PREDICTING COMBINED EFFECTS OF CHEMICAL STRESSORS: POPULATION-LEVEL EFFECTS OF ORGANIC CHEMICAL MIXTURES WITH DEB-IBM

ABSTRACT

Most regulatory ecological risk assessment frameworks largely disregard discrepancies between the laboratory, where effects of single substances are assessed on individual organisms, and the real environment, where organisms live together in populations and are often exposed to multiple simultaneously occurring substances. We assessed the capability of individual-based models with a foundation in the dynamic energy budget theory (DEB-IBM) to predict combined effects of chemical mixtures on populations when they are calibrated on toxicity data of single substances at the individual level only. We calibrated a DEB-IBM for *Daphnia magna* for four compounds (pyrene, dicofol, alfa-hexachlorocyclohexane, and endosulfan), covering different physiological modes of actions. We then performed a 17-week population experiment with *D. magna* (designed using DEB-IBM), in which we tested mixture combinations of these chemicals at relevant concentrations, in a constant exposure phase (7 weeks exposure and recovery), followed by a pulsed exposure phase (3-day pulse exposure and recovery). The DEB-IBM was validated by comparing blind predictions of mixture toxicity effects with the population data. The DEB-IBM accurately predicted mixture toxicity effects on population abundance in both phases when assuming independent action at the effect mechanism level. The population recovery after the constant exposure was well-predicted, but recovery after the pulse was not. The latter could be related to insufficient consideration of stochasticity in experimental design, model implementation, or both. Importantly, the mechanistic DEB-IBM performed better than conventional statistical mixture assessment methods. We conclude that DEB-IBM, calibrated using only single-substance individual-level toxicity data, produces accurate predictions of population-level mixture effects and can therefore provide meaningful contributions to ecological risk assessment of environmentally realistic mixture exposure scenarios.

Keywords

Mixture toxicity

Daphnia magna

Dynamic Energy Budget (DEB)

Individual Based Model (IBM)

Effect assessment

Ecological modelling

Mechanistic effect modelling

Population modelling

1 INTRODUCTION

The current risk assessment frameworks lead to discrepancies between protection goals stated in regulations and the information we gather for the risk assessment procedure (Galic et al., 2010). Ecological risk assessment (ERA) of chemicals nowadays is very well-standardized, with documented protocols, based on ecotoxicological tests with individual organisms exposed to pure substances. Additionally, the assessment is based on a set of well-defined apical endpoints, such as individual growth, reproduction, and mortality (Preuss et al., 2009). In contrast, protection goals in ERA regulations mainly target higher levels of organisation, such as the population or community level (Hommen et al., 2010). Results from individual-level toxicity tests are thus extrapolated to account for the population or community in the environment, using for instance assessment factors (AFs) to deal with the uncertainty on the extrapolation (Van Leeuwen and Vermeire 2007; ECHA 2017). More recently, to account for potential mixture effects, the implementation of a mixture assessment factor (MAF) has been suggested by the European Commission (EC 2020). However, the scientific validity of AFs and the MAF are a topic of strong debate in the scientific community (Chapman et al., 1998; Forbes et al., 2008; Forbes et al., 2009; van Boekhuizen et al., 2017; Kortenkamp and Faust 2018). The largely arbitrary nature of these AFs does not reduce the uncertainty about potential effects of chemical mixtures in a population context.

One of the potential issues are synergistic effects of mixtures, where the toxicity of the mixture is larger than expected based on the toxicities of the individual chemicals (when compared to a reference mixture model such as the independent action [IA] model or the concentration addition [CA] model) (Cedergreen 2014). Synergisms can happen when chemicals in a mixture interact and enhance each other's toxicity. This raises questions on what the significance of these effects is, how common they are, and do they pose a threat to the health of the

environment? Current mixture toxicity studies are generally performed at the individual level on apical endpoints (SCHER 2012). This still leaves uncertainty related to mixture effects on higher levels of biological organisation.

DEB-TKTD (Dynamic Energy Budget Toxicokinetic-Toxicodynamic models, i.e., updated from DEBtox, an extension of the Dynamic Energy Budget (DEB) theory; Sherborne et al., 2020) can offer mechanistic explanations on how a chemical stressor affects energy allocation within an organism (Jager 2017). Within individual-based model (IBM) frameworks, DEB can mechanistically predict effects of stressors on populations (Vlaeminck et al., 2019), and extensions to account for mixture toxicity have been proposed (Jager et al., 2010; Vlaeminck et al. 2021). In DEB-TKTD, a chemical stressor can interfere with one of four metabolic energy flows of an organism, the so-called Physiological Modes of Action (PMoAs): assimilation (1), costs of growth (2), reproduction efficiency (3) or maintenance costs (4) (Supplemental Information, Appendix G – Figure G1). Affecting one of these energy flows will cause sub-lethal effects to emerge at the individual level, such as a reduced growth rate or reproductive output. Modelling challenges arise when multiple stressors are present since multiple energy flows can be affected simultaneously. Because the energy flows are linked within DEB (Figure G1), multiple disruptions can interfere with each other, leading to stronger (or weaker) effects at the apical (or higher) level.

In a population context, exposure to a mixture is even harder to predict as we need to consider population-level mechanisms. Changes in population dynamics are usually caused by a change of either the input of new individuals (due to reproduction or immigration) or the output (loss) of individuals (i.e., death or emigration). The population can be considered a closed system with a specific energy balance: energy input due to feeding is translated to biomass (growth of individuals), reproduction (creation of new individuals), and energy loss (due to respiration,

faeces). Consequently, any imbalance in the energy allocation rates caused by a chemical stressor can affect population dynamics. In addition, population-related feedback mechanisms (e.g., relaxation of intra-specific competition for food; De Roos et al., 2013) can compensate some of the chemical's effects, but it is not straightforward to predict the ultimate outcome, unless with mechanistic population models (Vlaeminck et al., 2019; Pereira et al., 2019).

The standard *D. magna* reproduction test considers cumulative reproduction per female after 21 days as the relevant endpoint for effect assessment (OECD 2012). The derived EC50 represents the effective value where 50% decrease in reproduction is observed at the end of the 21-day exposure period. However, in DEB-TKTD a 50% decrease in reproductive output can be obtained by all four PMoAs, all of which directly or indirectly decrease energy allocation to reproduction (Jager 2017). An example of this is shown with a simulation (Figure 1), in which a 50% decrease in reproduction after 21 days in a standard test is predicted with a generic implementation of DEB-IBM for *D. magna* with each PMoA. However, each PMoA affects DEB energy flows differently, as some can affect ultimate size or individual growth (Figure 1). Moreover, the outcome at the population level will highly depend on the applicable PMoA. At the population level, a 50% decrease on individual-level reproduction predicted by the PMoA of increase in growth costs does not lead to the same result as a 50% effect by the PMoA of decreased reproduction efficiency. With DEB-IBM, the PMoA related to reproduction predicts a lower total number of individuals but only slight effects on total biomass. An increase in the cost of growth on the other hand has slightly positive effects on total abundance, while no effects on total biomass are predicted. In contrast, the other two PMoAs (decrease in assimilation and increase in maintenance costs) predict both a decrease in total abundance and total biomass. Details on how these simulations were performed can be found in the Supplemental Information (Appendix A).

The situation becomes even more complicated when two (or more) PMoAs are affected simultaneously. In addition, chemicals can cause lethal effects as well. As shown in Figure 1, the predicted effect at the population level is not always what is expected and will highly depend on the PMoA of the substance. Hansul et al. (2021) have shown that combined mixture effects (synergistic and antagonistic) at the population level can emerge from DEB-IBM, but depend on mixture model selection (i.e., independent action or concentration addition), the inclusion of intraspecific variation, and the PMoA that is selected for a substance to predict toxicity. The question is then: in a multiple chemical stressor environment, what patterns/effects are observed on populations exposed to mixtures? And can we predict those patterns/effects with DEB-IBM?

The general goal of our study was to assess if DEB-IBM can predict combined effects of binary and ternary mixtures on lab-exposed populations, based on data assessed at the individual level with individual substances only. Using the theoretical DEB-TKTD framework as foundation, we started from an IBM implementation of DEB to predict effects of binary combinations of PMoAs to *Daphnia magna* populations. We hypothesized that population-level effects of mixtures emerge from the linking of energy flows within DEB. We performed a population experiment to test mixture toxicity effects in a more realistic setting than what is currently assessed in ERA, with DEB-IBM as a guide for our experimental design (i.e., selecting mixture combinations based on PMoAs and setting relevant exposure concentrations). In a population experiment, laboratory *D. magna* populations (i.e., closed system without migration) were exposed to binary and ternary combinations of compounds selected based on their inferred PMoA (4 compounds were tested: pyrene, dicofol, α-hexachlorocyclohexane, and endosulfan). We wanted to assess specific combinations of different PMoAs. The patterns predicted by the DEB-IBM were tested against the population data. Based on the model validation with the population data, predictive performance of the DEB-IBM was assessed.

2 MATERIAL AND METHODS

2.1 Population experiment

2.1.1 Experimental design

In the population experiment, *D. magna* populations were exposed to mixtures of chemical compounds. The total duration of the experiment was 17 weeks and consisted of two main phases: a constant exposure period (1 week of set-up, 7 weeks exposure and 3 weeks recovery) and a pulsed exposure period (3-day pulsed exposure, followed by 5 weeks and 4 days of non-exposure) (Figure 2). In the first phase, the concentration of the substances remained constant (i.e., the constant phase). Exposure was initiated after one week and lasted 7 weeks in total. Thereafter, 3 weeks of recovery was imposed. In the second phase, a short 3-day pulsed exposure was applied (i.e., the acute pulse phase). After that, no exposure was applied to study the population recovery.

Four compounds were used for the experiment: pyrene, dicofol, alfa-hexachlorocyclohexane (a-HCH), and endosulfan. These four compounds were selected based on their inferred DEB-TKTD PMoA (details in Appendix B). Combinations of the compounds were set based on predicted DEB-IBM patterns with different PMoA combinations (see details on model-based design of the experiment in Appendix B). Interesting combinations of PMoAs were selected for the population experiment. Using the same model structure, the concentrations of the substances were selected based on an expected decrease in equilibrium population abundance (see Appendix B). For the constant exposure period, the concentration of each substance was selected based on an expected 40% decrease in equilibrium population abundance (single chemical exposure only), except for a-HCH where lower concentrations were used due to its low solubility and low 21d-EC10 values on individual growth (see Supplemental Information – Appendices B and H, Table H1). For the pulsed exposure period, the concentrations were

selected based on an expected 40% decrease in abundance after a 3-day pulse (single exposure only), except for a-HCH where lower concentrations were again used.

A total of 11 treatments were considered in the population experiment. Two control treatments were selected: a control treatment without exposure, and a solvent control treatment. The solvent-exposed treatment was exposed to similar solvent levels as the binary mixture treatments (ethanol + acetone ≤ 0.01%). A single substance treatment for each of the four compounds was considered. Four binary mixture treatments were applied to encompass different combinations of PMoAs: pyrene – dicofol (R-M), pyrene – HCH (R-A), dicofol – HCH (M-A), and HCH – endosulfan (A-A). One ternary mixture was assessed, combing pyrene, dicofol and HCH (R-M-A). All treatments were triplicated (total of 30 aquaria). Detailed explanation on concentration and combination selection can be found the Supplemental Information (Appendix B).

The initial population abundance at day 0 was the same in each of the treatments' aquaria: 10 juveniles (< 24h of age) and 5 adults (3 weeks of age). Population dynamics over time were assessed following an image analysis protocol (see Supplemental Information). Based on a series of photographs, the number of individuals and their size (based on pixel density) was calculated. Photographs were taken twice per week during medium renewal. Analysis of the photos gives information on the population abundance over time (distinction between total, juvenile and adult population), the size distribution, and the total population biomass (estimated from the number of individuals and their size).

For the medium, a modified M4 medium was used (for details see Pereira et al. 2019). Medium was renewed twice per week. During the constant exposure period, the organic compounds were spiked again with every renewal of medium (to keep the concentration as constant as

possible). The organic compounds were stored in a fridge (4°C) away from light and dissolved in their respective carrier solvent (dicofol in ethanol; pyrene, endosulfan, and a-HCH in acetone). The total solvent concentration in each aquarium did not exceed 0.01% (= 100 µL/L) (as advised by OECD 2019). The temperature was recorded daily and remained constant throughout the experiment (20.2 °C, SD = 1.4 °C, n = 403). Food was added daily with a density of 2.5 mg C per L of *Raphidocelis subcapitata* (formerly known as *Pseudokirchneriella subcapitata*). Dissolved organic carbon (DOC) of the medium was sampled at the start of every week (4.4 mg/L, SD = 1.4 mg/L, n = 34, analysed with a TOC-L CPH; Shimadzu). Samples of the old and freshly renewed medium were taken with every medium renewal (i.e., twice a week). The concentrations of the organic compounds were measured using gas-chromatography mass-spectrometry (GC-MS; DSQ, Thermo Finnigan; limits of quantification [LOQ]: 15 µg/L for pyrene, 16 µg/L for dicofol, 36 µg/L for α -HCH, and 40 µg/L for endosulfan [I+II]; limits of detection [LOD]: 5 µg/L for pyrene, 5 µg/L for dicofol, 12 µg/L for α-HCH, and 11 µg/L for endosulfan). The pH of the old and freshly renewed medium was also recorded in parallel with medium renewal (8.3, SD = 0.2, n = 408).

2.1.2 Statistical analysis

Statistical analysis was performed for various population-level endpoints to test for: significant differences between treatments and significant mixture toxicity interactions. The data was pooled in two phases: the constant phase (days 20 to 56) and the pulse phase (days 83 to 87). Three population-level endpoints were investigated: population abundance (number of individuals), population biomass, and average length. All statistical analyses were performed in R-Studio (RStudio Team 2020).

First, data normality was checked with the Shapiro-Wilk test (R-function: *shapiro.test*). For all endpoints, at least one treatment showed deviations from normality (see Appendix F, Table

F1). Hence, a non-parametric test, the Mann-Whitney U-test (i.e., Wilcoxon rank sum test), was performed to determine significant deviations between exposed treatments and the solvent control treatment (R-function: *wilcox.test*).

Second, an ANOVA analysis was performed to test for significant mixture toxicity interactions and deviations from additivity. A log-transformation was performed on the data, which allows to identify significant interactive effects relative to the independent action model with an ANOVA analysis (De Coninck et al., 2013). Performing an ANOVA analysis on log-transformed values mathematically imposes the independent action null model, which is also known as the multiplicative null model in ecological research (see derivation in the Supplemental Information of De Coninck et al., 2013; Faust et al., 2003; Schäfer and Piggot, 2018). The ANOVA analysis was performed in R, with 'day' as a categorical variable, and the 'nominal concentration' of each substance as continuous variable. Normality and heteroscedasticity were checked on the residuals, using QQ-plots and visualizing fitted values vs. the residuals (see Appendix F, file F3). For two treatments only, pyrene-HCH and endosulfan-HCH, the assumptions were violated for total population biomass and average length during the first phase and should be interpreted with caution. When the ANOVA revealed a statistically significant interaction, the interaction is classified as synergistic if a stronger effect is found than the effect calculated with independent action (IA; Equation 1). If a smaller effect is found, then the effect is classified as antagonistic.

IA was used to estimate the mixture effect at the population level based on single-substance population data (Equation 1). IA is commonly used for both lethal (quantal) and sublethal (continuous) data (SCHER, 2012; Nys et al., 2018), although the underlying assumptions of this multiplicative null-model are well-defined only for lethal data (Schäfer and Piggot, 2018). The IA reference model for mixture effects (Equation 1) estimates the percentage decrease of

a population endpoint under combined exposure to A and B (E_{mix} [%]) from the percentage effects caused by compound A (E_A [%]) and compound B (E_B [%]) (Bliss, 1939):

$$E_{mix,IA} = E_A + E_B - E_A * E_B$$
$$= 1 - (1 - E_A) * (1 - E_B) \quad (1)$$
$$\text{where } E_i = \frac{Y_{control} - Y_i}{Y_{control}}$$

Equation 1 can be rearranged and solved for the population abundance under exposure to the mixture under IA ($Y_{mix,IA}$ [#]). This gives Equation 2 where the population abundance under IA under exposure to a mixture of compounds A and B can be estimated based on observed abundance when exposed to compound A (Y_A [#]; single exposure) and compound B (Y_B [#]; single exposure) (see full deviation of Equation 2 in the Supplemental Information, Appendix D):

$$Y_{mix,IA} = \frac{Y_A * Y_B}{Y_{control}} \quad (2)$$

2.2 Simulating mixture toxicity with DEB-IBM

2.2.1 Model formulation

An individual-based model implementation based on the dynamic energy budget theory (DEB-IBM) was used to predict population-level of mixtures to *D. magna* populations. Details on the DEB-IBM model structure can be found in the TRACE document provided by Vlaeminck et al. (2021). The DEB model for *D. magna* is used as foundation. Toxicokinetics (TK) were implemented to describe the damage dynamics over time. Lethal and sub-lethal effects were implemented using an adapted version of GUTS-RED-SD and DEB-TKTD (Jager and Ashauer 2018; Sherborne et al., 2020). A schematic overview of the model structure is giving in the

Supplemental Information, Appendix G – Figure G2. Note that a more complete version of the DEB-IBM is reported here compared than what was used for the theoretical simulations in the introduction (Figure 1), and the experimental design (Section 2.1.1). The main differences are the inclusion of TK and a different shape for the relation between TK and effects (DEB-TD and GUTS-RED-SD) (see Supplemental Information, Appendix B). We refer to the two different models as: the initial model (i.e., used for experimental design; see Supplemental Information, Appendix B) and the complete model (used in the remainder of the study).

The first step in describing toxic effects of a compound is the toxicokinetics (TK). Here, we used a scaled one-compartment kinetic model which includes the effect of growth on the surface-area to volume ratio and dilution through growth (Equation 3). The change in scaled internal damage (D_w [µg/L]) will depend on the external (water) concentration (C_w [µg/L]), the size of the organism (length L [cm]), the ultimate size of the control (i.e., unexposed; L_m [cm]), and the dominant rate parameter (k_D [d^{-1}]). Note that no interactions between internal damages of chemicals were implemented at the TK level. The TK are determined as follows:

$$\frac{dD_w}{dt} = k_D \frac{L_m}{L}(C_w - D_w) - D_W \frac{3}{L}\frac{dL}{dt} \quad \text{with} \quad L_m = \frac{\dot{v}}{g * \dot{k_M}} \tag{3}$$

where the ultimate size of the control L_m [cm] is a function of different DEB parameters, including the energy conductance $\dot{v}$ [cm d^{-1}], the energy investment ratio g [-], and the specific somatic maintenance rate $\dot{k_M}$[d^{-1}].

The build-up of internal scaled damage leads to effects, i.e., the toxicodynamics (TD). Effects will either be lethal (i.e., mortality), sub-lethal (e.g., reduction in growth or reproduction), or both. To describe lethal effects, the General Unified Threshold model for Survival (GUTS) was used, whereas for sub-lethal effects DEB-TKTD was used. A difference with the approach used in Vlaeminck et al. (2021) (is the relationship used to relate the internal scaled damage

(D_w; Equation 3) to the effect mechanism (GUTS or DEB-TKTD). Commonly, a hockey stick relationship is used to relate the damage level to stress on the PMoA (in DEB-TKTD) or the hazard rate of dying (in GUTS) (Jager 2017; Jager and Ashauer 2018). Here, we used a log-logistic relationship to relate the scaled damage to stress on the PMoA (similarly as Vlaeminck et al., 2019) (see Equation 4). The value of the stress level can vary between 0 (no disruption of the DEB process) and 1 (complete disruption of the DEB process). Two parameters were considered: $ED50_{DEB}$ [µg/L], which is the half-maximal damage level (i.e., the damage level corresponding to 50% disruption at the PMoA level), and S_{DEB} [-], the slope of the log-logistic relationship between damage and stress:

$$s = \frac{1}{1 + \left(\frac{D_w}{ED50_{DEB}}\right)^{-S_{DEB}}} \quad (4)$$

Depending on the PMoA, different DEB parameters are affected and will either increase or decreases as the stress (Equation 4) increases. In DEB-TKTD, generally four main DEB-PMoAs are considered: decrease in energy assimilation (1), increase in maintenance costs (2), increase in cost of growth (3), decrease in reproduction efficiency (4) (see Supplemental Information, Appendix H – Table H2).

For lethal effects, a stochastic death (SD) approach was followed based on the reduced GUTS with stochastic death (GUTS-RED-SD). The GUTS-RED-SD model first calculates the scaled internal damage, similar as in Equation 3. Standard GUTS-RED does not use length to calculate the change in the internal damage (i.e., growth dilution and change in surface area to volume ratio). However, extensions including length (such as Equation 3) have been formulated for GUTS (Jager and Ashauer, 2018) and are commonly used in DEBtktd (Sherborne et al., 2020). Based on the internal scaled damage, the hazard rate is calculated. The hazard rate increases

linearly with the internal scaled damage, but only increases when damage is above a certain threshold (i.e., a hockey stick relationship). The hazard rate can be interpreted as the "instantaneous probability to die" (Jager and Ashauer 2018) within an infinitesimally small timestep. The survival probability of an individual is calculated by integrating its hazard rate over the exposure duration. Here we simplified the GUTS-RED-SD approach as follows: we directly related the probability to die within a timestep (= 1 hour) to the scaled internal damage of an individual without explicitly calculating a hazard rate. This stochastic probability of dying is calculated hourly and for each individual. A randomly generated number (between 0 and 1) then determines whether an individual dies or not. This still follows the principles of stochastic death as in the GUTS-RED-SD but within a reduced mathematical framework.

Similar as for the stress on the PMoA, a log-logistic shape was used to relate scaled internal damage to the probability of dying (p_m [-]) (Equation 5). This probability varies between 0 (no chance of dying in the simulated hourly timestep) and 1 (100% chance of dying the simulated timestep). A higher scaled internal damage leads to a higher probability of dying. This implementation is an adapted form of reduced GUTS with stochastic death (GUTS-RED-SD). Note that the same internal scaled damage, calculated with Equation 3, is used in Equations 4 and 5. The assumption is that damage between DEB-TD and SD is shared, meaning that a single damage pool leads to both lethal and sub-lethal effects. The probability of dying depends on two parameters: $ED50_{GUTS}$ [µg/L], the half maximal damage level (i.e., the damage level leading to 50% mortality within a given timestep), and S_{GUTS} [-], the slope of the log-logistic relationship:

$$p_m = \frac{1}{1 + \left(\frac{D_w}{ED50_{GUTS}}\right)^{-S_{GUTS}}} \quad (5)$$

2.2.2 Mixture toxicity in DEB-IBM

To predict mixture toxicity effects, two approaches at the damage level were considered as discussed by Vlaeminck et al. (2021): the independent action approach (IA) and the damage addition (DA). IA assumes that two compounds in a mixture exert an effect that is independent of the other compound. Both compounds will thus have their external concentration and their own internal scaled damage, and each will affect a PMoA and have an individual hazard rate, completely independent of the other compound. When two compounds have the same PMoA, but are assumed to act independently, the multiplicative mixture model is applied on the non-effects on the PMoA (as in Equation 1; see Table 1). For instance, for GUTS, this means the predicted survival probabilities are multiplied (i.e., survival multiplication; Equation 6). The probability of dying due to exposure to the mixture ($p_{m,mix}$ [-]) is calculated by multiplying the independent probabilities of dying due to compound A ($p_{m,A}$ [-]) and compound B ($p_{s,B}$ [-]). Equation 6 can also be expressed in terms of survival probabilities (p_s [-]), leading to the formula for survival multiplication as defined by Jager and Ashauer (2018):

$$\begin{aligned} p_{m,mix} &= p_{m,A} + p_{m,B} - p_{m,A} * p_{m,B} \\ &= 1 - (1 - p_{m,A}) * (1 - p_{m,B}) \end{aligned} \quad (6)$$

Or in terms of survival probabilities:

$$p_{s,mix} = p_{s,A} * p_{s,B} \quad \text{where} \quad p_s = 1 - p_m$$

For DEB-TD, IA means the relative effects (*RE*) on the affected parameters can be multiplied. For instance, for the PMoA related to reproduction, the reproduction efficiency κ_R [-] decreases with increasing stress. Each compounds causes its own stress on the PMoA. The effect on reproduction efficiency is then calculated according to Table 1 (left column).

DA assumes that both compounds in a mixture cause the same form of damage, indicating they share their damage pool. Consequently, they affect the same PMoA and lead to an integrated probability of dying. This implies that the form of damage caused by one compound is interchangeable with an equally effective form of damage caused by the other compound (same assumption as the concentration addition [CA] model that is generally used to assess mixture toxicity effects on the apical level). The effect caused by the mixture can be calculated by solving Equation 7. The combined effect of the mixture x [%], is related to the sum of the weighted damages of the mixture constituents (similar as CA; see Nys et al., 2018). The weighted damage of a compound is found by scaling the damage of a compound (e.g., $D_{w,A}$ [µg/L]) to its x% effective damage (i.e., the damage of compound A causing x% effect – $ED_{x,A}$ [µg/L]). For any mixture of compounds under DA, Equation 7 needs to hold. In DEB-IBM, Equation 7 is solved iteratively for every individual at each timestep. The mixture effect x [%] under DA is calculated based on the damage levels of each compound ($D_{w,A}$ [µg/L] and $D_{w,B}$ [µg/L]) and the corresponding effective damage at the x% level ($ED_{x,A}$ [µg/L] and $ED_{x,B}$ [µg/L]):

$$\frac{D_{w,A}}{ED_{x,A}} + \frac{D_{w,A}}{ED_{x,A}} = 1 \qquad (7)$$

2.2.3 Calibration of DEB-TKTD and GUTS-RED-SD parameters

We calibrated the complete model at the individual level using individual-level data for each of the selected compounds (without using mixture toxicity data at the individual level). Data on growth, reproduction and survival of *D. magna* was used to calibrate lethal and sub-lethal effects of each compound (data available in Appendix E). Calibration was performed using sequential Approximate Bayesian Computation (seqABC; Toni et al., 2009; van der Vaart et al., 2016). For each parameter, minimum and maximum boundaries were set (Supplemental

Information, Appendix H – Table H3). A total of 10 000 simulation iterations were performed with randomized values of each parameter drawn from a uniform distribution based on the boundaries (Table H3), i.e., the *prior* distribution. Based on the maximum log-likelihood (Equation 8) between data and model prediction, the 100 best iterations were selected, for which the *posterior* distribution was constructed. The *posterior* distribution was then set as the *prior*, repeating the process for a total of 5 iterations. The final *posterior* distributions were reported for each compound (median value and the 95% credibility intervals).

For calibration, the overall log-likelihood $\mathcal{L}$ of a parameter set θ given dataset Y is given by the sum of the likelihoods of each endpoint (growth, survival and reproduction; Equation 8). The likelihood of a specific endpoint $\ell(\theta|Y)$ is based on the sum of squares of the residuals (SSR) between data y_i and the average model prediction $\hat{y}_i$ on timepoint *i* for *n* datapoints. Plots of the *prior* and final *posterior* distributions of the likelihoods of each parameter can be found in the Supplemental Information (Appendix G, Figures G18 to G21).

$$\mathcal{L}(\theta|Y) = \ell_{growth} + \ell_{reproduction} + \ell_{survival}$$

$$\text{with} \quad \ell(\theta|Y) = -\frac{n}{2} * \ln(SSR) \quad \text{and} \quad SSR = \sum_{i=0}^{n} (y_i - \hat{y}_i)^2 \tag{8}$$

2.2.4 Additional correction at the population level

On top of the implementation of lethal and sub-lethal effects, the temperature correction, and a recalibration of the maximum surface-specific assimilation rate (methods in Pereira et al., 2019; see TRACE documentation in Vlaeminck et al., 2021), an additional correction was implemented for the control prediction to match the observations in the solvent control treatment. The DEB-IBM control prediction matched the observed peak abundance solvent control, but there was a mismatch later in the exposure. A gradual decrease in abundance was

observed over time (results shown further in paragraph 3.1), which was not reflected by the DEB-IBM predictions. We suspect there is a crowding-related process missing in the implementation, or potential changes in competition and/or starvation strategy, which induce a gradual decrease in population abundance. To compensate for this gradual decrease, we corrected the functional response (f [-]) in the DEB-IBM simulation. Every individual's functional response was multiplied with a correction factor *corr* [-].

A gradual, linear decrease in the functional response was implemented (see Supplemental Information, Appendix G – Figure G11). We assumed that the functional response is maximal up until peak abundance (i.e., on day 20). Hence, from day 0 to 20 the correction factor has a value of 1. We estimated a ratio of about 0.41 between the (uncorrected) predicted abundance and observed abundance in the solvent treatment at the end of the experiment (day 119) (see Figure G3). Therefore, the correction factor at the end of the experiment has a value of 0.41. We used linear interpolation to determine the value for the correction factor between day 20 (at peak abundance) and day 119 (end of the experiment) (see Figure G11).

2.2.5 Validation of the DEB-IBM

During DEB-IBM simulations, the conditions of the population experiment were mimicked to validate the observed population dynamics. The same exposure concentrations, initial population abundance, volume, food addition, temperature, etc., were implied as in the experiment. For the exposure concentrations, linear interpolation was used to interpolate between the concentration measurement after and before renewal (new and old medium). For the binary mixtures, the IA model was applied for both DEB-TD and GUTS to predict the mixture toxicity. For the a-HCH and endosulfan mixture, we also used the DA model to predict mixture effects. This was the only mixture where the inferred PMoAs were the same (a

decrease in energy assimilation). In this case, DA was applied for the DEB-TD and GUTS sub-models.

Population dynamics were visually validated by plotting the observed and predicted abundances over time against each other. In addition, the total biomass and the average length were also investigated. The relative trends in abundances were investigated as well by dividing the predictions and observations by their relative control. Quantitative validation was performed using the normalised root-mean-square error (NRMSE) between log-transformed data and predictions (Equation 9). The NRMSE was calculated by taking the root of the mean-square error (MSE) and normalizing it by the average (log-transformed) observed value ($\bar{Y}$). The MSE was calculated as the square error between (log-transformed) data ($y_{i,obs}$) and prediction ($y_{i,pred}$), averaged by the number of data points (n). The NRMSE was calculated for each treatment separately and we split it between the constant phase and the pulse phase:

$$NRMSE = \frac{RMSE}{\bar{Y}} = \frac{1}{\bar{Y}}\sqrt{MSE} \quad with \quad MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{i,obs} - y_{i,pred}\right)^2 \qquad (9)$$

3 RESULTS

3.1 Experimental results

The population experiment showed considerable effects of the organic compounds (single doses) to the *D. magna* population. Pyrene, dicofol and endosulfan induced clear effects in both the constant exposure phase and the pulse phase (Supplemental Information, Appendix G – Figure G3). Pyrene, dicofol and endosulfan caused a decrease in average abundance during the constant phase (days 20 to 56) of 43%, 61% and 43%, respectively (Table 4). In contrast, the trends in the a-HCH treatment followed more closely the solvent control treatment (Figure G3), with no significant decrease in abundance (Mann-Whitney U-test: $p = 0.296$; Supplemental Information, Appendix F – Table F2). After the 3-day pulse exposure (day 77 to 80), a clear decrease in abundance was observed in the pyrene, dicofol and endosulfan treatments, at 87%, 89% and 50%, respectively (Mann-Whitney U-test: $p < 0.001$, $p < 0.001$, $p < 0.001$; Table 4; Table F2). For a-HCH, there was no significant effect on population abundance after the pulse (Mann-Whitney U-test: $p = 0.59$; Table C2). Total population biomass and average length were not impacted significantly by any of the single substances during the constant phase (Table F2). During the pulse phase, both a-HCH and endosulfan significantly decreased the total biomass and average length (Table F2).

For all binary mixture treatments, clear effects were observed on total population abundance (Figure G3). During the constant exposure, a significant decrease in abundance was observed for all binary mixtures (ranging from 70% to 82%; Table 4). Recovery was observed in all binary treatments (Figure G3). An overshoot in population abundance was observed during the 3-week recovery after the constant phase. After the pulse, all treatments showed significant reductions in total abundance (Table F2). For the pyrene-dicofol and dicofol-HCH treatments, 2 of the 3 replicates became extinct, while one recovered. For the pyrene-HCH treatment, all

replicates became extinct. For endosulfan-HCH, all replicate populations persisted following exposure. Total population biomass was significantly impacted in the pyrene-HCH treatment (Table F2). In the HCH-endosulfan treatment, significant reduction in biomass and average size were observed in both phases (Table F2). None of the exposed populations in the ternary mixture treatments managed to persist following the constant phase period (Figure G3). Significant effects were observed on all tested endpoints (Figure 3 and Table F2).

The ANOVA analysis revealed significant interaction effects in the mixture treatments for several endpoints. A schematic overview of the ANOVA results is shown in Figure 3. Detailed results can be found in the Supplemental Information (Appendix F – Tables F3 to F9). In the constant exposure phase, significant deviations from additivity (IA) were observed for total population abundance for all mixtures. For some treatments, deviations from IA were also found on total population biomass and average size. According to the Mann-Whitney U-test results (indicated in red in Figure 3), total population abundance deviated significantly from the solvent control in all treatments. Based on comparison with IA calculations (see Table 4 for total population abundance; and Tables F3 and F4 for total population biomass and average length), the interactive effects were exclusively synergistic for total population abundance in the constant phase. For the HCH-endosulfan treatment and the ternary mixture, all endpoints showed synergistic interaction effects in the ANOVA. No interactive effects of the mixtures were found after the acute pulse phase, even though significant deviations with the solvent control were found based on the Mann-Whitney U-test (Figure 3).

3.2 Modelling results

3.2.1 Calibration results

The TK, DEB-TD and GUTS-RED-SD parameters calibrated for each of the four investigated compounds can be found in Table 2. Data of the individual-level effects can be downloaded

from Figshare (see Supplemental Information, Appendix I). Results of the individual-level calibration can be seen in the Supplemental Information (Appendix G, Figures G7 to G10). The established PMoAs were decreased reproduction efficiency for pyrene, increased maintenance costs for dicofol, and decreased energy assimilation for endosulfan and a-HCH.

3.2.2 Validation results

Independent population simulations of mixture toxicity with the calibrated parameters were performed with the DEB-IBM (see Supplemental Information, Appendix G – Figure G12). Based on the plots and the NRMSE (Table 3), a good correspondence between data and predictions is seen. During the constant exposure phase, the predicted abundance reached a stable equilibrium, which matches the observations. The predicted and observed equilibria of all but one treatment are situated at lower population abundances than those of the control. Only in the HCH treatment, no decrease of equilibrium abundance during the constant phase was observed nor predicted. After the constant exposure phase, an increase in abundance was observed and predicted, except in the HCH treatment. The pulse exposure induced strong effects in each of the exposed populations, and this strong effect was also predicted. After the pulse exposure, the model predictions deviate from the observations for some treatments. Deviation is apparent for dicofol, where recovery after the pulse is predicted to be slower than what was observed. For the pyrene-dicofol mixture treatment, one of three replicates was observed to persist and recover, whereas the model predicts population extinction. For pyrene-HCH, the opposite is seen, i.e., the model predicts persistence while extinction was observed in all replicates.

Model accuracy is expressed as the NRMSE (Table 3). For all treatments excluding the ternary mixture treatment, the NRMSE was lower than 0.5 in the constant phase of the experiment (Table 6). Generally, a value of 0.5 or lower for the NRMSE indicates a reasonable match

between the model and prediction (e.g., EFSA, 2018 and Focks et al., 2018). For the ternary mixture, population extinction was predicted, which was observed in the data as well. For the pulsed exposure phase, the NRMSE were generally higher than for the constant exposure phase. For the single substances, only for pyrene and dicofol an NRMSE larger than 0.5 was found in the pulsed phase. For the binary mixture treatments, NRMSE values were all higher than 0.5, except for the endosulfan-HCH treatment with IA.

The relative trends in total population abundance over time were observed (Figure 4). The DEB-IBM generally captured the observed absolute and relative trends in effects over time on population abundance for the single substance exposed treatments during the constant phase of the experiment. Specifically for pyrene, faster recovery was predicted after the pulse than what was observed. For dicofol the opposite was seen, where faster recovery was observed than what was predicted. For a-HCH, the non-effects during the constant phase were generally captured, with slight effects after the pulsed exposure. In the endosulfan treatment, the model captured the observed oscillations in population abundance.

The mixture effects were well predicted by the model using DEB-IBM with IA, at least for the constant exposure phase of the experiment (Figures 3 and 4). After the pulsed exposure, the DEB-IBM predicted differences in recovery compared to what is observed in the experiment. For pyrene-dicofol, the predicted recovery was much slower than what was observed. This delay in recovery was also visible in the single dicofol treatment, where faster recovery was observed than predicted. For pyrene-HCH on the other hand, the model predicted recovery whereas in the data, extinction was observed after the pulse. For dicofol-HCH, recovery after the pulse was observed, and the prediction was in line with one of the 3 replicates that recovered in the experiment (Figure 4). For the ternary mixture, extinction was predicted at the end of the constant exposed phase, which was observed in the experiment as well. Similar plots as Figure

4 were made for relative total biomass and relative average size over time (Supplemental Information, Appendix G – Figures G14 and G15).

Based on the general (relative) patterns from the validation (Figure 4) and the derived NRMSE (Table 3), the model performed best in predicting effects during the constant exposure phase, the recovery phase, and the acute effects after the pulse. During the recovery phase, the model generally captured the decrease in average abundance that was observed. The predicted delay in recovery after the 7-week constant exposure phase matched the observations quite well. The decrease in abundance after the 3-day pulsed exposure corresponded as well. However, the model seemed to perform considerably poorer in predicting the recovery after the acute phase. In some replicates of some treatments, recovery was predicted where extinction was observed, e.g., in the pyrene-HCH treatment. In others, recovery was observed where extinction was predicted, e.g., in the pyrene-dicofol treatment.

3.2.3 Mixture model comparison for HCH-endosulfan

In Figure 5, predictions are shown for the two different combinations of mixture toxicity approaches: IA (left) and DA (right). Based on the plot (Figure 5) and the NRMSE (Table 3), mechanistic IA leaned closest to the observations of the HCH-endosulfan treatment. Especially in the constant phase, there was a good correspondence between data and the predictions, with the lowest NRMSE of all mixture toxicity modes during this phase (NRMSE = 0.187; Table 3). IA also had the lowest NRMSE of all modes during the pulse phase (NRMSE = 0.498). Applying DA on both GUTS and DEB-TD predicted stronger effects compared to full IA. The recovery after the constant exposure phase was completely missing when DA is applied. With DA, the NRMSE in both phases were higher than with IA, indicating worse correspondence (i.e., NRMSE of 0.302 and 0.514 for the constant and pulse phase respectively).

3.2.4 Comparison between observations, DEB-IBM predictions, and statistical IA calculations

Table 4 shows the observed, DEB-IBM-predicted, and IA-estimated (mixtures only; see Section 2.1.2 for statistical IA calculation) percentage decrease of total population abundance. For the single exposed treatments, DEB-IBM performed well in predicting the percentage effects. For dicofol, there was a slight underestimation of the percentage effect during the constant phase (60% observed vs. 45% predicted with DEB-IBM), but overestimation after the pulse phase (89% observed vs. 100% predicted). For endosulfan, there was an overestimation during both phases: 42% observed vs. 61% predicted in the constant phase, and 50% observed vs. 90% predicted.

For the mixture treatments, the predicted decrease in abundance matched well with the observations, and this in both phases: the deviation is less than 11% (range: 1 – 11%) during the constant phase, and less than 4% (range: 1.4 – 3.5%) in the pulse phase. In contrast, the statistical IA approach considerably underestimated the effects observed in the experiment for the mixture treatments: the offset reached up to 33% (range: 5.4 – 33%) in the constant phase, and up to 30% (range: 0.1 – 30%) during the pulse phase. For all mixtures the DEB-IBM performed better than the statistical IA calculations, except for the pyrene-dicofol treatment (11% offset with DEB-IBM, and only 5% with statistical IA). Statistical IA generally underestimated the mixture effects observed in the experiment. Finally, the DEB-IBM predicted the observed extinction in the ternary mixture, whereas the statistical IA did not. Important to note is that statistical IA estimates the effect with population-level data, whereas DEB-IBM uses individual-level data from standard tests.

4 DISCUSSION

Overall, our work resulted in four key findings from the experiment and the model validation, each of which will be discussed in more detail in the following sections:

1. With DEB-IBM, effects of chemical mixtures on population dynamics could be predicted rather well, considering the fact we performed blind predictions at the population level with a model that was calibrated on single-substance, individual-level data only.
2. Our developed DEB-IBM was able to predict the mixture effects with great accuracy during the constant exposure phase, the recovery phase, and the acute pulse phase. However, after the acute pulse phase the model became unreliable.
3. Interactive mixture effects were observed on total population abundance in some of the mixture treatments. Considering that the mixture effects were well predicted with our DEB-IBM, which assumes no interactions between substances at TK or PMoA level, it appears that interactive mixture effects (i.e., synergisms) at population level can "emerge" from non-interactivity at physiological and TK levels.
4. Mechanistic models, such as DEB-IBM, calibrated without the use of population-level effect data perform at least similarly, but mostly better in predicting mixture toxicity effects than statistical models, such as (statistical) IA, which do use information from population experiments with single substances.

4.1 Blind predictions of mixture toxicity effects on *D. magna* populations

Visual inspection of the relative trends indicated a strong correlation between the (blind) DEB-IBM predictions (assuming IA for mixture toxicity) and the population data (Figures 4 and G12). The DEB-IBM was calibrated using data on apical endpoints (i.e., survival, growth, and reproduction over time) from a standard 21-day chronic reproduction test. Using individual-

level data only, the DEB-IBM predicted population-level dynamics under mixture toxicity exposure (Figure 4). No mixture toxicity data (at the individual or population level) was included in the model calibration. The validation of the DEB-IBM was thus performed with blind model predictions against independently generated data at the population level.

Predictions of the control

The control population data showed a gradual decrease in abundance that was not captured by the model (see Supplemental Information – Appendix G, Figure G16). We suspected there were a few processes at hand that could explain this observed decrease, either alone or in combination: a shift over time in food composition or food quality, a shift in the starvation coping mechanism (Kooijman 2010), or a crowding-related effect (Preuss et al., 2009). We implemented a change in the functional response (f [-]) (i.e., decrease in food availability) over time to match the observations (see paragraph 2.2.3). Even with this correction, the relative trends in predicted average size over time do not completely correlate with the observations (see Supplemental Information – Appendix G, Figure G13). The DEB-IBM predicted a constant average size throughout the experiment. However, in the control observations, there was a gradual shift from smaller to larger individuals. We suspect this is caused by how intra-specific competition for food is implemented. Intra-specific competition implemented here follows the approach of Preuss et al. (2009). The scaled functional response (f [-]) is adjusted for all individuals equally by considering the total amount of food that can maximally be eaten by all individuals and the total amount of available food. In real populations, however, larger-sized individuals may have a competitive advantage, i.e., the size-efficiency hypothesis (Brooks and Dodson 1965; Hall et al., 1976). Larger individuals can potentially swim greater distances and have larger intrinsic filtration rates (Burns 1969). The functional response could thus be dependent on the size of the individual. More research is needed to uncover the mechanisms that explain intra-specific competition between daphnids. Nonetheless, the

method of Preuss et al. (2009) is a simple and generic approach to model intra-specific competition between daphnids and proves sufficient in predicting the control population abundances over time.

Blind predictions of mixture effects at the population level

The DEB-IBM predicted the trends in the population experiment with great accuracy assuming mechanistic IA on the lethal and sub-lethal effect mechanisms (i.e., GUTS-RED-SD and DEB-TD respectively). These predictions of mixture effects at the population level were blind predictions, meaning they were performed at the population level without explicit inclusion of population-level effect data. Moreover, no data on mixture toxicity at the individual (or population) level were included. The predicted mixture toxicity effects at the population thus purely emerge from the effects of each individual substance calculated for every single individual in the population. Mechanistic (population) models thus seem fit for evaluating mixture toxicity at both the individual and population level. Mixture toxicity on substances is not always available, and only a limited number of combinations can be tested in the laboratory. The strength of our modelling approach here is that mixture toxicity data was not needed.

The applicability of DEB-IBM is also broader than the general statistical models used in risk assessment. Statistical IA and CA are commonly applied at the individual level to predict mixture effects on apical endpoints, e.g., survival after 21 days. These models can also be used to predict population-level effects of mixtures, although population-level data with the single substances is needed. Moreover, to apply CA at the population level, a dose-response relationship is needed for each substance in the mixture and the population endpoint of interest. This requires a significant amount of experimental work, as different concentrations need to be tested on populations. This limits the applicability of CA for population-level mixture toxicity. Additionally, statistical IA and CA are not useful for extrapolation. IA and CA can describe

the mixture effects (i.e., contribution of each substance to the mixture effect), and deviations from these models can be classified as synergistic or antagonistic. However, limited information is given on the mechanisms leading to toxicity (e.g., the mode of action). DEB-IBMs, on the other hand, are mechanistic models, based on fundamental laws in biology, that quantitatively describe energy allocation on the sub-organismal level (Kooijman 2010) and predict the state of the population (Martin et al., 2013a). Implemented mechanism can explain the processes leading up to toxicity at the population level. Mechanistic population models, such as DEB-IBMs, only need individual-level data to make an informed assessment of long-term (i.e., longer than what can be tested in the laboratory) effects at the population level (Preuss et al., 2009; Pereira et al., 2019).

Blind predictions of mixture effects at the population level with IA or DA

DA and IA for mixture toxicity can be applied at the lethal (GUTS) and sub-lethal (DEB-TKTD) toxicity sub-models, and different trends on population dynamics are predicted when comparing these. For all mixture treatments, the IA delivered accurate predictions of the population dynamics. IA was assumed as the most applicable mixture toxicity model in DEB-IBM since for most combinations of substances, the inferred DEB-TD PMoA was different (See Supplemental Information, Appendix H – Tables H4 and H5).

The endosulfan-HCH mixture treatment was the only mixture where a similar DEB-PMoA was inferred during calibration (Tables H4 and H5). Both endosulfan and HCH were determined to decrease energy assimilation with increasing concentrations. Both compounds have a similar toxicological pathway, as both pesticides generally affect the nervous system and can cause muscle spasms (DeLorenzo et al., 2002; Willet et al., 1998). Because of the similarity in toxic effects (i.e., neurotoxicity), it could be assumed that the toxicological pathways have a joined branch, and that some impacted processes along the chain of events are shared. Therefore, the

DA model was tested for this combination, in addition to IA. The assumption under the DA model is that the internal damage caused by each of these compounds is of the same form (i.e., the two forms of damage can be substituted and are considered dilutions of each other). However, based on the plots and the NRMSE, the combination of IA for both DEB-TD and GUTS predicted the data of the endosulfan-HCH treatment best (Figures 5 and G17; Table 3). Any model using DA at DEB-TD and GUTS yielded a visual mismatch with the data and worse values of the NRMSE. We can thus assume that IA is the correct implementation for this treatment.

Even though both compounds share a mode of action (i.e., affecting nerve and muscle functioning), this does not mean endosulfan and a-HCH share the exact same toxicological pathway or molecular initiating event. Endosulfan is known to block the gamma-amino butyric acid (GABA)-activated chloride (Cl^-) channels (IRAC 2021). Lindane, the gamma isomer of HCH, also is known to interfere with the GABA-receptor (Nagata and Narahashi 1995). However, a-HCH has been shown to have little to no effect on the GABA-receptor, although this was assessed *in vivo* on rats and not on *D. magna*. Additionally, the way the damage is initiated (i.e., the molecular reception) or the location of the damage inside the organism (i.e., the targeted organ), might differ between these compounds. Moreover, differences can be seen in the individual-level effect data (Supplemental Information, Appendix H – Table H1). Based on the dose-response curves at the individual, we see that for endosulfan both lethal and sub-lethal effects are important (LC10 = 169 µg/L, EC10 = 119 µg/L). In contrast, for a-HCH, the sub-lethal effects are most important (LC10 > 1.7 mg/L, EC10 = 643 µg/L). Pyrene and dicofol have different toxic modes of action. Pyrene has a toxicological pathway that is connected to the molting of D. magna (Olmstead et al., 2005). The pesticide dicofol is a known endocrine disruptor that induces sex changes in vertebrates (Vinggaard et al., 2000). No clear toxicological pathway of dicofol in D. magna has been established.

4.2 Accuracy and reliability of mixture toxicity with the DEB-IBM

Based on the visual validation of the relative effects (Figure 4) and quantitative validation with the NRMSE (Table 3), a good correspondence was found between the DEB-IBM predictions with IA and the experimental validation. The model performed very well in predicting the decrease in abundance during the constant exposure phase and the population recovery after the constant exposure phase (Figure 4). The strong decrease in abundance that was predicted when the acute pulse is applied, also matched the observation. The main deviations between model predictions and observations were present in the recovery phase after the acute pulse. After the acute shock, the model had become unreliable. We identify two main causes that explain this: stochasticity and toxicokinetics (TK).

The rate of population recovery of *D. magna* will mainly depend on the presence of (gravid or more tolerant) female adults that can quickly produce offspring after having survived the exposure pulse. Since death is assumed to be a stochastic process, the number of surviving adults will differ among replicates, and so will probability of population persistence (or extinction). One of the main advantages of IBMs is the inclusion of these stochastic processes (see details on stochasticity in the TRACE documentation in Vlaeminck et al., 2021), which makes each IBM simulation run unique. Therefore, multiple simulation iterations are usually performed, to get a general idea of the stochastic behaviour of the predictions with the associated persistence/exinction (Figure 4 and G12). We reran the simulations, increased the number of runs to 100, and calculated the persistence (and extinction) probability per treatment. In the data, one of three replicates persisted after the acute pulse phase for the dicofol-HCH mixture. By increasing the number of simulations for this treatment, the probability of persistence was estimated at 58% (n = 100; extinction probability is thus 42%). Out of 100 simulation runs, about 58 populations persisted after the acute pulse, compared to 1 in 3 populations in the data. This shows the importance of including stochasticity in effect

assessment, and IBMs deal particularly well with stochasticity compared to other types of population models (Acolla et al., 2020). Based on a Bernoulli trial experiment with a chance of success (= persistence) of 58%, we estimate an exact probability of 31% that one in three replicates is observed to persist for the dicofol-HCH treatment (Appendix H, Table H8). Therefore, the specific outcome of the experiment in this treatment in terms of persistence/extinction is not unlikely and not in disagreement with the DEB-IBM model predictions (with "unlikely" being arbitrarily defined as a probability <5%, in analogy with a p-value cut-off of 0.05 in statistical significance testing). Based on the Bernoulli trials, the outcome of the experiment is likely (>5%) for most treatments, except for the pyrene-dicofol (<0.03%) and the pyrene-HCH treatment (0.003%) (Appendix H, Table H8). For pyrene-HCH, the DEB-IBM clearly overpredicts the persistence probability, whereas for pyrene-dicofol, the model underpredicts the persistence probability. With DEB-IBM the population extinction probability can be estimated. The IBM can disentangle sources of variability and stochasticity, which is impossible with the data alone, as variability is inherent to experimental testing. Extinction probabilities could be investigated experimentally, but are extremely resource demanding to validate, since a high number of replicated populations is required. It is therefore also, in hindsight, not necessarily surprising that average observed population densities following acutely very toxic pulses (across replicates with both persistence and extinction as ultimate outcome) were not always accurately predicted, as only 3 replicate observations were available.

In addition, recovery will likely depend on TK processes as well. If internal damage recovers slowly, long-term effects can occur, affecting the rate of population recovery. The built-up damage after the pulse can linger, causing long-term effects to occur. Contrarily, *D. magna* might possess a damage repair mechanism where it could easily recover damage from specific substances through elimination or metabolisation (Jager et al., 2011). Additionally, damage

could be passed along from the mother individual to the neonates (i.e., maternal transfer; Jager 2017), a process which was not implemented. The current TK implementation only assumed a one-compartment model, where the uptake and elimination rate of the substance depended on one parameter: the dominant rate (Equation 3). This implementation is in line with common implementations of TK, as mentioned in Jager et al., (2011) and Sherborne et al. (2020). More complex implementations of TK exist and will lead to different trends in prediction, but would also require more experimental data, e.g., on bioaccumulation. Mainly, *ad hoc* implementations of TK are used because the form of TK is highly dependent on the complexity of the organism (e.g., vertebrates vs. invertebrates) and the route of exposure (i.e., dietary uptake via the gut vs. passive diffusion) (Jager and Ashauer 2018). We conclude that the current implementation of TK does not perform well in the recovery phase following exposures with very high lethality.

To improve model reliability of the DEB-IBM in predicting mixture toxicity effects after acute pulsed exposure, several refinement options could be explored: at the individual level, we could (1) generate additional data for TKTD calibration, (2) rethink the model structure that predicts toxic effects; and (3) at the population level we could include additional population-level processes that influence population dynamics. With the first option, additional individual-level data with pulsed exposure would improve the calibration of the TKTD-related parameters, better reflecting damage recovery and (measured) internal concentrations over time. More specifically, the addition of pulsed exposure data at concentration levels close to 100% mortality, would improve reliability of the DEB-IBM when assessing pulsed acute shock exposure. Additional data could potentially improve the fit. However, there will always be variability on test results, but also stochasticity. Variability can be reduced (e.g., working with *Daphnia* clones, keeping environmental conditions constant), but stochasticity cannot be reduced as it is inherent to the real environment. We have seen this in our test results as some replicates of certain treatments showed different trends, e.g., the dicofol-HCH treatment

(Figure 4). This stochasticity is, however, an inevitable consequence of complex ecological systems (Tilman et al., 1998). A complete assessment of stochasticity (e.g., with extinction probabilities) would require an extensive population-level experiment (i.e., multiple concentrations of acute pulses of mixture) to validate the DEB-IBM.

The second option is rethinking model structure and making adaptations to meet the model's needs. Currently, the DEB-IBM only considered a stochastic approach to mortality (i.e., stochastic death [SD]). However, in some cases individuals can show tolerance to specific concentrations of a compound (i.e., individual tolerance [IT] approach) (Jager et al., 2011). Testing out the IT approach (or a hybrid IT-SD model) could be worthwhile if model validation is unsatisfactory. The final option, adding additional processes, could improve the predicted trends as well. Crowding-related processes are known to govern population dynamics in daphnids (Burns 1995; Preuss et al., 2009). We partly included the crowding-related effects by correcting the functional response over time (Section 2.2.4). A more mechanistic implementation of crowding could increase the reliability of the model.

4.3 Emerging interactive effects with DEB-IBM

Based on the population-level effects, interactive effects were observed in some of the treatments and for some of the observed endpoints. For instance, in the a-HCH treatment, no statistically significant effects were observed on population abundance during the constant phase. However, in the pyrene-HCH treatment, the decrease during the constant phase was about 72% compared to the control. This was larger than expected based on the statistical IA calculation with the single-substance treatments: -6.0% for a-HCH and 42% for pyrene (Table 4). Additionally, the ANOVA analysis revealed a statistically significant interaction effect between HCH and pyrene in this mixture treatment for total population abundance in the constant phase (Figure 3). By performing an ANOVA, the data is explicitly statistically

compared with expectations based on a null-model, which, due to log-transformation of the data, is a multiplicative model, i.e., the IA model (see rationale in De Coninck et al., 2013, Faust et al., 2003; Schäfer and Piggot, 2018). The null-hypothesis is that the data follows the IA model. If the interaction term in the ANOVA is significant, it means there is enough information to conclude there is a deviation from the IA model, meaning there are interactive mixture effects. When comparing the observed to the expected effect, the direction of the effect can be classified as synergistic or antagonistic relative to the IA reference model (Table 4).

Interestingly, the DEB-IBM correctly predicted the non-additive effects on total population abundance. Combinations of affected PMoAs can thus lead to diverging effects at the population level, even when non-interactivity is assumed at the physiological (DEB) level, or effects could be enhanced due to interconnectedness of DEB energy flows (Galic et al., 2018). Any impact due to a toxic stressor will cause a redistribution of energy within the population (De Roos et al., 2013), changing population dynamics over time. During the constant phase, the offset between calculated (with statistical IA) and observed effects in the mixture treatments ranged from 5 to 33% (Table 4). However, the difference between predicted (with DEB-IBM) and observed effects was much smaller, ranging from 1 to 11% (Table 4). These observed synergisms on total population abundance were thus relatively accurately predicted with the DEB-IBM and emerge naturally from the implemented toxicity mechanisms and the assumption of independent action at DEB and TK levels.

Comparison to other mixture toxicity studies with mechanistic population models is hard since the number of comparable studies are limited. Applications of DEB-IBM for population-level effect assessment exist in literature, but mainly for single chemical stressors, e.g., Martin et al. (2013b) and Pereira et al. (2019). Mixture toxicity implementations in DEB-based models have first been postulated by Jager et al. (2010). Their analysis was focussed on sub-lethal effects

(growth and reproduction) at the individual-level. Our study went further and included lethal effects while also extrapolating them to the population level.

Bart et al. (2021) have developed a mixture toxicity approach for GUTS-RED, building on the theory suggested by Jager and Ashauer (2018). They specifically tested applicability of DA and IA for both the stochastic death (SD) and the individual tolerance (IA) approach in GUTS-RED, hence only testing effects on apical survival (i.e., mortality after 21 days). Our approach only considers the SD approach for lethal effects. Additionally, their assessment was based on apical effects, whereas here population-level effects were investigated. They also included mixture toxicity data at the apical level in their assessment, which were not used here. Bart et al. (2021) argued that low effect concentrations are important to derive the applicable mixture toxicity mode. This is true for their implementation (with the hockey stick approach) as low effect levels under the damage threshold will not produce toxic effects under IA. For our implementation (using the log-logistic curve), we argue that high effect levels are determining, since these will lead to stronger deviations between DA and IA at the population level (see predictions for the HCH-endosulfan; Figure 5).

In a recent study by Hansul et al. (2021), a similar modelling approach was considered to predict mixture toxicity effects of heavy metals (Cu, Ni, and Zn) to *D. magna* populations. The main difference with previous DEB-IBM approaches, such as Pereira et al., (2019) and Vlaeminck et al., (2021), is that Hansul et al. (2021) allowed single stressors to affect arbitrary combinations of PMoAs. This increases the complexity of the model, but overcomes the uncertainty involved in setting one specific PMoA, which can be hard for some substances if the data is lacking (e.g., no growth data). The calibrated DEB-IBM of Hansul et al. (2021), assuming IA for mixture toxicity at the physiological level, predicted observed synergistic and antagonistic population-level effects rather well. Their conclusions are in line with our

findings. The mixture toxicity effects at the population level emerge from the individual effects of the substances. Galic et al. (2018) have discussed emergence of synergisms and antagonisms with PMoA combinations in detail with a DEB-IBM of *Gammarus pseudolimnaeus*. Their findings highlight the inconsistencies between individual and population-level responses with combinations of PMoAs. Galic et al. (2018) have stated that using individual-level data as proxy for population or community-level effects of mixtures might lead to severe underestimation of the risk. Based on our results, and the findings of Hansul et al. (2021) and Galic et al. (2018), mechanistic (population) models can identify and explain discrepancies in mixture effects amongst biological levels, i.e., from the sub-organismal (PMoA) to the population level.

David et al. (2020) also used a DEB-IBM framework to predict mixture effects on populations, in their case for the three-spined stickleback (*Gasterosteus aculeatus*). They applied the DEB-IBM within a community context (mesocosms in the form of lotic artificial streams). Their community experiment included more complexity and sources of variability that require consideration, such as species interactions, temperature effects, inter-mesocosm variability, variations in food composition and concentration, and intra-specific variability. In addition, their approach considers a whole-mixture approach, where the mixture of the pharmaceutical compounds is assumed as "one" substance causing a single effect, in their case, increased mortality. In contrast, our approach considers each substance individually, while additionally considering the lethal and sub-lethal effects separately. As such, the predicted population effects purely emerge from the lethal and sub-lethal effects of the substances individually. The methods in the current study are more flexible when it comes to extrapolation of single substance effects to mixture effects. The model of David et al. (2020), however, can only be used for the mixture for which it was calibrated, and for the ratio at which the compounds are present in the mixture. Contrarily, the limitation of our approach is that it is only applicable for

isolated monoclonal populations of *D. magna* because it does not consider any community context. In addition, the model of David et al. (2020) was developed for lotic mesocosm (i.e., flowing water including sediment, multiple species, variable conditions), whereas our experimental conditions are different (i.e., semi-static renewal and constant environmental conditions).

There is still limited understanding of the joint effects of multiple stressors, both chemical and ecological in parallel (Goussen et al., 2020). Our study deals with combinations of PMoAs in a population context. Other studies, such as Pereira et al. (2019), have dealt with the combination of a chemical stressors (i.e., Ni) and an abiotic stressor (i.e., temperature). Yet, in the environment, multiple stressors (chemical and ecological) are acting simultaneously. Current ERA techniques generally lack ecological realism (Forbes et al., 2009; Forbes and Callow 2013; Goussen et al., 2016). DEB-based models provide a mechanistic basis for understanding how organisms or populations react to stressors or combinations of stressors (Goussen et al., 2020). Our approach can predict how *D. magna* populations respond to dynamic exposure to chemical mixtures. Extending the DEB-IBM framework to include ecological factors, such as temperature and food availability, or biotic factors such as competition or predation, would bring us even closer to a more ecologically relevant and realistic assessment of effects.

4.4 Predictive power of DEB-IBM

One of the main questions that needs addressing is: does a mechanistic model based on bioenergetics perform better than commonly used (additive) mixture toxicity models (based on statistical IA) in predicting mixture toxicity on populations? Purely based on effects, we could explicitly compare the percentage effect predicted by the DEB-IBM and the percentage effect estimated with IA based on the data. For most mixture treatments, the DEB-IBM predicted the

effects better than the IA estimations (Table 4). But it is not straightforward to assess which model performs better as for some cases the IA calculations leaned closer to the observed values, which was mainly the case for the pulse phase. Nonetheless, the DEB-IBM had a significantly smaller prediction error than statistical IA when comparing with the data for both phases. Additionally, it must be stressed that statistical IA to predict mixture population effects cannot be performed without population data with single substances. DEB-IBM on the other hand only uses single-substance toxicity data at the individual level to make accurate predictions.

The IA model is one of the two most commonly used reference models for mixture toxicity effects in ecotoxicology and chemicals risk assessment, along with CA (SCHER, 2012; Nys et al., 2018). IA was originally used to analyse mixture mortality data (Bliss, 1939), and other null models have been proposed to predict joint effects (Schäfer and Piggott, 2018). Classification of mixture effects into synergisms or antagonism depends on the reference model that is chosen. Here, it was our goal to compare mixture toxicity estimations obtained with a commonly used null model (IA) to those obtained with our DEB-IBM (Table 4). Relative to IA as the null-model, synergisms at population level were detected at the population level. However, when DEB-IBM was used as the predictive model, these “synergisms” at population level emerged - without assuming any explicit interactive effect at physiological, TK or TD level - from the interconnectedness of affected DEB energy flows (Vlaeminck et al., 2021). Thus, while the common reference null-model of IA failed to accurately predict the observed mixture effects at the population level (since observations deviated synergistically from predictions with IA), the DEB-IBM was able to more accurately predict what is occurring in populations when they are exposed to mixtures. Deviations between observations and the DEB-IBM predictions did, however, occur in some instance, but these could help formulate (mechanistically grounded) hypotheses about mixture toxicity as a first step towards

identifying potential shortcomings of the model and to improve the model. The latter is, in our opinion, not possible with the current statistical reference models, such as IA, which cannot tell from the null expectation. If the reference null model fails to predict the effects, the mixture can be classified as synergistic or antagonistic, or alternative null models could be considered (Schäfer and Piggott, 2018), but it cannot tell anything about potential causes of deviations of observations from that null model. Also, the choice of the null model used as a reference in mixture toxicity evaluation remains, however, largely arbitrary, although guidelines to assist this choice have recently been formulated (Schäfer and Piggott, 2018). A limitation of our study is that we only considered the multiplicative IA model in our comparison between statistical and DEB-IBM-based mixture assessment, while other reference null models, such as the simple effect addition model, could also have been considered. This is a worthwhile avenue of further research into DEB-IBM modelling of mixture toxicity.

In the future, implementation of mechanistic (population) models will be necessary to improve ecological realism of risk assessment. Since the European Commission is considering the implementation of the mixture assessment factor (MAF), estimation of the MAF will be critical to protect the environment against adverse mixture toxicity effects (RIVM 2017). DEB-IBMs offer an added value to the assessment, with minimal additional effort in toxicity testing (i.e., include size measurement), while predicting realistic mixture toxicity effects on populations. Especially for prospective (e.g., new substances) or predictive risk assessment (e.g., the impact of a certain mitigation strategy), mechanistic models have a clear advantage over general descriptive methods. To give an example, conventional dose-response analyses are not applicable in situations with dynamic exposure patterns. With the EC10-value derived from a dose-response curve, one cannot predict the population-level response under pulsed exposure. Statistical risk assessment, such as the dose-response curve, is descriptive, rather than predictive. Here, mechanistic models have the advantage due to the inclusion of TKTD

processes, while also extrapolating effects across biological levels. DEB-IBM mixture toxicity modelling offers an excellent opportunity to reduce resources and test animals needed for assessing population-level effects and risks of mixtures.

The number of published studies on mechanistic effect models for mixture toxicity is limited (Jager et al., 2010; Cedergreen et al., 2017; Bart et al., 2021) and the focus is mainly on individual-level assessment. Only recently, applications of mixture toxicity in mechanistic population models have been published (Vlaeminck et al., 2021; Hansul et al., 2021). To increase the acceptance of mechanistic population models in regulatory risk assessment, more examples are needed, such as testing of other substances and organisms, while also providing transparency in the modelling approaches (e.g., through TRACE documentation) and sufficient model validation. Mechanistic (population) models could for instance provide support in deriving margins of safety for chemicals or estimating (mixture) effects at the population level, as these are more accurate in predicting ecologically relevant effects than simple statistical models while not requiring much additional data (i.e., only growth measurements).

5 CONCLUSION

A population experiment was designed with *D. magna* population exposed to mixtures of organic substances (pyrene, dicofol, a-HCH, and endosulfan), both in constant and pulsed exposure regimes. Significant interactive (i.e., non-additive) effects were found in the constant phase on total population abundance for all mixture treatments, and on total population biomass and average length for the HCH-endosulfan and pyrene-dicofol-HCH treatment. Significant effects were also found in the acute pulse phase, but no significant interactions were found.

With DEB-IBM that were calibrated only with single-substance individual-level data, population-level effects of mixtures were predicted. The blind predictions of mixture effects *D. magna* population dynamics were validated based on the population data. Mechanistic IA predicted mixture effects most accurately for all mixture treatments. The DEB-IBM with IA predicted the observed trends in the constant phase, the recovery after the constant phase, and the effects in the acute pulse phase.

The model was, however, not able to reproduce the observations after the pulse exposure phase. The model could be improved by including additional data or rethinking model structure, especially in terms of how stochasticity is considered. In some treatments, replicates became extinct whereas others recovered. Stochasticity is inherent to complex ecological systems, making it hard to assess model predictiveness for acute pulse toxicity with the current population data.

The DEB-IBM accurately predicted the non-additive mixture effects observed on total population abundance in the mixture treatments during the constant exposed phase. At the

population level, the predicted synergisms emerged from the affected DEB energy flows from every individual, which are linked in the IBM framework.

DEB-IBM performed better in predicting population-level effects than statistical IA. The offset in effects was smaller with DEB-IBM (up to 11% difference between predicted and observed) compared to statistical IA (up to 33% difference). Moreover, statistical IA can only predict mixture effects if single-substance data at the population level is available. DEB-IBM predicts population-level effects by extrapolating individual-level data from standard toxicity tests. Our work adds to the growing number of examples that indicate that population models can have a meaningful role as supportive tools for predictive (mixture) effect assessment in regulatory risk assessment frameworks.

FIGURE CAPTIONS

Figure 1. Simulations of the four classical Physiological Modes of Action (PMoAs) of DEB-TKTD and their effect on standard individual- and population-level endpoints with DEB-IBM. Each PMoA is simulated to decrease individual reproduction in a standard test after 21 days with 50%.

Figure 2. Timeline of the population experiment. The experiment consists of two phases: a constant exposure phase (first 11 weeks) and a pulsed exposure phase (last 6 weeks).

Figure 3. Schematic overview of the results of Mann-Whitney U-test, the 2-way and 3-way ANOVA on the population data for different endpoints in both phases (constant and pulse phase). Significant interactions are classified as synergistic and antagonistic based on comparison with the independent action calculation. Detailed ANOVA results can be found in the Supplemental Information (Appendix F).

Figure 4. Validation of the DEB-IBM based on relative effects. Observed (n = 3) and predicted (n = 30) population abundance, relative to their respective control, is plotted over time. Observations and predictions are shown for the control (solvent), the single substance treatments, and the mixture treatments (assuming IA at the DEB-TD and GUTS-RED-SD level). Bands on the observation (in blue) represent the (relative) standard error on the mean. Bands on the prediction (in grey) represent the minimum and maximum DEB-IBM predicted trends. The orange line represents the independent action (IA) calculation based single-substance data at the population level.

Figure 5. Relative abundance over time for the HCH-endosulfan exposed treatment. Comparison between mixture toxicity modes for the HCH-endosulfan mixture treatment. Observed (n = 3) and predicted (n = 30) population abundance, relative to their respective control, is plotted over time. Predictions are shown for the independent action (IA) and damage addition (DA) (at the DEB-TD and the GUTS-RED-SD level). The orange line represents the independent action (IA) calculation based single-substance data at the population level.

TABLE CAPTIONS

Table 1. Independent action model for DEB-TKTD. A distinction is made between parameters decreasing (left) and increasing (right) with stress. RE is the relative effect caused by one substance.

Table 2. Calibrated TK, DEB-TD and GUTS-RED-SD parameters for the compounds of interest. Numbers between brackets indicate the 95% confidence intervals. PMoA = physiological mode of action, A = decrease in energy assimilation, R = decrease in reproduction efficiency, M = increase in maintenance costs.

Table 3. Normalized Root Mean Square Errors (NRMSE) on (log-transformed) total population abundance. NA = not applicable, IA = independent action, DA = damage addition.

Table 4. Comparison between observed effects (n = 3), DEB-IBM predicted effects (n = 30), and IA estimated effects (based on data). Percentage decrease is calculated relative to the control observation/prediction. NA = not applicable, * = significant effect based on the Mann-Whitney U-test, S = synergistic interaction based on ANOVA, / = no synergistic interaction based on ANOVA.

REFERENCES

Acolla C, Vaugeois M, Grimm V, Moore AP, Rueda-Cediel P, Schmolke A, and Forbes VE. 2020. A Review of Key Features and Their Implementation in Unstructured, Structured, and Agent-Based Population Models for Ecological Risk Assessment. *Integrated Environmental Assessment and Management*, 17(3):521-540.

Ashauer R, and Jager T. 2018. Physiological modes of action across species and toxicants: the key to predictive ecotoxicology. *Environmental Science: Processes and Impacts*, 20(8):48-57.

Bart S, Jager T, Robinson A, Lahive E, Spurgeon DJ, and Ashauer R. 2021. Predicting Mixture Effects over Time with Toxicokinetic-Toxicodynamic Models (GUTS): Assumptions, Experimental Testing and Predictive Power. *Environmental Science and Technology*, 2021, 55(4):2430-2439.

Bliss CI. 1939. The toxicity of poisons applied jointly. *Annals of applied biology*, 26(3):585-615.

Brooks JL, and Dodson SI. 1965. Predation, body size, and composition of plankton. *Science*, 150:28-35.

Burns CW. 1995. Effects of crowding and different food levels on growth and reproductive investment of *Daphnia*. *Oecologia*, 101:234-244.

Burns CW. 1969. Relation between filtering rate, temperature, and body size in four species of *Daphnia*. *Limnology and Oceanography*, 14(5):693-700.

Cedergreen N. 2014. Quantifying Synergy: A Systematic Review of Mixture Toxicity Studies within Environmental Toxicology. *PLoS ONE,* 9(5):e96580.
URL https://doi.org/10.1371/journal.pone.0096580

Cedergreen N, Dalhoff K, Li D, Gottardi M, and Kretschmann AC. 2017. Can Toxicokinetic and Toxicodynamic Modelling Be Used to Understand and Predict Synergistic Interactions between Chemicals? *Environmental Science and Technology*, 51:14379-14389.

Cedergreen N, Dalhoff K, Li D, Gottardi M, and Kretschmann AC. 2017. Can Toxicokinetic and Toxicodynamic Modelling Be Used to Understand and Predict Synergistic Interactions between Chemicals? *Environmental Science and Technology*, 51:14379-14389.

David V, Joachim S, Catteau A, Nott K, Ronkart S, Robert C, Gillard N, Bado-Nilles A, Chadili E, Palluel O, Turies C, Julian N, Castiglione J, Dedourge-Geffard O, Hani Y, Geffard A, Porcher J-M, and Beaudouin, R. 2020. Effects of chronic exposure to a pharmaceutical mixture on the three-spined stickleback (*Gasterosteus aculeatus*) population dynamics in lotic mesocosms. *Aquatic Toxicology*, 224:105499.

De Coninck DI, De Schamphelaere KAC, Jansen M, De Meester L, and Janssen CJ. 2013. Inter- active effects of a bacterial parasite and the insecticide carbaryl to life-history and physiology of two *Daphnia magna* clones differing in carbaryl sensitivity. *Aquatic toxicology*, 130:149–159.

DeLorenzo ME, Taylor LA, Lund SA, Pennington PA, Strozier ED, and Fulton MH. 2002. Toxicity and Bioconcentration Potential of the Agricultural Pesticide Endosulfan in Phytoplankton and Zooplankton. *Archives of Environmental Contamination and Toxicology*, 42:173-181.

De Roos AM, Metz J, and Persson L. 2013. Ontogenetic symmetry and asymmetry in energetics. *Journal of Mathematical Biology*, 66(4-5):889-914.

European Commission (EC). 2020. *Chemicals Strategy for Sustainability: Towards a Toxic-Free Environment*. Brussels, Belgium, October 14th, 2020.

ECHA (Environmental Chemicals Agency). 2017. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB Assessment.

EFSA (European Food and Safety Authority). 2018. Scientific Opinion on the state of the art of Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms. *EFSA Journal*, published by John Wiley and Sons ltd. on behalf of European Food and Safety Authority, 16(8):e05377.

Faust M, Altenburger R, Backhaus T, Blanck H, Boedeker W, Gramatica P, Hamer V, Schloze M, Vighi M, and Grimme, LH. 2003. Joint algal toxicity of 16 dissimilarly acting chemicals is predictable by the concept of independent action. *Aquatic toxicology*, 63(1):43-63.

Focks A, Belgers D, Boerwinkel MC, Buijse L, Roessink I, and Van den Brink PJ. (2018). Calibration and validation of toxicokinetic-toxicodynamic models for three neonicotinoids and some aquatic macroinvertebrates. *Ecotoxicology*, 27(7):992-1007.

Forbes VE, and Calow P. 2013. Developing Predictive Systems Models to Address Complexity and Relevance for Ecological Risk Assessment. *Integrated Environmental Assessment and Management*, 9(3):e75-e80.

Forbes VE, Hommen U, Thorbek P, Heimbach F, Van den Brink PJ, Wogram J, Thulke HH, and Grimm V. 2009. Ecological Models in Support of Regulatory Risk Assessments of Pesticides: Developing a Strategy for the Future. *Integrated Environmental Assessment and Management*, 5(1):167-172.

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

Galic N, Sullivan LL, Grimm V, and Forbes VE. When things don't add up: quantifying impacts of multiple stressors from individual metabolism to ecosystem processing. 2018. *Ecology Letters*, 21:568-577.

Goussen B, Rendal C, Sheffield D, Butler E, Price OR, and Ashauer R. 2020. Bioenergetics modelling to analyse and predict the joint effects of multiple stressors: Meta-analysis and model corroboration. *Science of the Total Environment*, 749:141509.

Goussen B, Price OR, Rendal C, and Ashauer R. 2016. Integrated presentation of ecological risk of multiple stressors. *Scientific Reports*, 6:36004.

Grimm V, Frank K, Jeltsch F, Brandl R, Uchmański J, and Wissel C. 1996. Pattern-oriented modelling in population ecology. *Science of the Total Environment*, 183(1-2):151-166.

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke H, Weiner J, Wiegand T, and DeAngelis DL. 2005. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science*, 310(5750):987-991.

Hall DJ, Threlkeld ST, Burns CW, and Crowley PH. 1976. The size-efficiency hypothesis and the size structure of zooplankton communities. *Annual Review of Ecology and Systematics*, 7:177-208.

Hansul S, Fettweis A, Smolders E, and De Schamphelaere KAC. 2021. Interactive Metal Mixture Toxicity to *Daphnia magna* Populations as an Emergent Property in a Dynamic Energy Budget Individual-Based Model. *Environmental Toxicology*,

Hommen U, Baveco JMH, Galic N, and 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. *Integrated Environmental Assessment and Management*, 6(3):325-337.

Insecticide Resistance Action Committee (IRAC). 2021. The IRAC mode of classification online.
URL https://irac-online.org/modes-of-action/
Last accessed: September 21st 2021

Jager T. 2017. *Applications of Dynamic Energy Budget Theory in Ecotoxicology and Stress Ecology*, version 1.3. *Leanpub*.
URL https://leanpub.com/debtox_book

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

Jager T, Albert C, Preuss TG, and Ashauer R. 2011. General Unified Threshold Model of Survival - a Toxicokinetic- Toxicodynamic Framework for Ecotoxicology. *Environmental Science and Technology*, 45:2529-2540.

Jager T and Ashauer R. 2018. *Modelling survival under chemical stress. A comprehensive guide to the GUTS framework*, version 1.0. *Leanpub*.
URL https://leanpub.com/guts_book.

Kooijman SALM. 2010. *Dynamic Energy Budget theory for metabolic organisation*, 3rd edition. *Cambridge University Press*, Cambridge, UK.

Kortenkamp A, and Faust M. 2018. Regulate to reduce chemical mixture risk. *Science*, 361(6399):224-226.

Liess M, and Foit K. 2010. Intraspecific competition delays recovery of population structure. *Aquatic Toxicology*, 97:15-22.

Martin BT, Jager T, Nisbet RM, Preuss TG, and Grimm V. 2013a. Predicting population dynamics from the properties of individuals: A cross-level test of dynamic energy budget theory. *The American Naturalist*, 181:506-519.

Martin BT, Jager T, Nisbet RM, Preuss TG, Hammers-Wirtz M, and Grimm V. 2013b. Extrapolating ecotoxicological effects from individuals to populations: a generic approach based on dynamic energy budget theory and individual-based modelling. *Ecotoxicology*, 22:574-583.

Nagata K and Narahasi T. 1995. Differential effects of hexachlorocyclohexane isomers on the GABA receptor-chloride channel complex in rat dorsal root ganglion neurons. *Brain Research*, 704(1):85-91.

Nys C, Van Regenmortel T, Janssen CR, Oorts K, Smolders E, and De Schamphelaere KAC. 2018. A Framework for Ecological Risk Assessment of Metal Mixtures in Aquatic Systems. *Environmental Toxicology and Chemistry*, 37(3):623-642.

OECD. 2012. OECD Guidelines for the Testing of Chemicals, Section 2. Test no. 211: *Daphnia magna* Reproduction Test. OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, https://doi.org/10.1787/9789264185203-en

OECD. 2019. OECD Environment, Health and Safety, Publications Series on Testing and Assessment, No. 23: Guidance Document on Aquatic Toxicity Testing of Difficult Substances. *OECD Publishing*, Paris, France.

Olmstead AW, and LeBlanc GA. 2005. Joint action of polycyclic aromatic hydrocarbons: Predictive modeling of sublethal toxicity. *Aquatic Toxicology*, 75:253-262.

Pavlíková N, Bláhová L, Klán P, Bathula SR, Sklenář V, Giesy JP, and Bláh L. 2012. Enantioselective effects of alpha-hexachlorocyclohexane (HCH) isomers on androgen receptor activity *in vitro*. *Chemosphere*, 86(1):65-69.

Pereira CMS, Vlaeminck K, Viaene K, De Schamphelaere KAC. 2019. The Unexpected Absence of Nickel Effects on a *Daphnia* Population at 3 Temperatures is Correctly Predicted by a Dynamic Energy Budget Individual-Based Model. *Environmental Toxicology and Chemistry*, 38(7):1423-1433.

Preuss T, Hammers-Wirtz M, Hommen U, Rubach MN, and Ratte HT. 2009. Development and validation of an individual based Daphnia magna population model: The influence of crowding on population dynamics. *Ecological Modelling*, 220(3):310-329.

Rijksinstituut voor Volksgezondheid en Milieu (RIVM). 2017. Addressing combined effects of chemicals in environmental safety assessment under REACH - A thought starter. RIVM report.

RStudio Team. 2020. RStudio: Integrated Development Environment for R. RStudio inc., Boston (MA, USA).
URL http://www.rstudio.com/

Schäfer RB, and Piggott JJ. 2018. Advancing understanding and prediction in multiple stressor research through a mechanistic basis for null models. *Global Change Biology*, 24(5):1817-1826.

SCHER (Scientific Committee on Health and Environmental Risks). 2012. Toxicity and Assessment of Chemical Mixtures. *Official Journal of the European Union*.
URL https://publications.europa.eu/en/publication-detail/-/publication/ffab4074-6ce5-4f87-89b7-fbd438943b54/language-en .

Sherborne N, Galic N, and Ashauer R. 2020. Sublethal effect modelling for environmental risk assessment of chemicals: Problem definition, model variants, application and challenges. *Science of the Total Environment*, 745:141027.

Threlkeld ST. 1976. Starvation and size structure of zooplankton communities. *Freshwater Biology*, 6:489-496.

Tilman D, Lehman CL, and Bristow CE. 1998. Diversity-Stability Relationships: Statistical Inevitability or Ecological Consequence? *The American Naturalist*, 151(3):277-282.

Toni T, Welch D, Strelkowa N, Ipsen A, and Stumpf MPH. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society*, 6:187-202.

van Broekhuizen FA, Posthuma L, and Traas TP. 2017. Addressing combined effects of chemicals in environmental safety assessment under REACH-A thought starter. RIVM report 2016-0216 (National Institute for Public Health and the Environment, 2017).
URL https://rivm.openrepository.com/handle/10029/620826

Van Leeuwen CJ and Vermeire TG. 2007. *Risk Assessment of Chemicals: An Introduction*, 2nd edition. *Springer*, Dordrecht, The Netherlands.

Van der Vaart E, Johnsont ASA, and Sibly RM. 2016. Predicting how many animals will be where: How to build, calibrate and evaluate individual-based models. *Ecological Modelling*, 326:113-123.

Vinggaard AM, Hnida C, Breinholt V, and Larsen JC. 2000. Screening of Selected Pesticides for Inhibition of CYP19 Aromatase Activity *In Vitro*. *Toxicology in Vitro*, 14:227-234.

Vlaeminck K, Viaene K, Van Sprang P, De Schamphelaere KAC. 2019. The Use of Mechanistic Population Models in Metal Risk Assessment: Combined Effects of Copper and Food Source on *Lymnaea stagnalis* Populations. *Environmental Toxicology and Chemistry*, 38(5):1105-1119.

Vlaeminck K, Viaene K, Van Sprang P, De Schamphelaere KAC. 2021. Development and Validation of a Mixture Toxicity Implementation in the Dynamic Energy Budget–Individual-Based Model: Effects of Copper and Zinc on *Daphnia magna* Populations. *Environmental Toxicology and Chemistry*, 40(2):513-527.

Willet KL, Ulrich EM, and Hites RA. 1998. Differential Toxicity and Environmental Fates of Hexachlorocyclohexane Isomers. *Environmental Science and Technology*, 32(15):2197-2207.