Critical Reviews in Toxicology

ISSN: 1040-8444 (Print) 1547-6898 (Online) Journal homepage: http://www.tandfonline.com/loi/itxc20

The pros and cons of ecological risk assessment based on data from different levels of biological organization

Jason R. Rohr, Christopher J. Salice & Roger M. Nisbet

To cite this article: Jason R. Rohr, Christopher J. Salice & Roger M. Nisbet (2016) The pros and cons of ecological risk assessment based on data from different levels of biological organization, Critical Reviews in Toxicology, 46:9, 756-784, DOI: 10.1080/10408444.2016.1190685

To link to this article: http://dx.doi.org/10.1080/10408444.2016.1190685

CRITICAL REVIEWS IN TOXICOLOGY, 2016
VOL. 46, NO. 9, 756–784
http://dx.doi.org/10.1080/10408444.2016.1190685

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

The pros and cons of ecological risk assessment based on data from different levels of biological organization

Jason R. Rohr[a], Christopher J. Salice[b] and Roger M. Nisbet[c]

[a]University of South Florida, Department of Integrative Biology, Tampa, FL, USA; [b]Towson University, Department of Biological Sciences, Towson, MD, USA; [c]University of California at Santa Barbara, Department of Ecology, Evolution, and Marine Biology, Santa Barbara, CA, USA

ABSTRACT

Ecological risk assessment (ERA) is the process used to evaluate the safety of manufactured chemicals to the environment. Here we review the pros and cons of ERA across levels of biological organization, including suborganismal (e.g., biomarkers), individual, population, community, ecosystem and landscapes levels. Our review revealed that level of biological organization is often related negatively with ease at assessing cause–effect relationships, ease of high-throughput screening of large numbers of chemicals (it is especially easier for suborganismal endpoints), and uncertainty of the ERA because low levels of biological organization tend to have a large distance between their measurement (what is quantified) and assessment endpoints (what is to be protected). In contrast, level of biological organization is often related positively with sensitivity to important negative and positive feedbacks and context dependencies within biological systems, and ease at capturing recovery from adverse contaminant effects. Some endpoints did not show obvious trends across levels of biological organization, such as the use of vertebrate animals in chemical testing and ease at screening large numbers of species, and other factors lacked sufficient data across levels of biological organization, such as repeatability, variability, cost per study and cost per species of effects assessment, the latter of which might be a more defensible way to compare costs of ERAs than cost per study. To compensate for weaknesses of ERA at any particular level of biological organization, we also review mathematical modeling approaches commonly used to extrapolate effects across levels of organization. Finally, we provide recommendations for next generation ERA, submitting that if there is an ideal level of biological organization to conduct ERA, it will only emerge if ERA is approached simultaneously from the bottom of biological organization up as well as from the top down, all while employing mathematical modeling approaches where possible to enhance ERA. Because top-down ERA is unconventional, we also offer some suggestions for how it might be implemented efficaciously. We hope this review helps researchers in the field of ERA fill key information gaps and helps risk assessors identify the best levels of biological organization to conduct ERAs with differing goals.

ARTICLE HISTORY
Received 31 January 2016
Revised 12 May 2016
Accepted 13 May 2016
Published online 22 June 2016

KEYWORDS
Adverse outcome pathways; assessment endpoint; communities; ecosystems; extrapolation; mathematical model; measurement endpoint; mechanistic effect models; mesocosms; multi-species systems; populations; scale

Table of contents

Introduction ... 757
Current challenges in ERA ... 759
Unclear linkage between ERA outcomes and ecological endpoints of concern ... 759
Need to screen large numbers of chemicals ... 759
Need to protect a wide variety of species ... 759
Multiple stressors and context dependencies ... 760
Include assessment of recovery ... 760
Transparency and defensibility ... 760
Reduction in animal testing ... 760
The pros and cons of ERA at different levels of biological organization ... 761
Sub-individual level: biochemical and molecular responses ... 761
Individual level ... 763
Population level ... 764
Community level ... 766
Ecosystem level ... 769
Landscape level ... 770
Models relating responses at different levels of organization ... 770
Toxicokinetic–toxicodynamic (TK–TD) models ... 771
Models of adverse outcome pathways (AOP) ... 772
Bioenergetic models ... 772
Population models deriving from individual physiology and life histories ... 773
Multi-species models ... 774
Where to go from here? Next generation ERA ... 775
Conclusions ... 776
Endnote ... 777
Acknowledgements ... 777
Declaration of interest ... 777
References ... 777

CONTACT Jason R. Rohr jasonrohr@gmail.com University of South Florida, Department of Integrative Biology, SCA 110, Tampa, FL 33647, USA

Introduction

Ecological risk assessment (ERA) is the process used to evaluate the impact of human activities on the environment and is an important part of the information portfolio that informs environmental policy (USEPA 1992, 1998; Suter et al. 2005; Suter 2007, 2008; Hommen et al. 2010). If done well, it can prevent damage to ecosystems and the need for costly ecosystem restoration (Rohr et al. 2016; Rohr, Johnson, et al. 2013). Although generally applicable to any type of potential (anthropogenic) stressor, ERA is central to the use and regulation of manufactured chemicals, including pesticides and industrial compounds. ERA (hereafter, "ERA") was formalized nearly three decades ago by the U.S Environment Protection Agency (USEPA 1992) and thousands of chemical ERAs have been conducted (Suter 2008). Although ERA methods are under continued discussion and are frequently fraught with controversy (Rohr & McCoy 2010a; Boone et al. 2014; Boone & Rohr 2015), new developments have been slow to take hold (Landis 2002) (but see EFSA 2013, which recently provided guidance on ERA from individuals to landscapes).

Internationally, the basic and essential elements of an ERA include characterization of chemical exposure and characterization of chemically-induced effects. In simplistic terms, a potential for adverse effects (risk) may occur when the exposure concentration exceeds (by some predetermined margin) the concentration at which effects have been observed from toxicity studies (USEPA 1992, 1998; Suter et al. 2005; Suter 2007; Hommen et al. 2010). The ERA process is often designed as Tiered, with lower Tiers representing simpler and less resource-intensive estimates of risk (Table 1). A screening-level ERA is commonly considered the first Tier and at this level, risk is essentially estimated by dividing an exposure estimate by an effect estimate to obtain a unitless risk (or hazard) quotient. The resulting quotient is then compared to a pre-determined value (e.g., Level of Concern) for determining whether adverse effects of a particular chemical may be expected (USEPA 1992, 1998; Suter et al. 2005; Suter 2007, 2008; Hommen et al. 2010). In some cases the risk quotient or Level of Concern may be adjusted by an uncertainty factor that functionally increases the estimated risk. The application of uncertainty factors is to account for a number of conditions with unknown influence on actual risk, including the fact that only a small number of species are ever tested for toxicity in comparison to the many that are likely exposed. Although not a true estimate of risk because it lacks a probabilistic component, the quotient-based approach is very common internationally (e.g., also used in the European Union; Hommen et al. 2010). Furthermore, although generally considered a lower-Tier assessment approach, the risk (or hazard) quotient is frequently used as a definitive indicator of risk for ecological systems associated with chemical use (USEPA 1992, 1998). Ideally, lower-Tier assessments generally inform whether higher tier assessments are needed, which are then used to refine risk estimates by incorporating additional data and/or extrapolation models (Hommen et al. 2010) (Table 1).

Although the ultimate objective in chemical ERA is to produce defensible estimates of "safe" concentrations based on estimates of the magnitude and probability of adverse effects to the ecological receptors or resources, the widely held view is that it is unlikely that current ERAs effectively meet this goal (Taub 1997a; Landis 2002; De Laender et al. 2008a; De Laender et al. 2009; Luttik et al. 2011; Beketov & Liess 2012). In large part, this is because the scientific studies that are used to estimate exposure and risk are few and highly controlled while the environmental systems of greatest interest (communities and ecosystems) are numerous, diverse, complex and highly variable. For example, commonly available toxicity data for freshwater invertebrates would include a 96-h acute and a 21-day life cycle study on the freshwater cladoceran, *Daphnia magna,* which, in some cases, may be the only available toxicity study for freshwater invertebrates. Data obtained from highly controlled laboratory studies on standardized species and test systems are generally assumed to be reproducible and of high quality but these data may not be applicable to scenarios in the real world and the linkage to higher levels of biological organization is difficult to discern (Forbes & Calow 2002a; Suter et al. 2005; Forbes et al. 2006; Martin et al. 2014).

The challenge in linking data to reality is addressed within the ERA framework through the selection and use of assessment and measurement endpoints. Assessment endpoints are the "explicit expressions of the actual environmental values

Table 1. Example refinement levels ("tiers") in the U.S. EPA's, Office of Pesticide Programs ecological risk assessment process (USEPA).

Level	Basic description	Risk metric	Example
I	Conservative analysis designed to "screen out" situations where there is reasonable certainty of no risk concerns. Relies upon conservative estimates of exposure and effect.	Quotient-based metric compared to a pre-determined level of concern.	Deterministic comparisons of exposure and toxicity (e.g., LC50's)
II	Refined analysis built upon data used in Tier I, with added consideration of available data to incorporate variability and uncertainty. May still be conservative and general in nature.	Estimate of the probability and magnitude of an adverse effect to an ecological receptor.	Probabilistic models
III	Refined probabilistic analysis, with exploration influence of uncertainty and variability associated with model parameters driving predictions. Moves away from general applications to incorporate more biologically and spatially explicit scenarios.	Estimate of the probability and magnitude of an adverse effect to an ecological receptor.	Probabilistic models
IV	Site-specific, environmentally relevant, species specific data generated under relevant pesticide use conditions.	Field studies, previous lines of evidence.	Multiple lines of evidence

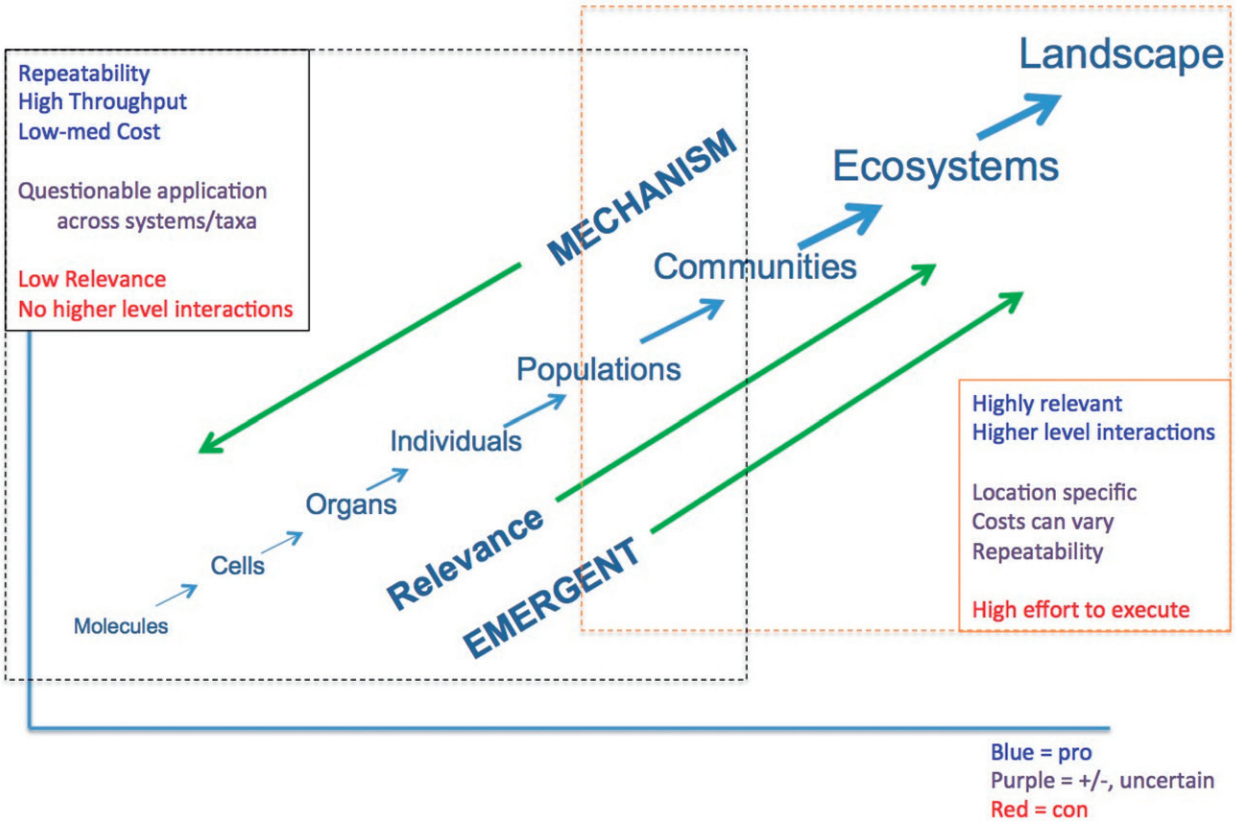

Figure 1. The pros, cons and uncertainties of ecological risk assessment based on data from different levels of biological organization. (This figure is presented in colour in the online version of the article.)

that are to be protected" (USEPA 1992, 1998) and are largely determined by what society (or involved stakeholders) perceives as ecosystem attributes worth protecting. Commonly, the public views the protection of vertebrate species and ecosystem services (ecosystem functions with specific values to humans) as a high priority but is unlikely to accept soil invertebrates as protection endpoints even though the latter can have very important roles in ecosystem function (USEPA 2003). For assessment endpoints to be useful, they should be specific, clearly defined and reflect management goals (see USEPA 2003 for guidance on selecting assessment endpoints). Measurement endpoints are "measurable responses to a stressor that are related to the valued characteristic chosen as the assessment endpoints" (Suter 1990; USEPA 1992; Suter 2007). It is the measurement endpoint that is used to infer effects on or protection of the assessment endpoints. Within the context of ERA of contaminated sites (a site-specific assessment), measurement endpoints might include actual measures of an assessment endpoint of interest. In most ERAs, however, the measurement endpoint is a measure of toxicity (e.g., LC50, NOAEC, etc.) obtained from a standard laboratory toxicity test using a model organism, whereas the assessment endpoint is often ecosystem function and biodiversity. However, occasionally the endpoint is more vaguely defined, such as pesticides shall have no unacceptable effects on the environment, as stated in EU pesticides legislation, which is why in the EU, higher tier testing using mesocosms has gained traction in ERA (Van den Brink 2006; Hommen et al. 2010; EFSA 2013; Van den Brink 2013).

If there is a disparity between what is generally measured (e.g., survival in laboratory toxicity tests) and the protection goal (e.g., ecosystem function and biodiversity), risk estimates that inform policy and environmental decisions may under- or over-estimate risk leading to environmental degradation or unnecessary remediation costs, respectively. Hence, a fundamental issue in ERA is the mismatch between the data available for use and the ecological systems that are the ultimate focus of protection (Figure 1).

Although the mismatch between measurement and assessment endpoints is well recognized, there are few reviews and syntheses of the advantages and disadvantages of ERA at different levels of biological organization, including mathematical modeling methods for ERA within and across levels of organization. The exception is the textbook Ecological Risk Assessment by Suter (2007), which offers a very good reference to various topics associated with ERA. Here our goal is to describe and highlight current experimental and extrapolation methods that are focused at different levels of organization as a means of improving our understanding of ERA across levels of biological organization and as a way to identify important gaps that warrant focused research attention. Hence, our targeted audiences are risk assessors, risk managers and researchers that can use this information to enhance ERA. The two essential elements of any ERA are estimates of exposure and effects, however, for this review we focus on effects estimation and extrapolation because it is the effects that manifest at different levels of biological organization. We first identify several key challenges in ERA that can be used to

frame a discussion around ERA at different levels of biological organization and how they may or may not address particular challenges. We further provide an overview of experimental approaches to obtaining data that reflect a specific level of organization and also provide an overview of extrapolation methods that are essentially used to make statements about effects at higher levels of organization based on data from lower levels of organization. We then highlight important uncertainties, data gaps and areas of future research.

Current challenges in ERA

As indicated above, a key challenge in ERA lies in linking data obtained from experiments or studies to estimates of risk that are meaningful to ecological receptors at different levels of organization. The following specific challenges can be useful in understanding how different levels of effects assessment can be used to address these challenges. Figure 1 provides an overview of the relationships between ERA at different levels of organization and the advantages and tradeoffs.

Unclear linkage between ERA outcomes and ecological endpoints of concern

A primary challenge in ecotoxicology lies in relating what we measure (e.g., acute toxicity study in *Daphnia magna*) to what society generally wishes to protect (e.g., ecosystem services delivered by freshwater systems). Or in ERA terms, relating measurement endpoints to assessment endpoints.

Ecotoxicological studies that closely follow published guidelines (OECD 2015) generally include endpoints such as survival, reproduction, and sometimes growth that relate to the potential fate of individuals of a species. While chemically-induced changes in these traits contribute directly to population characteristics, their impact on populations is not discernable in-and-of-themselves (unless there is 100% mortality or loss of reproduction). This is because the effects at the population level are dependent on the magnitude of the effect on a particular trait(s) but also on the relative influence of that trait on population dynamics (Forbes et al. 2010; Forbes et al. 2011; Rohr, Sager, et al. 2006), as well as other features of the environment. Further, it is possible that subtle, non-significant effects on different traits, when integrated or combined using population models, can generate significant effects at the population level (Luna et al. 2013). For studies in which surrogate communities are evaluated in mesocosms, the measurement endpoints (diversity, function) relate closely to system-level assessment endpoints. Thus, important points to consider when evaluating ecotoxicology and ERA across levels of organization are (1) how closely a given dataset or modeling approach tracks to assessment endpoints and (2) how clearly the linkages among data, models and assessment endpoints are described and whether there is a robust framework for describing those linkages.

Need to screen large numbers of chemicals

At the time of writing, there are about 100,000 chemicals registered for use in the European Union and over 92,000 registered in the U.S, additionally, there are over 100 million substances registered with the Chemical Abstracts Service (CAS), and approximately 15,000 new substances registered daily (PAN; CAS 2014; Stadnicka-Michalak et al. 2015). It is impossible to thoroughly study all chemicals that are in use and in development for future use (Rohr, Kerby, et al. 2006). The need for relevant toxicity data for a wide variety of existing and new chemicals is a significant challenge to ecotoxicology and ERA. Quantitative structure–activity models (QSARs) have been developed as a means of estimating toxicity of chemicals based on structure and/or chemical properties (e.g., octanol/water partition coefficient) and have been used for decades (Bradbury et al. 2003). QSARs continue to be developed and may hold promise for facilitating chemical assessments but are generally limited to providing insights into mode of action and organismal toxicity within chemical classes (Escher & Hermens 2002). With regard to actual experimentation to discern chemical toxicity, however, some study designs are more easily applied toward screening chemicals for toxicity in a relatively quick and cost effective manner. As indicated above, there is a tradeoff between study designs that lend themselves to screening chemicals versus those that produce data that provide insight into more ecologically complex scenarios. As an example, biochemical assays that are a key component of Endocrine Disruptor Screening Programs can be conducted in multi-well reaction plates. These assays generally test for activity of a test chemical towards a component of the endocrine system, such as vitellogenin levels (Nilsen et al. 2004) or steroidogenesis (Hecker & Giesy 2008). At the opposite extreme are studies that include more ecological complexity and, arguably, relate directly to potential risk in natural populations. For example, a seven-year study in which whole lakes were exposed to the endocrine disruptor, 17-alpha-ethanylestradiol, showed feminization of male fathead minnows at relatively low concentrations (Kidd et al. 2007). Studies of this nature are difficult to execute and thus clearly cannot be used routinely for ERA on different chemicals. Nonetheless, they provide evidence of potential chemical effects in natural systems (although this study admittedly did not have replication). While more ecologically complex studies are sensitive to important context dependencies in the wild, these same context dependencies might make it difficult to replicate the results across systems that might vary in biotic or abiotic traits.

Need to protect a wide variety of species

One significant challenge to ERA is that only a few species are normally tested compared to the vast diversity in real systems. As an example, there are approximately 1200 species of inland, freshwater fish in North America. Guideline toxicity studies, on the contrary, suggest toxicity testing on only a few commonly used species. Moreover, the species used in toxicity testing are frequently chosen because they are amenable to maintenance and use under controlled laboratory conditions and not because they have met some criteria with regard to their representativeness of species in the wild (Luttik et al. 2011). Unfortunately, there is considerable uncertainty regarding the amount of variability across species with

regard to responses to chemical stressors (Ibrahim et al. 2013, 2014). Nonetheless, an important consideration is how well certain approaches in ERA address variability among species (Luttik et al. 2011).

Multiple stressors and context dependencies

The vast majority of ERAs that have been conducted have focused on a single chemical or single active ingredient in isolation. Indeed, focus on the active ingredient is indicated in the U.S. Federal Insecticide, Fungicide and Rodenticide Act and in some EU directives (PAN; Jones et al. 2004; Hommen et al. 2010). In real environmental systems, however, multiple chemicals coexist (Halstead et al. 2014). As a blatant example, some pesticides are actually formulated as mixtures of multiple active and auxiliary ingredients and even under these conditions, the default approach in the U.S. is to assess risk as if the chemicals were in isolation (in the EU, tests of formulations are often required). In part, this might be justified in cases where the formulation includes pesticides with different targets (e.g., herbicide and insecticide) but it also applies when multiple active ingredients are combined and target the same pest (e.g., fungicides pyraclostrobin and fluxapyroxad are co-formulated in some products). Although there are multiple proposed methods for evaluating the toxicity and risk of mixtures, there is no global consensus on the optimal approach. Given the widespread and likely reality that ecological systems are exposed to chemical mixtures, this remains an important consideration for ERA.

In addition to mixtures of chemicals, other abiotic and biotic stressors in ecological systems can alter the toxicity and risk of a given chemical (Stampfli et al. 2011; Gergs et al. 2013). These "context dependencies" result in a range of sensitivities for a given species across a range of conditions that could include intra- or inter-specific competitors, predators, food availability, temperature, pH, disease, etc. In fact, abiotic stressors can increase toxicant sensitivity up to 30-fold in some population and community test systems (Liess et al. 2001; Liess & Beketov 2011). Importantly, responses over a gradient or combination of different contexts are infrequently monotonic or linear and are frequently difficult to predict. Nonetheless, these "other" stress factors are ubiquitously present in natural systems and have strong effects on the toxicity of chemicals. Although difficult to identify and apply across systems, species and habitats, context dependencies are widely recognized and demonstrated as important within ecotoxicology and ERA.

Include assessment of recovery

Current approaches in ERA essentially assume that ecological receptors and systems are static and that any estimated effects are presumed to result in nonreversible impacts to the receptor or system. Many species and systems, however, are capable of recovery. This is especially true for species or groups of species with high dispersal and short generation times. Also, within a landscape context, if risks are heterogeneously distributed, emigration from un-impacted areas can hasten the recovery of impacted areas. Recovery or perhaps more accurately, the capacity for recovery, is an important consideration in ERA because a system that recovers to pre-exposure conditions (or close to) may be considered to not be at risk. In this context, some level of biological effect may be permissible as long as it does not preclude a system from recovering to a desired level of the pre-exposure condition of a population or community (Kattwinkel et al. 2015). Although recovery is important, including it in ERA has been challenging and controversial partially because it depends on the traits of individuals investigated (e.g., generation time and dispersal abilities; Liess & von der Ohe 2005) and ecological contexts, such as competition (Liess et al. 2013).

Transparency and defensibility

Because multiple stakeholders and economic resources are frequently involved in environmental management decisions, ERA methods must be transparent and defensible. In part, the current methods were driven by a need for repeatability (low between study variance) and low variability (within study variance). Toxicity studies conducted following recognized guidelines (USEPA 1992, 1998; OECD 2015) benefit from having clear instructions and wide familiarity lending these toward easier acceptance by the regulatory community. Alternatively, studies and extrapolation procedures at higher levels of organization typically do not follow standardized methods and can be considerably "more complex", and thus it has been assumed that these studies are more challenging to repeat. As an example, the whole lake experiment evaluating the effects of EE2 on fish (Kidd et al. 2007) mentioned earlier would be challenging to repeat because factors will differ from lake to lake. However, repeatability and variability of effects assessment across levels of biological organization have not been systematically quantified. Additionally, some standardization in population, community, and ecosystem studies is certainly possible, which would enhance repeatability, transparency and defensibility. Moreover, standardized reporting procedures available for complex environmental modeling (Grimm et al. 2006, 2010; Forbes et al. 2011) increase transparency and defensibility and might offer an example that can be adopted for complex empirical designs. For instance, Schmolke et al. (2010) presented a standard format for documenting models and their analyses called TRACE (transparent and comprehensive ecological modeling), and Grimm et al. (2014) presented the first examples of TRACE documents, which are based on the idea of "model evaludation" (a fusion of "evaluation" and "validation") to describe the entire process of assessing a model's quality and reliability (Augusiak et al. 2014). TRACE has recently been incorporated into the European Food Safety Authority's opinion about good modeling practice (EFSA 2014).

Reduction in animal testing

Annually, more than a million fish are used for experimental and other scientific purposes in the European Union and between three and six million fish are used for whole effluent

testing in the United States (Scholz et al. 2013; Stadnicka-Michalak et al. 2015). These and other examples have created a movement in ecotoxicology and ERA to reduce the number of animals (primarily vertebrates) used for toxicity testing (Scholz et al. 2013; Burden et al. 2015). This is partly driven by a general desire to limit the suffering of living organisms. Additionally, there are economic benefits to reducing animal testing. Any proposed changes or additions to ERA methods must consider the impact on the use of animals (increase, decrease, no change) as the future is likely to see a stronger push toward limiting and justifying animal use in toxicity testing (Burden et al. 2015).

The pros and cons of ERA at different levels of biological organization

Below we discuss the pros, cons and unknowns of ERA across levels of biological organization, covering sub-individual, individual, population, community, ecosystem, and landscape levels of organization. At each focal level of organization, we first describe the types of studies, endpoints and mathematical modeling techniques that are regularly used. We do not discuss mathematical modeling techniques used to extrapolate across levels of organization because this content is relegated to the section below "Models relating responses at different levels of organization".

Sub-individual level: biochemical and molecular responses

Overview

Sub-individual responses to chemical stressors historically encompass biochemical and/or physiological measures that were collectively referred to as biomarkers. Although many biochemical and physiological biomarkers are theoretically possible, a few have emerged as more common across taxa, including metallothionein induction, acetylcholinesterase inhibition, cytochrome P450 induction, persoxisome proliferation and several indicators of oxidative stress (Cajaraville et al. 2000; Nel et al. 2013). While there has been a strong desire to apply biomarkers to ERA (e.g., Handy et al. 2003), they have also been widely criticized as unlikely to provide predictable insight into adverse effects at the level of the whole organism (Forbes et al. 2006). Instead, biomarkers can provide insight into potential mechanisms of toxicity or whether exposure to a contaminant or family of contaminants has occurred. With the advent of modern techniques in molecular biology, a wider range of sub-individual responses are now quantifiable and include measures of mRNA transcripts, proteins and metabolites. However, in many ways, the same criticisms levied for biomarkers can apply to any suborganismal level response whether physiological, biochemical or genetic – it is difficult to relate these responses to ecologically meaningful effects at higher levels of biological organization.

Recently, however, the potential utility of suborganismal responses has been rejuvenated by the development of the Adverse Outcome Pathway (AOP) concept. An AOP is a conceptual framework that portrays a sequential chain of causally linked events starting with a chemically-induced Molecular Initiating Event(s) (MIE) and culminating in an actual Adverse Outcome (AO) in a biological level of organization relevant to ERA (Ankley et al. 2010). The AOP construct has primarily presented AO's as disruption of physiological homeostasis at the cellular, tissue or organ levels (Ankley et al. 2010) with only tenuous linkages to higher levels of organization, such as at the population level (Kramer et al. 2011). Thus far, AOPs have predominantly focused on identifying MIEs and proximate downstream effects and have entailed the use of molecular techniques, such as tools to quantify gene expression (e.g., PCR, microarrays, RNA-seq) and hormone levels (e.g., ELISAs), to link MIEs to AOs and to enhance the interpretation of sub-individual responses (Ankley et al. 2010; LaLone, Villeneuve, Burgoon, et al. 2013; LaLone, Villeneuve, Cavallin, et al. 2013; Berninger et al. 2014; Martinovic-Weigelt et al. 2014). Once AOPs are well-described, regulatory agencies have encouraged scientists to make them accessible in an AOP knowledge-base and wiki. After which, researchers would ideally evaluate the conservation of AOPs across taxa. However, very little progress has been made on this latter goal given how nascent the AOP framework is (but see Ankley & Gray 2013; LaLone, Villeneuve, Burgoon, et al. 2013; LaLone, Villeneuve, Cavallin, et al. 2013).

Pros

There are several advantages of using suborganismal endpoints and the AOP construct for ERA (Table 2). First, AOPs explicitly emphasize cause–effect relationships by requiring the elucidation of causal links between an MIE and an AO. Because many biomarker and MIE tests are done with cell or tissue cultures or with blood or tissues samples (Ankley et al. 2010; LaLone, Villeneuve, Burgoon, et al. 2013; LaLone, Villeneuve, Cavallin, et al. 2013; Nel et al. 2013; Berninger et al. 2014; Martinovic-Weigelt et al. 2014), they can reduce the need for laboratory animals (Ankley et al. 2006). Additionally, these tests are presumed to have low variability and high repeatability within a species. However, this latter claim has not been thoroughly tested given the early stages of the AOP framework.

Given the hundreds of thousands of chemicals and chemical cocktails (EU 2001; Touart & Maciorowski 1997), a high-throughput screening (HTS) option for chemical risk assessment is very attractive, and thus, the real selling point to using biomarkers or MIE tests has been their potential value in screening large numbers of chemicals (Ankley et al. 2010; Ankley & Gray 2013; Nel et al. 2013). Tests targeted at quantifying biomarkers or MIEs regularly use assays conducted on 96- or 384-well plates, which can greatly speed-up risk assessment (Nel et al. 2013; Berninger et al. 2014; Martinovic-Weigelt et al. 2014). Adding to this is that advances in AOP research can, when combined with chemical structural information and physical–chemical properties, inform the development of QSARs (Ankley et al. 2010). In turn, QSARs could be used to identify chemicals for priority testing, which would support efforts to streamline ERA and reduce testing.

Unknowns

The cost of sub-individual level studies can be highly variable and thus is placed in the "unknown" category. If biomarker or

Table 2. Relationship between traits of ecological risk assessment (ERA) and levels of biological organization.

Trait of ERA	MIE*/biomarker	Individuals (conventional ERA)	Populations	Communities	Ecosystem	Landscapes
Ability to manipulate chemicals to address cause and effect	High	High	High	High	Med	Low
Need for animals	Low	High	High	High	High	High
Need for vertebrate animals	Low	High	Low	Med	Med	Med
Repeatability (between study variance)	High	High	Med	?	?	?
Variability (within study variance)	Low	Med	Med	?	?	?
Cost per study	Low/high	Med	Med	High	?	High
Cost per species tested (assuming species-specific information is attained)	Low/high	Med	Med	?	?	High
Ease at screening a large number of chemicals	High	Med	Low	Low	Low	Low
Endpoint relevant to ERA	Low	Low	Med	High	High	High
Results sensitive to ignored feedbacks (e.g., physiological, neg, density dependence, species interactions)	High	High	Med	Low	Low	Low
Ability to capture recovery associated with reproduction	Low	Low	High	High	High	High
Ability to capture recovery associated with dispersal	Low	Low	Low	Med	Med	High
Ability to screen a large number of species	Med,	Med	Low	High	High	?
Ability to address common context dependencies and multiple stressor effects	Low	Med	Med	High	High	Low

*Molecular initiating event.

MIE assays focus on one endpoint, they may be relatively inexpensive per study and per species tested. However, if transcriptomic-based approaches are required or encouraged (Berninger et al. 2014; Martinovic-Weigelt et al. 2014), then quantification of biomarker or MIEs can be quite costly both per study and per species. Hence, biomarker and MIE approaches have great potential for HTS (Ankley & Gray 2013; Nel et al. 2013), but their relevance for ERA and cost remain to be seen.

Cons

Even if biomarker and MIE tests can offer inexpensive, HTS, there are several concerns regarding the value of the data they produce for ERA (Table 2). Uncertainty in ERA is often reduced by minimizing the distance between the measurement and assessment endpoints (USEPA 1992, 1998; Suter et al. 2005; Suter 2007, 2008). However, biomarker and MIE tests presently have a greater distance between their measurement endpoints and ideal assessment endpoints for ERA than any of the other biological levels of organization addressed in this review. The measurement endpoints for biomarker and MIE tests are predominantly processes at the cellular or tissue levels (gene expression, hormone levels, protein abundance, fibrosis, etc.) (Ankley et al. 2010; Ankley & Gray 2013; Nel et al. 2013; Berninger et al. 2014; Martinovic-Weigelt et al. 2014), whereas the ideal assessment endpoints for ERA, as highlighted in the Introduction, are usually at the population level or above.

As a consequence of the level of their measurement endpoints, biomarker and MIE tests might not capture many of the realities of the natural world, such as important positive and negative feedbacks in biological systems. At the level of the whole organism, there are common physiological negative feedback mechanisms that can counter adverse effects at cellular, tissue or even organ levels, such as those that inhibit production of or responses to hormones (Meaney et al. 1996; Martin et al. 2010). There are also positive feedbacks; mitochondria experiencing oxidative stress may themselves enhance the production of reactive oxygen species (Park et al. 2011) and several positive feedbacks in the endocrine system that require interactions among multiple organs (Ewer et al. 1997; Bulun et al. 1999; Wintermantel et al. 2006). These feedbacks could be missed if biomarker and MIE tests are not repeatedly validated at the level of the whole organism. At the level of populations, there are also negative feedbacks (Ives 1995) that are missed by biomarker and MIE tests. For example, contaminant-induced losses of individuals do not always cause population-level declines because the survivors of the contaminant exposure might experience less competition for resources (negative density dependence) and thus might produce the same or even more offspring as were produced in the absence of the contaminant (Vonesh & De la Cruz 2002; Rohr, Sager, et al. 2006; Forbes et al. 2008, 2011; Thorbek et al. 2009; Rohr & Palmer 2013). Hence, even if a chemical kills individuals, it might not cause any reductions in population growth rates because of density-mediated compensatory responses (Moe et al. 2002; Rohr, Sager, et al. 2006; Salice, Rowe, et al. 2011). Similarly, at the community level, adverse effects on a species might be counteracted by greater or equally adverse effects on the species' natural enemies (Rohr & Crumrine 2005; Rohr, Raffel, et al. 2008; Rohr, Schotthoefer, et al. 2008; Rohr & McCoy 2010b; McMahon et al. 2013; Rohr, Raffel, et al. 2013), again resulting in stable population growth and perhaps stable delivery of ecosystem services (McMahon et al. 2012; Halstead et al. 2014). Similarly, species and populations can rapidly recover from short-term, adverse effects of contaminants, through either reproduction or dispersal (Stark et al. 2004; Rohr, Kerby, et al. 2006; Clements & Rohr 2009). By not studying at least individuals, recovery from contaminants is challenging to reliably assess using the biomarker and MIE tests.

Biomarker and MIE tests can also miss important positive feedbacks that can occur at individual, population and community levels (Ives 1995; Nisbet et al. 1996; Rohr, Kerby, et al. 2006; de Roos & Persson 2013). At the population level, Allee effects are positive density dependence (e.g., organismal fitness increases with conspecific densities) that facilitate mate finding and reproduction that can be crucial for population growth of small populations (Courchamp et al. 1999; Stephens & Sutherland 1999). Similarly, at the community level, a contaminant might have adverse effects on a mutualist or a

foundation species (a species that provides habitat for another species) resulting in secondary species declines or extinctions and thus greater adverse effects than would be predicted by biomarker, MIE, individual-, or even population-level studies (Ebenman & Jonsson 2005; Rohr, Kerby, et al. 2006; Beketov et al. 2013).

In addition to missing important feedbacks at the individual, population and community levels of biological systems, it is not entirely clear how biomarker and MIE tests at the subcellular to tissue or organ levels will adequately capture different abiotic and biotic conditions to which whole organisms are exposed (Boyle & Fairchild 1997). This is a significant limitation because the toxicity of many contaminants is often highly context dependent, changing with environmental conditions, such as light, pH, hydroperiod and temperature gradients (Barron et al. 2003; Rohr et al. 2004, 2011; Rohr & Palmer 2005; Noyes et al. 2009; Stampfli et al. 2011; Kimberly & Salice 2013; Rohr, Johnson, et al. 2013), and common intra- and inter-specific interactions (Relyea 2003; Relyea et al. 2005; Rohr & Crumrine 2005; Rohr, Raffel, et al. 2008). For example, warming up cells, tissues, or organs in isolation will not necessarily produce the same results as warming up an entire organism. Moreover, testing individuals in the absence of natural enemies and mutualists or testing social organisms in the absence of conspecifics could affect toxicity estimates (Relyea 2003; Relyea et al. 2005; Rohr & Crumrine 2005; Rohr, Raffel, et al. 2008; Salice & Kimberly 2013). For instance, several pesticides are more deadly to amphibians in the mere presence of chemical cues from predators (Relyea 2003; Gergs et al. 2013) and many contaminants are immunomodulators (Voccia et al. 1999; McMahon et al. 2011) or can have selective pressures that enhance the risk of infections (Salice & Roesijadi 2002; Rohr, Raffel, et al. 2008; Rohr, Schotthoefer, et al. 2008; Rohr & McCoy 2010b; Rohr, Raffel, et al. 2013). Thus, at least initially, it could be challenging for biomarker, MIE and even AOP approaches to avoid validating that these endpoints are strongly correlated with endpoints at the level of the individual, and even these studies will miss context dependencies that can only be observed at higher levels of biological organization.

Finally, although biomarker and MIE tests attempt to maximize control, standardization and repeatability and minimize complexity, there are presently few standardized methods for MIE- and AOP-based approaches. Additionally, biomarker, MIE and AOP approaches lack the standardized or semi-standardized mathematical models to facilitate risk assessment that are available for higher levels of biological organization, such as populations and communities (see below). However, both standardized methods and mathematical models are being developed (e.g., Kramer et al. 2011; Stevenson et al. 2013; Muller et al. 2014). For all of the reasons described above in this "Cons" section, uncertainty appears to be very high for biomarker, MIE and AOP approaches to ERA, perhaps requiring even larger uncertainty factors than used presently.

Individual level

Overview

Studies at the individual-level probably represent the most common level of biological organization used for ERA as they are the mainstay of ecotoxicological studies and are legally mandated in support of some chemical assessments (USEPA 1992, 1998; OECD 2015). These studies often call for the use individuals of a single species exposed to different concentrations and types of chemicals under otherwise benign rearing conditions (Nabholz et al. 1997; Touart & Maciorowski 1997). Individual-level studies are conducted for different durations to generate toxicity estimates for acute and/or chronic exposure durations; in acute studies, organisms are commonly grouped within replicates while for chronic studies, species can be housed individually or in groups within replicates. As a variant of guideline studies, more than one individual of a species can be placed in each replicate to specifically capture intraspecific interactions (Rohr et al. 2003, 2004; McMahon et al. 2011, 2013; Jennings et al. 2012), but these intraspecific interactions can affect the measurement endpoints and may therefore be more challenging to incorporate into ERA. Common measurement endpoints in acute individual-level studies are the estimated concentration or dose that kills 50% of individuals after a specified exposure duration (LC50 and LD50, respectively), growth, development, behavior and various physiological measurements (USEPA 1992, 1998; Nabholz et al. 1997; Touart & Maciorowski 1997; Halstead et al. 2015; OECD 2015). In longer duration, chronic studies, common endpoints include the No-Observed-Effect-Concentration (NOEC), the Lowest-Observed-Effect-Concentration (LOEC), as well as ECx estimates that describe the concentration at which a certain percentage effect on an endpoint occurs after a specific exposure duration (e.g., the EC10 for growth in *Daphnia* is the concentration at which there is a 10% decrement in growth compared to *Daphnia* in the control). There has been considerable criticism of NOEC-based endpoints, particularly because they are so sensitive to sample size (Crane & Newman 2000; Jager 2011, 2012; Landis & Chapman 2011). Although ECx approaches might be mild improvements over NOECs, more probabilistic risk assessments that also consider duration of exposure are preferable (Crane & Newman 2000; Jager 2011, 2012; Landis & Chapman 2011). Additionally, for most individual-level studies in which reproduction is an endpoint, adults are removed to fresh feeding suspensions, the young are counted and may be measured or weighed, but are not continued in the study protocol. Typically, individual-level studies are shorter in duration than the generation time of the study organism and, combined with removal of offspring, are thus unlikely to capture population dynamics that tend to occur over longer time periods (Walthall & Stark 1997; Stark 2005).

An important extension of single-species toxicity tests is the use of Species Sensitivity Distributions (SSDs) in ERA. SSDs are statistical models that include toxicity estimates from numerous species and are designed to identify chemical concentrations that are protective of an assemblage of species (Posthuma et al. 2001; Hose & Van den Brink 2004; Maltby et al. 2005). SSDs are beneficial because they begin to address, to some extent, the known variation in toxicant sensitivity across species and there is evidence that their use in ERA is protective (Hose & Van den Brink 2004; Maltby et al. 2005). However, they have also been criticized because the data used to populate SSDs are still fundamentally based on individual-level toxicity studies/endpoints, there can be

difficulties across species in finding common toxicity endpoints (especially for chronic exposures), the results are sensitive to the chosen modeling distribution, they lack explicit consideration of mechanisms, and the species toxicity data incorporated into SSDs are frequently not represented in specific or particular systems (Newman et al. 2000; Forbes & Calow 2002b; Baas & Kooijman 2015). Alternatively, some of their value lies in providing a sense of inter-species toxicant sensitivity and perhaps in finding generalities that can be used to improve overall understanding or predictive ability regarding toxicant effects (Baird & Van den Brink 2007). For example, Baas and Kooijman (2015) assessed the sensitivity of fifty species to four pesticides and found that high specific maintenance rate (metabolic rate) correlated with increased toxicant sensitivity. SSDs have clear limitations in some applications but hold promise as a way to explore the role that species traits play in predicting toxicant sensitivity.

Pros

Individual-level studies have several advantages (Table 2). First, there is a long-history of their use in toxicology and ERA and there are many established and standardized methodologies (USEPA 1992, 1998; EFSA 2013; OECD 2015). Individual-level tests generally attempt to maximize control and thus are effective at assessing cause–effect relationships between contaminant exposure and the measurement endpoint and typically are assumed to have low variability and high repeatability under the same contexts and sources of test organisms. Most acute toxicity tests only last for 96-h or less with survival as the only measurement endpoint. Thus, these tests offer medium ease at screening large numbers of chemicals. For very small organisms, some of these tests can even be conducted in 96- or 384-well plates (e.g., Parng et al. 2002; Chandler et al. 2004), but most individual-level tests require larger containers. Hence, on average, they probably are not as efficient as many 96- or 384-well biomarker or MIE plate assays, but they are certainly more feasible for HTS of chemicals than tests at higher levels of organization.

Unknowns

Costs of many toxicology and ecotoxicology studies are not well reported in the literature and thus there is some uncertainty regarding the costs of studies at various levels of biological organization. However, given their short duration and standardized methods, most assays at the individual level are probably of medium cost per study and per species (Table 2). We can imagine certain plate assays that could be cheaper, and many assays at higher levels of biological organization that could be more expensive. If there was a single or even a few species that were unilaterally the most sensitive to every chemical, then standardized LC50/LD50 assays would theoretically protect biodiversity. However, there is little evidence to support the "most sensitive species" concept (Cairns 1983, 1986; Cairns & Niederlehner 1987).

Cons

Like any single approach to ERA, individual-level studies have their weaknesses (Table 2). First, despite efforts to reduce animal testing especially on vertebrates, in several countries, ERA and chemical registration decisions are based on individual-level tests that require both vertebrate and invertebrate animals (Hofer et al. 2004; Ankley et al. 2006; Ahlers et al. 2008). Although guideline tests at the individual-level can capture physiological feedbacks within organisms, they cannot capture feedbacks at the population or community levels and generally do not assess recovery from contaminant exposure because these assays typically prevent reproduction and colonization and often do not include observations on organisms well after the chemical exposure period (to capture any recovery). Additionally, individual-level tests often do not assess common context dependencies in nature, such as temperature fluctuations that can affect physiology (Raffel et al. 2006, 2013; Kimberly & Salice 2014), other environmental conditions that can affect toxicity, or sublethal effects (Rohr et al. 2009; Ehrsam et al. in press). Moreover, they do not include intra- and inter-specific interactions that can increase or decrease the adverse effects of contaminants (Rohr, Kerby, et al. 2006; Clements & Rohr 2009). As an example, if there are no effects on a focal species at the sub-individual- or individual-levels but the contaminant decimates the focal species' natural enemies or prey that it requires to survive (Raffel et al. 2009; Staley et al. 2012), then the lack of effects at the sub-individual- or individual-level will produce an indirect effect at the community level (Rohr, Raffel, et al. 2008; Rohr, Schotthoefer, et al. 2008; Staley et al. 2010, 2011, 2014). For all the reasons described above, the individual-level of biological organization has the second greatest distance (behind biomarker and MIEs) between the measurement endpoint and ideal assessment endpoints, which produces considerable uncertainty in ERA.

Population level

Overview

Population-level studies vary considerably in their approaches to assessing risk from contaminants and can generally be characterized as (1) observation based, (2) model based or (3) both. Traditionally, a true population-level study would entail direct observations on the population of interest (time series data) and occur at a time scale long enough to capture salient populations dynamics, which typically translates to enough generations of an organism to assess trends in population growth rates (Hansen et al. 1999; Forbes et al. 2001; Forbes & Calow 2002a; Moe et al. 2002; Salice et al. 2009). Hence, the duration of these studies is highly dependent on the generation time and life span of the organism of interest. For viruses, bacteria, phytoplankton and protozoans, actual population-level tests can be completed on the time scale of hours to days. For many plants, gastropods, crustaceans and arthropods, population-level tests can last weeks to months. For many vertebrates and trees, however, population-level tests would require years to decades. As a consequence of this variation, many more population-level toxicity experiments have been conducted on organisms with short than long life spans. In fact, for very small organisms with short generation times (e.g., bacteria, phytoplankton), it is much easier to obtain population-level data than to obtain

individual- or sub-individual-level data. Hence, population-level data are often more abundant than individual- or sub-individual-level data for microbes.

For organisms with long life expectancies, observational studies can be particularly challenging and mathematical models can play a more prominent role in generating population-level estimates of risk. In this case, experiments on individual organisms that capture the effects of contaminants on key life history traits, such as rates of growth and development, fecundity, reproduction, or dispersal, and experiments on collections of organisms to estimate the strength of density dependence, can be valuable in estimating potential population-level effects (Forbes & Calow 2002a; Salice & Miller 2003; Stark 2005; Luna et al. 2013; Martin et al. 2013). Often, these experiments on key aspects of population dynamics are used to parameterize population-level models that can integrate these effects to more defensibly evaluate the consequences of a contaminant on populations (Salice & Miller 2003; Salice, Sample, et al. 2011; Luna et al. 2013; Martin et al. 2013; Erickson et al. 2014) than if any subset of these aspects were considered or if only a sub-individual- or individual-level approach was implemented. Although quantitative predictions from any mathematical model are only as good as its parameterization, model structure and validation (Augusiak et al. 2014; Grimm et al. 2014), population models can perform well in predicting the impacts of contaminants on the population dynamics of organisms with short generation times (e.g., Civitello et al. 2012; Martin et al. 2013; Stadnicka-Michalak et al. 2015). This suggests that, if adequately developed and validated, population-level models can be valuable for assessing risk to long-lived organisms (Forbes et al. 2010; Forbes et al. 2011; Muller et al. 2014). Additionally, in assessment scenarios where a single species is a focal point for an assessment endpoint, population-level approaches are particularly relevant. One such application is toward assessing chemical risk to threatened and endangered species (Forbes et al. 2010, 2011, 2015).

Pros

The pros and cons of population-level approaches to ERA often depend on the identified assessment endpoints, the availability of data, and on the life history (including longevity) of the organisms of interest. For short-lived organisms where actual population-level dynamics can be observed in experiments, the ability to generate data that convincingly demonstrates a cause–effect relationship between a contaminant and changes in population attributes (e.g., abundance, density, age- or size-structure, propensity to cycle) is high (Table 2). However, the strength of this inference declines as data become more difficult to obtain as life-span increases and only components of population dynamics can be observed. Hence, we have assigned in Table 2 a medium score for assessing cause–effect relationships between chemical exposure and population-level effects and for clearly linking measurement endpoints to assessment endpoints. An important caveat would be if the assessment endpoint was a population of a particular organism for which there were sufficient data to understand or project changes in population attributes resulting from chemical exposure. In this special case, the population-level approach would be scored high.

Given that it is not practical in most cases to conduct true population-level experiments on vertebrates, most population-level studies are conducted on invertebrates and thus a population-level approach to ERA often does not use many vertebrate animals (Stark & Banks 2003; Maron & Crone 2006). However, this also means that there is generally more uncertainty in assessing risk for vertebrates than invertebrates. The use of population-level models based on parameters extracted from the literature can also reduce the need for invertebrate and vertebrate animal testing, although this could add uncertainty.

The most important benefit of population-level studies is that their measurement endpoints are reasonably close to their assessment endpoints and this level of biological organization is among the most relevant and tractable for environmental management (Forbes et al. 2010, 2011). For most threatened, endangered, or at-risk species (TERS), population growth or abundance is the assessment and management endpoint, which is what is directly or indirectly being measured in population-level studies (Stark et al. 2004; Forbes et al. 2010, 2011). This should reduce uncertainty in the ERA. Moreover, most population-level studies capture some level of recovery from the contaminant exposure by allowing enough time for detoxification, density-mediated compensation, or dispersal (e.g., Salice & Miller 2003; Salice, Sample, et al. 2011; Civitello et al. 2012; Luna et al. 2013; Erickson et al. 2014) and thus capture more important features of the natural world than do sub-individual- or individual-level approaches.

Unknowns

Although costs for most toxicology studies are not well documented, we believe that there is likely a medium cost per study and per species for population-level tests (Table 2). Population-level modeling can be reasonably inexpensive because it only entails labor and a computer, whereas actual experiments have those costs in addition to costs for materials and supplies. Actual population-level experiments, however, are often much more long-term than traditional LC50/LD50 tests and thus will generally be more costly, especially for long-lived organisms where studies either have to be very long-term or multiple studies need to be conducted to capture different aspect of population dynamics. Importantly, however, models are often more cost-effective than experiments once models are developed and validated because they can be applied to new chemicals at a fraction of the cost, whereas repeating an experiment for each new chemical will incur the same cost every time (Table 2).

Like cost, the repeatability and variability of population-level toxicology studies is not well documented in the literature.

Cons

A substantial limitation of population-levels approaches to ERA is that they seldom allow screening large numbers of chemicals or species relative to sub-individual- and

individual-level approaches (Table 2). This is because population-level experiments can be moderately costly and time consuming. Even if mathematical models are employed, they generally will require some experiments on each chemical to properly parameterize the model (e.g., Salice & Miller 2003; Salice, Sample, et al. 2011; Civitello et al. 2012; Luna et al. 2013; Erickson et al. 2014). One approach around this might be to use exciting recent advances in the science of ERA that are based on the similar sensitivities and toxicities of related species and chemicals (Hammond et al. 2012). Guenard and colleagues (Guenard et al. 2011, 2014) developed and validated statistical models that reliably predict the toxicities of untested chemicals and species based on already available toxicity studies combined with species and chemical phylogenies. These statistical models, coupled with population parameter estimates from the literature, theoretically could be incorporated into mathematical models at the population level to identify particularly insidious chemicals to populations and species particularly sensitive to population-level declines given realistic exposure to specific contaminants (Rohr, Kerby, et al. 2006). However, these models admittedly rely on 96-h LC50 that generally ignore exposure durations and have the limitations described above.

Another major limitation of population-level approaches to ERA is that they generally do not consider meta-population or community-level dynamics. Metapopulation dynamics are driven by connections among populations associated with immigration and emigration (Hanski 1998; Spromberg et al. 1998). Dispersers from sites with stable or positive population growth can rescue populations from short-term negative population growth that might be associated with contaminant exposure (Hanski 1998; Spromberg et al. 1998). Likewise, nearby ecological traps that attract dispersers despite poor conditions can exacerbate negative population growth (Schlaepfer et al. 2002; Vonesh & Kraus 2009). Similarly, species interactions can improve or worsen the population-level impacts of contaminants, as described in previous sections.

For a handful of organisms that can be purchased and have standardized methodologies for their husbandry (e.g., *Daphnia* spp.) (Baird et al. 1990; Stark & Banks 2003; Heckmann et al. 2007), population-level experiments are likely moderately repeatable. Many species, however, cannot be purchased and thus source populations can differ in their traits that can affect toxicity (Semlitsch et al. 2000; Cothran et al. 2013; Hua et al. 2013). This, however, is a concern for any level of biological organization. Moreover, the lack of standardized methods for population-level experiments means that they often will differ in their resource levels, which can affect several key aspects of population dynamics (growth, reproduction, the strength of negative density dependence) (Forbes & Calow 2002a). Additionally, mathematical models at the population level can occasionally be sensitive to small differences in key parameters. Hence, for these reasons, we believe that there is a medium level of inter-assay variability and repeatability for population-level studies; however, this admittedly has not been well quantified, which is why we have conservatively placed these estimates in the "unknown" category (see "Unknowns" above).

Although population-level studies have measurement endpoints near assessment endpoints for TERS, in most other cases, the measurement and assessment endpoints are not very closely connected. This is because the assessment endpoints that are often the easiest "sell" to the public are ecosystem services, and most ecosystem services, such as decomposition, decontamination, biocontrol, disease control, clean water and pollination, are a product of the functions of species assemblages not the populations of single species (Hooper et al. 2005; MilleniumEcosystemAssessment 2005; Cardinale 2011; Cardinale et al. 2012; Civitello et al. 2015; Rohr et al. 2015) (there certainly are exceptions, such as certain species humans consume for food or emphasize for recreation, like particular fish species). Hence, even population-level endpoints might often have considerable uncertainty in assessing risk to endpoints most critical to the public.

Community level

Overview

Community-level studies quantify the effects of contaminants on interacting species, often capturing at least one of the following types of interactions: facilitative, competitive, predator-prey, or host-parasite interactions. Community-level toxicological studies can be manipulative or correlational. They also can capture different levels of ecological realism, including examinations of 1) simple species interactions under laboratory conditions, 2) complex communities in the laboratory, 3) complex communities in outdoor mesocosms and 4) compositional changes of species in nature. Thus, their pros and cons will, in part, depend on whether they can assess causation and the level of ecological realism they capture (Joern & Hoagland 1996; Taub 1997b; Levin et al. 1989).

The U.S. Environmental Protection Agency used to have four hierarchical tiers in their ERA protocols for pesticides, where the fourth tier required registrants to test the effects of the chemical outdoors, often on freshwater communities in mesocosms (Nabholz et al. 1997; Touart & Maciorowski 1997) (Table 1). As a consequence, in the 1980s and early 1990s, there was a push in the U.S. to develop standardized microcosm and mesocosm methodologies for ERA (Graney et al. 1994; Hill et al. 1994; Taub 1997b, 1988). In 1992, however, the US EPA did away with the fourth tier of testing, eliminating aquatic mesocosm and field studies as an obligatory tier of testing under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) (Nabholz et al. 1997; Touart & Maciorowski 1997). Personal interactions between the authors and EPA personnel, in addition to various memos (USEPA 1993), suggest that the primary reasons the EPA eliminated mesocosm testing and field study requirements was because they claimed that, relative to standard laboratory toxicity studies, mesocosm and field studies were more expensive, less repeatable and less expedient (Nabholz et al. 1997; Touart & Maciorowski 1997). Aquatic mesocosms remain an occasional component of ERA in the European Union (Van den Brink 2002, 2006, 2013; Traas et al. 2004; Hommen et al. 2010). Mesocosms represent the most common and arguably most defensible way to experimentally determine toxicant effects at

the community level; for this reason we emphasize mesocosm studies in this section.

The endpoints of community-level studies can include the abundance of many species in a community, the relative abundance of species (i.e., evenness), species dominance, species richness and/or diversity indices (e.g., Shannon Weiner) (Relyea 2005; McMahon et al. 2012; Halstead et al. 2014). In many cases, population dynamics (tracking populations through time) and ecosystem functions and services are also quantified in community-level studies (McMahon et al. 2012; Halstead et al. 2014), but quantifying these endpoints is not necessary for a study to be classified at the community level.

Like with population-level approaches to ERA, there are traditional mathematical modeling frameworks that can be useful in assessing the impacts of contaminants at the community level. Perhaps the most well-known are ordinary differential equations (ODEs). ODEs are the basis of classic Lotka-Volterra models of predator-prey dynamics (Holt & Polis 1997) and susceptible-infected-recovered models for host-parasite dynamics (Anderson & May 1991). In the context of ecotoxicology, these models can elucidate many indirect effects associated with competition, predation and trophic structure (Fleeger et al. 2003). Additionally, there are network-based modeling approaches based on food web topology (USEPA 2000; Dunne et al. 2002) and machine learning models (Van den Brink et al. 2002) that can estimate indirect effects and the risk of secondary extinctions. At the very least, these models can offer null expectations (Rohr, Kerby, et al. 2006) and in several cases they have successfully predicted the impacts of chemicals on species assemblages (USEPA 2000; Van den Brink et al. 2002).

Pros

One of the biggest advantages of community-level studies is that they can capture many potential feedbacks in biological systems (Joern & Hoagland 1996; Taub 1997b; Van den Brink 2006, 2013) (Table 2). For instance, for some species, outdoor community-level mesocosm studies can capture positive and negative density-dependence at the population level (for those species that can reproduce in the mesocosms during the experiment), recovery from contaminants based on detoxification, reproduction and dispersal (e.g., flying insects, some zooplankton, some algae, some microbes), and species interactions that can alter effects of contaminants (Staley et al. 2010, 2014; Douglas et al. 2015; Jayawardena et al. 2016). Additionally, outdoor mesocosm studies can do a good job of capturing cause–effect relationships between a contaminant and species densities. However, statistical approaches, such as structural equation models, or additional experiments are often necessary to determine whether the observed effects of the contaminant are direct or indirect (mediated by species interactions) (Rohr, Schotthoefer, et al. 2008; Schotthoefer et al. 2011; McMahon et al. 2012; Halstead et al. 2014). Effects of contaminants on species interactions in outdoor mesocosms studies have also been shown to match effects of contaminants in natural systems (Larsen et al. 1986; Pontasch et al. 1989; Stay et al. 1989; Niederlehner et al. 1990; Pontasch & Cairns 1991; Taub 1997b; Rohr, Schotthoefer, et al. 2008), demonstrating that outdoor mesocosm studies can capture the complexities of natural systems without sacrificing the ability to assess causal relationships. Studies of species interactions in the laboratory, of course, are less likely to capture the realities of communities in the wild and correlational studies at the community level cannot determine causality.

Community-level studies can also capture common and important context dependencies that can affect toxicity. These include the presence of predators and species interactions. Additionally, outdoor mesocosms and field surveys offer natural variation in abiotic conditions, such as temperature and ultraviolet radiation, and provide natural substrates upon which contaminants can adhere, realistically reducing their bioavailability. Because of strong evidence of a positive relationship between biodiversity and ecosystem functions and services (Hooper et al. 2005; MilleniumEcosystemAssessment 2005; Cardinale 2011; Cardinale et al. 2012), community-level studies allow scientists and regulators to simultaneously assess the effects of contaminants on the populations of species of conservation concern and to estimate the likely impacts of the contaminant on ecosystem functions and services. For all the reasons above, community-level studies, especially outdoor mesocosm experiments that can capture both cause–effect relationships and ecological realism, provide measurement endpoints that may be close to assessment endpoints and thus might offer lower uncertainty to ERA.

Microcosm and mesocosm studies also offer an efficient approach for screening effects of contaminants on large numbers of species. Many community-level studies quantify the effects of contaminants on tens of naturally co-occurring species (Taub 1997b; Relyea 2005; Van den Brink 2006, 2013; McMahon et al. 2012; Halstead et al. 2014). With advances in DNA sequencing that has made it easier and cheaper to quantify the abundance of microbial species, many community-level toxicological studies are now quantifying the effects of contaminants on hundreds of sympatric species (Engelen et al. 1998; Johnsen et al. 2001; Nielsen et al. 2014). Moreover, community-level studies can offer information on the risk of a contaminant to many species under ecologically-relevant conditions. Thus, mesocosm studies capture many ecological realities on tens to hundreds of species and might do so much more efficiently than conducting LC50 (or other standardized tests) tests on each species separately.

Although many community-level toxicological studies require animal testing, some approaches restrict the investigation to invertebrates and thus do not require vertebrate animals. Others involve solely plant or microbial communities. Additionally, Taub and colleagues (Taub 1997b) developed protocols for testing chemicals in standardized aquatic microcosms in the laboratory that only consist of primary producers and aquatic invertebrates. Excluding vertebrates that are common in most ecosystems increases uncertainty in ERA but it does help to deal with pressures to reduce vertebrate animal use (Burden et al. 2015). Taub and colleagues' (Taub 1997b) protocols are described in the American Society for Testing Materials (ASTM) E1366-11 Standardized Aquatic Microcosm: Fresh Water (ASTM 2011) and provide quality control guidelines for the control replicates. Moreover, the work of Taub and colleagues offers standardized methods for

community-level experiments, which can increase repeatability of results and reduce variability. In fact, inter-laboratory tests were conducted by three laboratories for a total of seven experiments that met quality-control standards. These tests revealed low variability and a high level of consistency in results within and across the laboratories (Taub 1997b). Similar repeatability was found for larger-scale outdoor cosm experiments (Van den Brink 2006). Additionally, Taub and colleagues provide standardized protocols to simulate natural levels of immigration to each replicate to capture dispersal-mediated recovery and a wealth of analytical tools, such as rapid statistical analyses, a mathematical model and artificial intelligence methodologies (Taub 1997b).

Unknowns

One of the criticisms of community-level mesocosm studies has been that they have high variability and low repeatability (Nabholz et al. 1997; Touart & Maciorowski 1997; Van den Brink 2013). However, we know of no systematic quantification of the variability and repeatability of community-level toxicological studies (Table 2) with the exception of the intra- and inter-laboratory tests associated with Taub and colleagues' standardized microcosms (Taub 1997b). For that matter, we know of no systematic quantification of the variability and repeatability of toxicological studies at any level of biological organization, but we assume that they are higher for studies at lower levels of organization. However, most published mesocosm and microcosm studies only have 3–6 replicates (Taub 1997b; Skelly & Kiesecker 2001; Relyea 2005; McMahon et al. 2012; Halstead et al. 2014). To detect effects of treatments with only three to six replicates, variability among replicates must be low or effect sizes must be very large. We doubt that effect sizes in toxicology are universally large, which suggests that variability in mesocosm studies might not be as large as previously assumed. We encourage a quantitative comparison of variance estimates of traditional LC50 studies and community-level micro- and mesocosm studies to better evaluate the validity of this criticism.

Another criticism of mesocosm studies is that they are more costly per study than standard laboratory toxicity tests (Nabholz et al. 1997; Touart & Maciorowski 1997; Van den Brink 2013). While it is true that an outdoor mesocosm study is almost certainly more expensive to conduct than an acute or chronic toxicity test (or suborganismal studies), mesocosm studies can often provide estimates of risk to many species (assuming that they present species-specific information rather than just total biomass, species richness and evenness, or ecosystem metrics) and do so under conditions that capture important ecological interactions that can never be addressed by studies at lower-levels of biological organization, such as catastrophic regime shifts that have occurred with exposure to some contaminants (Scheffer & Carpenter 2003; Scheffer et al. 2009). To more defensibly compare the cost effectiveness of mesocosm studies to studies at other levels of biological organization, it might be worthwhile to standardize the costs on a per species basis and to carefully consider the quality of the data that each study offers to risk assessors (i.e., the benefits; including ecological realism, diversity of taxa, and the value of the species tested). As an example, it is possible that a single mesocosm study offering dose–responses on 10 naturally interacting species could be cheaper than the sum of the costs of 10 sub-individual-, individual-, or population-level studies on each of the same species in isolation, but this cost has not yet been quantified. Even if mesocosm studies are not cheaper on a per study or per species basis, by capturing greater ecological realism and studying more species on average than studies at lower levels of biological organization, they might reduce uncertainty in ERA enough to justify any higher cost.

Another critique is that the results of mesocosm experiments can be hard to interpret (Van den Brink 2006). However, an opinion of a European Food Safety Authority Panel (EFSA 2006) showed that, when evaluated by experts, mesocosm experiments yield interpretable results and can provide unambiguous answers.

Cons

Perhaps the biggest justified criticism of community-level studies is that they do not offer an efficient approach for screening large numbers of chemicals (Table 2). Capturing population- and community-level processes takes time, which is why mesocosm studies often last four weeks or longer (Taub 1997b; Relyea 2005; McMahon et al. 2012; Halstead et al. 2014). Additionally, mesocosm studies require considerable effort on the part of the researchers to collect or culture all the species for the experiment. For these reasons, community-level studies certainly do not offer a HTS option for the thousands of chemicals that need to be assessed. However, uncertainty in the ecological relevance of a particular level of biological organization to ERA and whether a level of biological organization offers efficient high-throughput-screening options generally seem to be inversely correlated, such that HTS methods have high uncertainty and approaches that do not lend themselves well to HTS have lower uncertainty. Hence, risk assessors will likely have to tradeoff the level of certainty with the number of chemicals that are assessed.

Standardization of methods and thus repeatability of results across laboratories is another concern for community-level approaches. This is because the effects of contaminants can depend on species composition (Rohr & Crumrine 2005), which varies widely across space and time. For the standardized microcosms described above, all the organisms can be purchased but nutritional status and sources may still cause variation in sensitivity. For non-standardized community-level experiments, there are not suppliers of entire natural communities and thus laboratories often use different source populations and communities, which can theoretically make it challenging to repeat results across laboratories. It is partly this lack of suppliers that makes standardized methods difficult, but efforts certainly could be made to standardize other aspects of mesocosm studies, such as standardizing how mesocosms are initially established, the presence of sediment, endpoints to measure, and how they are measured. In fact, the US EPA was moving forward in the 1980s and early 1990s with cross laboratory standardization of community-level ecotoxicology studies (Taub 1997b; Touart 1988) before the US

EPA eliminated tier IV testing and thus field and mesocosm studies.

The criticism regarding the lack of standardization in mesocosm studies and different communities producing different results, however, is becoming less problematic as we better understand the functional roles and guilds of species and the strengths and directions of species interactions. In fact, a recent study suggests that we can use food web theory and models to predict both the effects of individual contaminants and contaminant mixtures on communities by knowing something about the direct toxicities of the contaminant to the species and the directions and strengths of species interactions (Halstead et al. 2014). This, coupled with evidence that species traits (Baird & Van den Brink 2007; Van den Brink et al. 2013) and phylogenies of chemicals and species can be used to predict the toxicities of untested chemicals and species (Guenard et al. 2011; Hammond et al. 2012; Guenard et al. 2014) and the development of mathematical network models for community dynamics (USEPA 2000; Dunne et al. 2002; Ebenman & Jonsson 2005), offers hope that this criticism will be less of a hurdle to predicting effects of contaminants on communities in the future (Rohr, Kerby, et al. 2006). Additionally, many of these same critiques levied at mesocosm studies, such as the critique that the results will depend on the species and environmental conditions tested, can be applied to all levels of biological organization.

Ecosystem level

Overview

Ecosystem-level toxicological studies quantify the effect of a contaminant on an ecosystem process, function, or service. Ecosystem processes and functions are the result of complex interactions between biotic (living organisms) and abiotic (chemical and physical) components of ecosystems through the universal driving forces of matter and energy (de Groot et al. 2002; Munns et al. 2015). Finally, ecosystem services are strictly those ecosystem attributes that provide direct or indirect value to humans (de Groot et al. 2002; Munns et al. 2015). Common ecosystem properties quantified in ecosystem-level toxicological studies are net primary productivity, whole system metabolism, dissolved oxygen, pH, turbidity, rates of decomposition and nutrient cycling (e.g., carbon, phosphorus, nitrogen) (Taub 1997b; Relyea 2005; McMahon et al. 2012; Halstead et al. 2014). Because there is considerable evidence that assemblages of species or biodiversity in general are responsible for ecosystem functions and services (Cardinale et al. 2012), many, but not all, ecosystem-level toxicological studies also quantify biodiversity (Taub 1997b; Relyea 2005; McMahon et al. 2012; Halstead et al. 2014), but quantifying biodiversity is not necessary for classification as an ecosystem-level study.

Pros

There are several benefits of ecosystem-level studies (Table 2). Most importantly, if they quantify an ecosystem service, they may have a measurement endpoint that matches an assessment endpoint desired by the public, such as pollination or decontamination of polluted water. Hence, defending toxicological studies at the ecosystem level to the general public should be easier than for many other levels of biological organization. Importantly, the results of ecosystem-level studies inherently include physiological feedbacks, population dynamics and species interactions. Thus, they can detect regime shifts that are well documented for certain contaminants such as excessive nutrients in aquatic ecosystems (Scheffer & Carpenter 2003; Scheffer et al. 2009). Additionally, they can assess recovery of some ecosystem processes associated with organisms that can recover through reproduction or dispersal (e.g., flying insects, some zooplankton, some algae, some microbes). Given that most ecosystem studies tend to have multiple species (because species assemblages often drive ecosystem processes), they have the potential to screen chemicals on large number of species; however, this is not often done. Ecosystem studies also might not require vertebrate animals as there are standardized aquatic microcosms with four trophic levels and only primary producers and invertebrates where ecosystem properties and processes can be quantified (Taub 1997b), and soil ecosystems, with or without plants, have been used commonly in toxicological studies (Ge et al. 2011, 2014; Priester et al. 2012).

Ecosystem-level studies, on average, have a medium level of ability to assess cause–effect relationships. Although there are some experimental ponds where chemicals have been applied and effects quantified (Larsen et al. 1986; Hanazato 1998; Fairchild & Sappington 2002; Boone et al. 2004), generally it is challenging to get approvals to apply chemicals to entire natural ecosystems and thus most effects on natural ecosystems tend to be correlational with a relatively low level of confidence in assessing cause–effect relationships. Hence, many ecosystem-level studies are natural experiments or correlational because they are conducted in nature. However, field or semi-field studies are often required for pesticide registration in the EU (Hommen et al. 2010). In contrast, manipulative mesocosm studies where ecosystem variables are quantified offer a high level of confidence in assessing cause–effect relationships.

Unknowns

The same unknowns that apply to community-level studies generally apply to ecosystem-level studies (Table 2). Additionally, because functional redundancies (two or more species having the same function or service in an ecosystem) can be common in communities (Walker 1992; Fairchild et al. 1994; Carlisle & Clements 2005; Ramsey et al. 2005), effects of contaminants on ecosystem processes should theoretically be less variable than effects on communities or even populations, suggesting that this level of biological organization might offer higher repeatability and lower variability than community or population-level studies. However, this has yet to be thoroughly tested. Additionally, if regime shifts (which can be context dependent) do occur, they could result in considerable variability among studies or locations.

Cons

Ecosystem-level studies offer a very limited potential for screening large numbers of chemicals because of the time and effort necessary to quantify most ecosystem processes (Table 2). Also, most manipulative ecosystem-level studies will entail studying at least some microbes and perhaps even macroscopic invertebrates and thus most will require animal exposure to chemicals. Results of ecosystem-level studies can be sensitive to initial community composition (Rohr & Crumrine 2005), but new data suggest that functional redundancies in communities might reduce the likelihood of this context-dependency from occurring widely (Halstead et al. 2014).

Finally, ecosystem-level studies that do not quantify biodiversity in addition to ecosystem variables run the risk of missing contaminant-induced population declines and extirpations of non-TER and TER species. This is because one species in an ecosystem can be replaced by other species that provide similar ecosystem functions. Thus, the functions or services offered by an ecosystem can be unchanged by a contaminant despite the contaminant altering species composition (Walker 1992; Fairchild et al. 1994; Carlisle & Clements 2005; Ramsey et al. 2005). For this reason, we encourage ERA on both functional (i.e., ecosystem) and structural (i.e., community) endpoints.

Landscape level

Overview

Cairns encouraged the development of landscape ecotoxicology in 1993 (Cairns 1993; Cairns & Niederlehner 1996). Although there has been a moderate output of papers on the topic since then (Landis & Wiegers 1997; Wiegers et al. 1998; Landis 2002, 2003a, 2003b; Hayes & Landis 2004), recently there seems to have been a resurgence of interest (Beketov & Liess 2012; Focks et al. 2014; Schafer 2014; Wendt-Rasch et al. 2014). Landscape toxicological studies generally quantify differences in biotic or abiotic variables at sites within a landscape that differ in their levels of contamination. Consequently, most landscape studies are correlational, at best taking advantage of natural experiments. The exception would be mathematical modeling studies that attempt to assess the effects of a contaminant within a metapopulation (connected populations) or metacommunity (connected communities) context (Leibold et al. 2004; Topping et al. 2014, 2015, 2016). Nevertheless, landscape-level ERA is becoming more feasible as countries have begun national water quality monitoring programs (e.g., US Geological Survey's National Water-Quality Assessment Program; Stone et al. 2014).

Pros

The results of landscape-level studies inherently include physiological feedbacks, population dynamics, species interactions and dispersal (Table 2). Because they include these feedbacks, like community- and ecosystem-level studies, they are able to detect regime shifts that have been well documented for certain contaminants (Scheffer & Carpenter 2003; Scheffer et al. 2009). Additionally, if landscape studies have a temporal component, they can assess the landscape-dependent recovery of toxicological endpoints (Hunsaker et al. 1990; Suter 1990). The fact that all landscape studies are conducted in nature means that they are ecologically relevant rather than contrived. Additionally, because they often cover broad spatial scales, they can detect the effects of biotic and abiotic factors that often depend on spatial scale (Cohen et al. in press). Given that most landscape-level toxicological studies are correlational or entail mathematical modeling, landscape studies generally do not entail exposing animals to contaminants beyond that which is occurring in nature. Given that landscape studies can have multiple species, they have the potential to screen chemicals on large number of species; however, most focus solely on one or a few species. If landscape-level studies quantify biodiversity or ecosystem services, then their measurement endpoints would be close to preferable assessment endpoints.

Unknown

Landscape-level toxicological studies have many of the same unknowns as community level studies (Table 2). For instance, there are very little data on the variability, repeatability and the cost per study or per species of landscape studies. Despite the lack of data on costs, we suspect that they are high because of the challenging logistics and high costs of sampling at the landscape scale. However, once landscape models have been developed and validated, the cost could decrease (e.g., Topping et al. 2014, 2015, 2016).

Cons

The most serious limitation of landscape-level toxicological field studies is their correlational nature and thus their inability to confidently establish cause–effect relationships between contaminant exposure and the response variable (Table 2). Establishment of dose–response relationships and weight-of-evidence approaches can improve the strength of causality of landscape studies. If hundreds of chemicals are tested at each field site, then landscape studies could be useful for screening large numbers of chemicals. But given the ephemeral nature of some chemicals and the associated logistical challenges of covering large spatial scales, we do not propose that landscape field studies are an effective or efficient approach for screening large numbers of chemicals. In contrast, it might be possible to develop and validate landscape models across multiple chemicals and then use these validated models to predict effects on untested chemicals (see Topping et al. 2014, 2015, 2016), but this remains to be seen. Because of these logistical challenges, landscape studies often only focus on one to a few species and thus often have measurement endpoints that far from the most defensible assessment endpoints.

Models relating responses at different levels of organization

Each empirical approach that focuses on single levels of biological organization has cons and uncertainties, so data at any level will have added value if their interpretation can take

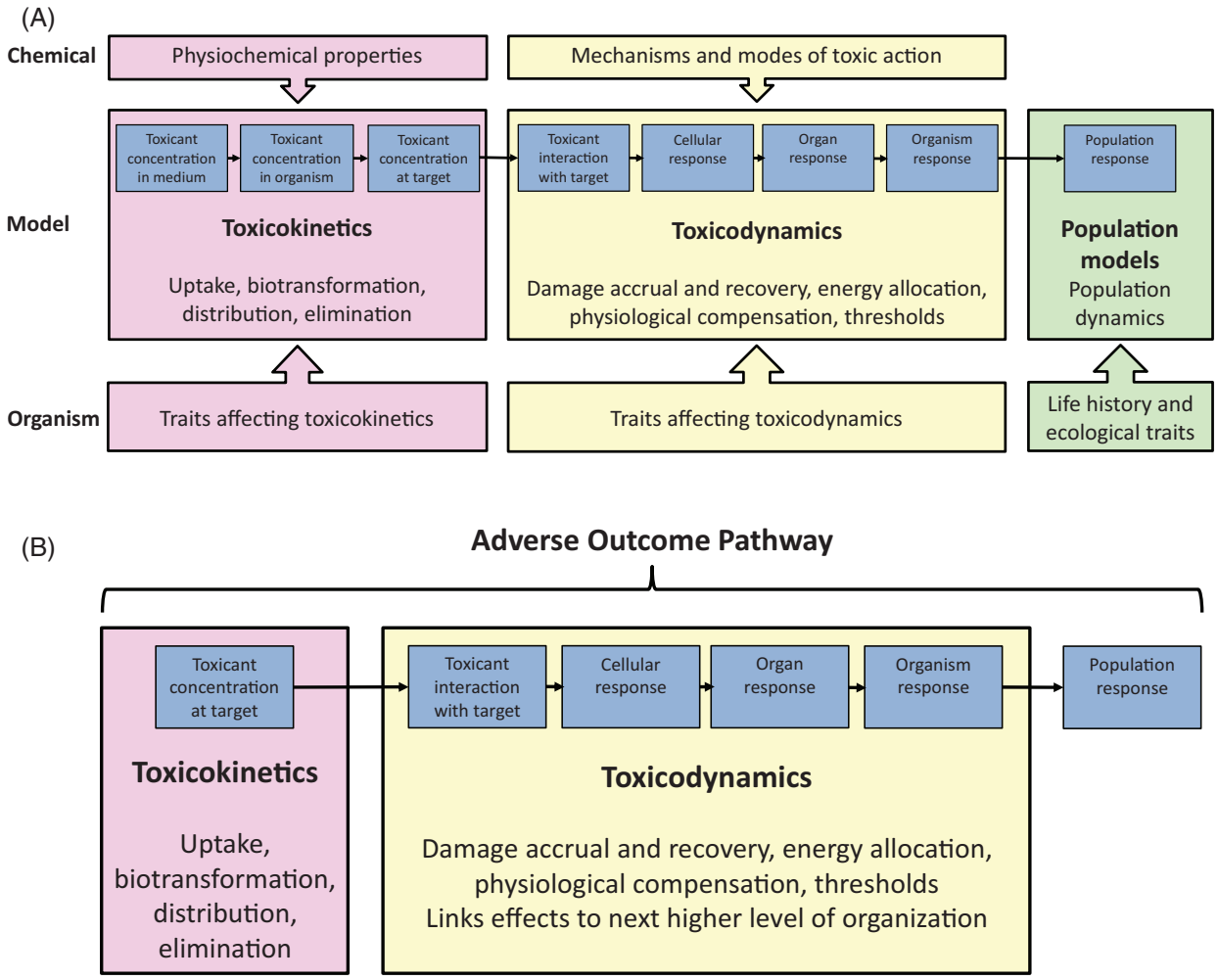

Figure 2. Relationships among toxicokinetic, toxicodynamic and population-level models (A). Components of these same models that make up adverse-outcome-pathway models (B). Modified with permission of Roman Ashauer.

account of information at other levels. In particular ERA requires that we relate biochemical, physiological and individual-level exposure endpoints to population, community and ecosystem implications. This requires mathematical models (Forbes & Calow 2013). In this section we give brief descriptions of some models used to make suborganismal-to-individual and individual-to-population connections. The current modeling state-of-the-art does not extend to predicting effects across multiple levels of organization, so we restrict our discussion largely to predicting up or down one level. We emphasize approaches that include the individual level as this is the natural starting point for ecological models. This is because population, community and ecosystem dynamics are simply the outcome of many individuals interacting with each other and with their environment, while the ecological impact of suborganismal (molecular, biochemical, physiological) responses to stress is expressed through changes at the individual level, and evolutionary change involves changes in the relative fitness of individual organisms. For additional information on mathematical models to inform ERA, we encourage readers to see Bartell et al. (2003), Pastorok et al. (2003), Hommen et al. (2010), Galic et al. (2010), Schmolke et al. (2010), Forbes et al. (2011), and EFSA (2014).

One immediate challenge in ecotoxicological modeling is that many standard protocols yield metrics such as LC_x or EC_x that depend strongly on the prescribed experimental conditions and duration (Baas et al. 2010), whereas ecologically relevant models require *biology-based* variables and parameters that are independent of experimental protocols (Jager et al. 2010). This implies that where possible we should use process-based dynamic models.

Toxicokinetic–toxicodynamic (TK–TD) models

Organismal response to contaminant exposure is commonly determined by *internal* concentrations. Toxicokinetic (TK) models describe the time course of accumulation, transformation and distribution of chemical contaminants within individuals (Figure 2(A)). Toxicodynamic (TD) models describe the subsequent response of an organ or whole organism (Figure 2(A)). Thus, TK–TD combinations underpin all models of ecotoxicity (Ashauer & Escher 2010; Ashauer et al. 2011; Jager 2015), for example, the generalized unified threshold model of survival (Jager et al. 2011) and the "DEBtox" models discussed below. *All* process-based models linking levels of organization require either an explicit TK sub-model or implicit assumption of a TK sub-model (such as equilibrium between internal and external concentrations) (Figure 2(A)). Thus, TK models have no *general* "cons", though of course *specific* TK models for particular

systems require validation in the contexts where they will be applied. For example, one-compartment TD models are sufficient for many situations, but inadequate in others.

TD models come at many levels of complexity. The next two subsections describe contrasting approaches: AOP-based models that describe the kinetics of impacted pathways and bioenergetic models that characterize the response of whole organisms to exposure using kinetic equations for a small number of more abstract variables that focus on physiological functions (Figure 2(B)).

Models of adverse outcome pathways (AOP)

The AOP paradigm (Ankley et al. 2010; Kramer et al. 2011), described earlier, involves tracing causal connections from one or more MIEs to ecologically relevant endpoints. Thus, in principle, modeling AOPs involves characterizing processes at many suborganismal levels of organization (gene expression, molecular interactions, cellular responses, organ responses, organism responses) – an intimidating challenge (Sturla et al. 2014). Also intimidating is that they require estimation of a very large number of parameters for every species studied, The state-of-the-art is changing rapidly, but it appears that for the most part, large statistical models (which may or may not involve dynamic equations) represent the most advanced approach for *inferring* a likely AOP from data (e.g., Antczak et al. 2015). Most dynamic models *assume* a known AOP and predict organismal endpoints of ecological importance, for example, those linking molecular mechanisms determining levels of steroid hormones or vitellogenin with well-understood connections to reproduction (Murphy et al. 2005; Watanabe et al. 2009; Li et al. 2011; Sundling et al. 2014).

Some detailed mechanistic AOP models consider biochemical processes within specific interacting organs. For example Watanabe et al. (2009) developed a systems model of hypothalamic-pituitary-gonadal axis in male fathead minnows with 123 parameters and 40 differential equations describing processes in brain, gonad, blood and liver. By contrast, an earlier model by Murphy et al. (2005) did not consider the organ level and described vitellogenin production using a system of eight differential equations with three further equations relating the model's state variables to target endpoints.

At all levels of model complexity, the primary challenge is parameter estimation. A common approach, also used with the bioenergetics models discussed below, is to first estimate the values of as many parameters as possible using published information, and to keep the values of these parameters fixed while "calibrating" the remaining model to target data sets by adjusting values of the remaining parameters. For example, the model of Watanabe et al. (2009) had 97 "fixed" parameters and 26 calibration parameters.

In short, the primary "pro" of AOP models is their clear connection to known biochemistry and physiology; the offsetting "con" is parameter richness.

Bioenergetic models

Many toxicants cause a reduction in the rates at which organisms feed or assimilate energy from food and an increase in respiration rate. This leads to a reduction in the energy available to support growth and reproduction and many toxicological studies have used this as an endpoint (e.g., Widdows et al. 1995). This in turn has motivated attempts to characterize sublethal effects of contaminants using bioenergetic models. By far, the most sophisticated implementation of this approach is based on Dynamic Energy Budget (DEB) theory[1] (Kooijman 2000, 2001, 2010; Nisbet et al. 2000) (but see Sibly et al. 2013 for an overview of other energy budget theories). DEB models assume that the combined dynamical properties of the large number of interacting biochemical networks in an organism can be described using a *small* number of variables describing *aggregates of compounds* such as "structural biomass", "reserve", or "reproductive material". The model equations describe ontogenetic growth, reproduction and survival in arbitrary, variable and potentially stressed environments (Figure 3(A)).

Application of DEB theory to ecotoxicology, sometimes called the "DEBtox" approach (Kooijman & Bedaux 1996; OECD 2006; Muller et al. 2009; Jager & Zimmer 2012), requires that a DEB model is coupled to TK and TD sub-models. Jager (2015) offers a very readable, almost "math-free" exposition of the concepts, with the practicalities covered in supplementary materials. Typically, the TD model assumes that body burden (determined by an explicit or implicit TK model) impacts one or more of the energy flows in the DEB model (different colored circles in Figure 3(A)). The DEBtox approach has been applied to both laboratory and field data from a broad range of target organisms [bacteria (Klanjscek et al. 2012) to whales (Klanjscek et al. 2007)] and toxicants, including engineered nanomaterials (Klanjscek et al. 2012, 2013; Holden et al. 2013; Muller et al. 2010, 2014).

There is a large body of literature on methods for estimating DEB model parameters, including routine multivariate, nonlinear regression (or analogous likelihood) methods (Kooijman et al. 2008), a computer-intensive state-space method (Fujiwara et al. 2005), and a recent Bayesian approach (Johnson et al. 2013). Lika et al. (2011) proposed a particularly innovative, heuristic "pseudo-Bayesian" approach, based on predicted scaling relations among DEB parameters [Chapter 8 of (Kooijman 2010)] that provides a route to a "first cut" at parameter values for a new species; code and a data compilation are curated by a team of five users of DEB models and accessible through a portal at the Vrije Universiteit, Amsterdam – http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/index.html. At the time of writing, that web site contained parameter estimates for over 300 species.

A common criticism of Kooijman's DEB models is that they are parameter-rich and hence "inefficient" (e.g., Marquet et al. 2014). The counter-argument recognizes that parameter count must take into account the number of processes explained. Thus, Kearney et al. (2015) noted that a DEB model of an organism's development, feeding, growth, maintenance, metabolic heating, reproduction and senescence under any sequence of environmental fluctuations in food and temperature requires approximately 1.5 parameters per process modeled. Furthermore, in particular applications, the model of growth collapses to a two-parameter (von Bertalanffy) form, but the theory offers unambiguous connections between

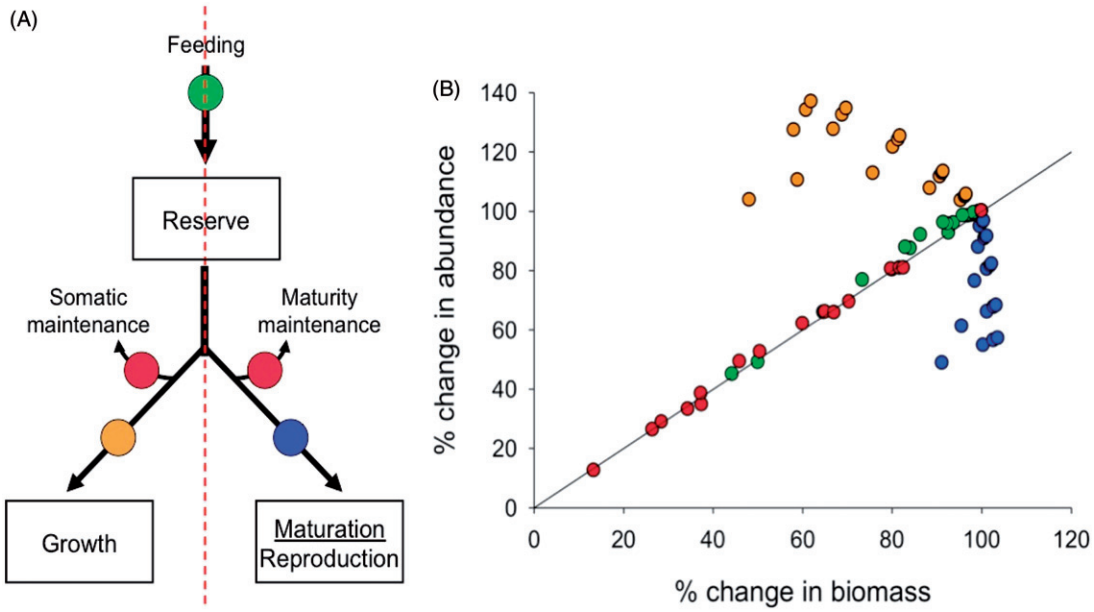

Figure 3. Population-level consequences of different physiological models of action (PMoA) as predicted by DEB theory. (A) The energy flows assumed by Kooijman's (2010) DEB theory. The red dashed line differentiates PMoAs that alter energy allocation to growth and reproduction symmetrically (Feeding, green; Maintenance Costs, red), and those that act asymmetrically (Growth Costs, orange; Embryonic Hazard, blue). (B) The relationship between changes in population abundance and biomass compared to control populations. Each point corresponds to the mean value of abundance and biomass with different levels of exposure to a hypothetical toxicant. The colors correspond with the PMoAs indicated in the left panel. Adapted from Martin et al. (2014). (This figure is presented in colour in the online version of the article.)

hypothesized physiological mode of toxicant action and the two parameters (e.g., Muller, Nisbet, et al. 2010; Muller, Osenberg, et al. 2010; Muller et al. 2014). Where such extreme simplification is not possible, a remarkably powerful simplified version of Kooijman's model with fewer parameters, known as "DEBkiss" is available (Jager et al. 2013).

Kooijman's DEB model is remarkably successful in simultaneously fitting diverse data sets within a coherent description of metabolic organization (Kooijman 2010). However, for detailed quantitative predictions for any particular organism, model modifications are often necessary; for example juvenile zebrafish (*Danio rerio*) develop faster than the standard model predicts (Augustine et al. 2011), and *Daphnia* spp. do not follow the precise rules for portioning energy to growth versus reproduction (Nisbet et al. 2010; Ananathasubramaniam et al. 2015). With these organism-specific bioenergetics models, it is possible to follow the strategy used in AOP models of working with a mix of "fixed" and "calibration" parameters [e.g., most calculations in the *Daphnia magna* meta-analysis of Ananathasubramaniam et al. (2015) had 22 fixed parameters, and 1 calibration parameter].

In summary, the pros and cons of DEB (and other bioenergetic) models are the obverse of those for AOP models: weak connection with known biochemistry offset by strength in ability to predict organismal responses due to parameter sparseness.

Population models deriving from individual physiology and life histories

Given quantitative information on organism's life history in different environments, together with a biologically-based representation of toxic effects, there are a number of mechanistic approaches to predicting population dynamic consequences of exposure to toxicants. Common to these approaches is a requirement that population response is an "emergent property" that can be derived from information or assumptions at the individual level (Grimm & Martin 2013). Conceptually, the simplest of these can be predicted using so called individual-based population models (IBMs) – computer simulations where the "state" (e.g., age, size, sex, body burden of toxicant) of each individual organism is tracked, along with concurrent changes in its environment (Grimm & Railsback 2005; Grimm & Railsback 2012). For example, if individual performance in response to toxicant exposure is described by a bioenergetic model, the IBM describes the growth, reproduction and mortality risk of individual organisms with a shared environment. Recently developed software, DEB-IBM (Martin et al. 2012), implemented in a free software platform NETLOGO (http://ccl.northwestern.edu/netlogo/), has greatly eased the practicalities of working with IBMs in both basic ecology (Martin et al. 2012) and ecotoxicology (Martin et al. 2013). A recent study on *Daphnia* using DEB-IBM demonstrated the importance of TD mechanisms for predicting population dynamics using data on individual responses (Martin et al. 2014); toxicity mediated by different physiological processes that led to the *same* outcome in a standard reproduction test may cause qualitatively different effects at the population level, ranging from almost no effect to extinction. Figure 3(B) shows that with increasing exposure to toxicant, different physiological modes of action can lead to (i) proportional decreases in population density and biomass, (ii) decreasing population density with little change in biomass, (iii) decreasing biomass with little change in population density.

IBMs can also describe the population-level implications of toxicity that changes organismal behavior. For example, Murphy et al. (2008) analyzed laboratory data on the response

of individual larval fish to artificial predator stimuli when exposed to different levels of methylmercury, and then used an IBM to predict the impact of exposure on the dynamics of a larval cohort.

IBMs have been criticized as being too parameter-rich to have predictive value, a complaint not unlike that discussed above for DEB models. The resolution of the problem is similar – to use a *single model* to predict *multiple patterns* in data. In the context of IBMs this approach to model formulation and parameterization is known as "pattern oriented modeling" (POM) (Grimm & Railsback 2005, 2012). Bayesian approaches, notably Approximate Bayesian computation (van der Vaart et al. 2015, 2016) offer additional routes to parameter estimation, again paralleling the state-of-the-art for DEB modeling.

An alternative individual-based approach to modeling effects of contaminants on populations uses "physiologically structured population models" (PSPMs) (Tuljapurkar & Caswell 2012). PSPMs assume a very large (formally infinite) population and that all individuals in a given state have deterministic responses to any given environment. These assumptions allow the bookkeeping to proceed through a series of mathematical steps that lead to partial differential or integral equations describing the population dynamics, with good software available that facilitates analysis (de Roos 2014).

If it is possible to neglect feedbacks from a focal population on its environments, PSPMs open the way to relating individual physiological responses to log-run population growth rate (r) or expected lifetime reproduction (R_0), metrics normally associated with life history theory (Kooijman & Metz 1984; Sibly & Calow 1989; Baird et al. 1990; Nisbet et al. 2000; Muller, Nisbet, et al. 2010; Muller et al. 2014). Although calculations of r and R_0 neglect the many positive and negative feedbacks whose importance we emphasized earlier in this article, they may still give useful qualitative insights, and often represent the only available approach with limited data. For example, Muller et al. (2014) estimated an EC50 for the expected lifetime production of reproductive matter in marine mussels exposed to ZnO nanoparticles that is less than 20% of the EC50 for feeding.

Another approach for predicting population responses from life history information uses a cohort-based approach expressed through matrix population models (Caswell 2001). Matrix models relate the population census in different ages, stages or size classes at successive, *discrete*, time points. Applications in ecotoxicology require careful explicit or implicit consideration of the *continuous time* TK/TD dynamics within a time step. For examples, see Klaniscek et al. (2006) or Billoir et al. (2007). In many cases, a superior, but more mathematically challenging representation of cohort dynamics is the "escalator boxcar train" approximation to continuous time PSPMs (De Roos et al. 1992).

Information on long-run population growth rates has also been aggregated in SSDs. SSD's typically use some standard metric from toxicity tests on individual organisms (e.g., LC50, NOEC), but Kamo & Naito (2008) proposed constructing SSDs using the concentrations that lead to predicted zero net population growth rate ($r = 0$) with an application to zinc toxicity, an approach subsequently applied to copper toxicity for 13 freshwater species (Kamo & Naito 2008).

In summary, the two most powerful tools for projecting from the individual-to-population levels are IBMs and PSPMs. User-friendly software is available to help practical implementation. Their greatest apparent "con" is demanding data requirements, but recent advancements in methodology, especially POM, offer a way forward, with further advances using Bayesian methodology likely. Life history metrics, such as long run population growth rate, can be estimated from information on individuals but neglect potentially important ecological feedbacks. Discrete time matrix models can be derived from life history information, but application to ecotoxicology requires careful consideration of the continuous time processes that occur within time steps.

Multi-species models

Several multi-species models have been developed to predict community and ecosystem responses to chemical contaminants to inform ERA. Of these models, AQUATOX is possibly the most comprehensive (Park et al. 2008) and well validated (USEPA 2000; De Laender et al. 2008b; Sourisseau et al. 2008), and is endorsed by a chemical regulatory agency (the USEPA). Hence, for brevity, we focus our discussion of multi-species models on AQUATOX and refer the reader to the literature for information on additional multi-species models (e.g., Bartell et al. 2003; Park et al. 2008; Preziosi & Pastorok 2008; Galic et al. 2010).

AQUATOX combines aquatic ecosystem, chemical fate and ecotoxicological submodels to evaluate past, present and future fate of nutrients, sediments and organic chemicals in water bodies, as well as their direct and indirect effects on the resident organisms, including members of periphyton, phytoplankton, macrophyte, invertebrate and fish communities (Park et al. 2008) (Figure 4). AQUATOX simulates the transfer of biomass and chemicals from one compartment of the ecosystem to another and can evaluate chemical effects for a variety of aquatic ecosystems, including vertically stratified lakes, reservoirs and ponds, rivers and streams, and estuaries. The chemical fate submodels of AQUATOX include chemodynamics of neutral and ionized organic chemicals, bioaccumulation as a function of sorption and bioenergetics, and biotransformation to daughter products (Figure 4). The ecological effects submodels include both sublethal and lethal effects of the chemicals, as well as interactions among species to capture indirect effects (Figure 4). Additionally, AQUATOX can model up to 20 organic chemicals simultaneously. The model has a very flexible structure and provides analytical tools useful for uncertainty analysis, nominal range sensitivity analysis and comparison of perturbed and control simulations (Park et al. 2008).

Importantly, AQUATOX has been well validated, demonstrating that it can reliably predict the effects of organic contaminants on community and ecosystem properties of aquatic systems (USEPA 2000; De Laender et al. 2008b). For example, De Laender et al. (2008b) showed that AQUATOX population-level NOECs were at least protective for 85% of all considered populations in aquatic mesocosm studies. Additionally,

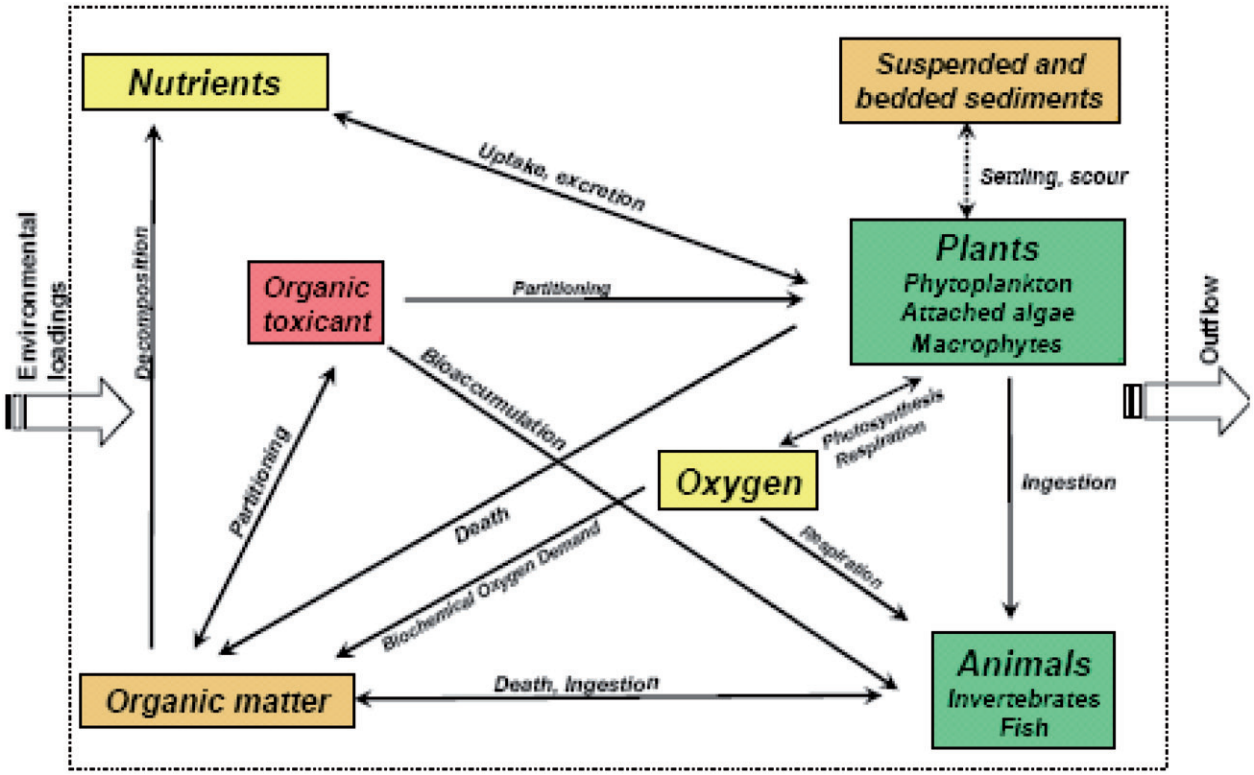

Figure 4. Schematic for how the AQUATOX model simulates ecological processes and effects in aquatic ecosystems over time. Reproduced with permission from USEPA and Dick Park.

AQUATOX generated an accurate or a moderately conservative ecosystem-level NOEC for seven and four of the 11 ecotoxicological mesocosm studies, respectively. Some admonishments, however, are that AQUATOX relies heavily on the use of NOECs and thus comes with the critiques mentioned above and in previous papers about these statistics. Additionally, it remains unclear how well AQUATOX performs in predicting effects at sites that deviate substantially from those used to parameterize the model. Nevertheless, because of the successes of AQUATOX and other community- and ecosystem-level ERA, they are being adopted in several countries in the EU, such as the Netherlands (Van den Brink et al. 2002; Traas et al. 2004).

In summary, multi-species models such as AQUATOX have proven useful in predicting effects of chemicals in ecological systems (Naito et al. 2002; De Laender et al. 2008b). Among the important pros of multi-species models is that they relate directly to levels of biological organization of high relevance for protection. Also, the models directly address and incorporate interactions among species. Because several multi-species models and guidance documents are publicly available, they do not require significant resources to acquire and use. Despite likely utility in ERA, there are still limitations to multi-species models. For example, modeled scenarios should closely match the system that is the focus of assessment or protection, necessitating sufficient knowledge of the system of interest (Naito et al. 2002). In some cases, models are limited to particular classes of chemicals; AQUATOX, for example, is limited to organic chemicals (Park et al. 2008). Additionally, when considering multiple, simultaneous chemical exposures, the default assumption is that they interact additively, which is not necessarily the case for many chemical combinations (PapeLindstrom & Lydy 1997) Also, the underlying concentration response functions used in toxicity submodels can have an important effects on model outcomes (De Laender et al. 2008b), pointing to the importance of high quality data for model input. Nevertheless, multi-species models provide useful and relevant information for ERA and can constitute and important part of a weight-of-evidence approach to risk estimation and characterization.

Where to go from here? Next generation ERA

Chemical risk assessors need to evaluate the safety of thousands of chemicals and to defend their assessments with repeatable assays that have low uncertainty and yet capture endpoints valued by the public. This dual requirement is challenging because the level of biological organization at which ERA studies are most easily conducted is lower than, and potentially disconnected from, the endpoints valued by the public and that relate more directly to ecosystem-level effects and services (Table 2, Figure 1). To address these issues, we propose that the next generation of ERAs include, from the start, both ends of the spectrum of levels of organization. We believe that if there is an ideal level of biological organization to conduct ERA, it will only emerge from a combination of ERAs that are simultaneously implemented from the bottom of biological organization up as well as from the top down. The details of such an approach require considerable further research; however the work reviewed above points to practical, cost-effective first steps, and to directions for further improvement.

We envision a tiered approach to ERA but one in which the screening level includes *both* low-level (molecular and organismal) endpoints and high-level (community and ecosystem) endpoints. Thus, we encourage next-generation ERA to include, *as part of the first assessment tier,* simple mesocosm tests. Consequently, determining optimal protocols for such tests should be a top research priority. For tier 1 studies, we propose exploring the use of mesocosms to first test multiple chemicals simultaneously, each at a *single and defensible* "worst case field concentration" (such as the highest concentration measured in nature or an estimated environmental concentration based on fate and transport models). We argue that because mesocosms simulate ecologically realistic scenarios and often are longer-term studies (typically last for weeks or months) than traditional 96-h LC50 tests, organisms are often exposed to a range of toxicant concentrations because of degradation and sorption processes; in fact, in many cases, they will be exposed all concentrations below the initial exposure concentration allowing for the detection of adverse effects at low concentrations that are not observed at higher concentrations (i.e., a non-monotonic dose response; e.g. McMahon et al. 2011, 2013). If no adverse effects are observed across the range of included taxa in these mesocosm tests, additional testing and establishment of dose–response relationships may not be warranted, thus increasing the efficiency of ERA and allowing for more chemicals to be screened. Many of the cost-related criticisms of mesocosm studies described above might be less applicable to such a simplified tier 1 test. If there are adverse effects at a worst case field concentration, we encourage concentration– or dose–response studies on the affected taxa to determine "safe" concentrations. The EU regularly employs mesocosm studies in their ERA and we encourage more widespread use of their guidelines (e.g., EFSA 2013), with the exception that we would not require establishing a mesocosm dose–response in the first tier. A significant advantage of the proposed approach would be the ability to screen many species, the inclusion of endpoints more closely related to biodiversity and ecosystem services quantified from the start of ERA in a manner that includes many natural feedbacks, and an improved ability for community- and ecosystem-level tests to screen chemicals.

For chemicals, or combinations of chemicals, that are identified by HTS measurements, organismal screening, or mesocosms as candidates for more thorough study, we propose two parallel approaches, still working from both ends of biological organization. First, suborganismal- and organismal-level studies need design advances that recognize the accelerating progress in the development of mechanistic effects models for ecotoxicology and risk assessment, as described earlier in this paper. A key feature of such models is the need to obtain some time-resolved data as these data open the way to fitting and testing models. For example, HTS measurements of cellular endpoints should be taken at a few time points rather than one. Likewise, standardized organism-level chronic toxicity tests (e.g., *Daphnia* reproduction tests) should routinely report time-resolved data and a limited number of auxiliary measurements (e.g., final size) wherever experimentally and economically practical (Martin et al. 2014). Modeling approaches are advancing rapidly, and with the more widespread use and apparent utility of bioenergetic models (especially DEB), there exists an opportunity to link effects across levels of organization using combinations of data and modeling platforms. We anticipate the emergence of a modeling framework that successfully links data obtained from "omics", AOP analyses, organismal-level effects, through population or multi-species models, and finally to the delivery of ecosystem services. While such a framework represents a lofty goal, many of the modeling components exist and it is a matter of ensuring fluid communication among modeling platforms and levels of biological organization. The extent to which the emerging framework will be "predictive" remains to be seen but we anticipate, at the very least, a clearer identification of data needs to generate defensible risk estimates for higher levels of organization.

Second, for chemicals that cause an adverse ecological effect as identified in the first tier, we advocate additional testing using mesocosm designs that specifically include species interactions and/or environmental factors hypothesized to be important for the eco-toxicity of a specific chemical (or mixture). Developing higher-tier mesocosm designs that identify and elucidate important context-dependencies should be a research priority. The aim of these higher-tier studies would be twofold. First, they can assist in more clearly identifying "other" factors, stressors or species that increase the vulnerability of certain systems to the effects of a particular toxicant. We realize that addressing the myriad of context-dependencies and species is daunting, but we believe that QSARs and SSDs can provide some direction to reduce this complexity. With the assistance of QSARs, risk assessors should be able to identify the abiotic factors already known to affect the toxicity of chemicals with similar structures as the chemical of interest. Likewise, they could use SSDs to identify taxa that are particularly sensitive to chemicals with similar structure. These efforts could serve to focus these higher-tier tests so that they can fine tune risk estimates and management. Second, these studies could include endpoints or components that more directly inform how particular chemicals impact the delivery of specific ecosystem services.

Conclusions

Here we discuss current challenges of ERA of manufactured chemicals and review the pros, cons and unknowns of ERA across levels of biological organization, from sub-individual to landscapes levels. Our review suggests that there is a positive relationship between level of biological organization and ease at screening large numbers of species, sensitivity to important negative and positive feedbacks and context dependencies within biological systems, and ease at capturing recovery from contaminant exposure. In contrast, there was generally a negative relationship between level of biological organization and ease at assessing cause–effect relationships, ease of HTS of large numbers of chemicals (greater for suborganismal endpoints), and uncertainty of ERA because low levels of biological organization tend to have a large distance between their measurement and assessment endpoints. The need for vertebrate animals in chemical testing did not show an obvious trend across levels of biological organization. Although it

is generally assumed that level of biological organization is associated negatively with the repeatability of chemical effects assessment and positively with the variability and cost of chemical effects assessment, we found no quantitative analysis in the literature to support these assumptions, representing an important knowledge gap for ERA. Additionally, we also noted that the assumption that the cost of chemical effects assessment increases with level of biological organization is based on the cost per study rather than the cost per species or some other metric that incorporates both the quantity and/or quality of information per dollar spent. Given that community-level studies often simultaneously gather chemical effects information on tens to hundreds of species, whereas suborganismal- and individual-level studies generally focus on a single species, the cost per species of ERA across levels of biological organization might show a very different pattern than the cost per study. Hence, this represents another important knowledge gap for ERA.

To compensate for weaknesses of ERA at any given level of biological organization, we also reviewed mathematical modeling approaches to extrapolate effects across levels of organization. We highlight toxicokinetic–toxicodynamic (TK–TD) models, models of AOP, bioenergetic models, individual-based population models and multi-species models. Models of AOPs have a clear connection to known biochemistry and physiology but this is offset by the "con" that they are often parameter rich and data "hungry". We noted that the primary "pros" and "cons" of bioenergetics models are opposite models of AOPs. Individual-based models represent a powerful tool for projecting from the individual-to-population levels and user-friendly software is available to help with their practical implementation. Multi-species models are becoming more accessible, do a reasonable job predicting outcomes of toxicity, and capture important biological interactions. The "cons" of multi-species models are that outputs can be sensitive to underlying submodels, they are relatively data intensive, and they require adequate knowledge of focal ecosystems.

Finally, we provide recommendations for next generation ERA. We advocate approaching ERA simultaneously from the bottom of biological organization up as well as from the top down because both have clear advantages, all while employing mathematical modeling approaches where possible to enhance ERA. Moreover, we suggest that by doing so, if there is an ideal level of biological organization to conduct ERA, it will emerge. Because top-down ERA is not customary, we offer guidelines for how it might be implemented efficaciously. By reviewing the pros and cons of ERA across levels of biological organization and the mathematical modeling approaches to extrapolate chemical effects across levels of organization, we hope we have identified key information gaps to conducting an informed ERA and have provided risk assessors with a road map to identify the best levels of biological organization to conduct ERAs with differing goals.

Endnote

1. The description "Dynamic Energy Budget model" is commonly used in the literature to refer to models based on Kooijman's theory, though other authors (e.g. Lika and Nisbet 2000, Nisbet et al. 2004) use the term to characterize any dynamic model of energy budgets. In this paper, we use the narrow definition and describe other models as "bioenergetics models".

Acknowledgements

The authors appreciated the constructive comments offered by seven reviewers who were selected by the Editor, six of which were anonymous to the authors and one was self-identified. Consideration of the comments helped improve the paper.

Declaration of interest

This research was supported by grants from the National Science Foundation (EF-1241889), National Institutes of Health (R01GM109499, R01TW010286), US Department of Agriculture (NRI 2006-01370, 2009-35102-0543) and US Environmental Protection Agency (CAREER 83518801) to J.R.R. RMN had support from the US National Science Foundation and the US Environmental Protection Agency under Cooperative Agreement Number EF-0830117 and from the US Environmental Protection Agency under grant 835797. CJS was supported by the U.S. Air Force Civil Engineer Center (contract number FA8903-12-C-008), the US Environmental Protection Agency (grant number 83580002) and the Strategic Environmental Research and Development Program (grant number 16 ER02-014/ER-2627). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the provided funding agencies. The affiliation of the authors is as shown on the cover page and includes universities only. CJS was a former employee of the USEPA.

References

Ahlers J, Stock F, Werschkun B. 2008. Integrated testing and intelligent assessment – new challenges under REACH. Environ Sci Pollut Res Int. 15:565–572.

Ananathasubramaniam B, McCauley E, Gust K, Kennedy AJ, Muller EB, Perkins EJ, Nisbet RM. 2015. Relating suborganismal processes to ecotoxicological and population level endpoints using a bioenergetic model. Ecol Appl. 25:1691–1710.

Anderson RM, May R. 1991. Infectious diseases of humans: dynamics and control. New York: Oxford University Press.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem. 29:730–741.

Ankley GT, Daston GP, Degitz SJ, Denslow ND, Hoke RA, Kennedy SW, Miracle AL, Perkins EJ, Snape J, Tillitt DE, et al. 2006. Toxicogenomics in regulatory ecotoxicology. Environ Sci Technol. 40:4055–4065.

Ankley GT, Gray LE. 2013. Cross-species conservation of endocrine pathways: a critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. Environ Toxicol Chem. 32:1084–1087.

Antczak P, White TA, Giri A, Michelangeli F, Viant MR, Cronin MTD, Vulpe C, Falciani F. 2015. Systems biology approach reveals a calcium-dependent mechanism for basal toxicity in *Daphnia magna*. Environ Sci Technol. 49:11132–11140.

Ashauer R, Agatz A, Albert C, Ducrot V, Galic N, Hendriks J, Jager T, Kretschmann A, O'Connor I, Rubach MN, et al. 2011. Toxicokinetic–toxicodynamic modeling of quantal and graded sublethal endpoints: a brief discussion of concepts. Environ Toxicol Chem. 30:2519–2524.

Ashauer R, Escher BI. 2010. Advantages of toxicokinetic and toxicodynamic modelling in aquatic ecotoxicology and risk assessment. J Environ Monit. 12:2056–2061.

ASTM. 2011. ASTM standard practice for standardized aquatic microcosm: fresh water. In: Annual book of ASTM standards. Philadelphia (PA): American Society for Testing Materials. p. 1048–1082.

Augusiak J, Van den Brink PJ, Grimm V. 2014. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. Ecol Model. 280:117–128.

Augustine S, Gagnaire B, Floriani M, Adam-Guillermin C, Kooijman SALM. 2011. Developmental energetics of zebrafish, *Danio rerio*. Comp Biochem Physiol A Mol Integr Physiol. 159:275–283.

Baas J, Jager T, Kooijman B. 2010. Understanding toxicity as processes in time. Sci Total Environ. 408:3735–3739.

Baas J, Kooijman SALM. 2015. Sensitivity of animals to chemical compounds links to metabolic rate. Ecotoxicology. 24:657–663.

Baird DJ, Barber I, Calow P. 1990. Clonal variation in general responses of Daphnia-Magna straus to toxic stress.1. Chronic life-history effects. Funct Ecol. 4:399–407.

Baird DJ, Van den Brink PJ. 2007. Using biological traits to predict species sensitivity to toxic substances. Ecotoxicol Environ Safe. 67:296–301.

Barron MG, Carls MG, Short JW, Rice SD. 2003. Photoenhanced toxicity of aqueous phase and chemically dispersed weathered Alaska North Slope crude oil to Pacific herring eggs and larvae. Environ Toxicol Chem. 22:650–660.

Bartell SM, Pastorok RA, Akcakaya HR, Regan H, Ferson S, Mackay C. 2003. Realism and relevance of ecological models used in chemical risk assessment. Hum Ecol Risk Assess. 9:907–938.

Beketov MA, Kefford BJ, Schafer RB, Liess M. 2013. Pesticides reduce regional biodiversity of stream invertebrates. Proc Natl Acad Sci USA. 110:11039–11043.

Beketov MA, Liess M. 2012. Ecotoxicology and macroecology – time for integration. Environ Pollut. 162:247–254.

Berninger JP, Martinovic-Weigelt D, Garcia-Reyero N, Escalon L, Perkins EJ, Ankley GT, Villeneuve DL. 2014. Using transcriptomic tools to evaluate biological effects across effluent gradients at a diverse set of study sites in Minnesota, USA. Environ Sci Technol. 48:2404–2412.

Billoir E, Pery ARR, Charles S. 2007. Integrating the lethal and sublethal effects of toxic compounds into the population dynamics of *Daphnia magna*: a combination of the DEBtox and matrix population models. Ecol Model. 203:204–214.

Boone MD, Bishop CA, Boswell LA, Brodman RD, Burger J, Davidson C, Gochfeld M, Hoverman JT, Neuman-Lee LA, Relyea RA, et al. 2014. Pesticide regulation amid the influence of industry. Bioscience. 64:917–922.

Boone MD, Rohr JR. 2015. The trouble with risk assessment lies at the foundation. Bioscience. 65:227–228.

Boone MD, Semlitsch RD, Fairchild JF, Rothermel BB. 2004. Effects of an insecticide on amphibians in large-scale experimental ponds. Ecol Appl. 14:685–691.

Boyle TP, Fairchild JF. 1997. The role of mesocosm studies in ecological risk analysis. Ecol Appl. 7:1099–1102.

Bradbury SP, Russom CL, Ankley GT, Schultz TW, Walker JD. 2003. Overview of data and conceptual approaches for derivation of quantitative structure–activity relationships for ecotoxicological effects of organic chemicals. Environ Toxicol Chem. 22:1789–1798.

Bulun SE, Zeitoun K, Takayama K, Noble L, Michael D, Simpson E, Johns A, Putman M, Sasano H. 1999. Estrogen production in endometriosis and use of aromatase inhibitors to treat endometriosis. Endocr-Relat Cancer. 6:293–301.

Burden N, Sewell F, Chapman K. 2015. Testing chemical safety: what is needed to ensure the widespread application of non-animal approaches? PLoS Biol. 13:e1002156.

Cairns J. 1986. The myth of the most sensitive species. Bioscience. 36:670–672.

Cairns J. Jr. 1983. Are single species toxicity tests alone adequate for estimating environmental hazards? Hydrobiologia. 100:47–57.

Cairns J. Jr. 1993. Will there ever be a field of landscape toxicology. Environ Toxicol Chem. 12:609–610.

Cairns J, Niederlehner BR. 1987. Problems associated with selecting the most sensitive species for toxicity testing. Hydrobiologia. 153:87–94.

Cairns J, Niederlehner BR. 1996. Developing a field of landscape ecotoxicology. Ecol Appl. 6:790–796.

Cajaraville MP, Bebianno MJ, Blasco J, Porte C, Sarasquete C, Viarengo A. 2000. The use of biomarkers to assess the impact of pollution in coastal environments of the Iberian Peninsula: a practical approach. Sci Total Environ. 247:295–311.

Cardinale BJ. 2011. Biodiversity improves water quality through niche partitioning. Nature. 472:86–U113.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, et al. 2012. Biodiversity loss and its impact on humanity. Nature. 486:59–67.

Carlisle DM, Clements WH. 2005. Leaf litter breakdown, microbial respiration and shredder production in metal-polluted streams. Freshw Biol. 50:380–390.

CAS. 2016. Chemical Abstracts Service Databases. Available from: www.cas.org/content/cas-databases

Caswell H. 2001. Matrix population models: construction, analysis, and interpretation. Sinauer Associates, Inc., Sunderland, MA.

Chandler GT, Cary TL, Volz DC, Walse SS, Ferry JL, Klosterhaus SL. 2004. Fipronil effects on estuarine copepod (*Amphiascus tenuiremis*) development, fertility, and reproduction: a rapid life-cycle assay in 96-well microplate format. Environ Toxicol Chem. 23:117–124.

Civitello DJ, Forys P, Johnson AP, Hall SR. 2012. Chronic contamination decreases disease spread: a Daphnia-fungus-copper case study. Proc R Soc Lond Ser B-Biol Sci. 279:3146–3153.

Civitello DJ, Cohen J, Fatima H, Halstead NT, Liriano J, McMahon TA, Ortega CN, Sauer EL, Sehgal T, Young S, et al. 2015. Biodiversity inhibits parasites: Broad evidence for the dilution effect. P Natl Acad Sci USA. Jul 14; 112:8667–8671.

Clements WH, Rohr JR. 2009. Community responses to contaminants: using basic ecological principles to predict ecotoxicological effects. Environ Toxicol Chem. 28:1789–1800.

Cohen J, Civitello DJ, Brace AJ, Feichtinger E, Ortega N, Richardson JC, Sauer EL, Liu X, Rohr JR. in press. Spatial scale modulates the strength of ecological processes driving disease distributions. P Natl Acad Sci USA.

Cothran RD, Brown JM, Relyea RA. 2013. Proximity to agriculture is correlated with pesticide tolerance: evidence for the evolution of amphibian resistance to modern pesticides. Evol Appl. 6:832–841.

Courchamp F, Clutton-Brock T, Grenfell B. 1999. Inverse density dependence and the Allee effect. Trends Ecol Evol. 14:405–410.

Crane M, Newman MC. 2000. What level of effect is a no observed effect? Environ Toxicol Chem. 19:516–519.

de Groot RS, Wilson MA, Boumans RMJ. 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol Econ. 41:393–408.

De Laender F, De Schamphelaere KAC, Vanrolleghem PA, Janssen CR. 2008a. Do we have to incorporate ecological interactions in the sensitivity assessment of ecosystems? An examination of a theoretical assumption underlying species sensitivity distribution models. Environ Int. 34:390–396.

De Laender F, De Schamphelaere KAC, Vanrolleghem PA, Janssen CR. 2008b. Validation of an ecosystem modelling approach as a tool for ecological effect assessments. Chemosphere. 71:529–545.

De Laender F, De Schamphelaere KAC, Vanrolleghem PA, Janssen CR. 2009. Comparing ecotoxicological effect concentrations of chemicals established in multi-species vs. single-species toxicity test systems. Ecotoxicol Environ Safe. 72:310–315.

de Roos AM. PSPManalysis: a Matlab/C Package for Numerical Analysisof Physiologically Structured Population Models. 2014. https://staff.fnwi.uva.nl/a.m.deroos/PSPManalysis/index.html.

de Roos AM, Diekmann O, Metz JAJ. 1992. Studying the dynamics of structured population models: a versatile technique and its application to *Daphnia*. Am Nat. 139:123–147.

de Roos AM, Persson L. 2013. Population and community ecology of ontogenetic development. Princeton (NJ): Princeton University Press.

Douglas MR, Rohr JR, Tooker JF. 2015. Neonicotinoid insecticide travels through a soil food chain, disrupting biological control of non-target pests and decreasing soya bean yield. J Appl Ecol. 52:250–260.

Dunne JA, Williams RJ, Martinez ND. 2002. Food-web structure and network theory: the role of connectance and size. Proc Natl Acad Sci USA. 99:12917–12922.

Ebenman B, Jonsson T. 2005. Using community viability analysis to identify fragile systems and keystone species. Trends Ecol Evol (Amst). 20:568–575.

EFSA. 2006. Opinion of the scientific panel on plant health, plant protection products and their residues on a request from the EFSA related to the aquatic risk assessment for cyprodinil and the use of a mesocosm study in particular. EFSA J. 329:1–77.

EFSA PoPPPatR 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA J. 3290:1–268.

EFSA PoPPPatR. 2014. Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA J. 3589:1–92.

Ehrsam M, Knutie SA, Rohr JR. in press. The herbicide atrazine induces hyperactivity and compromises tadpole detection of predator chemical cues. Environmental Toxicology and Chemistry. DOI: 10.1002/etc.3377.

Engelen B, Meinken K, von Wintzintgerode F, Heuer H, Malkomes HP, Backhaus H. 1998. Monitoring impact of a pesticide treatment on bacterial soil communities by metabolic and genetic fingerprinting in addition to conventional testing procedures. Appl Environ Microbiol. 64:2814–2821.

Erickson RA, Cox SB, Oates JL, Anderson TA, Salice CJ, Long KR. 2014. A Daphnia population model that considers pesticide exposure and demographic stochasticity. Ecol Model. 275:37–47.

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

EU. 2001. White Paper: strategy for a future chemicals policy. Brussels (27.2.2001, COM(2001) 88 final).

Ewer J, Gammie SC, Truman JW. 1997. Control of insect ecdysis by a positive-feedback endocrine system: roles of eclosion hormone and ecdysis triggering hormone. J Exp Biol. 200:869–881.

Fairchild JF, Lapoint TW, Schwartz TR. 1994. Effects of an herbicide and insecticide mixture in aquatic mesocosms. Arch Environ Contam Toxicol. 27:527–533.

Fairchild JF, Sappington LC. 2002. Fate and effects of the triazinone herbicide metribuzin in experimental pond mesocosms. Arch Environ Contam Toxicol. 43:198–202.

Fleeger JW, Carman KR, Nisbet RM. 2003. Indirect effects of contaminants in aquatic ecosystems. Sci Total Environ. 317:207–233.

Focks A, ter Horst M, van den Berg E, Baveco H, van den Brink PJ. 2014. Integrating chemical fate and population-level effect models for pesticides at landscape scale: new options for risk assessment. Ecol Model. 280:102–116.

Forbes VE, Brain R, Edwards D, Galic N, Hall T, Honegger J, Meyer C, Moore DRJ, Nacci D, Pastorok R, et al. 2015. Assessing pesticide risks to threatened and endangered species using population models: findings and recommendations from a CropLife America Science Forum. Integ Environ Assess Manage. 11:348–354.

Forbes VE, Calow P. 2002a. Population growth rate as a basis for ecological risk assessment of toxic chemicals. Philos Trans R Soc Lond B Biol Sci. 357:1299–1306.

Forbes VE, Calow P. 2002b. Species sensitivity distributions revisited: a critical appraisal. Hum Ecol Risk Assess. 8:473–492.

Forbes VE, Calow P. 2013. Developing predictive systems models to address complexity and relevance for ecological risk assessment. Integr Environ Assess. 9:E75–E80.

Forbes VE, Calow P, Grimm V, Hayashi T, Jager T, Palmqvist A, Pastorok R, Salvito D, Sibly R, Spromberg J, et al. 2010. Integrating population modeling into ecological risk assessment. Integr Environ Assess Manag. 6:191–193.

Forbes VE, Calow P, Grimm V, Hayashi TI, Jager T, Katholm A, Palmqvist A, Pastorok R, Salvito D, Sibly R, et al. 2011. Adding value to ecological risk assessment with population modeling. Hum Ecol Risk Assess. 17:287–299.

Forbes VE, Calow P, Sibly RM. 2008. The extrapolation problem and how population modeling can help. Environ Toxicol Chem. 27:1987–1994.

Forbes VE, Palmqvist A, Bach L. 2006. The use and misuse of biomarkers in ecotoxicology. Environ Toxicol Chem. 25:272–280.

Forbes VE, Sibly RM, Calow P. 2001. Toxicant impacts on density-limited populations: a critical review of theory, practice, and results. Ecol Appl. 11:1249–1257.

Fujiwara M, Kendall BE, Nisbet RM, Bennett WA. 2005. Analysis of size trajectory data using an energetic-based growth model. Ecology. 86:1441–1451.

Galic N, Hommen U, Boveco JM, van den Brink PJ. 2010. Potential application of ecological models in the European environmental risk assessment of chemicals II: review of models and their potential to address environmental protection aims. Integr Environ Assess Manage. 6:338–360.

Ge Y, Priester JH, Van de Werfhorst LC, Walker SL, Nisbet RM, An Y-J, Schimel JP, Gardea-Torresdey JL, Holden PA. 2014. Soybean plants modify metal oxide nanoparticle effects on soil bacterial communities. Environ Sci Technol. 48:13489–13496.

Ge Y, Schimel JP, Holden PA. 2011. Evidence for negative effects of TiO_2 and ZnO nanoparticles on soil bacterial communities. Environ Sci Technol. 45:1659–1664.

Gergs A, Zenker A, Grimm V, Preuss TG. 2013. Chemical and natural stressors combined: from cryptic effects to population extinction. Sci Rep. 3:2036.

Graney RL, Kennedy JH, Rodgers JH, editors. 1994. Aquatic mesosocm studies in ecological risk assessment. Boca Raton: CRC Publisher.

Grimm V, Augusiak J, Focks A, Frank BM, Gabsi F, Johnston ASA, Liu C, Martin BT, Meli M, Radchuk V, et al. 2014. Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. Ecol Model. 280:129–139.

Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz SK, Huse G, et al. 2006. A standard protocol for describing individual-based and agent-based models. Ecol Model. 198:115–126.

Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. 2010. The ODD protocol: a review and first update. Ecol Model. 221:2760–2768.

Grimm V, Martin BT. 2013. Mechanistic effect modeling for ecological risk assessment: where to go from here? Integr Environ Assess Manage. 9:e58–e63.

Grimm V, Railsback SF. 2005. Individual-based modeling and ecology. Princeton (NJ): Princeton University Press.

Grimm V, Railsback SF. 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. Philos Trans R Soc B Biol Sci. 367:298–310.

Guenard G, von der Ohe PC, de Zwart D, Legendre P, Lek S. 2011. Using phylogenetic information to predict species tolerances to toxic chemicals. Ecol Appl. 21:3178–3190.

Guenard G, von der Ohe PC, Walker SC, Lek S, Legendre P. 2014. Using phylogenetic information and chemical properties to predict species tolerances to pesticides. Proc R Soc Lond Ser B-Biol Sci. 281:20133239.

Halstead NT, Civitello DJ, Rohr JR. 2015. Comparative toxicities of organophosphate and pyrethroid insecticides to aquatic macroarthropods. Chemosphere. 135:265–271.

Halstead NT, McMahon TA, Johnson SA, Raffel TR, Romansic JM, Crumrine PW, Rohr JR. 2014. Community ecology theory predicts the effects of agrochemical mixtures on aquatic biodiversity and ecosystem properties. Ecol Lett. 17:932–941.

Hammond JI, Jones DK, Stephens PR, Relyea RA. 2012. Phylogeny meets ecotoxicology: evolutionary patterns of sensitivity to a common insecticide. Evol Appl. 5:593–606.

Hanazato T. 1998. Response of a zooplankton community to insecticide application in experimental ponds: a review and the implications of the effects of chemicals on the structure and functioning of freshwater communities. Environ Pollut. 101:361–373.

Handy RD, Galloway TS, Depledge MH. 2003. A proposal for the use of biomarkers for the assessment of chronic pollution and in regulatory toxicology. Ecotoxicology. 12:331–343.

Hansen F, Forbes VE, Forbes TL. 1999. Using elasticity analysis of demographic models to link toxicant effects on individuals to the population level: an example. Funct Ecol. 13:157–162.

Hanski I. 1998. Metapopulation dynamics. Nature. 396:41–49.

Hayes EH, Landis WG. 2004. Regional ecological risk assessment of a near shore marine environment: Cherry Point, WA. Hum Ecol Risk Assess. 10:299–325.

Hecker M, Giesy JP. 2008. Novel trends in endocrine disruptor testing: the H295R Steroidogenesis Assay for identification of inducers and inhibitors of hormone production. Anal Bioanal Chem. 390:287–291.

Heckmann L-H, Callaghan A, Hooper HL, Connon R, Hutchinson TH, Maund SJ, Sibly RM. 2007. Chronic toxicity of ibuprofen to *Daphnia magna*: effects on life history traits and population dynamics. Toxicol Lett. 172:137–145.

Hill IR, Heimbach F, Leeuwangh P, Matthiessen P, editors. 1994. Freshwater field test for hazard assessment of chemicals. Boca Raton: CRC Press, Inc.

Hofer T, Gerner I, Gundert-Remy U, Liebsch M, Schulte A, Spielmann H, Vogel R, Wettig K. 2004. Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. Arch Toxicol. 78:549–564.

Holden PA, Nisbet RM, Lenihan HS, Miller RJ, Cherr GN, Schimel JP, Gardea-Torresdey JL, California U. 2013. Ecological nanotoxicology: integrating nanomaterial hazard considerations across the subcellular, population, community, and ecosystems levels. Accounts Chem Res. 46:813–822.

Holt RD, Polis GA. 1997. A theoretical framework for intraguild predation. Am Nat. 149:745–764.

Hommen U, Baveco JM, Galic N, van den Brink PJ. 2010. Potential application of ecological models in the European environmental risk assessment of chemicals I: review of protection goals in EU directives and regulations. Integr Environ Assess Manage. 6:325–337.

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, et al. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecol Monogr. 75:3–35.

Hose GC, Van den Brink PJ. 2004. Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm, and field data. Arch Environ Contam Toxicol. 47:511–520.

Hua J, Morehouse NI, Relyea R. 2013. Pesticide tolerance in amphibians: induced tolerance in susceptible populations, constitutive tolerance in tolerant populations. Evol Appl. 6:1028–1040.

Hunsaker CT, Graham RL, Suter GW, Oneill RV, Barnthouse LW, Gardner RH. 1990. Assessing ecological risk on a regional scale. Environ Manage. 14:325–332.

Ibrahim L, Preuss TG, Ratte HT, Hommen U. 2013. A list of fish species that are potentially exposed to pesticides in edge-of-field water bodies in the European Union-a first step towards identifying vulnerable representatives for risk assessment. Environ Sci Pollut Res. 20:2679–2687.

Ibrahim L, Preuss TG, 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. Ecol Model. 280:65–75.

Ives AR. 1995. Predicting the response of populations to environmental change. Ecology. 76:926–941.

Jager T. 2011. Some good reasons to Ban ECx and related concepts in ecotoxicology. Environ Sci Technol. 45:8180–8181.

Jager T. 2012. Bad habits die hard: the NOEC's persistence reflects poorly on ecotoxicology. Environ Toxicol Chem. 31:228–229.

Jager T. 2015. Making sense of chemical stress: application of dynamic energy budget theory in ecotoxicology and stress ecology LeanPub. Available from: https://leanpub.com/debtox_book

Jager T, Albert C, Preuss TG, Ashauer R. 2011. General Unified Threshold Model of survival – a toxicokinetic–toxicodynamic framework for ecotoxicology. Environ Sci Technol. 45:2529–2540.

Jager T, Martin BT, Zimmer EI. 2013. DEBkiss or the quest for the simplest generic model of animal life history. J Theor Biol. 328:9–18.

Jager T, Vandenbrouck T, Baas J, De Coen WM, Kooijman SALM. 2010. A biology-based approach for mixture toxicity of multiple endpoints over the life cycle. Ecotoxicology. 19:351–361.

Jager T, Zimmer EI. 2012. Simplified Dynamic Energy Budget model for analysing ecotoxicity data. Ecol Model. 225:74–81.

Jayawardena UA, Rohr JR, Navaratne AN, Amerasinghe PH, Rajakaruna RS. 2016. Combined effects of pesticides and trematode infections on the hourglass tree frog Polypedates cruciger. EcoHealth. 13:111–122.

Jennings DE, Congelosi AM, Rohr JR. 2012. Insecticides reduce survival and the expression of traits associated with carnivory of carnivorous plants. Ecotoxicology. 21:569–575.

Joern A, Hoagland KD. 1996. In defense of whole-community bioassays for risk assessment. Environ Toxicol Chem. 15:407–409.

Johnsen K, Jacobsen CS, Torsvik V, Sorensen J. 2001. Pesticide effects on bacterial diversity in agricultural soils – a review. Biol Fertil Soils. 33:443–453.

Johnson LR, Pecquerie L, Nisbet RM. 2013. Bayesian inference for bioenergetic models. Ecology. 94:882–894.

Jones R, Leahy J, Mahoney M, Murray L, Odenkirchen E, Petrie R, Stangel C, Sunzenauer I, Vaituzis Z, Williams AJ. 2004. Overview of the ecological risk assessment process in the office of pesticide programs. Washington (DC): UEP Agency. Available from: http://www.epa.gov/sites/production/files/2014-11/documents/ecorisk-overview.pdf

Kamo M, Naito W. 2008. A novel approach to determining a population-level threshold in ecological risk assessment: a case study of zinc. Hum Ecol Risk Assess. 14:714–727.

Kattwinkel M, Liess M, Arena M, Bopp S, Streissl F, Roembke J. 2015. Recovery of aquatic and terrestrial populations in the context of European pesticide risk assessment. Environ Rev. 23:382–394.

Kearney MR, Domingos T, Nisbet R. 2015. Dynamic energy budget theory: an efficient and general theory for ecology. Bioscience. 65:341–341.

Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, Flick RW. 2007. Collapse of a fish population after exposure to a synthetic estrogen. Proc Natl Acad Sci USA. 104:8897–8901.

Kimberly DA, Salice CJ. 2013. Interactive effects of contaminants and climate-related stressors: high temperature increases sensitivity to cadmium. Environ Toxicol Chem. 32:1337–1343.

Kimberly DA, Salice CJ. 2014. Complex interactions between climate change and toxicants: evidence that temperature variability increases sensitivity to cadmium. Ecotoxicology. 23:809–817.

Klaniscek T, Caswell H, Neubert MG, Nisbet RM. 2006. Integrating dynamic energy budgets into matrix population models. Ecol Model. 196:407–420.

Klanjscek T, Nisbet RM, Caswell H, Neubert MG. 2007. A model for energetics and bioaccumulation in marine mammals with applications to the right whale. Ecol Appl. 17:2233–2250.

Klanjscek T, Nisbet RM, Priester JH, Holden PA. 2012. Modeling physiological processes that relate toxicant exposure and bacterial population dynamics. PLoS One. 7:e26955.

Klanjscek T, Nisbet RM, Priester JH, Holden PA. 2013. Dynamic energy budget approach to modeling mechanisms of CdSe quantum dot toxicity. Ecotoxicology. 22:319–330.

Kooijman SALM. 2000. Dynamic energy and mass budgets in biological systems. 2nd ed. New York: Cambridge University Press.

Kooijman SALM. 2001. Quantitative aspects of metabolic organization: a discussion of concepts. Philos Trans R Soc Lond B Biol Sci. 356:331–349.

Kooijman S. 2010. Dynamic energy budget theory for metabolic organization. Cambridge, UK: Cambridge University Press.

Kooijman SALM, Bedaux JJM. 1996. The analysis of aquatic toxicity data. Amsterdam: VU University Press.

Kooijman SALM, Metz JAJ. 1984. On the dynamics of chemically stressed populations: the deduction of population consequences from effects on individuals. Ecotox Environ Safe. 8:254–274.

Kooijman SALM, Sousa T, Pecquerie L, van der Meer J, Jager T. 2008. From food-dependent statistics to metabolic parameters, a practical guide to the use of dynamic energy budget theory. Biol Rev Camb Philos Soc. 83:533–552.

Kramer VJ, Etterson MA, Hecker M, Murphy CA, Roesijadi G, Spade DJ, Spromberg JA, Wang M, Ankley GT. 2011. Adverse outcome pathways and ecological risk assessment: bridging to population-level effects. Environ Toxicol Chem. 30:64–76.

LaLone CA, Villeneuve DL, Burgoon LD, Russom CL, Helgen HW, Berninger JP, Tietge JE, Severson MN, Cavallin JE, Ankley GT. 2013. Molecular target sequence similarity as a basis for species extrapolation to assess

the ecological risk of chemicals with known modes of action. Aquat Toxicol. 144:141–154.

LaLone CA, Villeneuve DL, Cavallin JE, Kahl MD, Durhan EJ, Makynen EA, Jensen KM, Stevens KE, Severson MN, Blanksma CA, et al. 2013. Cross-species sensitivity to a novel androgen receptor agonist of potential environmental concern, spironolactone. Environ Toxicol Chem. 32:2528–2541.

Landis W, Chapman PM. 2011. Well past time to stop using NOELs and LOELs. Integr Environ Assess Manage. 7:vi–viii.

Landis WG. 2002. Uncertainty in the extrapolation from individual effects to impacts upon landscapes. Hum Ecol Risk Assess. 8:193–204.

Landis WG. 2003a. The frontiers in ecological risk assessment at expanding spatial and temporal scales. Hum Ecol Risk Assess. 9:1415–1424.

Landis WG. 2003b. Twenty years before and hence; Ecological risk assessment at multiple scales with multiple stressors and multiple endpoints. Hum Ecol Risk Assess. 9:1317–1326.

Landis WG, Wiegers JA. 1997. Design considerations and a suggested approach for regional and comparative ecological risk assessment. Hum Ecol Risk Assess. 3:287–297.

Larsen DP, Denoyelles F, Stay F, Shiroyama T. 1986. Comparisons of single species, microcosm and experimental pond responses to atrazine exposure. Environ Toxicol Chem. 5:179–190.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett. 7:601–613.

Levin SA, Harwell MA, Kelly JR. Kimball DD, editors. 1989. Ecotoxicology: problems and approaches. New York: Springer-Verlag.

Li ZH, Kroll KJ, Jensen KM, Villeneuve DL, Ankley GT, Brian JV, Sepulveda MS, Orlando EF, Lazorchak JM, Kostich M, et al. 2011. A computational model of the hypothalamic-pituitary-gonadal axis in female fathead minnows (*Pimephales promelas*) exposed to 17 alpha-ethynylestradiol and 17 beta-trenbolone. BMC Syst Biol. 5:63.

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

Liess M, Champeau O, Riddle M, Schulz R, Duquesne S. 2001. Combined effects of ultraviolet-B radiation and food shortage on the sensitivity of the Antarctic amphipod *Paramoera walkeri* to copper. Environ Toxicol Chem. 20:2088–2092.

Liess M, Foit K, Becker A, Hassold E, Dolciotti I, Kattwinkel M, Duquesne S. 2013. Culmination of low-dose pesticide effects. Environ Sci Technol. 47:8862–8868.

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

Lika K, Kearney MR, Freitas V, van der Veer HW, van der Meer J, Wijsman JWM, Pecquerie L, Kooijman SALM. 2011. The "covariation method" for estimating the parameters of the standard Dynamic Energy Budget model I: philosophy and approach. J Sea Res. 66:270–277.

Luna TO, Plautz SC, Salice CJ. 2013. Effects of 17 alpha-ethynlestradiol, fluoxetine, and the mixture on life history traits and population growth rates in a freshwater gastropod. Environ Toxicol Chem. 32:2771–2778.

Luttik R, Hart A, Roelofs W, Craig P, Mineau P. 2011. Variation in the level of protection afforded to birds and crustaceans exposed to different pesticides under standard risk assessment procedures. Integr Environ Assess Manage. 7:459–465.

Maltby L, Blake N, Brock TCM, Van Den Brink PJ. 2005. Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems. Environ Toxicol Chem. 24:379–388.

Maron JL, Crone E. 2006. Herbivory: effects on plant abundance, distribution and population growth. Proc R Soc Lond Ser B Biol Sci. 273:2575–2584.

Marquet PA, Allen AP, Brown JH, Dunne JA, Enquist BJ, Gillooly JF, Gowaty PA, Green JL, Harte J, Hubbell SP, et al. 2014. On theory in ecology. Bioscience. 64:701–710.

Martin B, Jager T, Nisbet RM, Preuss TG, Grimm V. 2014. Limitations of extrapolating toxic effects on reproduction to the population level. Ecol Appl. 24:1972–1983.

Martin BT, Jager T, Nisbet RM, Preuss TG, Hammers-Wirtz M, Grimm V. 2013. Extrapolating ecotoxicological effects from individuals to populations: a generic approach based on Dynamic Energy Budget theory and individual-based modeling. Ecotoxicology. 22:574–583.

Martin BT, Zimmer EI, Grimm V, Jager T. 2012. Dynamic Energy Budget theory meets individual-based modelling: a generic and accessible implementation. Methods Ecol Evol. 3:445–449.

Martin LB, Hopkins WA, Mydlarz LD, Rohr JR. 2010. The effects of anthropogenic global changes on immune functions and disease resistance. In: Year in ecology and conservation biology. p. 129–148.

Martinovic-Weigelt D, Mehinto AC, Ankley GT, Denslow ND, Barber LB, Lee KE, King RJ, Schoenfuss HL, Schroeder AL, Villeneuve DL. 2014. Transcriptomic effects-based monitoring for endocrine active chemicals: assessing relative contribution of treated wastewater to downstream pollution. Environ Sci Technol. 48:2385–2394.

McMahon TA, Halstead NT, Johnson S, Raffel TR, Romansic JM, Crumrine PW, Boughton RK, Martin LB, Rohr JR. 2011. The fungicide chlorothalonil is nonlinearly associated with corticosterone levels, immunity, and mortality in amphibians. Environ Health Persp. 119:1098–1103.

McMahon TA, Halstead NT, Johnson S, Raffel TR, Romansic JM, Crumrine PW, Rohr JR. 2012. Fungicide-induced declines of freshwater biodiversity modify ecosystem functions and services. Ecol Lett. 15:714–722.

McMahon TA, Romansic JM, Rohr JR. 2013. Nonmonotonic and monotonic effects of pesticides on the pathogenic fungus *Batrachochytrium dendrobatidis* in culture and on tadpoles. Environ Sci Technol. 47:7958–7964.

Meaney MJ, Diorio J, Francis D, Widdowson J, LaPlante P, Caldji C, Sharma S, Seckl JR, Plotsky PM. 1996. Early environmental regulation of forebrain glucocorticoid receptor gene expression: implications for adrenocortical responses to stress. Dev Neurosci. 18:49–72.

MilleniumEcosystemAssessment. 2005. Ecosystems and human well-being: biodiversity synthesis. Washington, DC: WR Institute.

Moe SJ, Stenseth NC, Smith RH. 2002. Density-dependent compensation in blowfly populations give indirectly positive effects of a toxicant. Ecology. 83:1597–1603.

Muller EB, Hanna SK, Lenihan HS, Miller R, Nisbet RM. 2014. Impact of engineered zinc oxide nanoparticles on the energy budgets of *Mytilus galloprovincialis*. J Sea Res. 94:29–36.

Muller EB, Kooijman S, Edmunds PJ, Doyle FJ, Nisbet RM. 2009. Dynamic energy budgets in syntrophic symbiotic relationships between heterotrophic hosts and photoautotrophic symbionts. J Theor Biol. 259:44–57.

Muller EB, Nisbet RM, Berkley H. 2010. Sublethal toxicant effects with dynamic energy budget theory: model formulation. Ecotoxicology. 19:48–60.

Muller EB, Nisbet RM, Berkley HA. 2010. Sublethal toxicant effects with dynamic energy budget theory: model formulation. Ecotoxicology. 19:48–60.

Muller EB, Osenberg CW, Schmitt RJ, Holbrook SJ, Nisbet RM. 2010. Sublethal toxicant effects with dynamic energy budget theory: application to mussel outplants. Ecotoxicology. 19:38–47.

Munns WR, Rea AW, Mazzotta MJ, Wainger LA, Saterson K. 2015. Toward a standard lexicon for ecosystem services. Integr Environ Assess Manag. 11:666–673.

Murphy CA, Rose KA, Alvarez Mdel C, Fuiman LA. 2008. Modeling larval fish behavior: scaling the sublethal effects of methylmercury to population-relevant endpoints. Aquat Toxicol. 86:470–484.

Murphy CA, Rose KA, Thomas P. 2005. Modeling vitellogenesis in female fish exposed to environmental stressors: predicting the effects of endocrine disturbance due to exposure to a PCB mixture and cadmium. Reprod Toxicol. 19:395–409.

Nabholz JV, Clements RG, Zeeman MG. 1997. Information needs for risk assessment in EPA's office of pollution prevention and toxics. Ecol Appl. 7:1094–1098.

Naito W, Miyamoto K, Nakanishi J, Masunaga S, Bartell SM. 2002. Application of an ecosystem model for aquatic ecological risk assessment of chemicals for a Japanese lake. Water Res. 36:1–14.

Nel A, Xia T, Meng H, Wang X, Lin S, Ji Z, Zhang H. 2013. Nanomaterial toxicity testing in the 21st century: use of a predictive toxicological approach and high-throughput screening. Acc Chem Res. 46:607–621.

Newman MC, Ownby DR, Mezin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson BA. 2000. Applying species-sensitivity distributions in

ecological risk assessment: assumptions of distribution type and sufficient numbers of species. Environ Toxicol Chem. 19:508–515.
Niederlehner BR, Pontasch KW, Pratt JR, Cairns J. 1990. Field evaluation of predictions of environmental effects from a multispecies microcosm toxicity test. Arch Environ Contam Toxicol. 19:62–71.
Nielsen MB, Kjeldsen KU, Lever MA, Ingvorsen K. 2014. Survival of prokaryotes in a polluted waste dump during remediation by alkaline hydrolysis. Ecotoxicology. 23:404–418.
Nilsen BM, Berg K, Eidem JK, Kristiansen SI, Brion F, Porcher JM, Goksoyr A. 2004. Development of quantitative vitellogenin-ELISAs for fish test species used in endocrine disruptor screening. Anal Bioanal Chem. 378:621–633.
Nisbet RM, McCauley E, Johnson LR. 2010. Dynamic energy budget theory and population ecology: lessons from Daphnia. Philos Trans R Soc Lond B Biol Sci. 365:3541–3552.
Nisbet RM, Muller EB, Lika K, Kooijman S. 2000. From molecules to ecosystems through dynamic energy budget models. J Anim Ecol. 69:913–926.
Nisbet RM, Murdoch WW, Stewart-Oaten A. 1996. Consequences for adult fish stocks of human-induced mortality on immatures. In: Detecting ecological impacts. San Diego: Academic Press. p. 257–277.
Noyes PD, McElwee MK, Miller HD, Clark BW, Van Tiem LA, Walcott KC, Erwin KN, Levin ED. 2009. The toxicology of climate change: environmental contaminants in a warming world. Environ Int. 35:971–986.
OECD, editor. 2006. Current approaches in the statistical analysis of ecotoxicity data: a guidance to application. Paris.
OECD. 2015. OECD guidelines for the testing of chemicals: effects on biotic systems. Paris, France: OfEC-oa Development. Available from: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-2-effects-on-biotic-systems_20745761
PAN. Pesticides Database. Available from: http://www.pesticideinfo.org/Search_Products.jsp
PapeLindstrom PA, Lydy MJ. 1997. Synergistic toxicity of atrazine and organophosphate insecticides contravenes the response addition mixture model. Environ Toxicol Chem. 16:2415–2420.
Park J, Lee J, Choi C. 2011. Mitochondrial network determines intracellular ROS dynamics and sensitivity to oxidative stress through switching inter-mitochondrial messengers. PLoS One. 6:e23211.
Park RA, Clough JS, Wellman MC. 2008. AQUATOX: modeling environmental fate and ecological effects in aquatic ecosystems. Ecol Model. 213:1–15.
Parng C, Seng WL, Semino C, McGrath P. 2002. Zebrafish: a preclinical model for drug screening. Assay Drug Dev Technol. 1:41–48.
Pastorok RA, Akcakaya HR, Regan H, Ferson S, Bartell SM. 2003. Role of ecological modeling in risk assessment. Hum Ecol Risk Assess. 9:939–972.
Pontasch KW, Cairns J. 1991. Multispecies toxicity tests using indigenous organisms: predicting the effects of complex effluents in streams. Arch Environ Contam Toxicol. 20:103–112.
Pontasch KW, Niederlehner BR, Cairns J. 1989. Comparisons of single-species, microcosm and field responses to a complex effluent. Environ Toxicol Chem. 8:521–532.
Posthuma L, Suter IIGW, Traas TP. 2001. Species sensitivity distributions in ecotoxicology. Boca Raton, FL: CRC Press.
Preziosi DV, Pastorok RA. 2008. Ecological food web analysis for chemical risk assessment. Sci Total Environ. 406:491–502.
Priester JH, Ge Y, Mielke RE, Horst AM, Moritz SC, Espinosa K, Gelb J, Walker SL, Nisbet RM, An Y-J, et al. 2012. Soybean susceptibility to manufactured nanomaterials with evidence for food quality and soil fertility interruption. Proc Natl Acad Sci U S A. 109:E2451–E2456.
Raffel TR, Halstead NT, McMahon T, Romansic JM, Venesky MD, Rohr JR. 2013. Disease and thermal acclimation in a more variable and unpredictable climate. Nat Clim Chang. 3:146–151.
Raffel TR, Rohr JR, Kiesecker JM, Hudson PJ. 2006. Negative effects of changing temperature on amphibian immunity under field conditions. Funct Ecol. 20:819–828.
Raffel TR, Sheingold JL, Rohr JR. 2009. Lack of pesticide toxicity to *Echinostoma trivolvis* eggs and miracidia. J Parasitol. 95:1548–1551.
Ramsey PW, Rillig MC, Feris KP, Gordon NS, Moore JN, Holben WE, Gannon JE. 2005. Relationship between communities and processes; new insights from a field study of a contaminated ecosystem. Ecol Lett. 8:1201–1210.
Relyea R, Hoverman J. 2006. Assessing the ecology in ecotoxicology: a review and synthesis in freshwater systems. Ecol Lett. 9:1157–1171.
Relyea RA. 2003. Predator cues and pesticides: a double dose of danger for amphibians. Ecol Appl. 13:1515–1521.
Relyea RA. 2005. The impact of insecticides and herbicides on the biodiversity and productivity of aquatic communities. Ecol Appl. 15:618–627.
Relyea RA, Schoeppner NM, Hoverman JT. 2005. Pesticides and amphibians: the importance of community context. Ecol Appl. 15:1125–1134.
Rohr JR, Civitello DJ, Crumrine PW, Halstead NT, Miller AD, Schotthoefer AM, Stenoien C, Johnson LB, Beasley VR. 2015. Predator diversity, intraguild predation, and indirect effects drive parasite transmission. P Natl Acad Sci USA. Mar 10; 112:3008–3013.
Rohr JR, Crumrine PW. 2005. Effects of an herbicide and an insecticide on pond community structure and processes. Ecol Appl. 15:1135–1147.
Rohr JR, Elskus AA, Shepherd BS, Crowley PH, McCarthy TM, Niedzwiecki JH, Sager T, Sih A, Palmer BD. 2003. Lethal and sublethal effects of atrazine, carbaryl, endosulfan, and octylphenol on the streamside salamander, *Ambystoma barbouri*. Environ Toxicol Chem. 22:2385–2392.
Rohr JR, Elskus AA, Shepherd BS, Crowley PH, McCarthy TM, Niedzwiecki JH, Sager T, Sih A, Palmer BD. 2004. Multiple stressors and salamanders: effects of an herbicide, food limitation, and hydroperiod. Ecol Appl. 14:1028–1040.
Rohr JR, Farag AM, Cadotte MW, Clements WH, Smith JR, Ulrich CP, Woods R. 2016. Transforming ecosystems: when, where, and how to restore contaminated sites. Integr Environ Assess Manage. 12:273–283.
Rohr JR, Halstead NT, Raffel TR. 2012. The herbicide atrazine, algae, and snail populations. Environ Toxicol Chem. 31:973–974.
Rohr JR, Johnson P, Hickey CW, Helm RC, Fritz A, Brasfield S. 2013. Implications of global climate change for natural resource damage assessment, restoration, and rehabilitation. Environ Toxicol Chem. 32:93–101.
Rohr JR, Kerby JL, Sih A. 2006. Community ecology as a framework for predicting contaminant effects. Trends Ecol Evol (Amst). 21:606–613.
Rohr JR, McCoy KA, 2010a. Preserving environmental health and scientific credibility: a practical guide to reducing conflicts of interest. Conserv Lett. 3:143–150.
Rohr JR, McCoy KA, 2010b. A qualitative meta-analysis reveals consistent effects of atrazine on freshwater fish and amphibians. Environ Health Persp. 18:20–32.
Rohr JR, Palmer BD. 2005. Aquatic herbicide exposure increases salamander desiccation risk eight months later in a terrestrial environment. Environ Toxicol Chem. 24:1253–1258.
Rohr JR, Palmer BD. 2013. Climate change, multiple stressors, and the decline of ectotherms. Conserv Biol. 27:741–751.
Rohr JR, Raffel TR, Halstead NT, McMahon TA, Johnson SA, Boughton RK, Martin LB. 2013. Early-life exposure to a herbicide has enduring effects on pathogen-induced mortality. Proc R Soc Lond Ser B-Biol Sci. 280:20131502.
Rohr JR, Raffel TR, Sessions SK, Hudson PJ. 2008. Understanding the net effects of pesticides on amphibian trematode infections. Ecol Appl. 18:1743–1753.
Rohr JR, Sager T, Sesterhenn TM, Palmer BD. 2006. Exposure, postexposure, and density-mediated effects of atrazine on amphibians: breaking down net effects into their parts. Environ Health Persp. 114:46–50.
Rohr JR, Schotthoefer AM, Raffel TR, Carrick HJ, Halstead N, Hoverman JT, Johnson CM, Johnson LB, Lieske C, Piwoni MD, et al. 2008. Agrochemicals increase trematode infections in a declining amphibian species. Nature. 455:1235–1239.
Rohr JR, Sesterhenn TM, Stieha C. 2011. Will climate change reduce the effects of a pesticide on amphibians? Partitioning effects on exposure and susceptibility to pollution. Glob Change Biol. 17:657–666.
Rohr JR, Swan A, Raffel TR, Hudson PJ. 2009. Parasites, info-disruption, and the ecology of fear. Oecologia. 159:447–454.
Salice CJ, Kimberly DA. 2013. Environmentally relevant concentrations of a common insecticide increase predation risk in a freshwater gastropod. Ecotoxicology. 22:42–49.

Salice CJ, Miller TJ. 2003. Population-level responses to long-term cadmium exposure in two strains of the freshwater gastropod Biomphalaria glabrata: results from a life-table response experiment. Environ Toxicol Chem. 22:678–688.

Salice CJ, Miller TJ, Roesijadi G. 2009. Demographic responses to multigeneration cadmium exposure in two strains of the freshwater gastropod, *Biomphalaria glabrata.* Arch Environ Contam Toxicol. 56:785–795.

Salice CJ, Roesijadi G. 2002. Resistance to cadmium and parasite infection are inversely related in two strains of a freshwater gastropod. Environ Toxicol Chem. 21:1398–1403.

Salice CJ, Rowe CL, Pechmann JHK, Hopkins WA. 2011. Multiple stressors amd complex life cycles: insights from a population-level assessment of breeding site contamination and terrestrial habitat loss in an amphibian. Environ Toxicol Chem. 30:2874–2882.

Salice CJ, Sample BE, Neilan RM, Rose KA, Sable S. 2011. Evaluation of alternative PCB clean-up strategies using an individual-based population model of mink. Environ Pollut. 159:3334–3343.

Schafer RB. 2014. ET&C perspectives. Environ Toxicol Chem. 33:1193–1194.

Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, van Nes EH, Rietkerk M, Sugihara G. 2009. Early-warning signals for critical transitions. Nature. 461:53–59.

Scheffer M, Carpenter SR. 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. Trends Ecol Evol. 18:648–656.

Schlaepfer MA, Runge MC, Sherman PW. 2002. Ecological and evolutionary traps. Trends Ecol Evol. 17:474–480.

Schmolke A, Thorbek P, Chapman P, Grimm V. 2010. Ecological models and pesticide risk assessment: current modeling practice. Environ Toxicol Chem. 29:1006–1012.

Schmolke A, Thorbek P, DeAngelis DL, Grimm V. 2010. Ecological models supporting environmental decision making: a strategy for the future. Trends Ecol Evol. 25:479–486.

Scholz S, Sela E, Blaha L, Braunbeck T, Galay-Burgos M, Garcia-Franco M, Guinea J, Kluever N, Schirmer K, Tanneberger K, et al. 2013. A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment. Regul Toxicol Pharmacol. 67:506–530.

Schotthoefer AM, Rohr JR, Cole RA, Koehler AV, Johnson CM, Johnson LB, Beasley VR. 2011. Effects of wetland vs. landscape variables on parasite communities of *Rana pipiens*: links to anthropogenic factors. Ecol Appl. 21:1257–1271.

Semlitsch RD, Bridges CM, Welch AM. 2000. Genetic variation and a fitness tradeoff in the tolerance of gray treefrog (*Hyla versicolor*) tadpoles to the insecticide carbaryl. Oecologia. 125:179–185.

Sibly RM, Calow P. 1989. A life-cycle theory of responses to stress. Biol J Linn Soc. 37:101–116.

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

Skelly DK, Kiesecker JM. 2001. Venue and outcome in ecological experiments: manipulations of larval anurans. Oikos. 94:198–208.

Sourisseau S, Basseres A, Perie F, Caquet T. 2008. Calibration, validation and sensitivity analysis of an ecosystem model applied to artificial streams. Water Res. 42:1167–1181.

Spromberg JA, John BM, Landis WG. 1998. Metapopulation dynamics: indirect effects and multiple distinct outcomes in ecological risk assessment. Environ Toxicol Chem. 17:1640–1649.

Stadnicka-Michalak J, Schirmer K, Ashauer R. 2015. Toxicology across scales: cell population growth in vitro predicts reduced fish growth. Sci Adv. 1:e1500302.

Staley ZR, Rohr JR, Harwood VJ. 2010. The effect of agrochemicals on indicator bacteria densities in outdoor mesocosms. Environ Microbiol. 12:3150–3158.

Staley ZR, Rohr JR, Harwood VJ. 2011. Test of direct and indirect effects of agrochemicals on the survival of fecal indicator bacteria. Appl Environ Microbiol. 77:8765–8774.

Staley ZR, Rohr JR, Senkbeil JK, Harwood VJ. 2014. Agrochemicals indirectly increase survival of E. coli O157:H7 and indicator bacteria by reducing ecosystem services. Ecol Appl. 24:1945–1953.

Staley ZR, Senkbeil JK, Rohr JR, Harwood VJ. 2012. Lack of direct effects of agrochemicals on zoonotic pathogens and fecal indicator bacteria. Appl Environ Microbiol. 78:8146–8150.

Stampfli NC, Knillmann S, Liess M, Beketov MA. 2011. Environmental context determines community sensitivity of freshwater zooplankton to a pesticide. Aquat Toxicol. 104:116–124.

Stark JD. 2005. How closely to acute lethal concentration estimates predict effects of toxicants on populations? Integr Environ Assess Manage. 1:109–113.

Stark JD, Banks JE. 2003. Population-level effects of pesticides and other toxicants on arthropods. Annu Rev Entomol. 2003. 48:505–519.

Stark JD, Banks JE, Vargas R. 2004. How risky is risk assessment: the role that life history strategies play in susceptibility of species to stress. Proc Natl Acad Sci USA. 101:732–736.

Stay FS, Katko A, Rohm CM, Fix MA, Larsen DP. 1989. The effects of atrazine on microcosms developed from four natural plankton communities. Arch Environ Contam Toxicol. 18:866–875.

Stephens PA, Sutherland WJ. 1999. Consequences of the Allee effect for behaviour, ecology and conservation. Trends Ecol Evol (Amst). 14:401–405.

Stevenson LM, Dickson H, Klanjscek T, Keller AA, McCauley E, Nisbet RM. 2013. Environmental feedbacks and engineered nanoparticles: mitigation of silver nanoparticle toxicity to *Chlamydomonas reinhardtii* by algal-produced organic compounds. PLoS One. 8:e74456.

Stone WW, Gilliom RJ, Ryberg KR. 2014. Pesticides in U.S. streams and rivers: occurrence and trends during 1992–2011. Environ Sci Technol. 48:11025–11030.

Sturla SJ, Boobis AR, FitzGerald RE, Hoeng J, Kavlock RJ, Schirmer K, Whelan M, Wilks MF, Peitsch MC. 2014. Systems toxicology: from basic research to risk assessment. Chem Res Toxicol. 27:314–329.

Sundling K, Craciun G, Schultz I, Hook S, Nagler J, Cavileer T, Verducci J, Liu YS, Kim J, Hayton W. 2014. Modeling the endocrine control of vitellogenin production in female rainbow trout. Math Biosci Eng. 11:621–639.

Suter GW. 1990. Endpoint for regional ecological risk assessments. Environ Manage. 14:9–23.

Suter GW. 2007. Ecoological risk assessment. 2cd ed. Boca Raton (FL): CRC Press.

Suter GW. 2008. Ecological risk assessment in the United States Environmental Protection Agency: a historical overview. Integr Environ Assess Manage. 4:285–289.

Suter GW, Norton SB, Fairbrother A. 2005. Individuals versus organisms versus populations in the definition of ecological assessment endpoint. Integr Environ Assess Manage. 1:397–400.

Taub FB. 1997a. Are ecological studies relevant to pesticide registration decisions? Ecol Appl. 7:1083–1085.

Taub FB. 1997b. Unique information contributed by multispecies systems: examples from the standardized aquatic microcosm. Ecol Appl. 7:1103–1110.

Thorbek P, Forbes VE, Heimbach F, Hommen U, Thulke HH, van den Brink P, Wogram J, Grimm V, editors. 2009. Ecological models for regulatory risk assessments of pesticides: developing a strategy for the future. Boca Raton, FL: CRC Press.

Topping CJ, Craig PS, de Jong F, Klein M, Laskowski R, Manachini B, Pieper S, Smith R, Sousa JP, Streissl F, et al. 2015. Towards a landscape scale management of pesticides: ERA using changes in modelled occupancy and abundance to assess long-term population impacts of pesticides. Sci Total Environ. 537:159–169.

Topping CJ, Dalby L, Skov F. 2016. Landscape structure and management alter the outcome of a pesticide ERA: evaluating impacts of endocrine disruption using the ALMaSS European Brown Hare model. Sci Total Environ. 541:1477–1488.

Topping CJ, Kjaer LJ, Hommen U, Hoye TT, Preuss TG, Sibly RM, van Vliet P. 2014. REcovery based on plot experiments is a poor predictor of landscape-level population impacts of agricultural pesticides. Environ Toxicol Chem. 33:1499–1507.

Touart LW. 1988. Aquatic mesocosm tests to support pesticide registrations. Washington (DC): HED U.S. Environmental Protection Agency. Availbale from: http://nepis.epa.gov/Adobe/PDF/9101ALWK.PDF

Touart LW, Maciorowski AF. 1997. Information needs for pesticide registration in the United States. Ecol Appl. 7:1086–1093.

Traas TP, Janse JH, Van den Brink PJ, Brock TCM, Aldenberg T. 2004. A freshwater food web model for the combined effects of nutrients and insecticide stress and subsequent recovery. Environ Toxicol Chem. 23:521–529.

Tuljapurkar S, Caswell H. 2012. Structured-population models in marine, terrestrial, and freshwater systems. New York, NY: Springer Science & Business Media.

TIM Version 3.0 beta Technical Description and User Guide – Appendix I – Overview and History of Tiered Risk Assessment Framework. Available from: https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/tim-version-30-beta-technical-description-and-user-0

USEPA. 1992. Framework for ecological risk assessment. Washington (DC): UEP Agency. (EPA/630/R-92/001).

USEPA. 1993. Pesticide reregistration progress report.

USEPA. 1998. Guidelines for ecological risk assessment. Washington (DC): UEP Agency. (EPA/630/R-95/002F).

USEPA. 2000. AQUATOX for Windows: model validation reports. Washington (DC). (EPA-823-R-00-008).

USEPA. 2003. Generic Ecological Assessment Endpoints (GEAEs) for ecological risk assessment. Washington (DC): U.S. Environmental Protection Agency. p. 67.

Van den Brink PJ. 2006. Letter to the editor: response to recent criticism on aquatic semifield experiments: opportunities for new developments in ecological risk assessment of pesticides. Integr Environ Assess Manage. 2:202–203.

Van den Brink PJ. 2013. Assessing aquatic population and community-level risks of pesticides. Environ Toxicol Chem. 32:972–973.

Van den Brink PJ, Baird DJ, Baveco H, Focks A. 2013. The use of traits-based approaches and eco(toxico)logical models to advance the ecological risk assessment framework for chemicals. Integr Environ Assess Manage. 9:e47–e57.

Van den Brink PJ, Roelsma J, Van Nes EH, Scheffer M, Brock TCM. 2002. PERPEST model, a case-based reasoning approach to predict ecological risks of pesticides. Environ Toxicol Chem. 21:2500–2506.

van der Vaart E, Beaumont MA, Johnston ASA, Sibly RM. 2015. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. Ecol Model. 312:182–190.

van der Vaart E, Johnston ASA, Sibly RM. 2016. Predicting how many animals will be where: how to build, calibrate and evaluate individual-based models. Ecol Model. 326:113–123.

Voccia I, Blakley B, Brousseau P, Fournier M., 1999 Immunotoxicity of pesticides: a review. Toxicol Ind Health. 15:119–132.

Vonesh JR, De la Cruz O. 2002. Complex life cycles and density dependence: assessing the contribution of egg mortality to amphibian declines. Oecologia. 133:325–333.

Vonesh JR, Kraus JM. 2009. Pesticide alters habitat selection and aquatic community composition. Oecologia. 160:379–385.

Walker BH. 1992. Biodiversity and ecological redundancy. Conserv Biol. 6:18–23.

Walthall WK, Stark JD. 1997. Comparison of two population-level ecotoxicological endpoints: the intrinsic (r(m)) and instantaneous (r(i)) rates of increase. Environ Toxicol Chem. 16:1068–1073.

Watanabe KH, Li Z, Kroll KJ, Villeneuve DL, Garcia-Reyero N, Orlando EF, Sepulveda MS, Collette TW, Ekman DR, Ankley GT, et al. 2009. A computational model of the hypothalamic-pituitary-gonadal axis in male fathead minnows exposed to 17 alpha-ethinylestradiol and 17 beta-estradiol. Toxicol Sci. 109:180–192.

Wendt-Rasch L, Poulsen V, Duquesne S. 2014. In response: regulatory risk assessment and landscape ecotoxicology – a governmental perspective. Environ Toxicol Chem. 33:1196–1197.

Widdows J, Donkin P, Brinsley MD, Evans SV, Salkeld PN, Franklin A, Law RJ, Waldock MJ. 1995. Scope for growth and contaminant levels in North-Sea mussels Mytilus-Edulis. Mar Ecol Prog Ser. 127:131–148.

Wiegers JK, Feder HM, Mortensen LS, Shaw DG, Wilson VJ, Landis WG. 1998. A regional multiple-stressor rank-based ecological risk assessment for the fjord of Port Valdez, Alaska. Hum Ecol Risk Assess. 4:1125–1173.

Wintermantel TM, Campbell RE, Porteous R, Bock D, Grone HJ, Todman MG, Korach KS, Greiner E, Perez CA, Schutz G, et al. 2006. Definition of estrogen receptor pathway critical for estrogen positive feedback to gonadotropin-releasing hormone neurons and fertility. Neuron. 52:271–280.