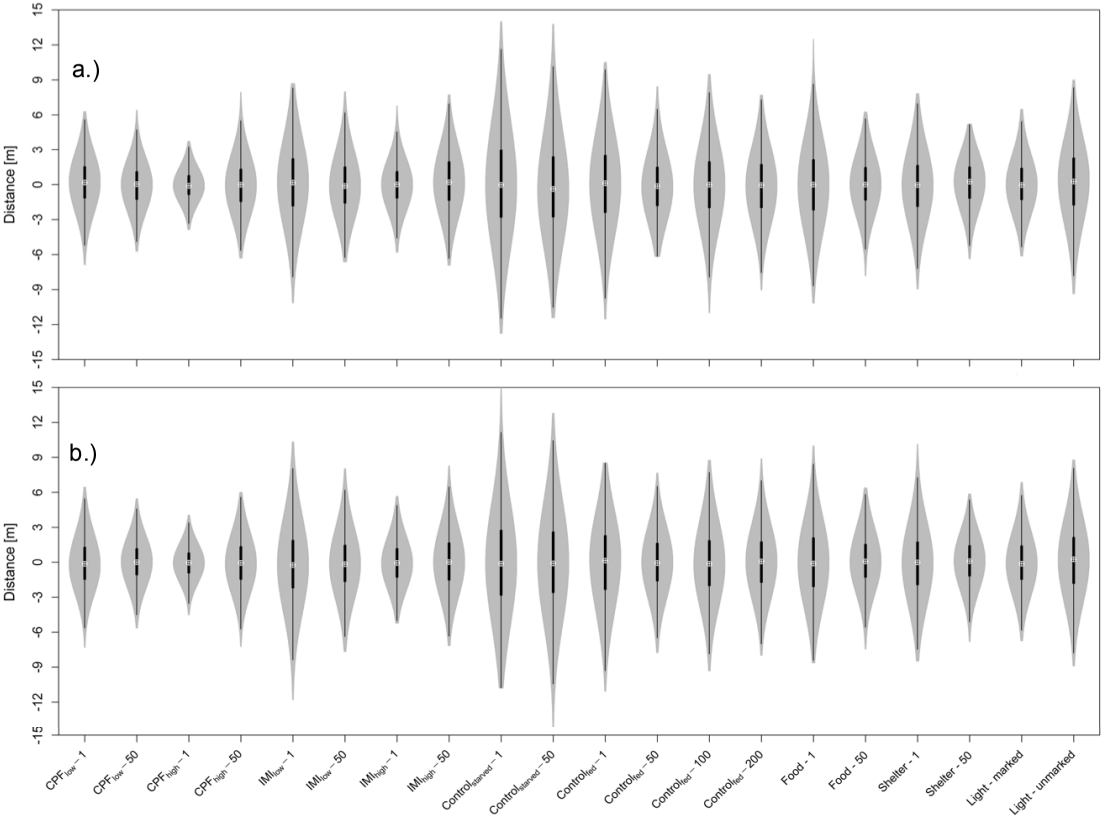


Figure S5.1: Pearson correlation coefficients observed in the different experimental setups.

Figure S5.2: Estimated 24-hour dispersal distances using parameters obtained from experimental observations. a.) Spatially unrestricted movement, fixed pairs of turning angles and step lengths drawn from experimental data sets; b.) Spatially restricted movement in a ditch scenario, fixed pairs of turning angles and step lengths drawn from experimental data sets; c.) Spatially unrestricted movement, random pairs of turning angles and step lengths drawn from experimental data sets; d.) Spatially unrestricted movement, fixed pairs of turning angles and step lengths drawn from experimental data sets, whereby individuals were assumed to be juveniles capable of covering only 10% of the step length of an adult.



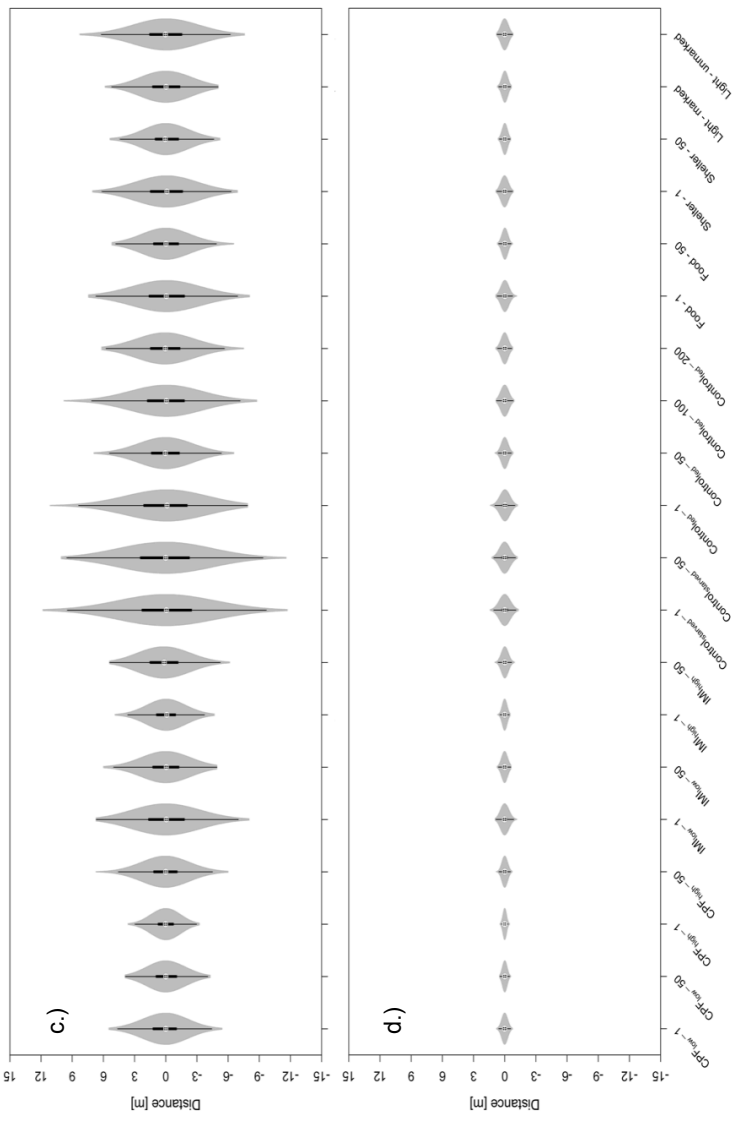


Figure S5-2 (continued)

Table S5.1: Parameters for the obtained 24h dispersal distributions assuming different environmental conditions (unit: meters)

	Unrestricted walk, no physical limitation, fixed angle-step length pairs		**Ditch scenario, no physical limitation, fixed angle-step length pairs**		**Unrestricted walk, physical limitation, fixed angle-step length pairs**		**Unrestricted walk, no physical limitation, random angle-step length pairs**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
CPF low - 1	0.15	2.03	-0.08	2.07	0.007	0.21	0.08	1.68
CPF low - 50	-0.03	1.75	0.01	1.71	0.001	0.17	-0.04	1.46
CPF high - 1	-0.05	1.25	-0.06	1.29	0.005	0.13	0.02	1.07
CPF high - 50	-0.02	2.10	-0.06	2.07	0.002	0.20	0.03	1.72
Imi low - 1	0.17	3.10	-0.21	3.14	-0.004	0.31	-0.09	2.52
Imi low - 50	-0.03	2.32	-0.12	2.41	0.010	0.23	0.00	1.88
Imi high - 1	-0.03	1.74	-0.01	1.85	0.010	0.18	-0.07	1.43
Imi high - 50	0.24	2.41	0.09	2.38	-0.010	0.24	0.15	1.89
Control (starved) - 1	0.10	4.21	-0.09	4.09	-0.002	0.40	-0.06	3.49
Control (starved) - 50	-0.13	3.98	-0.02	3.86	0.001	0.38	0.02	3.37
Control (fed) - 1	0.04	3.48	-0.06	3.41	-0.010	0.36	0.01	2.98
Control (fed) - 50	-0.08	2.47	-0.04	2.39	0.001	0.24	0.04	1.96
Control (fed) - 100	-0.01	3.01	-0.06	2.93	0.008	0.30	-0.02	2.61
Control (fed) - 200	-0.10	2.60	0.09	2.49	0.008	0.26	0.02	2.16
Food - 1	-0.05	3.03	-0.02	2.99	-0.014	0.29	-0.06	2.57
Food - 50	0.04	2.11	0.12	2.15	-0.007	0.22	-0.04	1.81
Shelter - 1	-0.05	2.68	-0.11	2.82	0.017	0.27	-0.05	2.26
Shelter - 50	0.14	1.98	0.08	1.93	0.002	0.20	-0.05	1.68
Light - marked	0.06	2.07	-0.06	2.16	-0.04	1.87	-0.006	0.22
Light - unmarked	0.19	2.92	0.18	2.84	0.02	2.38	0.013	0.28

LIST OF REFERENCES

Aarseth, K.A. & Schram, T.A. (1999). Wavelength-specific behaviour in Lepeophtheirus salmonis and Calanus finmarchicus to ultraviolet and visible light in laboratory experiments (Crustacea: Copepoda). *Marine Ecology Progress Series*, **186**, 211–217.

Aber, J.D. (1997). Why Don't We Believe the Models? *Bulletin of the Ecological Society of America*, **78**, 232–233.

Åbjörnsson, K., Wagner, B.M.A., Axelsson, A., Bjerselius, R. & Olsén, K.H. (1997). Responses of Acilius sulcatus (Coleoptera: Dytiscidae) to chemical cues from perch (Perca fluviatilis). *Oecologia*, **111**, 166–171.

Abramoff, M.D., Magalhães, P.J. & Ram, S.J. (2004). Image processing with ImageJ. *Biophotonics International*, **11**, 36–42.

Abuzaid, A.H., Mohamed, I.B. & Hussin, A.G. (2011). Boxplot for circular variables. *Computational Statistics*, **27**, 381–392.

Adcock, J.A. (1979). Energetics of a population of the isopod Asellus aquaticus: Life history and production. *Freshwater Biology*, **9**, 343–355.

Agatz, A., Ashauer, R. & Brown, C.D. (2014). Imidacloprid perturbs feeding of Gammarus pulex at environmentally relevant concentrations. *Environmental Toxicology and Chemistry*, **33**, 648–653.

Aiken, R.B. & Roughley, R.E. (1985). An Effective Trapping and Marking Method for Aquatic Beetles. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **137**, 5–7.

Albanese, B., Angermeier, P.L. & Peterson, J.T. (2009). Does mobility explain variation in colonisation and population recovery among stream fishes? *Freshwater Biology*, **54**, 1444–1460.

Allema, A.B., Rossing, W.A.H., Van der Werf, W., Heusinkveld, B.G., Bukovinszky, T., Steingröver, E. & Van Lenteren, J.C. (2012). Effect of light quality on movement of Pterostichus melanarius (Coleoptera: Carabidae). *Journal of Applied Entomology*, **136**, 793–800.

Amorim, M.J.B., Novais, S., Römbke, J. & Soares, A.M.V.M. (2008). Avoidance test with Enchytraeus albidus (Enchytraeidae): Effects of different exposure time and soil properties. *Environmental Toxicology*, **155**, 112–116.

Andrikovics, S. (1981). Further data to the daily migration of the larvae of aquatic insects. *Opuscula Zoologica*, **17**, 49–56.

Arakelova, K.S. (2001). The evaluation of individual production and scope for growth in aquatic sow bugs (Asellus aquaticus). *Aquatic Ecology*, **35**, 31–42.

Argent, R.M. (2004). An overview of model integration for environmental applications - components, frameworks and semantics. *Environmental Modelling & Software*, **19**, 219–234.

Arthur, J.D., Gröner, M.K., Hayhurst, K.J. & Holloway, C.M. (1999). Evaluating the effectiveness of independent verification and validation. *Computer*, **32**, 79–83.

Ashauer, R., Agatz, A., Albert, C., Ducrot, V., Galic, N., Hendriks, A.J., Jager, T., Kretschmann, A., O'Connor, I., Rubach, M.N., Nyman, A.-M., Schmitt, W., Stadnicka, J., van den Brink, P.J. & Preuss, T.G. (2011). Toxicokinetic-toxicodynamic modelling of quantal and graded sub-lethal endpoints - a brief discussion of concepts. *Environmental Toxicology and Chemistry*, **30**, 2519–2524.

Ashauer, R., Caravatti, I., Hintermeister, A. & Escher, B.I. (2010). Bioaccumulation kinetics of organic xenobiotic pollutants in the freshwater invertebrate Gammarus pulex modeled with prediction intervals. *Environmental Toxicology and Chemistry*, **29**, 1625–1636.

Augusiak, J. & Van den Brink, P.J. (2015). Studying the movement behavior of benthic macroinvertebrates with automated video tracking. *Ecology and Evolution*, **5**, 1563–1575.

Augusiak, J. & Van den Brink, P.J. (2016). The influence of insecticide exposure and environmental stimuli on the movement behaviour and dispersal of a freshwater isopod. *Ecotoxicology*.

Augusiak, J., Van den Brink, P.J. & Grimm, V. (2014). Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach. *Ecological Modelling*, **280**, 117–128.

Aumann, C.A. (2007). A methodology for developing simulation models of complex systems. *Ecological Modelling*, **202**, 385–396.

Azevedo-Pereira, H.M.V.S., Lemos, M.F.L. & Soares, A.M.V.M. (2011). Effects of imidacloprid exposure on Chironomus riparius Meigen larvae: Linking acetylcholinesterase activity to behaviour. *Ecotoxicology and Environmental Safety*, **74**, 1210–1215.

Baatrup, E. & Bayley, M. (1993). Effects of the Pyrethroid Insecticide Cypermethrin on the Locomotor Activity of the Wolf Spider Pardosa amentata: Quantitative Analysis Employing Computer-Automated Video Tracking. *Ecotoxicology and Environmental Safety*, **26**, 138–152.

Bacsi, Z. & Zemankovics, F. (1995). Validation: an objective or a tool? Results on a winter wheat simulation model application. *Ecological Modelling*, **81**, 251–263.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, **12**, 183–210.
Barlas, Y. & Carpenter, S. (1990). Philosophical roots of model validation: Two paradigms. *System Dynamics Review*, **6**, 148–166.
Bart, J. (1995). Acceptance Criteria for Using Individual-Based Models to Make Management Decisions. *Ecological Applications*, **5**, 411–420.
Batschelet, E. (1981). *Circular Statistics in Biology*. Academic Press.
Bayley, M., Baatrup, E. & Bjerregaard, P. (1997). Woodlouse locomotor behavior in the assessment of clean and contaminated field sites. *Environmental Toxicology and Chemistry*, **16**, 2309–2314.
Becher, M.A., Osborne, J.L., Thorbek, P., Kennedy, P.J. & Grimm, V. (2013). REVIEW: Towards a systems approach for understanding honeybee decline: a stocktaking and synthesis of existing models (I. Steffan-Dewenter, Ed.). *Journal of Applied Ecology*, **50**, 868–880.
Beck, M.B., Ravetz, J.R., Mulkey, L.A. & Barnwell, T. (1997). On the problem of model validation for predictive exposure assessments. *Stochastic Hydrology and Hydraulics*, **11**, 229–254.
Beketov, M.A. & Liess, M. (2008). Potential of 11 pesticides to initiate downstream drift of stream macroinvertebrates. *Archives of Environmental Contamination and Toxicology*, **55**, 247–253.
Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A. & Fath, B.D. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, **40**, 1–20.
Von Bertalanffy, L. (1957). Quantitative laws in metabolism and growth. *The Quarterly review of biology*, **32**, 217–231.
Beven, K.J. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, **320**, 18–36.
Beven, K. (2002). Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system. *Hydrological Processes*, **16**, 189–206.
Blockwell, S.J., Maund, S.J. & Pascoe, D. (1998). The Acute Toxicity of Lindane to Hyalella azteca and the Development of a Sublethal Bioassay Based on Precopulatory Guarding Behavior. *Archives of Environmental Contamination and Toxicology*, **35**, 432–440.
Boesten, J.J.T.I. (2000). Modeller subjectivity in estimating pesticide parameters for leaching models using the same laboratory data set. *Agricultural Water Management*, **44**, 389–409.
Boesten, J.J.T.I., Businelli, M., Delmas, A., Edwards, V., Helweg, A., Jones, R., Klein, M., Kloskowski, R., Layton, R., Marcher, S., Schäfer, H., Smeets, L., Russell, M., Styzcen, M., Travis, K., Walker, A. & Yon, D. (1995). *Leaching models and EU registration*.
Bogoch, I.I., Brady, O.J., Kraemer, M.U.G., German, M., Creatore, M.I., Kulkarni, M.A., Brownstein, J.S., Mekaru, S.R., Hay, S.I., Groot, E., Watts, A. & Khan, K. (2016). Anticipating the international spread of Zika virus from Brazil. *The Lancet*, **387**, 335–336.
Borenstein, D. (1998). Towards a practical method to validate decision support systems. *Decision Support Systems*, **23**, 227–239.
Boström, C., Törnroos, A. & Bonsdorff, E. (2010). Invertebrate dispersal and habitat heterogeneity: Expression of biological traits in a seagrass landscape. *Journal of Experimental Marine Biology and Ecology*, **390**, 106–117.
Botkin, D.B. (1993). *Forest Dynamics: An Ecological Model*. Oxford University Press, USA, New York, NY.
Böttger, R., Feibicke, M., Schaller, J. & Dudel, G. (2013). Effects of low-dosed imidacloprid pulses on the functional role of the caged amphipod Gammarus roeseli in stream mesocosms. *Ecotoxicology and Environmental Safety*, **93**, 93–100.
Brederveld, R.J., Jähnig, S.C., Lorenz, A.W., Brunzel, S. & Soons, M.B. (2011). Dispersal as a limiting factor in the colonization of restored mountain streams by plants and macroinvertebrates. *Journal of Applied Ecology*, **48**, 1241–1250.
Van den Brink, P.J., Baird, D.J., Baveco, J.M. (Hans) & Focks, A. (2013). The use of traits-based approaches and eco(toxico)logical models to advance the ecological risk assessment framework for chemicals. *Integrated Environmental Assessment and Management*, **9**, e47–e57.
Van den Brink, P.J., Baveco, J.M. (Hans), Verboom, J. & Heimbach, F. (2007). An individual-based approach to model spatial population dynamics of invertebrates in aquatic ecosystems after pesticide contamination. *Environmental Toxicology and Chemistry*, **26**, 2226–2236.
Van den Brink, P.J., Van Smeden, J.M., Bekele, R.S., Dierick, W., De Gelder, D., Noteboom, M. & Roessink, I. (2015). Acute and chronic toxicity of neonicotinoids to nymphs of a mayfly species and some notes on seasonal differences. *Environmental Toxicology and Chemistry*.
Brock, T.C.M., Arts, G.H.P., Maltby, L. & Van den Brink, P.J. (2006). Aquatic risks of pesticides, ecological protection goals, and common aims in European Union legislation. *Integrated Environmental Assessment and Management*, **2**, 20–46.

Brooks, A.C., Gaskell, P.N. & Maltby, L. (2009). Sublethal effects and predator-prey interactions: implications for ecological risk assessment. *Environmental Toxicology and Chemistry*, **28**, 2449–2457.

Brown, A.R., Whale, G., Jackson, M., Marshall, S., Hamer, M., Solga, A., Kabouw, P., Galay-Burgos, M., Woods, R., Nadzialek, S. & Maltby, L. (2016). Towards the Definition of Specific Protection Goals for the Environmental Risk Assessment of Chemicals: A Perspective on Environmental Regulation in Europe. *Integrated Environmental Assessment and Management*.

Bundschuh, M., Appeltauer, A., Dabrunz, A. & Schulz, R. (2012). Combined effect of invertebrate predation and sublethal pesticide exposure on the behavior and survival of Asellus aquaticus (crustacea; isopoda). *Archives of Environmental Contamination and Toxicology*, **63**, 77–85.

Cailleaud, K., Michalec, F.-G., Forget-Leray, J., Budzinski, H., Hwang, J.-S., Schmitt, F.G. & Souissi, S. (2011). Changes in the swimming behavior of Eurytemora affinis (Copepoda, Calanoida) in response to a sub-lethal exposure to nonylphenols. *Aquatic Toxicology*, **102**, 228–231.

Cain, M.L. (1989). The analysis of angular data in ecological field studies. *Ecology*, **70**, 1540–1543.

Calenge, C. (2006). The package 'adehabitat' for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling*, **197**, 516–519.

Cant, E.T., Smith, A.D.M., Reynolds, D.R. & Osborne, J.L. (2005). Tracking butterfly flight paths across the landscape with harmonic radar. *Proceedings of the Royal Society - B: Biological Sciences*, **272**, 785–790.

Caswell, H. (1976). The validation problem. *Systems Analysis and Simulation in Ecology* (ed B. Patten), pp. 313–325. Academic Press, New York, NY.

Statistics Netherlands (CBS), 2008. Plant Protection Use Survey The Hague, Netherlands. http://www.cbs.nl/nl-NL/menu/themas/natuur-milieu/cijfers/ bestrijdingsmiddelen/default.htm. Swift,

Céréghino, R., Giraudel, J.L. & Compin, A. (2001). Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. *Ecological Modelling*, **146**, 167–180.

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition*, **25**, 203–226.

Chambers, M.R. (1977). A comparison of the population ecology of Asellus aquaticus (L.) and Asellus meridianus rac. in the reed beds of the Tjeukemeer. *Hydrobiologia*, **53**, 147–154.

Charoy, C.P. (1995). Modification of the swimming behaviour of Brachionus calyciflorus (Pallas) according to food environment and individual nutritive state. *Hydrobiologia*, **313-314**, 197–204.

Charoy, C.P. & Janssen, C.R. (1999). The swimming behaviour of Brachionus calyciflorus (rotifer) under toxic stress. *Chemosphere*, **38**, 3247–3260.

Charoy, C.P., Janssen, C.R., Persoone, G. & Clément, P. (1995). The swimming behaviour of Brachionus calyciflorus (rotifer) under toxic stress. I. The use of automated trajectometry for determining sublethal effects of chemicals. *Aquatic Toxicology*, **32**, 271–282.

Chen, J., Wang, Z., Li, G. & Guo, R. (2014). The swimming speed alteration of two freshwater rotifers Brachionus calyciflorus and Asplanchna brightwelli under dimethoate stress. *Chemosphere*, **95**, 256–60.

Chevalier, J., Harscoët, E., Keller, M., Pandard, P., Cachot, J. & Grote, M. (2015). Exploration of Daphnia behavioral effect profiles induced by a broad range of toxicants with different modes of action. *Environmental Toxicology and Chemistry*, **34**, 1760–1769.

Clausing, P. (2015). The EFSA Conclusion on the Peer Review of the Glyphosate Risk Assessment A Reality Check Background.

Collin, S.P. & Marshall, N.J. (2003). *Sensory Processing in Aquatic Environments* (S.P. Collin & N.J. Marshall, Eds.). Springer New York, New York, NY.

Committee on Models in the Regulatory Decision Process. (2007). *Models in Environmental Regulatory Decision Making*. National Academies Press, Washington, D.C.

Conroy, M.J., Cohen, Y., James, F.C., Matsinos, Y.G. & Maurer, B.A. (2010). Parameter Estimation, Reliability, and Model Improvement for Spatially Explicit Models of Animal Populations. *Ecological Applications*, **5**, 17–19.

Council Directive 91/414/EC of 15 July 1991 concerning the placing of plant protection products on the market. (1991). European Parliament and The Council, European Union.

Courbaud, B., Lafond, V., Lagarrigues, G., Vieilledent, G., Cordonnier, T., Jabot, F. & De Coligny, F. (2015). Applying ecological model evaludation: Lessons learned with the forest dynamics model Samsara2. *Ecological Modelling*, **314**, 1–14.

Cramer, W., Bondeau, A., Woodward, F.I., Prentice, I.C., Betts, R.A., Brovkin, V., Cox, P.M., Fisher, V., Foley, J.A., Friend, A.D., Kucharik, C., Lomas, M.R., Ramankutty, N., Sitch, S., Smith, B., White, A. & Young-Molling, C. (2001). Global response of terrestrial ecosystem structure and function to CO 2 and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357–373.

Creed, R.P. & Miller, J.R. (1990). Interpreting animal wall-following behavior. *Experientia*, **46**, 758–761.

Crout, N., Kokkonen, T., Jakeman, A.J., Norton, J.P., Anderson, R., Assaf, H., Gaber, N., Gibbons, J., Holzworth, D., Mysiak, J., Reichl, J., Seppelt, R., Wagener, T. & Whitfield, P. (2008). Good Modelling Practice. *Environmental Modelling & Software*, **3**, 15–32.
Davy-Bowker, J. (2002). A mark and recapture study of water beetles (Coleoptera: Dytiscidae) in a group of semi-permanent and temporary ponds. *Aquatic Ecology*, **36**, 435–446.
Dawidowicz, P., Pijanowska, J. & Ciechomski, K. (1990). Vertical migration of Chaoborus larvae is induced by the presence of fish. *Limnology and Oceanography*, **35**, 1631–1637.
Delgado, M. del M., Penteriani, V., Morales, J.M., Gurarie, E. & Ovaskainen, O. (2013). A Statistical Framework for Inferring the Influence of Conspecifics on Movement Behaviour. *Methods in Ecology and Evolution*, n/a–n/a.
Dell'Omo, G. (2002). *Behavioural Ecotoxicology*. John Wiley & Sons.
Denoël, M., Libon, S., Kestemont, P., Brasseur, C., Focant, J.-F. & De Pauw, E. (2013). Effects of a sublethal pesticide exposure on locomotor behavior: A video-tracking analysis in larval amphibians. *Chemosphere*, **90**, 945–951.
Van Dijk, T.C., Van Staalduinen, M.A. & Van der Sluijs, J.P. (2013). Macro-invertebrate decline in surface water polluted with imidacloprid (N. Desneux, Ed.). *PLoS ONE*, **8**, e62374.
Dodson, S.I., Hanazato, T. & Gorski, P.R. (1995). Behavioral responses of Daphnia pulex exposed to carbaryl and Chaoborus kairomone. *Environmental Toxicology and Chemistry*, **14**, 43–50.
Dohmen, G.P., Preuss, T.G., Hamer, M., Galic, N., Strauss, T., Van den Brink, P.J., De Laender, F. & Bopp, S. (2015). Population-level effects and recovery of aquatic invertebrates after multiple applications of an insecticide. *Integrated Environmental Assessment and Management*, **9999**, 1–15.
Dray, S., Royer-Carenzi, M. & Calenge, C. (2010). The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, **25**, 673–681.
Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J. & Sullivan, C.A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, **81**, 163–182.
Van Dyck, H. & Baguette, M. (2005). Dispersal behaviour in fragmented landscapes: Routine or special movements? *Basic and Applied Ecology*, **6**, 535–545.
(EC) European Commission. 1991. Council Directive 91/414/EEC of 15 July 1991 concerning the placing of plant protection products on the market. Official Journal L 230; p 1–32
(EC) European Parliament and The Council. (2009). REGULATION (EC) No 1107/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 October 2009. *EU Parliament*, **309**, 1–50.
Edwards, A.M., Phillips, R.A., Watkins, N.W., Freeman, M.P., Murphy, E.J., Afanasyev, V., Buldyrev, S. V., Da Luz, M.G.E., Raposo, E.P., Stanley, H.E. & Viswanathan, G.M. (2007). Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, **449**, 1044–1048.
Englund, G. & Hambäck, P.A. (2004). Scale-dependence of movement rates in stream invertebrates. *Oikos*, **105**, 31–40.
Ensminger, M.P., Budd, R., Kelley, K.C. & Goh, K.S. (2013). Pesticide occurrence and aquatic benchmark exceedances in urban surface waters and sediments in three urban areas of California, USA, 2008-2011. *Environmental Monitoring and Assessment*, **185**, 3697–3710.
European Parliament and The Council. (2009). REGULATION (EC) No 1107/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 October 2009. *EU Parliament*, **309**, 1–50.
Fagerstrom, T. (1987). On theory, data and mathematics in ecology. *Oikos*, **50**, 258–261.
Faimali, M., Garaventa, F., Piazza, V., Greco, G., Corrà, C., Magillo, F., Pittore, M., Giacco, E., Gallus, L., Falugi, C. & Tagliafierro, G. (2006). Swimming speed alteration of larvae of Balanus amphitrite as a behavioural end-point for laboratory toxicological bioassays. *Marine Biology*, **149**, 87–96.
Ferland-Raymond, B., March, R.E., Metcalfe, C.D. & Murray, D.L. (2010). Prey detection of aquatic predators: Assessing the identity of chemical cues eliciting prey behavioral plasticity. *Biochemical Systematics and Ecology*, **38**, 169–177.
Ferson, S. (1996). Automated quality assurance checks on model structure in ecological risk assessments. *Human and Ecological Risk Assessment*, **2**, 558–569.
Ferson, S., Oberkampf, W.L. & Ginzburg, L. (2008). Model validation and predictive capability for the thermal challenge problem. *Computer Methods in Applied Mechanics and Engineering*, **197**, 2408–2430.
Focks, A., Ter Horst, M.M.S., Van den Berg, E., Baveco, J.M. (Hans) & Van den Brink, P.J. (2014a). Integrating chemical fate and population-level effect models for pesticides at landscape scale: New options for risk assessment. *Ecological Modelling*, **280**, 102–116.
Focks, A., ter Horst, M., van den Berg, E., Baveco, H. & van den Brink, P.J. (2014b). Integrating chemical fate and population-level effect models for pesticides at landscape scale: New options for risk assessment. *Ecological Modelling*, **280**, 102–116.

Forbes, V.E. & Calow, P. (2012). Promises and problems for the new paradigm for risk assessment and an alternative approach involving predictive systems models. *Environmental Toxicology and Chemistry*, **31**, 2663–71.

Forbes, V.E., Calow, P., Grimm, V., Hayashi, T.I., Jager, T., Katholm, A., Palmqvist, A., Pastorok, R., Salvito, D., Sibly, R., Spromberg, J., Stark, J. & Stillman, R. a. (2011). Adding Value to Ecological Risk Assessment with Population Modeling. *Human and Ecological Risk Assessment: An International Journal*, **17**, 287–299.

Forbes, V.E., Calow, P., Grimm, V., Hayashi, T., Jager, T., Palmqvist, A., Pastorok, R., Salvito, D., Sibly, R., Spromberg, J., Stark, J. & Stillman, R.A. (2010). Integrating population modeling into ecological risk assessment. *Integrated Environmental Assessment and Management*, **6**, 191–193.

Forbes, V.E., Galic, N., Schmolke, A., Vavra, J., Pastorok, R. & Thorbek, P. (2016). Assessing the risks of pesticides to threatened and endangered species using population modeling: A critical review and recommendations for future work. *Environmental Toxicology and Chemistry*.

Forbes, V.E., Hommen, U., Thorbek, P., Heimbach, F., van den Brink, P.J., Wogram, J., Thulke, H.-H. & Grimm, V. (2009). Ecological models in support of regulatory risk assessments of pesticides: developing a strategy for the future. *Integrated Environmental Assessment and Management*, **5**, 167–172.

Forman, R.T.T. & Alexander, L.E. (1998). Roads and Their Major Ecological Effects. **29**, 207–231.

Frank, T.M. & Widder, E.A. (1994). Evidence for behavioral sensitivity to near-UV light in the deep-sea crustacean Systellaspis debilis. *Marine Biology*, **118**, 279–284.

Freilich, J.E. (1989). A Method for Tagging Individual Benthic Macroinvertebrates. *Journal of the North American Benthological Society*, **8**, 351–354.

Galic, N., Baveco, J.M. (Hans), Hengeveld, G.M., Thorbek, P., Bruns, E. & Van den Brink, P.J. (2012). Simulating population recovery of an aquatic isopod: Effects of timing of stress and landscape structure. *Environmental Pollution*, **163**, 91–99.

Galic, N., Hengeveld, G.M., Van den Brink, P.J., Schmolke, A., Thorbek, P., Bruns, E. & Baveco, J.M. (Hans). (2013). Persistence of aquatic insects across managed landscapes: effects of landscape permeability on re-colonization and population recovery (C.G. de Leaniz, Ed.). *PloS one*, **8**, e54584.

Galic, N., Hommen, U., Baveco, J.M. (Hans) & Van den Brink, P.J. (2010). Potential application of population models in the European ecological risk assessment of chemicals II: Review of models and their potential to address environmental protection aims. *Integrated Environmental Assessment and Management*, **6**, 338–360.

Garaventa, F., Gambardella, C., Di Fino, A., Pittore, M. & Faimali, M. (2010). Swimming speed alteration of Artemia sp. and Brachionus plicatilis as a sub-lethal behavioural end-point for ecotoxicological surveys. *Ecotoxicology*, **19**, 512–519.

García-de la Parra, L.M., Bautista-Covarrubias, J.C., Rivera-de la Rosa, N., Betancourt-Lozano, M. & Guilhermino, L. (2006). Effects of methamidophos on acetylcholinesterase activity, behavior, and feeding rate of the white shrimp (Litopenaeus vannamei). *Ecotoxicology and Environmental Safety*, **65**, 372–380.

Gass, S.I. (1983). Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Operations Research*, **31**, 603–631.

Getz, W.M. & Saltz, D. (2008). A framework for generating and analyzing movement paths on ecological landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19066–19071.

Goldsmith, T.H. & Fernandez, H.R. (1968). Comparative studies of crustacean spectral sensitivity. *Zeitschrift für Vergleichende Physiologie*, **60**, 156–175.

Goodall, D.W. (1972). Building and testing ecosystem models. *Mathematical models in ecology* pp. 173–194.

Graça, M.A.S., Maltby, L. & Calow, P. (1993). Importance of Fungi in the Diet of Gammarus-Pulex and Asellus-Aquaticus: II. Effects on Growth, Reproduction and Physiology. *Oecologia*, **96**, 304–309.

Grimm, V. (2008). Individual-Based Models. *Encyclopedia of Ecology* (eds S.E. Jorgenson & B. Fath), pp. 1959–1968. Academic Press, Oxford.

Grimm, V., Ashauer, R., Forbes, V.E., Hommen, U., Preuss, T.G., Schmidt, A., Van den Brink, P.J., Wogram, J. & Thorbek, P. (2009). CREAM: A European project on mechanistic effect models for ecological risk assessment of chemicals. *Environmental Science and Pollution Research*, **16**, 614–617.

Grimm, V., Augusiak, J., Focks, A., Frank, B.M., Gabsi, F., Johnston, A.S.A., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P. & Railsback, S.F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, **280**, 129–139.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U. & DeAngelis, D.L. (2006). A standard protocol for describing

individual-based and agent-based models. *Ecological Modelling*, **198**, 115–126.
Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J. & Railsback, S.F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, **221**, 2760–2768.
Grimm, V. & Martin, B.T. (2013). Mechanistic effect modeling for ecological risk assessment: where to go from here? *Integrated environmental assessment and management*.
Grimm, V. & Railsback, S.F. (2005). *Individual-based Modeling and Ecology*. Princeton University Press, Princeton, NJ (USA).
Grimm, V. & Railsback, S.F. (2012). Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 298–310.
Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T. & DeAngelis, D.L. (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science*, **310**, 987–991.
Grimm, V. & Wissel, C. (1997). Babel, or the ecological stability discussions: an inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia*, **109**, 323–334.
Guo, R., Ren, X. & Ren, H. (2012). Assessment the toxic effects of dimethoate to rotifer using swimming behavior. *Bulletin of Environmental Contamination and Toxicology*, **89**, 568–571.
Gurarie, E., Andrews, R.D. & Laidre, K.L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, **12**, 395–408.
Hagler, J.R. & Jackson, C.G. (2001). Methods for marking insects: current techniques and future prospects. *Annual Review of Entomology*, **46**, 511–543.
Hamda, N.T., Forbes, V.E., Stark, J.D. & Laskowski, R. (2014). Stochastic density-dependent matrix model for extrapolating individual-level effects of chemicals to the population: Case study on effects of Cd on Folsomia candida. *Ecological Modelling*, **280**, 53–64.
Hawkes, C. (2009). Linking movement behaviour, dispersal and population processes: Is individual variation a key? *Journal of Animal Ecology*, **78**, 894–906.
Hintze, J.L. & Nelson, R.D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, **52**, 181–184.
Hodges, K.E. (2008). Defining the problem: terminology and progress in ecology. *Frontiers in Ecology and the Environment*, **6**, 35–42.
Hodges, J.S. (1991). Six (Or So) Things You Can Do with a Bad Model. *Operations Research*, **39**, 355–365.
Hodges, J.S. & Dewar, J.A. (1992). *Is It You or Your Model Talking? - A Framework for Model Validation*. Santa Monica, CA: Rand.
Holdo, R.M. & Roach, R.R. (2013). Inferring animal population distributions from individual tracking data: theoretical insights and potential pitfalls. *Journal of Animal Ecology*, **82**, 175–81.
Holling, C.S. (1978). *Adaptive Environmental Assessment and Management*. John Wiley & Sons, New York, NY.
Holling, C.S. (1966). The Functional Response of Invertebrate Predators to Prey Density. *Memoirs of the Entomological Society of Canada*, **98**, 5–86.
Holomuzki, J.R. & Biggs, B.J.F. (2000). Taxon-Specific Responses to High-Flow Disturbance in Streams: Implications for Population Persistence. *Journal of the North American Benthological Society*, **19**, 670–679.
Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E. & Spiegel, O. (2008). Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19060–19065.
Hommen, U., Baveco, J.M. (Hans), Galic, N. & Van den Brink, P.J. (2010). Potential application of ecological models in the European environmental risk assessment of chemicals I: Review of protection goals in EU directives and regulations. *Integrated Environmental Assessment and Management*, **6**, 325–337.
Hommen, U., Forbes, V., Grimm, V., Preuss, T.G., Thorbek, P. & Ducrot, V. (2016). How to use mechanistic effect models in environmental risk assessment of pesticides: Case studies and recommendations from the SETAC workshop MODELINK. *Integrated Environmental Assessment and Management*, **12**, 21–31.
Hoover, S. & Perry, R.F. (1989). *Simulation: a problem-solving approach*. Addison-Wesley, Reading, MA.
Humphries, N.E., Weimerskirch, H., Queiroz, N., Southall, E.J. & Sims, D.W. (2012). Foraging success of biological Lévy flights recorded in situ. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 7169–7174.
Hunka, A.D., Meli, M., Thit, A., Palmqvist, A., Thorbek, P. & Forbes, V.E. (2013a). Stakeholders' perspective on ecological modeling in environmental risk assessment of pesticides: challenges and opportunities. *Risk Analysis*, **33**, 68–79.
Hunka, A.D., Palmqvist, A., Thorbek, P. & Forbes, V.E. (2013b). Risk communication discourse among ecological risk assessment professionals and its implications for communication with nonexperts. *Integrated Environmental Assessment and Management*, **9**, 616–622.

Hynes, H.B.N. & Williams, W.D. (1965). Experiments on competition between two Asellus species (Isopoda Crustacea). *Hydrobiologia*, **26**, 203–210.

Ibrahim, L., Preuss, T.G., Schaeffer, A. & Hommen, U. (2014). A contribution to the identification of representative vulnerable fish species for pesticide risk assessment in Europe-A comparison of population resilience using matrix models. *Ecological Modelling*, **280**, 65–75.

Iversen, T.M. & Thorup, J. (1988). A Three Years' Study of Life Cycle, Population Dynamics and Production ofAsellus aquaticus L. in a Macrophyte Rich Stream. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, **73**, 73–94.

Jager, T., Crommentuijn, T., Van Gestel, C.A.M. & Kooijman, S.A.L.M. (2004). Simultaneous Modeling of Multiple End Points in Life-Cycle Toxicity Tests. *Environmental Science & Technology*, **38**, 2894–2900.

De Jager, M., Bartumeus, F., Kölzsch, A., Weissing, F.J., Hengeveld, G.M., Nolet, B.A., Herman, P.M.J. & Van de Koppel, J. (2013). How superdiffusion gets arrested: ecological encounters explain shift from Lévy to Brownian movement. *Proceedings of the Royal Society - B: Biological Sciences*, **281**, 20132605.

Jakeman, A.J., Letcher, R.A. & Norton, J.P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, **21**, 602–614.

Janssen, C.R., Ferrando, M.D. & Persoone, G. (1994). Ecotoxicological studies with the freshwater rotifer Brachionus calyciflorus. IV. Rotifer behavior as a sensitive and rapid sublethal test criterion. *Ecotoxicology and Environmental Safety*, **28**, 244–255.

Jax, K. (2008). Concepts, not terms. *Frontiers in Ecology and the Environment*, **6**, 178–179.

Jin, D., Hoagland, P. & Morin Dalton, T. (2003). Linking economic and ecological models for a marine ecosystem. *Ecological Economics*, **46**, 367–385.

Johnston, A.S.A., Hodson, M.E., Thorbek, P., Alvarez, T. & Sibly, R.M. (2014). An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides. *Ecological Modelling*, **280**, 5–17.

Jones, T.T., Van Houtan, K.S., Bostrom, B.L., Ostafichuk, P., Mikkelsen, J., Tezcan, E., Carey, M., Imlach, B. & Seminoff, J.A. (2013). Calculating the ecological impacts of animal-borne instruments on aquatic organisms (S. Rands, Ed.). *Methods in Ecology and Evolution*, **4**, 1178–1186.

Källén, A., Arcuri, P. & Murray, J.D. (1985). A simple model for the spatial spread and control of rabies. *Journal of Theoretical Biology*, **116**, 377–393.

Kanare, H.M. (1985). *Writing the Laboratory Notebook.* American Chemical Society, 1155 16th St., N.W., Washington, DC 20036 (paperback--ISBN-0-8412-0933-2, $14.95; clothbound--ISBN-0-8412-0906-5, $24.95).

Kareiva, P.M. & Shigesada, N. (1983). Analyzing Insect Movement as a Correlated Random Walk. *Oecologia*, **56**, 234–238.

Kirchner, J.W., Hooper, R.P., Kendall, C., Neal, C. & Leavesley, G. (1996). Testing and validating environmental models. *Science of the Total Environment*, **183**, 33–47.

Klügl, F. (2008). A validation methodology for agent-based simulations. *Proceedings of the 2008 ACM symposium on Applied computing - SAC '08* p. 39. ACM Press, New York, New York, USA.

Van de Koppel, J., Gascoigne, J.C., Theraulaz, G., Rietkerk, M., Mooij, W.M. & Herman, P.M.J. (2008). Experimental evidence for spatial self-organization and its emergent effects in mussel bed ecosystems. *Science*, **322**, 739–742.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. University of Chicago Press, Chicago.

Kulakowska, K.A., Kulakowski, T.M., Inglis, I.R., Smith, G.C., Haynes, P.J., Prosser, P., Thorbek, P. & Sibly, R.M. (2014). Using an individual-based model to select among alternative foraging strategies of woodpigeons: Data support a memory-based model with a flocking mechanism. *Ecological Modelling*, **280**, 89–101.

Kulkarni, D., Hommen, U., Schäffer, A. & Preuss, T.G. (2014). Ecological interactions affecting population-level responses to chemical stress in Mesocyclops leuckarti. *Chemosphere*, **112**, 340–347.

Landry, M., Malouin, J.-L. & Oral, M. (1983). Model validation in operations research. *European Journal of Operational Research*, **14**, 207–220.

De Lange, H.J., Noordoven, W., Murk, A.J., Lürling, M. & Peeters, E.T.H.M. (2006a). Behavioural responses of Gammarus pulex (Crustacea, Amphipoda) to low concentrations of pharmaceuticals. *Aquatic Toxicology*, **78**, 209–216.

De Lange, H.J., Peeters, E.T.H.M. & Lürling, M. (2009). Changes in Ventilation and Locomotion of Gammarus pulex (Crustacea, Amphipoda) in Response to Low Concentrations of Pharmaceuticals. *Human and Ecological Risk Assessment*, **15**, 111–120.

De Lange, H.J., Sperber, V. & Peeters, E.T.H.M. (2006b). Avoidance of polycyclic aromatic hydrocarbon-contaminated by the freshwater invertebrates Gammarus pulex and Asellus aquaticus. *Environmental Toxicology and Chemistry*, **25**, 452–457.

Latombe, G., Parrott, L. & Fortin, D. (2011). Levels of emergence in individual based models: Coping with scarcity of data and pattern redundancy. *Ecological Modelling*, **222**, 1557–1568.
Lev-Ari, S. & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, **46**, 1093–1096.
Li, X. (1988). Effects of Contextual Cues on Inferring and Remembering Meanings of New Words. *Applied Linguistics*, **9**, 402–413.
Liu, C., Sibly, R.M., Grimm, V. & Thorbek, P. (2013). Linking pesticide exposure and spatial dynamics: An individual-based model of wood mouse (Apodemus sylvaticus) populations in agricultural landscapes. *Ecological Modelling*, **248**, 92–102.
Loizou, G., Spendiff, M., Barton, H.A., Bessems, J., Bois, F.Y., D'Yvoire, M.B., Buist, H., Clewell, H.J., Meek, B., Gundert-Remy, U., Goerlitz, G. & Schmitt, W. (2008). Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps. *Regulatory Toxicology and Pharmacology: RTP*, **50**, 400–411.
Lukančič, S., Žibrat, U., Mezek, T., Jerebic, A., Simčič, T. & Brancelj, A. (2010). A new method for early assessment of effects of exposing two non-target crustacean species, Asellus aquaticus and Gammarus fossarum, to pesticides, a laboratory study. *Toxicology and Industrial Health*, **26**, 217–228.
Macal, C.M. (2005). Model Verification and Validation.
Macedo-Sousa, J.A., Gerhardt, A., Brett, C.M.A., Nogueira, A.J.A. & Soares, A.M.V.M. (2008). Behavioural responses of indigenous benthic invertebrates (Echinogammarus meridionalis, Hydropsyche pellucidula and Choroterpes picteti) to a pulse of Acid Mine Drainage: a laboratorial study. *Environmental Pollution*, **156**, 966–973.
MacGillivray, B.H. & Richards, K. (2015). Approaches to evaluating model quality across different regime types in environmental and public health governance. *Global Environmental Change*, **33**, 23–31.
Malmqvist, B. (2002). Aquatic invertebrates in riverine landscapes. *Freshwater Biology*, **47**, 679–694.
Mandelbrot, B.B. (1967). How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science*, **156**, 636–638.
Mankin, J.B., O'Neill, R.V., Shugart, H.H. & Rust, B.W. (1977). The importance of validation in ecosystem analysis. *New Directions in the Analysis of Ecological Systems, Part 1* (ed G. Innis), pp. 63–71. The Society for Computer Simulation, La Jolla, CA.
Marcus, J.H., SUTCLIFFE, D.W. & WILLOUGHBY, L.G. (1978). Feeding and growth of Asellus aquaticus (Isopoda) on food items from the littoral of Windermere, including green leaves of Elodea canadensis. *Freshwater Biology*, **8**, 505–519.
Marino, D. & Ronco, A. (2005). Cypermethrin and Chlorpyrifos Concentration Levels in Surface Water Bodies of the Pampa Ondulada, Argentina. *Bulletin of Environmental Contamination and Toxicology*, **75**, 820–826.
Marsh, L.M. & Jones, R.E. (1988). The form and consequences of random walk movement models. *Journal of Theoretical Biology*, **133**, 113–131.
Martin, J.-R. (2004). A portrait of locomotor behaviour in Drosophila determined by a video-tracking paradigm. *Behavioural Processes*, **67**, 207–219.
Martin, B.T., Jager, T., Nisbet, R.M., Preuss, T.G. & Grimm, V. (2013). Predicting population dynamics from the properties of individuals: a cross-level test of dynamic energy budget theory. *The American naturalist*, **181**, 506–519.
Martin, B.T., Zimmer, E.I., Grimm, V. & Jager, T. (2012). Dynamic Energy Budget theory meets individual-based modelling: a generic and accessible implementation. *Methods in Ecology and Evolution*, **3**, 445–449.
Matott, L.S., Babendreier, J.E. & Purucker, S.T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research*, **45**, W06421.
Matsuda, K., Buckingham, S.D., Kleier, D., Rauh, J.J., Grauso, M. & Sattelle, D.B. (2001). Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors. *Trends in Pharmacological Sciences*, **22**, 573–580.
May, R.M. (1973). *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, NJ (USA).
Mayer, D.G. & Butler, D.G. (1993). Statistical validation. *Ecological Modelling*, **68**, 21–32.
Meli, M., Auclerc, A., Palmqvist, A., Forbes, V.E. & Grimm, V. (2013). Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: An individual-based model of springtails. *Ecological Modelling*, **250**, 338–351.
Meli, M., Palmqvist, A., Forbes, V.E. & Groeneveld, J. (2014). Two pairs of eyes are better than one: Combining individual-based and matrix models for ecological risk assessment of chemicals. *Ecological Modelling*, **280**, 40–52.
Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*.

Moon, H.P. (1968). The Colonization of Esthwaite Water and Ullswater, English Lake District, by Asellus (Crustacea, Isopoda). *Journal of Animal Ecology*, **37**, 405.

Muschal, M. & Warne, M.S.J. (2003). Risk Posed by Pesticides to Aquatic Organisms in Rivers of Northern Inland New South Wales, Australia. *Human and Ecological Risk Assessment: An International Journal*, **9**, 1765–1787.

Nams, V.O. (1996). The VFractal: a new estimator for fractal dimension of animal movement paths. *Landscape Ecology*, **11**, 289–297.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19052–19059.

Nautiyal, S., Rao, K.S., Kaechele, H. & Zander, P. (2010). Conceptual model development for landscape management in the mountains of the Indian Himalayan region : an approach for sustainable. *Online*, 1–19.

Neuert, C., Rademacher, C., Grundmann, V., Wissel, C. & Grimm, V. (2001). Struktur und Dynamik von Buchenurwäldern: Ergebnisse des regelbasierten Modells BEFORE. *Naturschutz und Landschaftsplanung*, **33**, 173–183.

Niemi, G.J., DeVore, P., Detenbeck, N., Taylor, D., Lima, A., Pastor, J., Yount, J.D. & Naiman, R.J. (1990). Overview of case studies on recovery of aquatic systems from disturbance. *Environmental Management*, **14**, 571–587.

Nørum, U., Frederiksen, M.A.T. & Bjerregaard, P. (2011). Locomotory behaviour in the freshwater amphipod Gammarus pulex exposed to the pyrethroid cypermethrin. *Chemistry and Ecology*, **27**, 569–577.

Nørum, U., Friberg, N., Jensen, M.R., Pedersen, J.M. & Bjerregaard, P. (2010). Behavioural changes in three species of freshwater macroinvertebrates exposed to the pyrethroid lambda-cyhalothrin: laboratory and stream microcosm studies. *Aquatic Toxicology*, **98**, 328–335.

Nyman, A.M., Schirmer, K. & Ashauer, R. (2014). Importance of toxicokinetics for interspecies variation in sensitivity to chemicals. *Environmental Science and Technology*, **48**, 5946–5954.

Obermüller, B., Puntarulo, S. & Abele, D. (2007). UV-tolerance and instantaneous physiological stress responses of two Antarctic amphipod species Gondogeneia antarctica and Djerboa furcipes during exposure to UV radiation. *Marine Environmental Research*, **64**, 267–285.

Økland, K.A. (1978). Life history and growth of Asellus aquaticus (L.) in relation to environment in a eutrophic lake in Norway. *Hydrobiologia*, **59**, 243–259.

Oreskes, N. & Belitz, K. (2001). Philosophical issues in model assessment. *Model validation: Perspectives in Hydrological Science*, **23**, 23.

Oreskes, N., Belitz, K., Shrader-Frechette, K., Sterman, J.D. & Rykiel, E.J.J. (1994a). The Meaning of Models. *Science*, **264**, 331.

Oreskes, N., Shrader-Frechette, K. & Belitz, K. (1994b). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science*, **263**, 641–646.

Oriade, C.A. & Dillon, C.R. (1997). Developments in biophysical and bioeconomic simulation of agricultural systems: a review. *Agricultural Economics*, **17**, 45–58.

Ormerod, P. & Rosewell, B. (2009). Validation and Verification of Agent-Based Models in the Social Sciences. *Epistemological Aspects of Computer Simulation in the Social Sciences* (ed F. Squazzoni), pp. 130–140. Lecture Notes in Artificial Intelligence. Springer Berlin Heidelberg.

Ovaskainen, O. (2004). Habitat-specific movement parameters estimated using mark-recapture data and diffusion model. *Ecology*, **85**, 242–257.

Overton, S. (1977). A strategy of model construction. *Ecosystem Modeling in Theory and Practice: An Introduction with Case Histories* (eds C. Hall & J. Day), pp. 49–73. John Wiley & Sons, New York, NY.

Papadakis, E.-N., Tsaboula, A., Kotopoulou, A., Kintzikoglou, K., Vryzas, Z. & Papadopoulou-Mourkidou, E. (2015). Pesticides in the surface waters of Lake Vistonis Basin, Greece: Occurrence and environmental risk assessment. *Science of the Total Environment*, **536**, 793–802.

Pastorok, R.A. (2002). *Ecological Modeling in Risk Assessment: Chemical Effects on Populations, Ecosystems, and Landscapes*. Lewis Publishers, Boca Raton, FL, USA.

Peeters, E.T.H.M., De Lange, H.J. & Lürling, M. (2009). Variation in the Behavior of the Amphipod Gammarus pulex. *Human and Ecological Risk Assessment*, **15**, 41–52.

Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R. & Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science (New York, N.Y.)*, **330**, 1496–501.

Pilkey, O.H. & Pilkey-Jarvis, L. (2007). *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future*. Columbia University Press, New York.

Pope, C.N. (2010). Organophosphorus pesticides: do they all have the same mechanism of toxicity? *Journal of Toxicology and Environmental Health. Part B: Critical Reviews*, **2**, 161–181.
Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchingson & Co, London.
Preuss, T.G., Hommen, U., Alix, A., Ashauer, R., Van den Brink, P.J., Chapman, P., Ducrot, V., Forbes, V.E., Grimm, V., Schäfer, D., Streissl, F. & Thorbek, P. (2009). Mechanistic effect models for ecological risk assessment of chemicals (MEMoRisk) - a new SETAC-Europe Advisory Group. *Environmental Science and Pollution Research*, **16**, 250–252.
R Core Team. (2013). R: A language and environment for statistical computing.
Rademacher, C., Neuert, C., Grundmann, V., Wissel, C. & Grimm, V. (2004). Reconstructing spatiotemporal dynamics of Central European natural beech forests: the rule-based forest model BEFORE. *Forest Ecology and Management*, **194**, 349–368.
Rademacher, C., Neuert, C., Grundmann, V., Wissel, C. & Grimm, V. (2001). Was charakterisiert Buchenurwälder? Untersuchungen der Altersstruktur des Kronendachs und der räumlichen Verteilung der Baumriesen in einem Modellwald mit Hilfe des Simulationsmodells BEFORE. *Forstwissenschaftliches Centralblatt*, **120**, 288–302.
Railsback, S.F. & Grimm, V. (2012). *Agent-based and individual-based modeling : a practical introduction*. Princeton University Press.
Railsback, S.F. & Harvey, B.C. (2002). ANALYSIS OF HABITAT-SELECTION RULES USING AN INDIVIDUAL-BASED MODEL. *Ecology*, **83**, 1817–1830.
Refsgaard, J.C. & Henriksen, H.J. (2004). Modelling guidelines - terminology and guiding principles. *Advances in Water Resources*, **27**, 71–82.
Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholten, H. & Kassahun, A. (2005). Quality assurance in model based water management – review of existing practice and outline of new approaches. *Environmental Modelling & Software*, **20**, 1201–1215.
Refsgaard, J.C., Van der Sluijs, J.P., Brown, J. & Van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, **29**, 1586–1597.
Ren, Z.-M., Li, Z.-L., Zha, J.-M., Rao, K., Ma, M., Wang, Z. & Fu, R.-S. (2008). The Avoidance Responses of Daphnia magna to the Exposure of Organophosphorus Pesticides in an On-Line Biomonitoring System. *Environmental Modeling & Assessment*, **14**, 405–410.
Ren, Z., Zha, J., Ma, M., Wang, Z. & Gerhardt, A. (2007). The early warning of aquatic organophosphorus pesticide contamination by on-line monitoring behavioral changes of Daphnia magna. *Environmental Monitoring and Assessment*, **134**, 373–383.
Rice, P.J., Drewes, C.D., Klubertanz, T.M., Bradbury, S.P. & Coats, J.R. (1997). Acute toxicity and behavioral effects of chlorpyrifos, permethrin, phenol, strychnine, and 2,4-dinitrophenol to 30-day-old Japanese medaka (Oryzias latipes). *Environmental Toxicology and Chemistry*, **16**, 696–704.
Rico, A. & Van den Brink, P.J. (2015). Evaluating aquatic invertebrate vulnerability to insecticides based on intrinsic sensitivity, biological traits, and toxic mode of action. *Environmental Toxicology and Chemistry*, **34**, 1907–1917.
Ridley, M. & Thompson, D.J. (1979). Size and Mating in Asellus aquaticus (Crustacea: Isopoda). *Z. Tierpsychol.*, **51**, 380–397.
Robinson, E.J.H., Richardson, T.O., Sendova-Franks, A.B., Feinerman, O. & Franks, N.R. (2008). Radio tagging reveals the roles of corpulence, experience and social information in ant decision making. *Behavioral Ecology and Sociobiology*, **63**, 627–636.
Rodrigues, A.C.M., Henriques, J.F., Domingues, I., Golovko, O., Žlábek, V., Barata, C., Soares, A.M.V.M. & Pestana, J.L.T. (2016). Behavioural responses of freshwater planarians after short-term exposure to the insecticide chlorantraniliprole. *Aquatic Toxicology*, **170**, 371–376.
Roessink, I., Merga, L.B., Zweers, H.J. & Van den Brink, P.J. (2013). The neonicotinoid imidacloprid shows high chronic toxicity to mayfly nymphs. *Environmental Toxicology and Chemistry*, **32**, 1096–1100.
Rubach, M.N., Ashauer, R., Maund, S.J., Baird, D.J. & Van den Brink, P.J. (2010). Toxicokinetic variation in 15 freshwater arthropod species exposed to the insecticide chlorpyrifos. *Environmental Toxicology and Chemistry*, **29**, 2225–2234.
Rubach, M.N., Crum, S.J.H. & Van den Brink, P.J. (2011). Variability in the dynamics of mortality and immobility responses of freshwater arthropods exposed to chlorpyrifos. *Archives of Environmental Contamination and Toxicology*, **60**, 708–721.
Rykiel, E.J.J. (1984). Modelling agroecosystems: lessons from ecology. *Agricultural ecosystems: unifying concepts* (eds R. Lowrance, B.R. Stinner & G.J. House), pp. 157–178. John Wiley & Sons, New York, NY.
Rykiel, E.J.J. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling*, **90**, 229–244.
Saltelli, A., Tarantola, S. & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling. *Statistical Science*, **15**, 377–395.

Sargent, R.G. (2005). Verification and validation of simulation models. *Proceedings of the 2005 Winter Simulation Conference* Proceedings of the 37th conference on Winter simulation. Syracuse, NY, USA, Orlando, Florida.

Scheller, R.M. & Mladenoff, D.J. (2006). An ecological classification of forest landscape simulation models: tools and strategies for understanding broad-scale forested ecosystems. *Landscape Ecology*, **22**, 491–505.

Scheller, R.M., Sturtevant, B.R., Gustafson, E.J., Ward, B.C. & Mladenoff, D.J. (2010). Increasing the reliability of ecological models using modern software engineering techniques. *Frontiers in Ecology and the Environment*, **8**, 253–260.

Schick, R.S., Loarie, S.R., Colchero, F., Best, B.D., Boustany, A., Conde, D.A., Halpin, P.N., Joppa, L.N., McClellan, C.M. & Clark, J.S. (2008). Understanding movement data and movement processes: current and emerging directions. *Ecology Letters*, **11**, 1338–1350.

Schlesinger, S. (1979). Terminology for model credibility. *Simulation*, **32**, 103–104.

Schmolke, A., Thorbek, P., Chapman, P. & Grimm, V. (2010a). Ecological models and pesticide risk assessment: current modeling practice. *Environmental Toxicology and Chemistry*, **29**, 1006–1012.

Schmolke, A., Thorbek, P., DeAngelis, D.L. & Grimm, V. (2010b). Ecological models supporting environmental decision making: A strategy for the future. *Trends in Ecology & Evolution*, **25**, 479–486.

Seuront, L., Brewer, M.C. & Strickler, J.R. (2004a). Quantifying Zooplankton Swimming Behavior: The Question of Scale. *Handbook of scaling methods in aquatic ecology: measurement, analysis, simulation* (eds L. Seuront & P.G. Strutton), pp. 333–360. CRC Press, Boca Raton, FL, USA.

Seuront, L., Hwang, J.-S., Tseng, L.-C., Schmitt, F.G., Souissi, S. & Wong, C.-K. (2004b). Individual variability in the swimming behavior of the sub-tropical copepod Oncaea venusta (Copepoda: Poecilostomatoida). *Marine Ecology Progress Series*, **283**, 199–217.

Seuront, L., Schmitt, F.G., Brewer, M.C., Strickler, J.R. & Souissi, S. (2004c). From random walk to multifractal random walk in zooplankton swimming behavior. *Zoological Studies*, **43**, 498–510.

Shugart, H.H. (1984). *A Theory of Forest Dynamics: The Ecological Implications of Forest Succession Models*. Springer-Verlag, New York, NY.

Skelsey, P., With, K.A. & Garrett, K.A. (2012). Why dispersal should be maximized at intermediate scales of heterogeneity. *Theoretical Ecology*, **6**, 203–211.

Van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J.R. & Risbey, J. (2005). Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. *Risk analysis an official publication of the Society for Risk Analysis*, **25**, 481–492.

Smith, M.J., Sherratt, J.A. & Lambin, X. (2008). The effects of density-dependent dispersal on the spatiotemporal dynamics of cyclic populations. *Journal of Theoretical Biology*, **254**, 264–274.

Song, M.Y., Stark, J.D. & Brown, J.J. (1997). Comparative toxicity of four insecticides, including imidacloprid and tebufenozide, to four aquatic arthropods. *Environmental Toxicology and Chemistry*, **16**, 2494–2500.

Stadnicka, J., Schirmer, K. & Ashauer, R. (2012). Predicting concentrations of organic chemicals in fish by using toxicokinetic models. *Environmental science & technology*, **46**, 3273–80.

Starfield, A.M. (1997). A Pragmatic Approach to Modeling for Wildlife Management. *The Journal of Wildlife Management*, **61**, 261–270.

Stark, J.D., Banks, J.E.J.E. & Acheampong, S. (2004). Estimating susceptibility of biological control agents to pesticides: influence of life history strategies and population structure. *Biological Control*, **29**, 392–398.

Stegeman, A., Bouma, A., Elbers, A.R.W., de Jong, M.C.M., Nodelijk, G., de Klerk, F., Koch, G. & van Boven, M. (2004). Avian Influenza A Virus (H7N7) Epidemic in The Netherlands in 2003: Course of the Epidemic and Effectiveness of Control Measures. *The Journal of Infectious Diseases*, **190**, 2088–2095.

Stillman, R.A., Railsback, S.F., Giske, J., Berger, U. & Grimm, V. (2015). Making Predictions in a Changing World: The Benefits of Individual-Based Ecology. *Bioscience*, **65**, 140–150.

Suchail, S., Guez, D. & Belzunces, L.P. (2001). Discrepancy between acute and chronic toxicity induced by imidacloprid and its metabolites in Apis mellifera. *Environmental Toxicology and Chemistry*, **20**, 2482–2486.

Tadini, G.V., Fano, E.A. & Colangelo, M.A. (1988). The life-history evolution of Asellus aquaticus L. explains its geographical distribution. *Internationale Vereinigung für Theoretische und Angewandte Limnologie Verhandlungen*, **23**, 2099–2106.

Taleb, N.N. (2010). *The Black Swan: The Impact of the Highly Improbable*. Random House, New York, NY.

Thorbek, P., Forbes, V.E., Heimbach, F., Hommen, U., Thulke, H.-H. & Van den Brink, P.J. (2009). *Ecological Models for Regulatory Risk Assessments of Pesticides: Developing a Strategy for the Future* (P. Thorbek, V.E. Forbes, F. Heimbach, U. Hommen, H.-H. Thulke, P.J. Van den Brink, J. Wogram & V. Grimm, Eds.). Society of Environmental Toxicology and Chemistry (SETAC) and CRC Press, Pensacola and Boca Raton, FL, USA.

Tolba, M.R. & Holdich, D.M. (1981). The effect of water quality on the size and fecundity of Asellus aquaticus (Crustacea:Isopoda). *Aquatic Toxicology*, **1**, 101–112.

Topping, C.J., Alrøe, H.F., Farrell, K.N. & Grimm, V. (2015). Per Aspera ad Astra: Through Complex Population Modeling to Predictive Theory. *The American Naturalist*, **186**, 669–674.
Topping, C.J., Dalkvist, T. & Grimm, V. (2012). Post-hoc pattern-oriented testing and tuning of an existing large model: lessons from the field vole (L.-F. Bersier, Ed.). *PloS one*, **7**, e45872.
Troldborg, L., Refsgaard, J.C., Jensen, K.H. & Engesgaard, P. (2007). The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system. *Hydrogeology Journal*, **15**, 843–860.
Turchin, P. (1998). *Quantitative Analysis of Movement*. Sinauer.
US - EPA, C. (2009). Guidance on the Development, Evaluation, and Application of Environmental Models. *Environmental Protection*, **EPA/100/K-**, 99.
Vaughn, C.C. (2010). Biodiversity Losses and Ecosystem Function in Freshwaters: Emerging Conclusions and Research Directions. *BioScience*, **60**, 25–35.
Vitagliano, G., Fano, E.A., Marchetti, E., Colangelo, M.A. & Vitagliano, E. (1991). Importance of longevity, growth, and diapause in the evolution of Asellus aquaticus. *Bolletino di zoologia*, **58**, 113–117.
Wallace, R.R., Hynes, H.B.N. & Kaushik, N.K. (1975). Laboratory experiments on factors affecting the activity of Gammarus pseudolimnaeus Bousfield. *Freshwater Biology*, **5**, 533–546.
Wallace, J.B. & Webster, J.R. (1996). The role of macroinvertebrates in stream ecosystem function. *Annual Review of Entomology*, **41**, 115–139.
Wang, M. & Luttik, R. (2012). Population level risk assessment: practical considerations for evaluation of population models from a risk assessor's perspective. *Environmental Sciences Europe*, **24**.
Watson, James D. "The Double Helix: Being a Personal Account of the Discovery of the Structure of DNA." *New York, Atheneum* (1968).
Van Waveren, R.H., Groot, S., Scholten, H., Van Geer, F.C., Wösten, J.H.M., Koeze, R.D. & Noort, J.J. (1999). Good Modelling Practice Handbook. *Fourth International Conference Hydro Informatics 2000* p. 165.
Weed, D.L. (2005). Weight of Evidence: A Review of Concept and Methods. *Risk Analysis*, **25**, 1545–1557.
Westerberg, L., Lindström, T., Nilsson, E. & Wennergren, U. (2008). The effect on dispersal from complex correlations in small-scale movement. *Ecological Modelling*, **213**, 263–272.
Whiles, M.R. & Wallace, J.B. (1992). First-year benthic recovery of a headwater stream following a 3-year insecticide-induced disturbance. *Freshwater Biology*, **28**, 81–91.
WHO Technical Report Series, No. 937, 2006. Annex 4. Appendix 5. 2006
Van Wijngaarden, R.P.A., Van den Brink, P.J., Crum, S.J.H., Oude Voshaar, J.H., Brock, T.C.M. & Leeuwangh, P. (1996). Effects of the insecticide Dursban® 4E (active ingredient chlorpyrifos) in outdoor experimental ditches: I. Comparison of short-term toxicity between the laboratory and the field. *Environmental Toxicology and Chemistry*, **15**, 1133–1142.
Wikipedia Contributers. (2013). Good Laboratory Practice. *Wikipedia, The Free Encyclopedia*. Retrieved December 7, 2013, from http://en.wikipedia.org/w/index.php?title=Good Laboratory Practice&oldid=583104663
Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. *Center for Connected Learning and ComputerBased Modeling Northwestern University Evanston IL*, **2009**, Evanston, IL. Retrieved from http://ccl.northwestern.edu/netlogo/
Williams, W.D. (1962). Notes on the ecological similarities of Asellus aquaticus (L.) and A. meridianus Rac. (Crust., Isopoda). *Hydrobiologia*, **20**, 1–30.
Wogram, J. & Liess, M. (2001). Rank Ordering of Macroinvertebrate Species Sensitivityto Toxic Compounds by Comparison with That of Daphnia magna. *Bulletin of Environmental Contamination and Toxicology*, **67**, 360–367.
Zein, M.A., McElmurry, S.P., Kashian, D.R., Savolainen, P.T. & Pitts, D.K. (2014). Optical bioassay for measuring sublethal toxicity of insecticides in Daphnia pulex. *Environmental Toxicology*, **33**, 144–151.
Zimmer, E.I., Jager, T., Ducrot, V., Lagadic, L. & Kooijman, S.A.L.M. (2012). Juvenile food limitation in standardized tests: A warning to ecotoxicologists. *Ecotoxicology*, **21**, 2195–2204.
Zurell, D., Grimm, V., Rossmanith, E., Zbinden, N., Zimmermann, N.E. & Schröder, B. (2012). Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps. *Ecography*, **35**, 590–603.

SUMMARY

In recent years, ecological effect models have been put forward as tools for supporting environmental decision-making. Often they are the only way to take the relevant spatial and temporal scales and the multitude of processes characteristic to ecological systems into account. Particularly for environmental risk assessments of pesticides the potential benefits of including modelling studies were recognized and a dialogue between different stakeholder groups was opened. Representatives from academia, pesticide-producing industries, and regulators are nowadays discussing their needs, possibilities, and ways of implementation for improving the use and usefulness of such models. However, it quickly became evident that not all involved parties possess the same background knowledge in regards to modelling terminology and model quality understanding. Different modelling approaches are already available and target different aspects of the complex processes leading to observable effects of pesticide exposure on population and ecosystem dynamics. Proper communication of a given model's structure, robustness, and soundness is crucial to render a model of real use to the decision-making. Doubts about a model's quality and mode of operation may lead to an immediate rejection of the conclusions drawn from its estimations. Accordingly, people with a limited understanding of modelling and the associated terminology regularly ask for a model to be "validated" before feeling they can rely on its output and let it inform a decision.

The first aim of this thesis is to improve the understanding on different sides of the discourse concerning trust, credibility and acceptance of ecological effect models in policy-making processes, particularly in the field of environmental risk assessment of pesticides. The second aim of this work is to propose, based on this improved understanding, ways to overcome obstacles.

In **Chapter 2**, I addressed different aspects surrounding quality assessment, validation, and communication of models in a literature review. Confusion about model validation is one of the main challenges in using ecological models for decision support, such as the regulation of pesticides. Decision makers need to know whether a model is a sufficiently good representation of its real counterpart and what criteria can be used to answer this question. Unclear terminology is one of the main obstacles to a good understanding of what model validation is, how it works, and what it can deliver. Therefore, we performed a literature review and derived a standard set of terms. “Validation” was identified as a catch-all term, which is thus useless for any practical purpose. We introduced the term “evaludation”, a fusion of “evaluation” and “validation”, to describe the entire process of assessing a model’s quality and reliability. Considering the iterative nature of model development, the modelling cycle, we identified six essential elements of evaludation: (i) “data evaluation” for scrutinising the quality of numerical and qualitative data used for model development and testing; (ii) “conceptual model evaluation” for examining the simplifying assumptions underlying a model’s design; (iii) “implementation verification” for testing the model’s implementation in equations and as a computer program; (iv) “model output verification” for comparing model output to data and patterns that guided model design and were possibly used for calibration; (v) “model analysis” for exploring the

model's sensitivity to changes in parameters and process formulations; this step is to ensure that the mechanistic basis of the main behaviours of the model has been well understood; and (vi) "model output corroboration" for comparing model output to new data and patterns that were not used for model development and parameterisation. Currently, most decision makers require "validating" a model by testing its predictions with new experiments or data. Despite being desirable, this is neither sufficient nor necessary for a model to be useful for decision support. We believe that the proposed set of terms and its relation to the modelling cycle can help to make quality assessments and reality checks of ecological models more comprehensive and transparent.

In **Chapter 3**, I used the evaludation framework to re-evaluate and adjust the documentation framework TRACE (TRAnsparent and Comprehensive Ecological modelling; Schmolke et al. 2010), a general framework for documenting a model's rationale, design, and testing. TRACE documents should provide convincing evidence that a model was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose. TRACE documents link the science underlying a model to its application, thereby also linking modellers and model users, for example stakeholders, decision makers, and developers of policies. TRACE thus becomes a tool for planning, documenting, and assessing model evaludation, which includes understanding the rationale behind a model and its envisaged use. Originally TRACE was aimed at documenting good modelling practice. However, the word "documentation" does not convey TRACE's urgency. Therefore, we re-define TRACE as a tool for planning, performing, and documenting *good modelling practice*. We report on first experiences in producing TRACE documents in which we found that the original idea underlying TRACE was valid, but to make its use more coherent and efficient, an update of its structure and more specific guidance for its use are needed.

In **Chapters 4** to **6**, I turn to an example of the measures that can be taken to increase general trust in a model's design and output. I chose MASTEP (Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides) for a case study. MASTEP is an individual-based model used to describe the effects on and recovery of the water louse *Asellus aquaticus* after exposure to an insecticide in pond, ditch, and stream scenarios. The model includes processes of mortality of *A. aquaticus*, life history, random walk between cells and density dependence of population regulation. One of the submodels receiving particular criticism was the random walk procedure and the uncertainty attached to the parameters used. The parameters were estimated based on experimental studies performed under very limiting conditions.

In **Chapter 4**, I present an experimental method to collect more precise parameter estimates and to better understand the movement behaviour of the crawling isopod *Asellus aquaticus* and the swimming fresh water amphipod *Gammarus pulex* under different conditions. Quantifying and understanding movement is critical for a wide range of questions in basic and applied ecology, including modelling of dispersal within the context

of mechanistic effect models for pesticide risk assessment. Movement ecology is also fostered by technological advances that allow automated tracking for a wide range of animal species. However, for aquatic macroinvertebrates such detailed methods do not yet exist. The experimental procedure that I developed employed video tracking of marked individuals that were introduced alone or as part of a group of unmarked individuals into arenas of approximately $1m^2$ in size. I recorded the paths of the marked individuals under a set of different conditions, i.e. presence or absence of food or shelter, population density, and after sublethal pesticide exposure. I tested the effects of different light sources and marking techniques on their movement behaviour first to establish the possibilities and limitations of the experimental protocol. Secondly, this approach aimed to ensure that the basic handling of test specimens would not bias conclusions drawn from movement path analyses. To demonstrate the versatility of the method, I described the influence of varying population densities on different movement parameters related to resting behaviour, directionality, and step lengths.

The developed protocol allows studying species with different modes of dispersal and under different conditions. For example, I found that Gammarids spend more time moving at higher population densities, while Asellids rest more under similar conditions. At the same time, in response to higher densities, Gammarids mostly decreased average step lengths, whereas Asellids hardly did. Gammarids, however, were also more sensitive to general handling and marking than Asellids. The protocol for marking and video tracking can be easily adopted for other species of aquatic macroinvertebrates or testing conditions, for example presence or absence of food sources, shelter, or predator cues. Nevertheless, limitations with regard to the marking protocol, material, and a species' physical build need to be considered and tested before a wider application, particularly for swimming species.

In **Chapter 5**, I provided further results gathered with the experimental protocol developed in **Chapter 4**. In this study, I aimed to determine the potential effects of exposure to two neurotoxic pesticides with different modes of action at different concentrations (chlorpyrifos and imidacloprid) on the locomotion behaviour of the water louse *Asellus aquaticus*. I contrasted the effects of the different exposure regimes on the behaviour of *Asellus* with the effects of the presence of food and shelter to estimate the ecological relevance of behavioural changes. Sublethal pesticide exposure reduced dispersal distances compared to controls, whereby exposure to chlorpyrifos affected not only animal activity but also step lengths. In contrast, imidacloprid only slightly affected step lengths. The presence of natural cues such as food or shelter induced only minor changes in behaviour, which hardly translated to changes in dispersal potential.

Behaviour links physiological functions with ecological processes and can be very sensitive towards environmental stimuli and chemical exposure. As such, behavioural indicators of toxicity are well suited for assessing impacts of pesticides at sublethal concentrations found in the environment. My findings illustrate that movement behaviour can serve as a sensitive endpoint in toxicity assessments. However, under natural conditions, depending on the

exposure concentration, the actual impacts might be outweighed by environmental conditions that an organism is subjected to. It is, therefore, of importance that the assessment of toxicity on behaviour is done under relevant environmental conditions.

Based on the experimental findings of **Chapter 4** and **5**, I refined the movement modelling procedure used in MASTEP to derive more realistic dispersal estimates in **Chapter 6**. In this study, I revisited the MASTEP model and tested the effects of pesticide application timing on population dynamics and recovery times. I compared the recovery time estimations from previous versions with the updated version and was able to integrate an increased level of environmental complexity that could not be addressed before due to a lack of data. Subsequently, I implemented density dependence of movement rates, heterogeneously spread food availability and assumptions concerning restrained movement rates for life stages as younger individuals cannot cover as much distance as adults may. Compared to former versions of the population model, recovery times did not change significantly when "adult only" movement was assumed for all simulated individuals. This indicates that the previous assumptions already yielded robust estimations. Accounting for life stage dependent movement restraints, though, delayed recovery when exposure occurred shortly before a reproduction cycle.

In **Chapter 7**, I reflected on the strengths and weaknesses of the effort that went into this thesis and the insights it provided. Moreover, I aimed to propose possible ways forward to increase the use of ecological effect modelling in environmental decision-making processes.

Based on my findings, I argue that an increase of assumptions regarding the implementation of ever more realism and environmental complexity needs to be judged carefully on a case-by-case basis. Increased realism in models can introduce an unwarranted increase in model complexity and uncertainty, which is not always supporting an improved credibility level of a model. However, such judgments can be only done in the light of available data, which are not always accessible or impossible to gather. Despite the need for basic ecological research to develop en parameterize more comprehensive ecological models, I further argue that a modelling study in general can benefit greatly from an improved plan that considers communication needs from the start. Considering such needs early on can help to developing a time- and cost-saving strategy for model testing and data collection, while providing a thorough understanding of a model's underlying mechanisms across several layers of stakeholder groups.

ACKNOWLEDGEMENTS

"Some people grumble that roses have thorns;
I am grateful that thorns have roses."

Alphonse Karr

Writing this section of the book surely is one of the most satisfying experiences in my life so far, but is also one that raises a lot of queer feelings of nostalgia. Writing this section means that PhD life is coming to a close, that another period in life is coming to an end. An end that, at times, I would not have been able or willing to see. The last few years held a number of challenges in store and yet turned into shiny memories thanks to so many wonderful people that I got a chance to meet during this journey. Writing this section means, I got out of this as a better version of myself and that I get to choose and to continue with the better parts when setting up the next stage of my life. And I hope that there will be a lot to take with me.

First and foremost, I would like to express my gratitude to my supervisors, Paul and Volker for giving me the opportunity to do my PhD as part of CREAM and for not only allowing me the freedom to develop the thesis to its final shape but also for actively helping me get there.

Paul, I still remember the day I received your email offering me the opportunity to come to Wageningen. At that time, I woke up back at Stanford, where I was visiting some friends from my Master thesis. It was 5 o'clock in the morning after I had just arrived and the jetlag got the better of me. To not wake up anyone else, I skimmed through my emails and came across that one. The quiet of the house found a sudden end then and there. Little did I know what was waiting for me in the coming years. Your support, ideas and pragmatism helped create a space in which I was challenged but that also enabled me to grow and learn - a lot - and to eventually reach this point, where I am finally able to close this chapter. Thank you for your patience with me.

Volker, your wisdom and support were an integral aspect to keep me in the game when I hit a wall. I will always treasure our conversations, your honesty and the lessons you taught me. Your advice on getting past the writer's block may continue being used in the future; if not for academic writing then in other forms of *Literaturergüssen*. Who knows what book I may ask you next to proofread for me? And without you and your enthusiasm, CREAM definitely would have been a less homely, less inspiring environment for teaching and raising all those fellows.

Further thanks go to all the CREAMers. This project was not only a great way to meet and juggle ideas with a diverse group of scientists. It was a place, where friendships were formed, some closer than others, but none of them replaceable. Alpi, Anni, Annika, Bea, Ben, Chun, Devdutt, Dragan, Elke, Faten, Ida, Jürgen, Julitta, Kasia, Lara, Mattia, Natnael, Tomek, Viola, thank you for embarking on this venture together with me and for sharing your spirits during the PhD days. Agnieszka (x2), Andreas, your next-level experience was

a big help to many of us. Dear supervisors, associated and visiting partners and scientists, thank you for the courses, discussions and encouragements that pushed us all to new levels. Valery, thanks for the desk and hospitality in Nebraska and to Pernille for those words of advice here and there that came at exactly the right time, although you might not even be aware of it.

My gratitude also goes to my opponents. Prof. Naguib and Dr van Loon, thank you to have taken the time to read and evaluate my thesis. Udo and Ryszard, thank you for your advice and insights throughout the CREAM days. I very much appreciate you agreeing to be my opponents.

Despite the beauty of the CREAM network, Wageningen as a home base had its own share of colleagues, peers and friends to offer.

Dear (Ex)Stressors, thank you for all the fun, the technical and non-technical support and the travels together. All your different energies, strengths and insights gave me much food for thought. Mascha, Mazhar, Nika, Andreu, I will never forget that first SETAC in Seville. What a fun trip! Not that the following congresses were less enjoyable but this first conference experience holds a special place in my mind. Mauricio, Jugk, Berhan, Fengjiao, Zhang, Concilia, Michael, it was great to have you in the group. Thank you for your company from far and near.

But what would a PhD be without the backbone of the Department that houses them? Dear AEW, thank you for being this inspirational group of people. Thank you for those stimulating Thursday lunches, discussions, chats and coffees in the hallways, PhD weekends, Sinterklaas events and uitjes. And thank you, John, Frits and Wendy for your help in the lab. To all you PhDs and PostDocs, thank you for numerous nights at the pub, company in the lab and office, chocolate, cakes, new food experiences, cultural insights, friendship and more. So many joined the group and many left again already. Annelies, Darya, Ingrid, Kristina, thanks for being there pretty much from the start and sharing your stories and opinions with me. Ana, Andrea, Ariadna, Arie, Babak, Bastiaan, Bernardo, Bregje, Ellen, Els, Irene, Jelle, Jochem, Joris, Maira, Marlies, Paula, Rafael, Sarian, Sebastian, Vasilis, I appreciate to have had a chance to meet you and I hope to remain in touch with at least most of you!

I was fortunate enough to be part of a second group, the ERA team at Alterra. Thank you everybody for allowing me glimpses at the practical side of environmental risk assessments, working with pesticides and different organism groups, and the ins and outs around those topics. I appreciate the talks and the fun we shared at different uitjes and afternoons at the pub. Thank you Hans and Andreas, for your patience and always being ready and available for conversations, answers and helping me with keeping an overview when things got tricky. Thank you Theo, Marie-Claire, Dick, René, Steven and Laura for the help with the practical and theoretical issues around my experiments.

Big thanks also go to my paranymphs, Ivo and Noël. Thank you for being there for almost all of my PhD days. Thank you for always having my back and helping me improve my work ever since I started. Ivo, your resources for trouble-shooting were pivotal to the whole endeavour, as were those interventional meetings here and there. Noël, your creativity and critical thinking pushed me to new approaches. Aside from work, both of you helped me develop and grow as only few people did. I am lucky to have people like you around that not only helped me but also introduced me to many more ideas and creatures.

This thesis may also not have come to a close at this point of time without the support and motivation through my new colleagues at Charles River, or should I still call it WIL Research? Thank you for accepting me into Eco and many more thanks for allowing me to learn and grow even further and for the patience while I finished the writing of this thesis.

As much as I enjoyed the time with colleagues and peers, I also needed some distraction from the PhD environment at times.

Thank you, former and current housemates and neighbours to make life in the *Afronding* even better. Kristin, Guillaume, Bart, Viola, Silja, Suzanne, Astrid, Jasper, Ariadna, Anna, Micha, Daniel, thanks for all the (pan)cakes, dinners, beers and company. Bart, thank you for not only getting me back on the motorbike or introducing me to the higher levels of coding but also for your lasting friendship.

And my dear friends from all over the world! Jule, Tila, Lea, Jenny, thank you for keeping my German alive even though we got scattered around the Netherlands and beyond. It felt good to share some of the disorientation in this country. I feel immensely grateful that there are people out there who are as prone as I am to loosing things that cause paths to cross. Julia, Caro, Malik, Helen, Alex, Haimil, my lovely group of unicorns, what would I have been without your never-ending supply of positive thinking? Thank for helping with all those wine bottles, reality escapes, and love.

Then ther are those great ladies I met through my attempts at overcoming my clumsiness. What would life have turned out to be without the inspiration from you? Raphaella, VeeNiz, Crystal, Kirilee, Gina, Esy, Suus, Linde, Ilona, Gemma, thank you for all those lessons and shared bruises in Barcelona, Wageningen and Arnhem! Your drive, power and passion was what I needed to discover other sides of my being. May the next adventures together be waiting just around the next spin to turn life upside down again.

Last but not least, those people who were there throughout my pre-PhD life and without who I wouldn't have gotten here in the first place. Sandy, Maike, thanks for your friendship and not letting life drift us apart despite those distances between you and Wageningen. My lovely friends from those days in the US, it always a pleasure to reconnect and feel like it was only yesterday that we last hung out and shared some giggles. The occasional escapes made possible through the bonds established way back in the days added to the fun parts of life. It is amazing how time may change but cannot harm true friendships.

To my lovely family, I cannot possibly express what your support and love means to me. Mum, Paps, Brüderchen, I will only say (write) one thing just so it finally is stated black on white: Ich hab euch mehr lieb! Danke, dass ihr nie aufgehört habt mich anzufeuern oder zu trösten, wenn es nötig war - oder einfach auf ein kroketje vorbeizukommen. Louis, José bedankt voor jullie bijstand toen ik mijn eerste gesprek had in Wageningen. Zonder jullie rust en advies de avond ervoor zou het wel anders gegaan kunnen zijn. Wat ben ik blij nu ook enkele lijntjes op Nederlands toevoegen te kunnen (ook als ze niet perfect zijn)?

Dear supervisors, colleagues, peers, friends and family, so many people contributed to the success (=completion) of this thesis in so many ways, big and small, practically, inspiringly, instructionally, emotionally. There are many who I may not have included by name. You still carry a special place in my heart. I wish you all the best in your professional and personal lives in the years ahead.

"We can only be said to be alive in those moments
when our hearts are conscious of our treasures."

Thornton Wilder

ABOUT THE AUTHOR

Jacqueline A. Augusiak was born on 26th of September in Offenbach a. M., Germany. In 2003 she started to follow the BSc. and MSc. study program *Water Science* at the University of Duisburg in Germany. This interdisciplinary program majoring in chemistry, microbiology and analytical chemistry allowed her to pursue her diverse interests in Sciences, chemistry and microbiology in particular. For her BSc she studied the effects of biofilm formation on the corrosion vulnerability of different types of alloys at the University of Delaware, USA. As part of an MSc research internship, on the other hand, she studied the membrane degradation in fuel cells. During her MSc time, she also worked as a student assistant in the Department of Geomicrobiology at the University of Duisburg. She eventually developed a deep appreciation for ecotoxicology during her MSc thesis at Stanford University when studying the effects of activated carbon mixing and in situ stabilization of PCBs in sediments to reduce the uptake of PCBs (polychlorinated biphenyls) into the food chain. She went on to start a PhD project in the Aquatic Ecology and Water Quality Management group at Wageningen University and the Environmental Risk Assessment team at Alterra as part of the Marie Curie initial training network CREAM (Mechanistic Effect Models for Ecological Risk Assessment of Chemicals). Under the supervision of Paul van den Brink and Volker Grimm, she researched how ecological effect models can contribute to regulatory risk assessments of pesticides and how they can be communicated more efficiently among the various stakeholder groups.

LIST OF PUBLICATIONS

PUBLISHED

Augusiak, J., Van den Brink, P.J. & Grimm, V. (2014). Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach. *Ecological Modelling*, **280**, 117–128.

Grimm, V., Augusiak, J., Focks, A., Frank, B.M., Gabsi, F., Johnston, A.S.A., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P. & Railsback, S.F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, **280**, 129–139.

Augusiak, J. & Van den Brink, P.J. (2015). Studying the movement behavior of benthic macroinvertebrates with automated video tracking. *Ecology and Evolution*, **5**, 1563–1575.

Augusiak, J. & Van den Brink, P.J. (2016). The influence of insecticide exposure and environmental stimuli on the movement behaviour and dispersal of a freshwater isopod. *Ecotoxicology*.

IN PREPARATION

Augusiak J. et al., Effects of externally induced changes of dispersal on aquatic macroinvertebrate recovery after pesticide exposure.

OVERVIEW OF COMPLETED TRAINING ACTIVITIES

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

For specialised PhD training

The Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Jacqueline Alexandra Augusiak

born on 26 September 1982 in Offenbach am Main, Germany

has successfully fulfilled all requirements of the
Educational Programme of SENSE.

Wageningen, 6 October 2016

the Chairman of the SENSE board	the SENSE Director of Education
Prof. dr. Huub Rijnaarts	Dr. Ad van Dommelen

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Ms Jacqueline Augusiak** has successfully fulfilled all requirements of the Educational PhD Programme of SENSE with a work load of 70.8 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2012)
- Research in context activity: Co-organising SENSE Symposium "Ecosystems under stress: assessing the impact of chemical and physical disturbances on ecological processes and ecosystem structure", Wageningen (2012)
- Basic statistics (2011)
- SENSE writing week (2013)

Other PhD and Advanced MSc Courses

- Ecological modelling, Helmholtz Centre for Environmental Research – UFZ (2010)
- Ecotoxicology and risk assessment, University of York & Reading University (2010)
- Matlab fundamentals, MathWorks®, Aachen (2010)
- Tools *(statistics in ecotoxicology, analyses of spatial and temporal patterns in data, GIS, and an introduction to database management)*, NERI & Fraunhofer Gesellschaft (2011)
- Complementary skills *(preparation of oral presentations, writing of grant proposals, scientific writing and poster presentations)*, Roskilde University & INRA (2011)
- Ecological models and data in R, Wageningen University (2012)
- Voice matters, Wageningen University (2012)
- Career Development, RWTH Aachen & Wageningen University (2014)

External training at a foreign research institute

- Training in use of NetLogo, Helmholtz Center for Environmental Research, Germany (2012)
- Ecological Modelling, University of Nebraska Lincoln, USA (2013)

Management and Didactic Skills Training

- Supervising three MSc students (2010, 2011, 2012)
- Co-organisation Open Midterm Workshop of CREAM project (2011)
- Member of the SENSE PhD Council and WIMEK PhD Council (2011-2013)
- Teaching in the MSc course 'Chemical stress ecology and risk assessment' (2013-2014)

Oral Presentations

- *Validation of mechanistic effect models: A practical approach*, SETAC Europe Annual Meeting, 12-16 May 2013, Glasgow, United Kingdom
- *Validation of mechanistic effect models for ecological risk assessments*, CREAM Open Conference , 10-13 June 2013, Leipzig, Germany

SENSE Coordinator PhD Education

Dr. ing. Monique Gulickx

This work was financially supported by the European Union under the 7th Framework Programme project acronym CREAM contract number PITN-GA- 2009-238148.

Financial support from the Aquatic Ecology and Water Quality Management department at Wageningen University for the printing of this thesis is gratefully acknowledged.