This document is the accepted manuscript version of the following article:
Kattwinkel, M., Reichert, P., Rüegg, J., Liess, M., & Schuwirth, N. (2016). Modeling macroinvertebrate community dynamics in stream mesocosms contaminated with a pesticide. Environmental Science and Technology, 50(6), 3165-3173. <https://doi.org/10.1021/acs.est.5b04068>

Modelling macroinvertebrate community dynamics in stream mesocosms contaminated with a pesticide

Mira Kattwinkel ^{,a}, Peter Reichert ^a, Johanna Rüegg ^a, Matthias Liess ^b, Nele Schuwirth ^a*

^a Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department Systems
Analysis, Integrated Assessment and Modelling, Überlandstrasse 133, Dübendorf, Switzerland

^b UFZ - Helmholtz Centre for Environmental Research, Department of System-Ecotoxicology,
Permoserstraße 188, Leipzig, Germany

* Department Systems Analysis, Integrated Assessment and Modelling, Eawag: Swiss Federal
Institute of Aquatic Science and Technology, Überlandstrasse 133, Dübendorf, Switzerland,
Phone: +41 (0)58 765 5262, Fax: +41 (0)58 765 5802, mira.kattwinkel@gmx.net

Word count: 5491 + 1 small figure (300) + 2 large figures (1200) = 6991

14 TOC ART:

18 ABSTRACT

19 Modelling community dynamics of aquatic invertebrates is an important but challenging task, in
20 particular in ecotoxicological risk assessment. Systematic parameter estimation and rigorous
21 assessment of model uncertainty are often lacking in such applications.

22 We applied the mechanistic food web model Streambugs to investigate the temporal
23 development of the macroinvertebrate community in an ecotoxicological mesocosm experiment
24 with pulsed contaminations with the insecticide thiacloprid. We used Bayesian inference to
25 estimate parameters and their uncertainty.

26 Approx. 85% of all experimental observations lie within the 90% uncertainty intervals indicating
27 reasonably good fits of the calibrated model. However, a validation with independent data was
28 not possible due to lacking data. Investigation of vital rates and limiting factors in the model
29 yielded insights into recovery dynamics. Inclusion of the emergence process and sub-lethal
30 effects turned out to be potentially relevant model extensions. Measurements of food source
31 dynamics, individual body size (classes), and additional knowledge on sub-lethal effects would
32 support more accurate modelling.

33 This application of a process-based, ecotoxicological community model with uncertainty
34 assessment by Bayesian inference increased our process understanding of toxicant effects in
35 macroinvertebrate communities and helped identifying potential improvements in model structure
36 and experimental design.

37 KEYWORDS: Dynamic food web model; model calibration; parameter estimation; Bayesian
38 inference; plant protection product; pesticides; environmental risk assessment

Introduction

Ecological modelling offers a powerful tool for formulating quantitative hypotheses about the processes that drive population and community dynamics, and for improving ecological understanding by comparing the consequences of these hypotheses with observed data. Moreover, ecological modelling is becoming increasingly important in both research and risk assessment applications in ecotoxicology.^{1, 2}

In the ecotoxicological context, so far single species population models have most often been applied (Grimm et al. 2013 and references therein)³, which can yield valuable insights into direct effects of toxicants such as pesticides. When compared to experimental data, such models can support a better understanding of the processes involved.^{4, 5}

However, single species population models are not capable of reflecting interspecific interactions, including competition and predation, that may modify toxicant effects on both individuals and populations.^{6, 7} Examples of the relevance of such biological interactions include the prolongation of long-term effects by competition and the culmination of successive toxicant pulses.^{8, 9} Such interactions can be investigated using two-species models.^{10, 11} Yet, in natural systems usually many more than two species interact. Hence, community models based on food webs consisting of several different taxa can yield additional insights into the effects of toxicants on communities, although at the cost of increased model complexity. The few examples of such food web models in the ecotoxicological context so far often focus on investigating model behaviour rather than on direct comparison of model output to observed community dynamics.^{12,}

¹³

The greater complexity of community models might be one reason why comparison of modelling results to measured data often yields moderate fits at best.^{14, 15} Moreover, particularly in the ecotoxicological context, both population and community models often lack systematic

parameter estimation as well as a rigorous assessment of model uncertainty and the consequences thereof. However, when model results are used for predictions, for instance in environmental risk assessment, the acknowledgement of the sources of uncertainty and their assessment with respect to the influence on model results is very important.² Bayesian parameter inference^{16, 17} offers a consistent framework for parameter estimation by combining prior knowledge about parameter values (e.g. from the literature, see also Reichert et al. (2015)¹⁸) with information gained from new data. Additionally, it allows parameter uncertainty to be propagated to the model results.

An iterative cycle of experimental work and modelling will support the understanding of the complex toxicant effects on communities. In this study we strove for progressing one step further in this cycle by using the process-based, mechanistic food web model Streambugs to investigate pesticide effects on the dynamics of the community composition of aquatic invertebrates.^{19, 20} The model combines knowledge from theoretical food web modelling, the metabolic theory of ecology, ecological stoichiometry and functional trait data. So far, Streambugs has been used to estimate the presence/absence or the probability of occurrence of individual taxa in rivers from steady state biomass under constant environmental conditions. In this study, we made use of Streambugs' universality and applied the model to time series data of a mesocosm experiment with the neonicotinoid insecticide thiacloprid applied as pulses.^{21, 22} We incorporated temporal dynamics of the contamination, concentration dependence of the resulting acute lethal toxic effects and taxon-specific sensitivities into the model. We used Bayesian inference to estimate model parameters from experimental observations, to assess model output uncertainty and the capability of the model to reproduce observed dynamics. A model validation with data not used for parameter inference was beyond the scope of this study due to lacking data. Because we are currently at a stage of model testing and calibration we did not use the model for predictions under untested conditions, but rather chose the approach to gain valuable insights into

assumptions, limitations and potential improvements of Streambugs for its applicability in ecotoxicology.

The study aims were (1) to assess the performance of Streambugs for modelling community dynamics under toxicant stress, (2) to increase the understanding of the processes underlying population dynamics in the mesocosm experiment, (3) to identify potential model improvements, and (4) to identify how to optimize experimental and sampling design to gain additional mechanistic knowledge. This is, to our knowledge, the first application of a process-based model to analyse the dynamics of pesticide effects on invertebrate communities that also includes the assessment of model uncertainty.

Methods

Experimental data

For model parameter estimation (model calibration), we used abundance data of an experiment in artificial outdoor streams that were contaminated with the neonicotinoid insecticide thiacloprid.^{21, 22} The experimental system consisted of 16 mesocosm streams (length 20 m, width at water surface 0.32 (± 0.03) m, average depth 0.25 (± 0.11) m, discharge 160 (± 9) L/min) with closed water circulation, each. The macroinvertebrate communities in the streams were established two years before the beginning of the experiment. A pulse contamination was applied in May 2006 and May 2007, respectively, without water exchange. For the first contamination, ten streams remained as untreated controls, while two streams were treated with a nominal thiacloprid concentration of 0.1, 3.2, and 100 $\mu\text{g/l}$, respectively. For the second contamination, two control streams per concentration were converted to treated ones, resulting in four replicates for each concentration and the control.

Macroinvertebrate sampling took place on 17 dates from September 2005 to December 2007. In addition to abundance measurements, emergence traps were installed to catch emerging adult

111 insects. We selected 16 out of a total 34 observed taxa, excluding those with very low
 112 abundances (≤ 1 individual per sampling area) and missing in several consecutive observations
 113 (SI, Table S1).

114 Thiachloprid concentrations in water were observed to decline rapidly over time²¹, presumably
 115 mainly due to microbial degradation.²³ Based on measured water concentrations in the
 116 mesocosms on hour 4, 10, 48, 120, 216, 312, 480 after contamination we estimated daily
 117 concentrations for model input using nonlinear least square regression (R function nls)²⁴ (SI,
 118 Figure S1).

119 Streambugs model

120 The Streambugs model predicts biomass densities of invertebrate taxa, algae and organic matter
 121 in an aquatic food web using differential equations.¹⁹ It is based on the metabolic theory of
 122 ecology, which allows for a parsimonious description of the dependence of the vital rates (i.e.
 123 biomass growth, respiration and mortality) on body mass and temperature.²⁵ Descriptions of all
 124 parameters are given in Table S2 in the SI. All processes scale with the basal metabolic rate
 125 ($r_{\text{basal metab}}$):

$$r_{\text{basal metab}} = \frac{n}{A} \cdot i_0 \cdot \left(\frac{M}{M_0} \right)^b \cdot e^{-\frac{E}{k_B T}} \cdot \frac{1}{EC} \quad , \quad (1)$$

126 where n is the number of individuals, A is area [m^2], i_0 a normalization constant [W], M the
 127 biomass of an individual [g], b an allometric scaling exponent, E the activation energy [eV], k_B
 128 Boltzmann's constant 8.617343×10^{-5} eV/K, T absolute temperature [K], EC the energy content
 129 of invertebrates or algae [J/gDM; DM = dry mass], and M_0 (= 1 g) is a mass to make the fraction
 130 non-dimensional.

Taxon dependent factors ($f_{\text{basal tax}}$, $f_{\text{gro tax}}$) and universal factors (f_{resp} , f_{death}) are used to describe respiration (r_{resp}), growth on different food sources j ($r_{\text{gro on } j}^{\text{cons}}$) and death (r_{death}) based on the basal metabolic rate. The respiration rate is thus given as:

$$r_{\text{resp}} = f_{\text{resp}} \cdot f_{\text{basal tax}} \cdot r_{\text{basal metab}} \quad (2)$$

The algae growth rate depends on light and nutrient availability. Invertebrate growth rates on various food sources are modified by food availability ($f_{\text{lim food}}$), food preferences ($f_{\text{pref } j}$) and self-inhibition accounting for density dependent intraspecific competition ($f_{\text{self inh}}$). Self-inhibition depends on habitat conditions described by current, temperature and microhabitat structure and on autecological, taxon-specific preferences (f_{current} , f_{temp} , f_{microhab}).²⁶ Such modification factors based on autecological traits incorporate a site and taxon specific habitat capacity depending on environmental conditions and taxa tolerances to these conditions.

$$r_{\text{gro on } j}^{\text{cons}} = f_{\text{cons}} \cdot f_{\text{gro tax}} \cdot f_{\text{lim food}} \cdot f_{\text{pref } j} \cdot f_{\text{self inh}} \cdot f_{\text{basal tax}} \cdot r_{\text{basal metab}} \quad (3)$$

$$f_{\text{lim food}} = \frac{D_{\text{food}}}{K_{\text{food}} + D_{\text{food}}} \quad (4)$$

$$f_{\text{pref } j} = \frac{D_j \cdot p_j}{\sum_f D_f \cdot p_f} \quad (5)$$

$$f_{\text{self inh}} = \frac{K_{\text{dens}}}{K_{\text{dens}} + D} \quad (6)$$

$$K_{\text{dens}} = h_{\text{dens}} \cdot f_{\text{current}} \cdot f_{\text{temp}} \cdot f_{\text{microhab}} \quad (7)$$

with D_{food} : sum of food densities, K_{food} : halfsaturation constant for sum of food sources, D_j : density of food j , p_j : preference for food j , D : biomass density of the taxon, h_{dens} : halfsaturation density of biomass regarding self-inhibition.

Death rates are modified by the saprobic conditions (f_{saproby}) and by pollution ($f_{\text{org pollut}}$). In this study, we replace $f_{\text{org pollut}}$ with f_{cont} , a taxon and concentration dependent factor (section toxic effects):

$$r_{\text{death}} = f_{\text{death}} \cdot f_{\text{cont}} \cdot f_{\text{saproby}} \cdot f_{\text{basal tax}} \cdot r_{\text{basal metab}} \quad (8)$$

A food web is constructed based on feeding preferences²⁶ and average body mass of the taxa (based on measurements and literature values²⁷), with larger predators feeding on smaller taxa (SI, Figure S2). Such a food web approach allows for analysing the effects of food-related species interactions.

Streambugs computes biomass densities $B_i(\theta)$ (θ = vector of model parameters), not abundances as observed in the experiment, and does not take emergence into account. Hence, the Streambugs output is translated into the number of individuals per area based on the taxon's average body mass (M_i). Biomass loss due to emergence is incorporated by a taxon and season specific factor ($f_{\text{emergence } i}$ between 0 and 1; SI, Table S3). Additionally, we assume that only a fraction of the individuals present are caught at the sampling (p_{catch}) and that there is a small proportion of individuals that are erroneously observed due to misclassification ($n_{\text{misc } i}$). The deterministic model results for observed abundances ($N_{\text{det } i}(\theta)$) are therefore calculated as:

$$N_{\text{det } i}(\theta) = \left(\frac{B_i(\theta)}{M_i} + n_{\text{misc } i} \right) \cdot p_{\text{catch}} \cdot f_{\text{emergence } i} \quad (9)$$

Toxic effects

In Streambugs, the effect of pollution was originally modelled as a universal increase in the death rate at polluted sites for taxa sensitive to organic pollution. To investigate the effects of a specific pesticide with temporal dynamics that affects each taxon specifically, we modified the original implementation as follows: Baseline death rates are multiplied by the concentration and taxon dependent factor f_{cont} to model acute lethal effects. Below a certain threshold concentration,

we assume no pesticide effect, hence mortality is not modified ($f_{\text{cont}} = 1$). Above this taxon-dependent threshold, mortality increases linearly with the logarithm of the concentration. We define this linear increase based on two supporting points: The first one is given by $f_{\text{cont}} = 1$ and the no effect threshold concentration that we set to $0.5 \log_{10}$ units below the taxon specific model parameter LC_{50} . This model assumption is based on estimates using NOEC and LC_{50} concentrations from data of the structurally similar substance imidacloprid.²⁸ As the second supporting point we determine the value of f_{cont} at the LC_{50} . Therefore, we solve the Streambugs equations for f_{cont} under conditions of an LC_{50} experiment (single taxa, no food limitation, 20°C, 50% mortality in 24h) (SI, Figure S3).

Statistical data analysis

Basics of Bayesian parameter inference

Bayesian parameter inference describes the current state of scientific knowledge about model parameters using probability distributions, and updates this knowledge with observed data. It is based on three elements: (1) The likelihood function of the model describes the probability distribution of observed data given the model parameters. Its formulation depends on the model and the assumed observation error distribution. (2) The prior probability distribution of the parameters describes the *a priori* information about the parameter values, often constructed as independent marginal distributions of the individual parameters (i.e. a probability distribution for each parameter without reference to the other parameters). This knowledge should be acquired by compiling all available information from the literature, from experts, and from previous studies. Particularly in complex models with many parameters, realistic and constraint priors reduce the risk that estimated parameters take unrealistic values and allow for detection of prior-data conflicts. (3) The observed data are needed to evaluate the likelihood.

The *a priori* information from the likelihood and the prior of its parameters is combined with the observed data to derive a posterior distribution of the parameters, the target of the inference, describing the probability distribution of the parameters in light of the actual observations. Except for very simple models, it is not possible to calculate the posterior distribution analytically. However, it is possible to sample from the posterior to derive numerical approximations of its properties based on the Markov Chain Monte Carlo approach (MCMC) using, for instance, a Metropolis-Hastings algorithm.^{16, 29} This approach constructs samples of parameters and allows the posterior distribution and its marginals to be approximated. For detailed general introductions to Bayesian inference and examples see e.g. Van Oijen et al. (2005)³⁰ and Gelman et al. (2004)¹⁶; for a justification of this approach in the context of environmental modelling see Reichert et al. (2015)¹⁸.

Formulation of the likelihood

We assume that in the mesocosms individual organisms are not evenly distributed in space but occur clumped (e.g. due to the heterogenic environment). This leads to an “overdispersed” distribution of the observation probability of abundance of taxon *i*, N_i , often described by a negative binomial distribution.^{31, 32} When parameterizing this distribution by its mean μ and standard deviation σ , the overdispersed case is characterized by $\sigma^2/\mu > 1$:

$$N_i \sim \text{nbinom}(\mu_i, \sigma) \quad . \quad (10)$$

We parameterize this distribution using the dispersion parameter k , after Elliot (1971)³¹, with

$$k = \frac{\mu(\theta)^2}{\sigma^2 - \mu(\theta)} \quad , \quad (11)$$

and the deterministic prediction, $N_{\text{det } i}(\theta)$ (eq. 9), as its mean, yielding the following probability distribution of observed abundance of taxon *i*, $P(N_i|\theta)$:

$$P(N_i|\theta) = \frac{\Gamma(N_i+k)}{\Gamma(k) \cdot N_i!} \cdot \left(\frac{k}{k+N_{\text{det } i}(\theta)}\right)^k \cdot \left(\frac{N_{\text{det } i}(\theta)}{N_{\text{det } i}(\theta)+k}\right)^{N_i} \quad (12)$$

See Figure 1 for examples. The parameters of the likelihood function, θ , consist of the Streambugs parameters, those describing the sampling and emergence process, and those of the negative binomial distribution. We assume the observation probabilities of the different taxa to be independent of each other. Therefore, the joint probability for the abundances of all taxa and all observation time points is the product of the probabilities of the individual taxa.

Figure 1. Examples of probability distributions for observing N_i individuals of single taxa for three results of the deterministic model $N_{\text{det } i}(\theta)$ and $k = 20$. The dot and dashed lines depict the (discrete) probability ($= 0.0263$) of a certain observation $N_{\text{obs } i} = 40$ given $N_{\text{det } i}(\theta) = 50$.

Prior distributions

Prior distributions describe the knowledge on model parameters and the uncertainty of this knowledge. For all parameters except the toxicological sensitivity to thiacloprid they were taken from Schuwirth et al. (2015)²⁰ (SI, Table S2).

Measurements of toxicological sensitivity to thiacloprid were only available for a small number of the investigated taxa and these values varied considerably between different experiments. We

224 dealt with this uncertainty by including the taxon specific LC_{50} in the inference process.
225 Relatively wide prior distributions were constructed based on laboratory LC_{50} values for
226 thiacloprid and for the more commonly tested and structurally similar imidacloprid.^{28, 33} The SI
227 gives a detailed description of the prior construction (Figures S4, S5, Table S4).
228

Parameter inference

The abundance data of replicate experimental streams was averaged to reduce variability and to account for the fact that in a deterministic model only a variation in initial conditions (D_{ini}) can results in variations in the results, if all other conditions are identical (as assumed in the replicates). Altogether, we used observations of 16 taxa in the four different experimental settings (control and three concentrations) at 16 points in time (using the first observation as the initial taxon density) amounting to 1024 data points. We inferred posterior distributions for 109 parameters consisting of five taxon specific invertebrate parameters ($f_{\text{basal tax}}$, $f_{\text{gro tax}}$, h_{dens} , K_{food} , LC_{50}) and 29 other (SI, Table S2, S4). Taxon specific mass, stoichiometric and trait based parameters were kept fixed at measured or literature values.

The likelihood of a given observed abundance for a certain parameter set (i.e. the probability of the actual observation given the Streambugs results derived with these parameter values and input variables) is calculated by inserting the actual observations $N_{\text{obs } i}$ for N_i in eq. 12, k was set to 20, p_{catch} to 0.8 and $n_{\text{misc } i}$ to 1% of the mean observed abundance per taxon. Additionally, we combined data from emergence traps with literature values to estimate $f_{\text{emergence}}$ (SI, Table S3). Furthermore we allow for a small input of biomass (equivalent to one individual per year for each taxon).

MCMC samples of the parameters should consist of many iterations to reach an adequate coverage of the posterior distribution by the chain of points sampled. For the proposal function (also called jump distribution), which draws a new sample based on the previous parameter set²⁹, we used a multivariate normal distribution and updated its variance and correlation approximately every 30000 steps based on the samples constructed so far. To test for convergence and improve the numerical accuracy of the results, we run 10 chains in parallel. The first 600000 iterations were omitted to exclude the burn-in phase. We constructed the marginal

posterior distributions (Figure 2; SI Figure S6) using a local smoother (R function density) on final samples of more than 600000 iterations per chain thinned by a factor of 100.

Model simulations

We show (Figure 3; SI Figure S8) the simulation results of the deterministic model for the parameter set with the maximum posterior density, $\theta_{\text{max post}}$,

$$N_{\text{det } i}(\theta_{\text{max post}}) \quad , \tag{13}$$

the results of the deterministic model with propagated posterior parameter uncertainty, θ_{post} ,

$$N_{\text{det } i}(\theta_{\text{post}}) \quad , \tag{14}$$

and the predicted results for new observations (considering posterior parameter and sampling uncertainty)

$$\int P(N_i|\theta) f_{\text{post}}(\theta) d\theta \quad . \tag{15}$$

Numerically, we approximate the distributions in eq. 14 and 15 by sub-sampling from the posterior parameter sample obtained from the MCMC chains (n=1000), and using the 5% and 95% percentiles as lower and upper limits, respectively. For the interval given in eq. 15 one would expect approximately 90% or more of the observations within the interval as it covers 90% of the observation error (described by the negative binomial distribution) plus the parameter uncertainty.

Results and Discussion

Parameter inference

First, we compare posterior with prior marginal distributions to investigate about which parameters we learned from the data. A gain in knowledge is reflected in a narrower and/or shifted marginal posterior distribution compared to the prior. A posterior that is located far from the high probability regions of the prior indicates a prior-data conflict which could indicate a

misuse of the parameter to compensate for model structure deficits or inadequate prior information.

Many parameters could be identified well by the inference (see SI for a comparative display of all prior and marginal posterior distributions, Figure S6). For example, the marginal posteriors of the taxon-specific parameters $f_{\text{basal tax}}$ and $f_{\text{gro tax}}$, modifying the basal metabolic rate and the growth rate, were most often identifiable, unequal and different from the prior with a mean of 1 (e.g., Figure 2 A, B). Hence, the modelled vital rates were not only determined by mass and temperature, as implied by the metabolic theory of ecology, but there was some taxon-specific deviation from this general pattern.^{25, 34} Additionally, in the model some taxa turned out to be clearly density mediated, with small h_{dens} values (e.g. *Sympetrum striolatum*, Figure 2 D) while others with high h_{dens} values were not limited by self-inhibition (e.g. *Asellus aquaticus*, Figure 2 C).

For some parameters we did not learn from the data (e.g. $f_{\text{e shredders}}$, the fraction of food lost during shredding by excretion and sloppy feeding; Figure 2 E). For those parameters, the posterior is similar to the prior distribution.

A potential prior-data conflict was present for the initial density of CPOM (coarse particulate organic matter) and the mineralization constant of FPOM (fine particulate organic matter), where the marginal posterior was shifted toward substantially larger values compared to the prior (Figure 2 F, G). Similarly, the estimates of other parameters determining algae and POM dynamics resulted in high initial densities (D_{ini}) followed by declines (k_{miner} , $h_{\text{dens algae}}$). The main reason for this was the observed population dynamics of *A. aquaticus*, which first rises and then declines (Figure 3). It has a strong influence on the likelihood estimation due to its relatively large abundance. To match its pattern in the simulations, all food sources of *A. aquaticus* (algae, CPOM and FPOM) needed to start at high levels and to subsequently decrease due to self-

inhibition (algae) or mineralization (CPOM and FPOM). However, as no measurements on algae density and organic matter were available, it was impossible to verify the posterior estimates for these parameters. Hence, including measurement of food sources in the experiment may be very important for process understanding. Still, the total abundance of all observed taxa increased at the beginning of the experiment and decreased toward the end.^{21, 22} This indicates that overall food resources could indeed start at ample amounts and decrease toward the end, which supports confidence in the parameter estimates.

Some of the model parameters were strongly correlated in the posterior (Spearman's $\rho > 0.7$; SI, Figure S7). In the case of such a strong correlation, the marginals should not be interpreted independently of each other. For 8 out of 16 taxa, the taxon-specific factor modifying the growth rate, $f_{\text{gro tax}}$, correlated negatively with the parameter describing self-inhibition, h_{dens} . The direction of this correlation was as would be expected from eq. 3, 6 and 7. This correlation was often present for taxa that were self-inhibited in the model. Such strong correlations could be reduced by reparameterizing the model. However, to preserve the biological or physical meaning of the parameters and to maintain Streambug's universality we did not consider a reparametrization.³⁵

Regarding the LC_{50} values (Figure 2 H-W), for some taxa we could not learn much from the data, probably because their dynamics were generally difficult to simulate due to strongly fluctuating observations (*Hydrometra stagnorum*) or very small population size (*Libellula quadrimaculata*). For others, the marginal posteriors were shifted toward higher values (*Ischnura elegans*, *Orthetrum coerulescens*, *Sympetrum striolatum*), often with a sharp lower limit in the posterior at $\log_{10}(LC_{50}) = 2$, corresponding to the highest concentration tested of 100 $\mu\text{g/L}$. This resulted in no or small simulated direct lethal effects of the tested thiacloprid concentrations. For another group of taxa, the marginal LC_{50} posteriors were markedly delimited to a small range

(*Asellus aquaticus*, Chironomidae, Oligochaeta, *Simulium latigonium*, *Succinea*), meaning that the model is quite sensitive to these parameters in both directions. For some taxa, the LC_{50} posterior was shifted toward markedly smaller values compared to the prior (e.g. Chironomidae, *Lymnea*, *S. latigonium*, *Succinea*). Hence, when modelling the effects observed in the experiment, these taxa were considerably more sensitive than expected from LC_{50} experiments. In these cases, acute lethal effects might be misused to explain sub-lethal effects, which are not included in the model. However, the strong acute decline of some of the taxa under concern in the observed abundances suggests that acute lethal effects are likely. Hence, this finding may rather indicate that acute lethal sensitivities were higher in the mesocosm community than can be expected from laboratory experiments without competitors, predators and food limitation. Finally, for some taxa the LC_{50} posterior distribution was multimodal, around the second highest and the highest concentration tested (*Notidobia ciliaris*, Figure 2 R), or even all three concentrations tested (*Limnephilus lunatus*, Figure 2 P). This can be explained by the way we implemented the contamination effect: The lower limit for a simulated acute lethal effect lay $0.5 \log_{10}$ units below the LC_{50} concentration and then increased linearly, resulting in a relatively steep increase in mortality. Because there was a difference of $1.5 \log_{10}$ units between the concentrations tested, the inference algorithm traded off matching the observations of one of the concentrations, respectively, while not fitting those of the other concentrations, either due to no effect or extinction caused by very high mortality. Hence, smaller differences between the concentrations tested might be relevant. .

Some parameters had to be fixed, omitting thereby the related uncertainty. We defined the initial densities of invertebrates, D_{ini} , using the first observed abundances. This implies that we assume no observation error in those measurements while for all other observations we described the sampling error by means of the likelihood formulation. However, a preliminary trial to infer

345 the D_{ini} together with the other parameters resulted in parameter values that completely dampen
346 the population dynamics so that the simulations just represented the average population sizes
347 without much dynamics. Therefore, it was not meaningful to include them in the inference.
348 Likewise, we fixed the value of the average body mass of the taxa using estimates from measured
349 body mass data. As the observed abundances contain no information on the mass of the sampled
350 individuals, there is no point in inferring the average body mass from this data. Moreover, the
351 average mass and the taxon-specific factors are highly correlated (eq. 1 and 2) and therefore
352 inferring all of them would hardly be possible.

353 Altogether, the parameter inference was quite resource intensive. Even when optimized starting
354 values for the parameters were used and the four streams and ten chains were processed in
355 parallel using CPUs with 2.3 – 2.53 GHz, it took approx. eight weeks to estimate the posterior
356 distributions.

Figure 2. Examples of marginal prior probability distributions (grey area), marginal posterior probability distributions (black line) and the parameter value with the overall maximum posterior probability (dashed line).

Simulating population dynamics and model fit

In general, it was possible to simulate community dynamics under time-varying environmental conditions, including pesticide exposure with the calibrated Streambugs model (examples in Figure 3; SI, Figure S8 all taxa and all concentrations). Maintaining coexistence was achieved, which can be difficult in a dynamic model with many interacting taxa. The uncertainty intervals were fairly narrow with a mean width in relation to the mean simulated abundance of 1.57. When comparing simulated and observed abundances, the calibrated model fitted the observations reasonably well, with approx. 85% of the observation points lying within the 90% uncertainty interval covering parameter and observation uncertainty (eq. 15). This is only slightly less than what would be expected if the observation error was correctly described by our assumptions. Estimating k in the negative binomial distribution from observational data using several replicates of individually collected and identified samples could improve the description of overdispersion.

For some taxa, the simulations matched the overall dynamics well. For instance, there was a strong increase in abundance of *A. aquaticus* simulated from September 2005 to May 2006, which is consistent in all observations (Figure 3). Additionally, the observed sharp decline after contamination in the highest concentration was simulated correctly. Likewise, the overall dynamics of *N. ciliaris* were matched well and all observations lay within the uncertainty interval (Figure 3). However, consistently in all control streams no individuals of this taxon were observed from mid-May to mid-June in both years. This may indicate that these zero observations resulted from the emergence of *N. ciliaris* during that time or another process that is not captured in Streambugs (and not from sampling errors, which are included in the intervals).

Streambugs allows for investigating the dynamics of vital rates and taxon-specific factors (SI, Figure S8), facilitating the analysis of recovery patterns. After the contamination effect in the highest concentration, there was a rapid increase in simulated Chironomidae abundance also

present in the observations. This can be explained by the release of strong self-inhibition present before contamination. Food sources (organic matter and algae) were available in high quantities to support the strong growth. The subsequent relatively rapid decline was due to a sharp increase in self-inhibition and due to predation, the latter also partly driving the fluctuating nature of the abundance. In contrast, *A. aquaticus*, which feeds on the same food sources (except from SusPOM (suspended particulate organic matter)) but was not self-inhibited before the contamination, recovered much slower. Likewise, *N. ciliaris*, not being self-inhibited, recovered slowly after the first and did not recover at all after the second contamination. The reason for this extinction (also present in all other concentrations and in the control streams in both observations and simulations) was the strong decrease in CPOM concentration, the only food source for *N. ciliaris* in the model. Hence, if there are no other limitations present preventing rapid population growth, populations that are self-inhibited before the contamination effects may recover faster compared to non-self-inhibited ones.^{36,37}

For some taxa, at most the average abundance was matched by the simulations while the dynamics of observations and simulations differed substantially. E.g. for *S. striolatum* (Figure 3) only the average abundance and a general upward trend was modelled correctly. Due to its high average body mass, *S. striolatum* was the second top predator in the system, according to the model that infers feeding links from feeding types and body mass. Hence, its population dynamics may have strong influence on several other taxa. Therefore, there might be a trade-off between fitting its population dynamics and that of other, more abundant taxa, which have a stronger influence on the likelihood. In general, the population dynamics of rare taxa with only up to four individuals per observation time point were not simulated well (e.g. *Hydropsyche angustipennis*, *Ischnura elegans*). However, the total difference between observations and simulations were small and, therefore, the influence of rare taxa on the likelihood was also small.

410 Furthermore, small populations might behave more stochastically compared to larger ones, which
411 makes simulation by a deterministic model more difficult. Additionally, they may require special
412 monitoring schemes.

413 These difficulties to simulate certain taxa reveal the challenges when increasing realism in
414 ecotoxicological modelling by simulating communities under semi-natural conditions in contrast
415 to single populations in laboratory systems where good model fits are generally easier to
416 achieve.^{5, 38} Additionally, it emphasizes that knowledge is still missing for straightforward
417 predictions to real world situations.³⁹

418

Figure 3. Population dynamics over time of selected taxa and concentrations. Blue dots: mean observed abundance (used for parameter inference ($N_{\text{obs } i}$) except the first observation data point); black line: simulation run with the parameter set at maximum posterior density ($N_{\text{det } i}(\theta_{\text{max post}})$); dark grey area: 90th percentile of simulation results sampled from the posterior ($n=1000$, $N_{\text{det } i}(\theta_{\text{post}})$); light grey area: predicted results for new observations ($\int P(N_i|\theta) f_{\text{post}}(\theta) d\theta$ constructed by adding a random number drawn from the respective negative binomial distribution to each simulation result); arrows: contamination events.

One of the aims of this study was to identify how the Streambugs model could be improved for future applications. Additionally, we aimed at determining what supplementary experimental and observational data would support a better process understanding.

Following the parsimony principle and accounting for availability of prior information, we chose a relatively simple way of modelling toxicant effects. This could be improved if data on full dose-response relationships at best also including the effect of exposure time was available for all taxa. Unfortunately, such data is rarely published and already collecting prior information on LC_{50} s (which are severely criticized for describing sensitivity⁴⁰) was challenging. A more sophisticated approach, e.g. based on toxicokinetics and toxicodynamics (TK/TD), might be even more realistic.⁴¹ Of particular relevance might be that we only considered acute lethal effects, although sub-lethal as well as chronic effects can play a substantial role, especially at smaller concentrations.⁴² Including sub-lethal effects might also prevent the multimodal distribution in the LC_{50} posteriors. Rapid sensitivity tests might be an option to produce prior information for both lethal and sub-lethal sensitivities.⁴³ This approach tests organisms directly after collection from the field and thereby increases both the number of tests compared to standard laboratory tests in a given time as well as the number of tested species substantially. The results might not be as precise as standard tests, but they would allow prior distributions for all taxa in the mesocosms to be estimated.

We included emergence by reducing the simulated abundance by means of a time dependent, taxon-specific factor in the likelihood calculation. However, this approach does not account for the fact that small abundances during emergence may also influence community dynamics as the emerging taxon is then missing in biological interactions (i.e. lacking as a food source for predators, exerting smaller competitive and predatory pressure). Therefore, including emergence in a process based way in Streambugs might be worthwhile, e.g. by explicitly modelling different life stages. This could also help to improve the food web relationships that are currently based on average body masses, which can be unrealistic if different life stages have considerably different sizes. Distributed delays models⁴⁴ or matrix models⁴⁵ might be a starting point for model

extension in this regard. Another option are individual based models (IBMs), which are currently popular in modelling ecotoxicological effects particularly in single populations.⁴⁶ Such a stochastic modelling approach would also allow incorporating the variability among the stream replicates. Nevertheless, all these approaches come at the cost of higher model complexity, adding additional uncertainty on model structure and parameters. Modelling different life stages also calls for observational data differentiating between them, further increasing the observation effort.

Furthermore, the food web, particularly the predator-prey relationships, could be improved by using stomach content or isotopic analysis. However, this information is not yet available for all taxa modelled.

In summary, on the modelling side, in this study, inclusion of the emergence process and sub-lethal effects turned out to be potentially relevant extensions. On the experimental side, measurements of the food sources and additional knowledge on sub-lethal effects would particularly support more accurate modelling. Additionally, smaller differences between the tested concentrations, special monitoring of rare species, measurement of size classes, and estimates of overdispersion from replicated samples to increase knowledge on the sampling error may be helpful.

Acknowledgements

We thank Marion Junghans from the Oekotoxzentrum for ecotoxicological information on thiacloprid and imidacloprid.

Supporting Information Available

Additional information on the model, prior and posterior distributions and modelling results can be found in the supporting information.

References

- Schmolke, A.; Thorbek, P.; Chapman, P.; Grimm, V. Ecological models and pesticide risk assessment: current modeling practice. *Environ. Toxicol. Chem.* **2010**, *29* (4), 1006-1012.
- EFSA PPR Panel Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products *EFSA Journal* **2014**, *12* (3), 92 pp.
- Grimm, V.; Thorbek, P. Population models for ecological risk assessment of chemicals: Short introduction and summary of a special issue. *Ecol. Model.* **2014**, *280*, 1-4.
- Agatz, A.; Cole, T. A.; Preuss, T. G.; Zimmer, E.; Brown, C. D. Feeding Inhibition Explains Effects of Imidacloprid on the Growth, Maturation, Reproduction, and Survival of *Daphnia magna*. *Environ. Sci. Technol.* **2013**, *47* (6), 2909-2917.
- Johnston, A. S. A.; Hodson, M. E.; Thorbek, P.; Alvarez, T.; Sibly, R. M. An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides. *Ecol. Model.* **2014**, *280*, 5-17.
- Qin, G. Q.; Presley, S. M.; Anderson, T. A.; Gao, W. M.; Maul, J. D. Effect of predator cues on pesticide toxicity: Toward an understanding of the mechanism of the interaction. *Environ. Toxicol. Chem.* **2011**, *30* (8), 1926-1934.
- Knillmann, S.; Stampfli, N. C.; Noskov, Y. A.; Beketov, M. A.; Liess, M. Interspecific competition delays recovery of *Daphnia* spp. populations from pesticide stress. *Ecotoxicology* **2012**, *21*, 1039-1049.
- Kattwinkel, M.; Liess, M. Competition matters: Species interactions prolong the long-term effects of pulsed toxicant stress on populations. *Environ. Toxicol. Chem.* **2014**, *33* (7), 1458-1465.
- Liess, M.; Foit, K.; Becker, A.; Hassold, E.; Dolciotti, I.; Kattwinkel, M.; Duquesne, S. Culmination of low-dose pesticide effects. *Environ. Sci. Technol.* **2013**, *47* (15), 8862-8868.
- Gergs, A.; Zenker, A.; Grimm, V.; Preuss, T. G. Chemical and natural stressors combined: from cryptic effects to population extinction. *Sci. Rep.* **2013**, *3*, 8 pp.
- Gabsi, F.; Schaffer, A.; Preuss, T. G. Predicting the sensitivity of populations from individual exposure to chemicals: The role of ecological interactions. *Environ. Toxicol. Chem.* **2014**, *33* (7), 1449-1457.
- Kooi, B. W.; Bontje, D.; Liebig, M. Model analysis of a simple aquatic ecosystem with sublethal toxic effects. *Math. Biosci. Eng.* **2008**, *5* (4), 771-787.
- De Laender, F.; Morselli, M.; Baveco, H.; Van den Brink, P. J.; Di Guardo, A. Theoretically exploring direct and indirect chemical effects across ecological and exposure scenarios using mechanistic fate and effects modelling. *Environ. Int.* **2015**, *74*, 181-190.
- Traas, T. P.; van Wezel, A. P.; Hermens, J. L. M.; Zorn, M.; van Hattum, A. G. M.; van Leeuwen, C. J. Environmental quality criteria for organic chemicals predicted from internal effect concentrations and a food web model. *Environ. Toxicol. Chem.* **2004**, *23* (10), 2518-2527.
- Sourisseau, S.; Basseres, A.; Perie, F.; Caquet, T. Calibration, validation and sensitivity analysis of an ecosystem model applied to artificial streams. *Water Res.* **2008**, *42* (4-5), 1167-1181.
- Gelman, A.; Carlin, J.; Stern, H.; Dunson, D.; Vehtari, A.; Rubin, D. *Bayesian Data Analysis*. Third ed.; Chapman & Hall: U.K., 2014.

17. Ellison, A. M. Bayesian inference in ecology. *Ecol. Lett.* **2004**, 7 (6), 509-520.
18. Reichert, P.; Langhans, S. D.; Lienert, J.; Schuwirth, N. The conceptual foundation of environmental decision support. *J. Environ. Manage.* **2015**, 154, 316-332.
19. Schuwirth, N.; Reichert, P. Bridging the gap between theoretical ecology and real ecosystems: modeling invertebrate community composition in streams. *Ecology* **2013**, 94 (2), 368-379.
20. Schuwirth, N.; Dietzel, A.; Reichert, P. The importance of biotic interactions for the prediction of macroinvertebrate communities under multiple stressors. *Funct. Ecol.* **2015**, doi: 10.1111/1365-2435.12605.
21. Beketov, M.; Schäfer, R. B.; Marwitz, A.; Paschke, A.; Liess, M. Long-term stream invertebrate community alterations induced by the insecticide thiacloprid: Effect concentrations and recovery dynamics. *Sci. Total Environ.* **2008**, 405, 96-108.
22. Liess, M.; Beketov, M. Traits and stress: keys to identify community effects of low levels of toxicants in test systems. *Ecotoxicology* **2011**, 20 (6), 1328-1340.
23. Krohn, J. Behaviour of thiacloprid in the environment. *Pflanzenschutz Nachrichten - Bayer - English Edition* **2001**, 54, 281-290.
24. R Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
25. Brown, J. H.; Gillooly, J. F.; Allen, A. P.; Savage, V. M.; West, G. B. Toward a metabolic theory of ecology. *Ecology* **2004**, 85 (7), 1771-1789.
26. Schmidt-Kloiber, A.; Hering, D. www.freshwaterecology.info - An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. *Ecol. Indic.* **2015**, 53, 271-282.
27. Bis, B.; Usseglio-Polatera, P. *Species Traits Analysis*; 2004; 134 pp.
28. Oekotoxzentrum EQS - Vorschlag des Oekotoxzentrums für: Imidacloprid (*Suggestion of the Oekotoxzentrum for an environmental quality standard for imidacloprid*); Oekotoxzentrum, Schweizerischen Zentrum für angewandte Ökotoxikologie, Eawag - EPFL: 2013; 39 pp.
29. Gamerman, D.; Lopes, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Taylor & Francis Inc.: U.K., 2006.
30. Van Oijen, M.; Rougier, J.; Smith, R. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol.* **2005**, 25 (7), 915-927.
31. Elliot, J. M. *Some methods for the statistical analysis of samples of benthic invertebrates*. Freshwater Biological Association: Kendal, 1971; 144 pp.
32. Gray, B. R. Selecting a distributional assumption for modelling relative densities of benthic macroinvertebrates. *Ecol. Model.* **2005**, 185 (1), 1-12.
33. U.S. Environmental Protection Agency (EPA) ECOTOX User Guide: ECOTOXicology Database System. Version 4.0. <http://www.epa.gov/ecotox/>
34. Peters, R. H. *The Ecological Implications of Body Size*. Cambridge University Press: 1983.
35. Reichert, P.; Omlin, M. On the usefulness of overparameterized ecological models. *Ecol. Model.* **1997**, 95, 289-299.
36. Forbes, V. E.; Sibly, R. M.; Linke-Gamenick, I. Joint effects of population density and toxicant exposure on population dynamics of *Capitella* sp. I. *Ecol. Appl.* **2003**, 13 (4), 1094-1103.
37. Moe, S. J.; Stenseth, N. C.; Smith, R. H. Density-dependent compensation in blowfly populations give indirectly positive effects of a toxicant. *Ecology* **2002**, 83 (6), 1597-1603.
38. Belleza, E. L.; Brinkmann, M.; Preuss, T. G.; Breitholtz, M. Population-level effects in *Amphiascus tenuiremis*: Contrasting matrix- and individual-based population models. *Aquat. Toxicol.* **2014**, 157, 207-214.
39. Segner, H. Ecotoxicology - How to assess the impact of toxicants in a multi-factorial environment? In *Multiple Stressors: A Challenge for the Future*, Mothersill, C.; Mosse, I.; Seymour, C., Eds. 2007; pp 39-56.

- 568 40. Jager, T. Some Good Reasons to Ban ECx and Related Concepts in Ecotoxicology. *Environ. Sci.*
569 *Technol.* **2011**, 45 (19), 8180-8181.
- 570 41. Ashauer, R.; Agatz, A.; Albert, C.; Ducrot, V.; Galic, N.; Hendriks, J.; Jager, T.; Kretschmann, A.;
571 O'Connor, I.; Rubach, M. N.; Nyman, A. M.; Schmitt, W.; Stadnicka, J.; van den Brink, P. J.; Preuss, T. G.
572 Toxicokinetic-toxicodynamic modeling of quantal and graded sublethal endpoints: A brief discussion of
573 concepts. *Environ. Toxicol. Chem.* **2011**, 30 (11), 2519-2524.
- 574 42. Agatz, A.; Brown, C. D. Evidence for Links between Feeding Inhibition, Population Characteristics,
575 and Sensitivity to Acute Toxicity for *Daphnia magna*. *Environ. Sci. Technol.* **2013**, 47 (16), 9461-9469.
- 576 43. Kefford, B. J.; Palmer, C. G.; Warne, M. S.; Nuggeoda, D. T. What is meant by "95% of species"?
577 An argument for the inclusion of rapid tolerance testing. *Hum. Ecol. Risk Assess.* **2005**, 11 (5), 1025-1046.
- 578 44. Kjaer, C.; Elmegaard, N.; Axelsen, J. A.; Andersen, P. N.; Seidelin, N. The impact of phenology,
579 exposure and instar susceptibility on insecticide effects on a chrysomelid beetle population. *Pesticide*
580 *Science* **1998**, 52 (4), 361-371.
- 581 45. Stark, J. D. Demography and Modeling To Improve Pesticide Risk Assessment of Endangered
582 Species. In *Pesticide Regulation and the Endangered Species Act*, Racke, K. D.; McGaughey, B. D.; Cowles,
583 J. L.; Hall, A. T.; Jackson, S. H.; Jenkins, J. J.; Johnston, J. J., Eds. 2012; Vol. 1111, pp 259-270.
- 584 46. Galic, N.; Forbes, V. Ecological models in ecotoxicology and ecological risk assessment: an
585 introduction to the special section. *Environ. Toxicol. Chem.* **2014**, 33 (7), 1446-1448.

586