Peer Community In
Ecotoxicology & Environmental Chemistry

RESEARCH ARTICLE

Open Access

Open Peer-Review

Open Data

Open Code

Chemical effects on ecological interactions within a model-experiment loop

Dominique Lamonica[1,2], Sandrine Charles[1], Bernard Clément[2], & Christelle Lopes[1]

[1] Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

[2] Université de Lyon, F-69000, Lyon; Université Lyon 1; ENTPE; CNRS, UMR 5023, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés; 3, rue Maurice Audin, 69518 Vaulx-en-Velin, France

Cite as:

Posted: ddth Month yyyy

Recommender:

Reviewers:

Correspondence:

dominique.lamonica@univ-lyon1.fr

This article has been peer-reviewed and recommended by

Peer Community In Ecology (https://doi.org/)

Abstract

We propose in this paper a method to assess the effects of a contaminant on a micro-ecosystem, integrating the population dynamics and the interactions between species. For that, we developed a dynamic model to describe the functioning of a microcosm exposed to a contaminant and to discriminate direct and indirect effects. Then, we get back from modelling to experimentation in order to identify which of the collected data have really been necessary and sufficient to estimate model parameters in order to propose a more efficient experimental design for further investigations. We illustrated our approach using a 2-L laboratory microcosm involving three species (the duckweed Lemna minor, the microalgae Pseudokirchneriella subcapitata and the daphnids Daphnia magna) exposed to cadmium contamination. We modelled the dynamics of the three species and their interactions using a mechanistic model based on coupled ordinary differential equations. The main processes occurring in this three-species microcosm were thus formalized, including growth and settling of algae, growth of duckweeds, interspecific competition between algae and duckweeds, growth, survival and grazing of daphnids, as well as cadmium effects. We estimated model parameters by Bayesian inference, using simultaneously all the data issued from multiple laboratory experiments specifically conducted for this study. Cadmium concentrations ranged between 0 and 50 μg.L^{-1}. For all parameters of our model, we obtained biologically realistic values and reasonable uncertainties. The cascade of cadmium effects, both direct and indirect, was identified. Critical effect concentrations were provided for the life history traits of each species. An example of experimental design adapted to this kind a microcosm was also proposed. This approach appears promising when studying contaminant effects on ecosystem functioning.

Keywords: Bayesian inference ; microcosm ; experimental design optimisation ; cadmium ; daphnid ; duckweed ; microalgae

1 Introduction

The toxic effects of contaminants are most often studied at the individual level, since it is easier to study life history traits of an isolated organism, studying for example its survival, development or capacity to reproduce. Moreover, monospecific bioassays are easy to implement and perform, and observation data at the individual level are straightforward to analyse, since they depict the direct effects of contaminants. Nevertheless effects can also be measured at other levels of biological organisation using various experimental devices that are chosen related to the level of interest, by adapting several characteristics uch as size, duration, number of species, abiotic compartment, etc (**Calow1993**). Multispecies devices, like microcosms and mesocosms, allow to study organisation levels from populations to ecosystem, by integrating population dynamics and interactions between species (**Forbes1997**; **Kimball1985**; **Ramade2002**). However, extrapolating toxic effects from one biological level to the next based on observation data remains a challenge. In particular, going from individual to population levels, or from population to community levels, implies taking into account intra- and inter-specific interactions, which are of major importance in the functioning of ecosystems, while it is necessary to integrate these interactions for a better assessment of the ecotoxicological risk (**Cairns(1984); De Laender et al.(2008); Preston(2002)**). Modelling tools have proven their utility to analyse ecotoxicological data, by highlighting the underlying mechanisms leading to observations at each level of biological organisation. But modelling appears particularly helpful when extrapolation of contaminant effects from biological levels reveals necessary.

Ecotoxicology relies on experimental data, while being concerned by the Replacement, Refinement and Reduction of Animals in Research (3Rs) program **Kilkenny2009a** and by difficulties linked to collection of field data. Taking the most of experimental data and reduc-

ing the amount of experiments to perform in general is a key issue the ecotoxicology field faces. Formal optimisation of experimental design can be applied to standard tests (namely monospecific bioassays): they have been questioned in terms of test duration and measured endpoints **Charles2016** or regarding the tested concentration range **Forfait-Dubuc2012**. Yet, more complex experimental designs, like microcosms or mesocosms may resist to formal optimisation. Some attempts have been made in simple cases to deal with standard dose-responses curves **Chevre2008a**; **Holland2015a**; **Keddig2015a**; **Khinkis2003a**; **Sitter1995a**; **Wang2006a** but to our knowledge, nothing similar exist for multi-species models. Nevertheless, when modelling has been integrated to the experimental framework, it can easily be used to evaluate *a posteriori* the relevance of the data, as a pragmatic and case-by-case method to analyse the information provided by data and possibly improve the experimental design.

The aim of our paper is to illustrate (1) how to use modelling to describe the functioning of a three-species microcosm exposed to a contaminant and to discriminate direct and indirect effects; and (2) how to get back from modelling to experimentation in order to identify which of the collected data have really been necessary and sufficient to estimate model parameters in order to propose a more efficient experimental design for further investigations.

Different steps have been set up to achieve our objectives, as summarised in Figure 1. We performed experiments to collect data on the microcosm species populations at different cadmium concentrations. In parallel, we formulated a model of the microcosm functioning under a chemical stressor based on coupled ordinary differential equations (ODE) and effect functions. First, using all data we estimated model parameters, in particular those related to effect functions (Figure 1, black boxes). This permitted to identify direct and indirect effects of cadmium on the population dynamics of the different species. We then globally analysed

the perturbations of our small community (objective 1). We also extracted EC_{50} for the different processes (Figure 1, orange boxes). In order to assess the relevance of certain data, we removed those data from the complete dataset to build partial datasets. Then, we estimated function parameters with the partial datasets. The newly estimated effect functions were then compared to the reference ones obtained with the complete dataset (Figure 1, green boxes). This allowed us to evaluate the added values of only considering partial datasets instead of the complete original one (objective 2).

2 Experiments and observed data

Experimental design Microcosms were identically prepared for all experiments according to **Lamonica2015**. Algae, duckweeds and daphnids were cultivated at the laboratory according to internal protocols **Clement2014**. According to the experiment, beakers were inoculated with one, two or three species at the start of the experiment (day 0). When algae were present, 4.10^7 cells of *P. subcapitata* were introduced into beakers. When daphnids were present, 10 daphnids (*Daphnia magna* neonates aged 24 ± 12 h) were introduced into beakers. When duckweeds were present, 8 fronds of duckweeds were introduced into beakers. The algal density in the water column was measured every two to three days with a particle counter **Lamonica2015**. The algal density at the bottom of the beakers was measured once during the experiment **Lamonica2016**. Daphnids neonates were removed from the microcosm every two days, meaning that reproduction was considered as an independent process in the microcosm functioning **Lamonica2015**. The number of daphnids in each beaker was counted (after neonate removal if necessary) and their size measured (from the centre of the eye to the caudal base of spine) twice or thrice per week. The duckweed fronds were counted every

two to three days. The experiments lasted between 13 and 21 days.

Experiments without cadmium Experiment 1 involved algae and daphnids as detailed in **Lamonica2015** (referred as "Experiment without sediment" in section 2.3.2.). Experiment 2 involved algae alone as detailed in **Lamonica2016** (referred as "Experiment 2" in section 2.2.1.). Experiment 3 involved algae and duckweeds as detailed in **Lamonica2016** (referred as "Experiment 3" in section 2.2.2.).

Experiments with cadmium Experiment 4 involved algae and duckweeds, with two conditions in species composition: duckweeds alone, and algae and duckweeds together. We tested five different cadmium concentrations (0, 11.1, 20.2, 35.5 and 51.1 μg/L) in triplicate for each condition. Three additional control beakers were inoculated with algae alone. The duration of this experiment was 14 days. From this experiment, we obtained different types of data under contaminant exposure: "monospecific data, duckweeds", "two species data, duckweeds" and "two species data, algae".

Experiment 5 involved the three species, with three conditions in species composition: duckweeds alone, algae and duckweeds, and algae, duckweeds and daphnids . We tested five different cadmium concentrations in triplicate for each condition (0, 2.25, 4.50, 6.88 and 9.09 μg/L). The duration of this experiment was 21 days. From this experiment, we obtained the following data under contaminant exposure: "monospecific data, duckweeds"; "two species data, duckweeds" and "two species data, algae"; "complete microcosm data, duckweeds", "complete microcosm data, algae" and "complete microcosm data, daphnids".

Experiment 6 involved algae alone. We tested five different cadmium concentrations (0, 26.2, 36.4, 40.8 and 43.6 μg/L) in triplicate. From this experiment, we obtained "monospecific data, algae".

As mentioned in **Lamonica2016**, we used measured cadmium concentrations in the medium instead of nominal ones. For that purpose, we measured dissolved cadmium concentrations as described by Clement et al. **Clement2014** at days 2, 7, 14 (and day 21 for Experiment 5) in each beaker. We then calculated the arithmetic mean of all the measurements. In total, for Experiments 4, 5 and 6, we thus obtained 13 concentrations (0, 2.25, 4.50, 6.88, 9.09, 11.1, 20.2, 35.5, 51.1, 26.2, 36.4, 40.8 and 43.6 μg/L) denoted by C_j, $j \in [0, 12]$ hereafter. The concentration in the controls of Experiments 4, 5 and 6 (that is with no contaminant) is denoted by C_0, corresponding to index $j = 0$. This is also the case in Experiments 1 to 3, that were conducted without contaminant.

3 Dynamic modelling

The description of the model follows the Overview, Design concepts and Details (ODD) protocol originally used for describing individual and agent-based models **Grimm2010** but adapted here for a dynamic model based on Ordinary Differential Equations (ODE). Below an overview of the model, with details on cadmium effects. The complete description can be found in Supplementary Information.

3.1 Processes and assumptions

Two processes are related to intrinsic algal dynamics: settling of suspended algae, described by an exponential decay of algal cells in the water column, and logistic growth of both suspended and settled algae **Lamonica2015**. Duckweed growth is related to the intrinsic duckweed dynamics through a logistic function **Lamonica2016**. The algae-duckweed interaction

is an interspecific competition, described with a Lotka-Volterra type I model **Lamonica2016**. Two processes are related to daphnid life history traits: survival (see section 3.3.1) and individual growth, described by a Von Bertalanffy model **Lamonica2015**. Two processes are related to algae-daphnid interaction: location of daphnid for grazing and ingestion of algae by daphnids **Lamonica2015**.

All assumptions are based on the experimental design described in "Experiments and observed data" section. As mentioned in **Lamonica2015**, we assume that algae are uniformly distributed in the water column and at the bottom of the beaker at each time step. We also assume that the settling speed of suspended algae is constant throughout the water column. Therefore, the water volume occupied by the suspended algae is supposed to decrease at the same speed as algal settling. As mentioned in **Lamonica2016**, we assume that algae and duckweeds are competing only for nutrients in the medium. We then assume that settled algae are too distant from duckweeds to interact with them, so that the interspecific competition only involves suspended algae. Finally, nterspecific competition has no effect on algae, as shown in **Lamonica2016**.

3.2 Variables and stochasticity

The deterministic part of algal dynamics in both compartments, as well as duckweed dynamics over time t (in days) is described with three coupled ODE. The deterministic part of daphnid survival and size are described with two other ODE. The model thus involves five state variables. The two first ones refer to the numbers of algal cells per beaker in the two compartments of the microcosm at time t and cadmium concentration C_j: the suspended algae in the water column (Compartment 1), and the settled algae at the bottom of the beaker (Compartment 2). The third state variable is the number of duckweed fronds per beaker at

time t and cadmium concentration C_j. The two other state variables refer to the number of alive daphnids in the microcosm through survival rate at time t and cadmium concentration C_j, and to the daphnid size at time t and cadmium concentration C_j.

The model is run on 21 days, corresponding to the duration of the longest experiment. We use stochasticity to describe variability on state variables, which sum up both uncertainties and variability sources within the processes.

3.3 Cadmium effect

As mentioned in **Lamonica2016**, we assume that cadmium affects growth rates of all species, as well as competition intensity parameters and daphnid survival. Cadmium is supposed not to affect the carrying capacities of both algae and duckweeds nor the algal settling rate.

3.3.1 Daphnid survival

Survival rate at time t and cadmium concentration C_j, denoted $S(t, C_j)$, is described by an exponential decay with an instantaneous mortality rate, m_0 (day^{-1}), which is assumed to be time-independent **Forfait-Dubuc2012**:

$$S(t, C_j) = \exp(-(m_0 + k_s \times \max(0, C_j - NEC)) \times t) \tag{1}$$

where k_S (μg^{-1}.L.day^{-1}) represents the cadium effect intensity and NEC (No Effect Concentration) (μg.L^{-1}) is the concentration from which the contaminant has an effect on survival. When concentration Cj is below the NEC, $\max(0, C_j - NEC)$ is equal to 0, thus there is no effect on survival rate which only depends on natural mortality m_0 and time t. However, when concentration Cj is greater than the NEC, $\max(0, C_j - NEC)$ is equal to the surplus of concentration and mortality due to cadmium is added to the natural mortality.

We consider a conditional binomial stochastic model for $D_s(t, C_j)$, the number of alive daphnids at time t and cadmium concentration C_j in the system **Forfait-Dubuc2012**:

$$D_s(t, C_j) \sim \mathcal{B}\left(\frac{S(t, C_j)}{S(t-1, C_j)}, D_s(t-1, C_j)\right) \tag{2}$$

where $\mathcal{B}$ stands for the binomial law. For each concentration C_j, the number of alive daphnids at time t depends on the number of alive daphnids at time $t-1$ and on the survival probability between $t-1$ and t, represented by $\frac{S(t)}{S(t-1)}$. We make here the implicit assumption that contaminant toxicokinetics is fast (which means that the internal concentration in the organism is supposed to be equal to the external concentration in the water C_j).

3.3.2 Growth rates and parameters of competition intensity

We choose a three-parameter log-logistic function to describe the effect of cadmium at concentration C_j on each parameter p susceptible to be modified as a function of the concentration:

$$p(C_j) = \frac{p_0}{1 + \left(\frac{C_j}{E_p}\right)^{b_p}} \tag{3}$$

where p_0 is the value of parameter p in the control (without contaminant), E_p is the cadmium concentration at which $p(C_j) = \frac{p_0}{2}$, which is equivalent to a 50% Effective Concentration (EC_{50}), and b_p is a curvature coefficient of the log-logistic function whose value is all the greatest that the effect intensity is high. We refer hereafter to Eq.(3) as "stress functions".

4 Statistical inference

In order to check if our model satisfactorily described the microcosm functioning, we used Bayesian inference to fit the model simultaneously to all our experimental data from the six above mentioned experiments. Estimates obtained for all the parameters are called "reference estimates" hereafter.

4.1 Parameter prior distributions

We defined prior distributions summarising all information on each parameter available in advance (SI Table S1). Some of the prior distributions described the decimal logarithm of the parameter because of an expected large range of possible values (for instance, b_{r_d}, k_S, and β_0) or extreme orders of magnitude (*e.g.*, large or small). Other prior distributions were defined based on previous experiments that were conducted using the same experimental device **Billoir2011**; **Billoir2012b**; **Delhaye2011a**; **Lamonica2015** (r_{a_0}), or based on additional experiments specifically conducted in the laboratory (E_{r_a}, b_{r_a}, r_{d_0}, β_0). At last, the remaining distributions were based on literature values **Billoir2008a**; **Biron2012**; **demott1982**; **Egloff1971** (k_0), except for parameters on which we had very vague information (E_β, b_β, b_{r_d}), so that their prior distributions were chosen as flat.

4.2 Computation

Monte Carlo Markov Chain (MCMC) computations were performed using the JAGS software via the *rjags* R package **Plummer2009**; **R**, after the model was discretised using the Euler method with a time step equal to 0.1 as stated in **Lamonica2015**. Three chains were run. A total of 20000 iterations was performed as a burn-in phase and inference was based on

100000 additional iterations for each of the three chains. To check the convergence of the estimation process, we used the Gelman and Rubin convergence diagnostic **Gelman1992**.

4.3 Posterior Predictive Check

To check posterior predictions of the model, we simulated new date at all experienced time steps and tested concentrations taking into account parameter uncertainties and stochasticity of the model **Lamonica2015**. 95% of the observed data are expected to be contained in the 95% credibility band of the predicted data, got from 2.5% and 97.5% percentiles of the predictions.

5 Look-back on the experimental design

We aim at determining which types of data could be sufficient to accurately (in terms of mode of the posterior distribution) and precisely (in terms of dispersion of the posterior distribution) estimate parameters of stress functions for the different species (Table 1). Our reference was the posterior distributions obtained when estimating parameters from the whole dataset, considering them as the "best possible estimates in the present case study in view of the model and all available data". To evaluate the information provided by certain data, we built partial datasets by removing these data from the whole dataset. Then, we re-estimated the model parameters using these partial datasets and compared the newly obtained estimates with the reference ones.

Four partial datasets (numbered A to D) were used to estimate the stress function parameters. They are summarized in Table 2. We evaluated only information provided by data collected under contaminated conditions, removing them successively to build the partial

datasets. Regarding data without contaminant, including the controls in Experiments 4 to 6, they were kept in all partial datasets.

Dataset A included all data except the "monospecific data" with contaminant. This corresponds to exclude data collected from beakers containing only one species. Thus, data from Experiment 6 and data "duckweeds alone" from Experiments 4 and 5 were not included in dataset A. Dataset B included all data except the "two species data" with contaminant. This corresponds to exclude data collected from beakers with both duckweeds and algae. Thus, duckweed and algae data collected from beakers containing both duckweeds and algae from Experiments 4 and 5 were not included in dataset B. In dataset C, we only included the "complete microcosm data", which corresponds to include only data collected from Experiment 5 with the three species. Dataset D was used to evaluate the information provided by data collected at concentrations lower than E_{r_a} and E_{r_d} (EC_{50} values) for algae and duckweeds, respectively. Thus, dataset D included all data except those related to duckweeds and algae exposed to the lowest cadmium concentrations ("monospecific data" and "two species data") collected from Experiment 5. All in all, we tested the minimum necessary dataset in terms of species combinations (one, two, or three species) from datasets A, B and C, and of concentration range via dataset D.

6 Results

6.1 Model fit and parameter estimates

Our MCMC algorithm always consistently converged according to Gelman and Rubin diagnostics for each simulation. The corresponding 2.5%, 50% and 97.5% quantiles of the posteriors for parameters of interest are summarised in Table 1. To keep results clear enough, only

fitting results from data of Experiment 5 (measured cadmium concentrations of 0, 2.25, 4.50, 6.88 and 9.09 μg/L) as medians of the credibility band for predicted data on algae dynamics and daphnid survival (Figure 2). On a general point of view, data were satisfactorily described by the model, with between 91% and 98% of observed data encompassed in the 95% credibility band of the predictions for the different species.

Marginal posterior distributions of the estimated parameters are shown in SI (Figures S3). We obtained narrow posterior distributions for almost all parameters, in particular parameters of interest, with the exception of parameters related to algae-daphnids interaction (grazing). The narrowness of posterior distributions indicates that sufficient information was available in the data to get posterior distributions of model parameters that are more precise than their priors. Such a gain of knowledge makes us confident in our fitting process.

6.2 Algae dynamics and daphnid survival

In the presence of duckweeds only (Figure 2(a), control), the number of algal cells per beaker in the water column increased during the first seven days when growth is higher than settling, and then decreased as growth declined while settling was continuing. In the presence of daphnids plus duckweeds, the global algal dynamics in the control was similar to the one without daphnids; however the number of algal cells per beaker was lower, due to daphnid grazing (Figure 2(b), control). There was additionally no effect of cadmium on the algal dynamics when duckweeds were present (Figure 2(a)). However, with daphnids, differences between tested concentrations appeared from the sixth day of experiment: the higher the cadmium concentration, the higher the number of algal cells (Figure 2(b)). This may be due to the decrease in daphnid number (Figure 2(c)), daphnid survival being highly affected by cadmium, particularly at the two highest tested concentrations C_3 and C_4.

6.3 Look-back on the experimental design

To take into account the potential effect of correlations between parameters, we compared 95% credibility intervals of the four stress functions predicted from the joint posterior distributions obtained for each partial dataset (A to D) to the one obtained from the whole dataset (Figure 3). The 95% credibility intervals of the predicted stress functions on daphnid growth rate and survival for each partial dataset appear superimposed to the 95% credibility intervals of the predicted stress functions for all the data. This result was expected since daphnid's survival and size data were included in all datasets, the whole and partial ones.

When "monospecific" data were removed (dataset A), the predicted stress functions were different from the reference ones with larger 95% credibility bands, particularly for competition parameter β. When "two species" data were removed (dataset B) the predicted stress functions for both algal and duckweed growth rates (r_a and r_d) were very close to the reference ones. On the contrary, the predicted stress function for competition parameter β was overestimated, and showed more uncertainty. When "monospecific" and "two species" data were removed (dataset C), we obtained very large 95% credibility intervals for predicted stress functions for both growth rates and the competition parameter. At last, dataset D (without data related to duckweeds and algae exposed to the lowest concentrations C_1 to C_4) led to very similar predicted stress functions for both algal and duckweed growth rates (r_a and r_d) compared to the reference ones, while for competition parameter β the predicted stress function was overestimated with a greater uncertainty.

7 Discussion

7.1 Cadmium effect

7.1.1 On parameters and processes

For parameters related to cadmium effect on algae and duckweeds, we obtained similar estimates to the ones obtained in a previous study involving only these two species **Lamonica2016**. However, posterior distributions were narrower in the present study for some of the parameters, e.g. E_{r_a}, b_{r_a}, E_β and b_β, mainly thanks to additional data we considered for fitting.

Parameters of the stress function on daphnid survival showed narrow posterior distributions. Parameter NEC was estimated at 4.47 $[2.95, 5.75]$ μg/L while other authors estimated either higher NEC values (8.6 μg/L **Nebeker1986**) or lower ones (0.720 $[0.0427, 1.78]$ **Forfait-Dubuc2012**) from monospecific studies. Our NEC estimate was high compared to the one obtained with data from the same microcosm, but including five species and sediment, namely 1.8 $[1.2, 2.3]$ μg/L **Billoir2012b**. Nevertheless, the NEC estimated by Billoir et al. was only based on survival data, and thus did not include data related to the dynamics of the other species. In particular, the algae dynamics link to the number of surviving daphnids was ignored. We thus also estimated the NEC only using survival data to compare 95% credible interval to the one of **Billoir2012b**. We obtained a NEC value of 3.47 $[0.030, 5.50]$ μg/L. This credible interval is quite large because the number of alive daphnids per beaker was highly variable, but it contains the credible interval obtained by **Billoir2012b**. The number of alive daphnids per beaker revealed difficult to describe because of the high inter-replicate variability between the tested cadmium concentrations.

In the literature, the effects of cadmium on daphnid growth may vary a lot from one study to another: $EC_{10} = 7.3$ μg/L for 17 days in**Knops2001** (monospecific bioassay conditions),

$NEC = 0.15$ μg/L in **Billoir2012b** (five-species microcosm conditions) or $EC_{50} = 2.7$ μg/L for 21 days in **Clement2014** (five-species microcosm conditions). In the present study, such effects are expressed through both parameters b_k and E_k (*i.e.*, EC_{50}). We obtained a very high but imprecise estimate for E_k (47.9 $[18.2, 2042]$ μg/L) indicating that daphnids were less sensitive in our experiment than in the one conducted by **Clement2014**. However, the narrow posterior distribution for curvature coefficient b_k (0.56 $[0.21, 1.04]$ μg/L) indicated that daphnid growth rate was already affected at the lowest concentrations, as also mentioned by **Billoir2012b**.

7.1.2 On the functioning of the microcosm

The microcosm functioning with cadmium or not is summarised in Figure 4. Cadmium effects on daphnid processes corresponded to a negative direct effect on survival, in particular at concentrations C_3 and C_4, as well as to a slight negative direct effect on growth. These direct effects of cadmium on daphnid processes impact both algae and duckweeds. Indeed they induced a decrease in daphnid grazing that led to an increase in algal density with increasing cadmium concentrations. Such a result was supported by the absence of a cadmium effect on algal growth at concentrations below 10 μg/L, that was a positive indirect effect of cadmium on algae below 10 μg/L. In addition, there was a negative direct effect of cadmium on the competition intensity. This latter did not compensate the negative direct effect of cadmium on growth of duckweeds, especially since algal density became higher, leading to a decrease in duckweed density.

Deciphering the cascade of cadmium effects on the three species is finally possible thanks to our modelling approach coupled with experiments. Unravelling the chemical direct and indirect effects as well as the interactions between species is necessary to correctly interpret the

global effect of a chemical substance on the functioning of a species community. Nevertheless, it is much more challenging when the number of species is increasing **Lamonica2016**. Hence, the model fits of the present study were overall satisfactory although some predicted data were overestimated compared to observed ones. In particular, the number of algal cells per beaker in the presence of both duckweeds and daphnids were overestimated by our model, as well as the number of duckweed fronds per beaker with algae and daphnids (SI, Figure S4 (b)).

7.2 Look-back on the experimental design

In ecotoxicology, optimising the experimental designs is not a recent concern **Albert2012**; **Andersen2000**; **Forfait-Dubuc2012**; **Wright2006**, but today mainly relates to the increasing use of concentration-response or effect models **Chevre2008a**; **Forfait-Dubuc2012**; **Holland2015a**; **Keddig2015a**; **Khinkis2003a**; **Sitter1995a**; **Wang2006a**. As formal optimisation first applied to monospecific bioassays, it mainly focused on the tested concentrations (range and number of concentrations) and on the number of tested individuals (per tested concentration and in total) **Forfait-Dubuc2012**. Formal optimisation is less suitable for microcosm experiments because microcosms are more complex devices. Moreover, microcosms are not standardised experimental tools since they are usually set up on a case-by-case basis, according to the specific objectives of the study **Cairns1993**; **Crossland1992**. Thanks to our modelling approach, we were able to question the relevance of using certain data to estimate the chosen parameters. Hence, using dataset B provided stress functions similar to the reference ones based on all available data. We also showed that stress function on the interspecific competition parameter (β) from partial dataset D differed from the reference one, while the stress functions on the processes related to growth (for both algae and duckweeds) remained un-

changed. Such results suggest that two-species data are not fully necessary, while a larger range of tested concentrations would be strongly recommended to estimate parameters related to effects on interactions between species.

Even if it remains difficult to know *a priori* which types of data would be absolutely necessary to best estimate stress function parameters, the pragmatic look-back we performed using modelling may guide further experiments with microcosms, for instance for other contaminants effects, dynamics under modified abiotic conditions, or even other species interactions. In particular, our study suggests that experiments can be specifically selected to gain knowledge on a three-species microcosm. In the end, we could make the following recommendations for further ecotoxicological studies with a microcosm device: a first experiment with the complete microcosm only (*i.e.*, with the three species) and a tested concentration range limited by the sensitivity of the most sensitive species; then a second experiment with monospecific microcosms only and a tested concentration range limited by the sensitivities of the two less sensitive species. If needed, additional experiments without contaminant may involve different combinations of species depending on their connections to each others.

8 Conclusion and perspectives

We provided EC_{50} values for the different processes affected by cadmium. Thanks to the understanding of the underlying processes that occurred in the microcosm functioning, we also managed to identify the cascade of cadmium effects induced by the interactions between species. In addition, we got back from modelling to experiments in order to determine which of the collected data were necessary and sufficient to precisely estimate model parameters, leading us to suggest a more efficient experimental design. Finally, we (1) highlighted the im-

portance of interactions by identifying the effect cascade occurring within a small ecosystem under chemical pressure; and (2) showed that alternative use of experimental data can help conceiving experimental designs for a microcosm study.

From a general perspective, it would be particularly interesting to integrate the method we developed with the microcosm into a broader conceptual framework to account for several indicators in environmental risk assessment. The weight of evidence (WOE) (**Piva2011**; **Weed2005**) approach combines several indicators to assess the risks associated with contaminated sediments. These indicators, including the bioavailability of the contaminant linked to the sediment chemistry and sub-lethal effects on biomarkers for two species (**Benedetti2014**), are integrated into an overall value summarizing the risk associated with the studied contaminant. Such an approach could provide a basis for formalizing the integration of quantitative characteristics extracted from our model in order to globally evaluate the risk associated with the tested contaminant in a microcosm.

Our method also permitted to assess which data to include when estimating parameters of interest in a dynamic ecosystem model from a laboratory based microcosm ecotoxicity study. Such an approach could be enhanced to better foresee further experiments with microcosms based a similar model. Beyond this, if parameters are simultaneously estimated a whole dataset, this makes possible to compare these reference estimates with those obtained with partial datasets. This gives knowledge on the data dependency in the modelling results. Last but not least, such a retrospective and descriptive sensitivity analysis puts light in the fact that data quality and design are more beneficial for modelling purpose than quantity. Ideally, as the use of models and big data in ecology increases **Vandenbrink2016**, modellers and experimenters could collaboratively and profitably elaborate model-guided experiments.

9 Figures

9.1 Figure 1

9.2 Figure 2 (a) and (b)

9.3 Figure 2 (c)

9.4 Figure 3

9.5 Figure 4

Figure 1. Workflow followed in the present study: (1) in black, development of the model of the microcosm functioning under cadmium pressure, data collection, and estimation of model parameters with the whole dataset, then extraction of stress functions, namely "reference stress functions"; (2) in orange, comparison of population dynamics with and without cadmium to analyse the global perturbation of the community, extraction of EC_{50} for the different processes; (3) in green, build of partial datasets by removing some parts of the complete dataset, estimation of parameters using those partial datasets, comparison of the newly estimated stress functions to the reference ones to assess how much information was provided by the removed data.

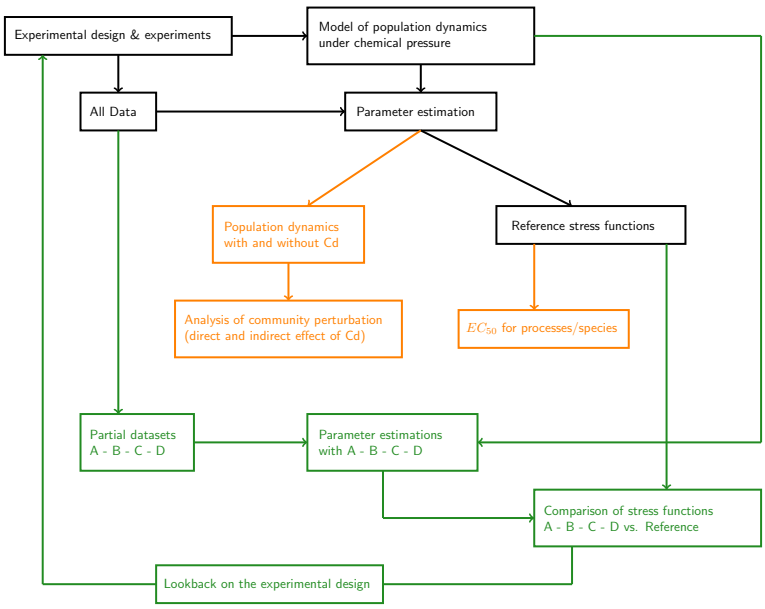


Figure 2. Model fitting for state variables related to algae dynamics: (a) with duckweeds; (b) with duckweeds and daphnids. Symbols refer to the different tested cadmium concentrations of Experiment 5. Plain lines stand for the fitted model with each parameter equal to its median value at each concentration C_j. The light grey area corresponds to the 95% credible band of the predicted data in the control.

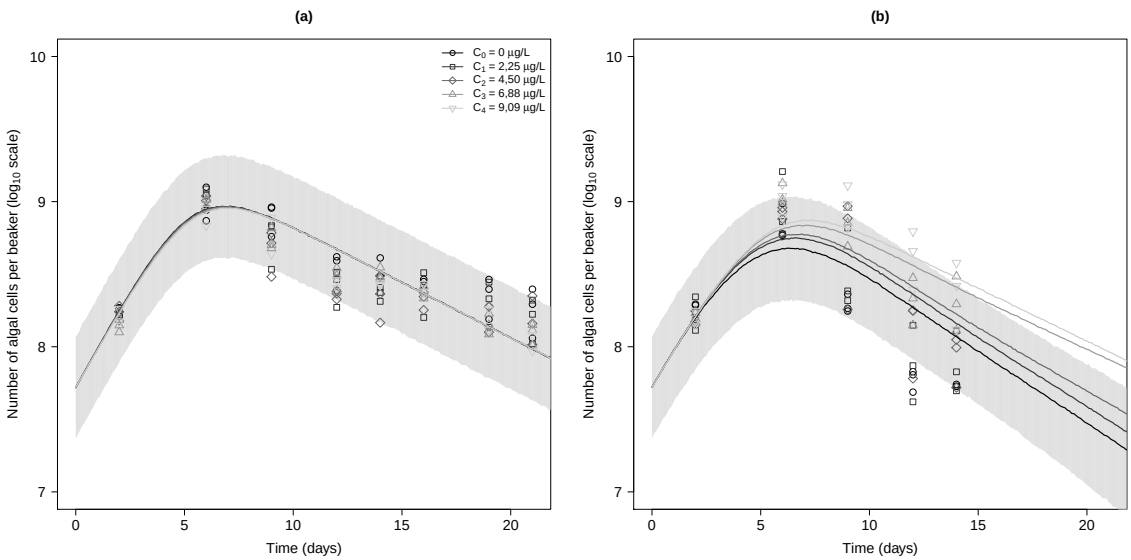


Figure 3. Fit plot for the number of alive daphnids per beaker. Symbols refer to the different tested cadmium concentrations of Experiment 5. Plain lines stand for the fitted model with each parameter equal to its median value at each concentration C_j. The light grey area corresponds to the 95% credible band of the predicted data in the control.

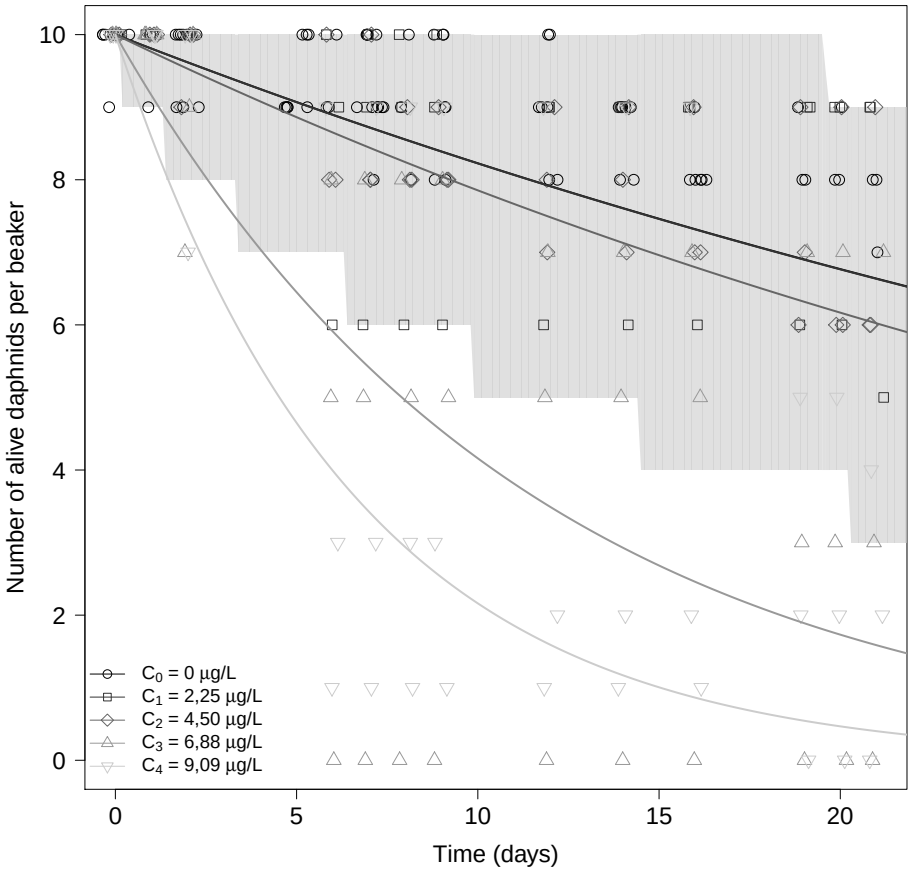


Figure 4. Comparison of stress functions obtained with complete and partial datasets. Plain black lines delimit the 95% credible bands for the stress functions obtained with the partial dataset, while red dotted lines delimit the 95% credible bands for the stress functions obtained with the complete dataset.

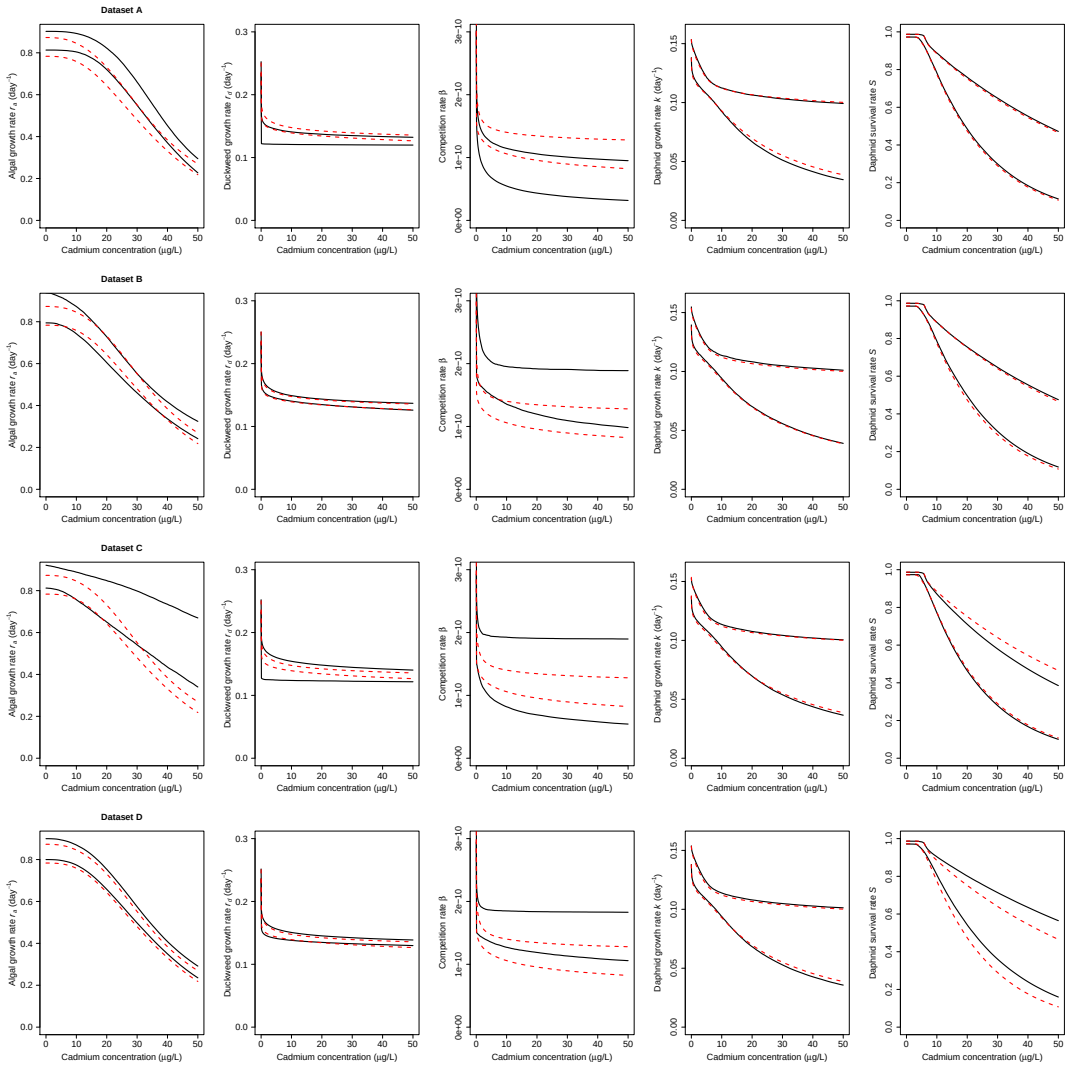
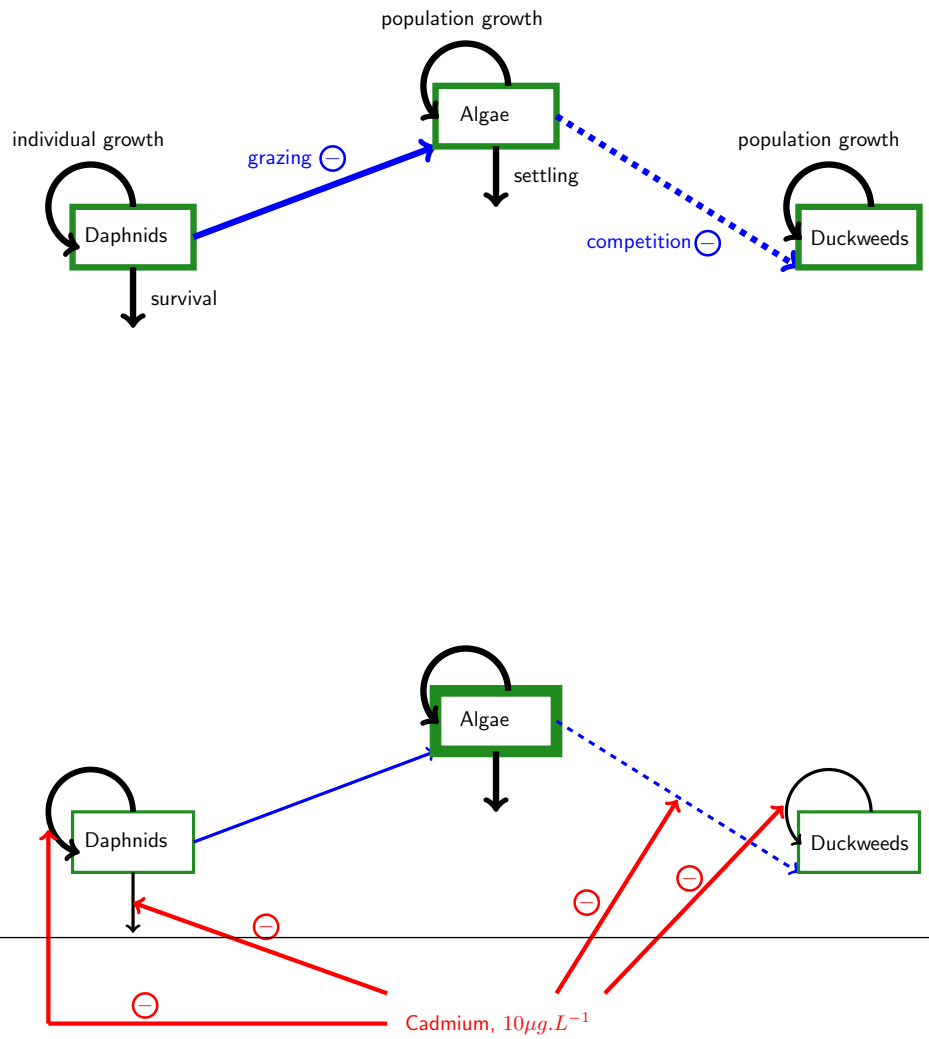


Figure 5. Global perturbation of the microcosm functioning without cadmium and with cadmium ($10\ \mu$g/L). Green boxes represent species population, black arrows represent species processes, blue arrows represent interactions between species, red arrows represent cadmium effect on the different processes. Thickness of boxes lines is proportional to population size and thickness of arrows is proportional to process intensity.



10 Tables

Table 1. Parameters of interest.

Symbol	Definition	Unit	Prior distribution* or value	Sources**	2.5, 50, 97.5 % posterior quantiles		
r_{a_0}	Intrinsic algal growth rate in the control	day^{-1}	$\mathcal{U}$ (0,2)	[1]	0.78	0.83	0.87
E_{r_a}	The concentration at which r_{a_0} is reduced by 50 %	μg.L^{-1}	$\log_{10}(E_{r_a}) \sim \mathcal{N}$ (1.78,0.1)	[2]	1.54	1.56	1.58
b_{r_a}	Curvature coefficient	-	$\log_{10}(b_{r_a}) \sim \mathcal{N}$ (0.24,0.1)	[2]	0.37	0.43	0.49
β_0	Competition intensity of algae on duckweeds in the control	# of cells per beaker^{-1}.day^{-1}	$\log_{10}(\beta_0) \sim \mathcal{U}$ (-11,-9)	[2]	-9.53	-9.46	-9.40
E_β	The concentration at which β_0 is reduced by 50 %	μg.L^{-1}	$\log_{10}(E_\beta) \sim \mathcal{U}$ (-1,4)	Vague	-0.99	-0.77	-0.15
b_β	Curvature coefficient	-	$\log_{10}(b_\beta) \sim \mathcal{U}$ (-3,3)	Vague	-1.11	-0.82	-0.62
m_0	Daphnid mortality rate in the control	daphnid.day^{-1}	$\log_{10}(m_0) \sim \mathcal{U}$ (-4,0)	Vague	-1.89	-1.71	-1.56
k_S	Slope of survival stress function	μg^{-1}.L.day^{-1}	$\log_{10}(k_S) \sim \mathcal{U}$ (-4,4)	Vague	-1.80	-1.54	-1.30
NEC	No Effect Concentration on daphnid survival	μg.L^{-1}	$\log_{10}(NEC) \sim \mathcal{U}$ (-2,1)	Vague	0.47	0.65	0.76
k_0	Daphnid growth rate in the control	day^{-1}	$\mathcal{N}$ (0.11,0.030)	[3]	0.138	0.146	0.154

Symbo	Definition	Unit	Prior distribution* or value	Sources**	2.5, 50, 97.5 % posterior quantiles		
E_k	The concentration at which k_0 is reduced by 50 %	μg.L^{-1}	$\log_{10}(E_k) \sim \mathcal{U}$ (-1,4)		1.26	1.68	3.31
b_k	Curvature coefficient	-	$\log_{10}(b_k) \sim \mathcal{U}$ (-3,3)		-0.68	-0.25	0.015
r_{d_0}	Intrinsic duckweed growth rate in the control	day^{-1}	$\mathcal{U}$ (0,2)	[2]	0.23	0.24	0.25
E_{r_d}	The concentration at which r_{d_0} is reduced by 50 %	μg.L^{-1}	$\log_{10}(E_{r_d}) \sim \mathcal{N}$ (2.44,0.2)	[2]	1.96	2.30	2.67
b_{r_d}	Curvature coefficient	-	$\log_{10}(b_{r_d}) \sim \mathcal{U}$ (-3,3)	Vague	-1.05	-0.91	-0.77

*Prior distribution: $\mathcal{N}$ stands for the normal law, $\mathcal{U}$ stands for the uniform law. **Sources: [1] Lamonica et al. 2016a, [2] Lamonica et al. 2016b, [3] Billoir et al. 2008.

Table 2. Experimental design and partial datasets.

Cadmium concentrations	Monospecific data	Two species data	Complete microcosm data
C_0	Algae Duckweeds	Algae, duckweeds Algae, daphnids	Algae, duckweeds, daphnids
C_{1-4}	Algae Duckweeds	Algae, duckweeds	Algae, duckweeds, daphnids
C_{5-8}	Algae Duckweeds	Algae, duckweeds	
C_{9-12}	Algae		

The black dashed, gray dashed, black and gray rectangles refer to the data that have been removed from the entire dataset to build partial dataset A, B, C and D, respectively.

Data accessibility

Data are available online: 10.5281/zenodo.6575576

Supplementary material

R scripts are available online: 10.5281/zenodo.6575576

Acknowledgements

Authors thank the ENTPE (École Nationale des Travaux Publics de l'État) and IXXI institute (Institut Rhônalpin des Systèmes Complexes) for their financial support. Authors also thank Pauline Le Quellec et Ludovik Hauduroy for their contribution to the experiments.

Conflict of interest disclosure

The authors of this preprint declare no financial conflict of interest with the content of this article.

References

[Albert et al. 2012] C. Albert, R. Ashauer, H. R. K′unsch, and P. Reichert. Bayesian experimental design for a toxicokinetic-toxicodynamic model (english). *Journal of statistical planning and inference*, 142(1):263 – 275, 0001.

[Billoir et al. 2008] E. Billoir, M.L. Delignette-Muller, A.R.R. Péry, and S. Charles. A Bayesian Approach to Analyzing Ecotoxicological Data. *Environ Sci Technol*, 42(23):8978–8984, 2008.

[Billoir et al. 2011] E. Billoir, H. Delhaye, B. Clément, M.L. Delignette-Muller, and S. Charles. Bayesian modelling of daphnid responses to time-varying cadmium exposure in laboratory aquatic microcosms. *Ecotoxicology and environmental safety*, 74(4):693–702, May 2011.

[Billoir et al. 2012] E. Billoir, H. Delhaye, C. Forfait, B. Clément, G. Triffault-Bouchet, S. Charles, and M.L. Delignette-Muller. Comparison of bioassays with different exposure time patterns: the added value of dynamic modelling in predictive ecotoxicology. *Ecotoxicology and environmental safety*, 75(1):80–6, January 2012.

[Biron et al. 2012] P-A. Biron, S. Massarin, F. Alonzo, L. Garcia-Sanchez, S. Charles, and E. Billoir. Population-level modeling to account for multigenerational effects of uranium in daphnia magna. *Environmental Science and Technology*, 46:1136–1143, 2012.

[Bullock et al. 2006] J. M. Bullock, K. Shea, and O. Skarpaas. Measuring plant dispersal: An introduction to field methods and experimental design. *Plant Ecology*, 186(2):217–234, 2006.

[Cairns and Cherry 1993] J. Cairns and D. S. Cherry. *Freshwater Multi-Species Test Systems*, pages 101–116. Blackwell Publishing Ltd., 1993.

[Charles et al. 2016] S. Charles, V. Ducrot, D. Azam, R. Benstead, D. Brettschneider, K. De Schamphelaere, S. Filipe Goncalves, J. W. Green, H. Holbech, T. H. Hutchinson, D. Faber, F. Laranjeiro, P. Matthiessen, L. Norrgren, J. Oehlmann, E. Reategui-Zirena, A. Seeland-Fremer, M. Teigeler, J. P. Thome, M. Tobor Kaplon, L. Weltje, and L. Lagadic. Optimizing the design of a reproduction toxicity test with the pond snail Lymnaea stagnalis. *Regulatory Toxicology and Pharmacology*, 81:47–56, 2016.

[Chèvre and Brazzale 2008] N. Chèvre and A. R. Brazzale. Cost-effective experimental design to support modeling of concentration-response functions. *Chemosphere*, 72(5):803–810, 2008.

[Clément et al. 2014] B. Clément, H. Delhaye, and G. Triffault-Bouchet. Comparison of laboratory batch and flow-through microcosm bioassays. *Ecotoxicology and environmental safety*, 108C:217–223, July 2014.

[Crossland and La Point 1992] N. O. Crossland and T. W. La Point. The design of mesocosm experiments. *Environmental toxicology and chemistry*, 11(1):1–4, 1992.

[Delhaye 2012] H. Delhaye. *Développement d'un bioessai de toxicité chronique en microcosme aquatique de laboratoire et évaluation de l'outil au travers de létude des effets du cadmium*. PhD thesis, Institut National des Sciences Appliquées de Lyon, 2012.

[DeMott 1982] W.R. DeMott. Feeding selectivities and relative ingestion rates of Daphnia and Bosmina. *Limnol. Oceanogr*, 27(3):518–527, 1982.

[Egloff and Palmer 1971] D.A. Egloff and D.S. Palmer. Size relations of the filtering area of two Daphnia species. *Limnol. Oceanogr*, 16(November):900–905, 1971.

[Forfait-Dubuc et al. 2012] C. Forfait-Dubuc, S. Charles, E. Billoir, and M.L. Delignette-Muller. Survival data analyses in ecotoxicology: critical effect concentrations, methods and models. What should we use? *Ecotoxicology (London, England)*, 21(4):1072–83, May 2012.

[Gelman and Rubin 1992] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–511, 1992.

[Grimm and Railsback 2005] V. Grimm, S. Railsback. *Individual-based Modeling and Ecology. Princeton University Press, Princeton, NJ.* 2005

[Grimm et al. 2010] V. Grimm, U. Berger, D.L. DeAngelis, J.G. Polhill, J. Giske, and S.F. Railsback. The ODD protocol: A review and first update. *Ecological Modelling*, 221(23):2760–2768, 2010.

[Guidetti 2002] P. Guidetti. The importance of experimental design in detecting the effects of protection measures on fish in Mediterranean MPAs. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 12(6):619–634, 2002.

[Holland-Letz and Kopp-Schneider 2015] T. Holland-Letz and A. Kopp-Schneider. Optimal experimental designs for dose-response studies with continuous endpoints. *Archives of Toxicology*, 89(11):2059–2068, 2015.

[Keddig et al. 2015] N. Keddig, S. Schubert, and W. Wosniok. Optimal test design for binary response data: the example of the fish embryo toxicity test. *Environmental Sciences Europe*, 27(1):15, 2015.

[Khinkis et al. 2003] L. A. Khinkis, L. Levasseur, H. Faessel, and W. R. Greco. Optimal design for estimating parameters of the 4-parameter hill model. *Nonlinearity in biology, toxicology, medicine*, 1(3):363–77, 2003.

[Kilkenny et al. 2009] C. Kilkenny, N. Parsons, E. Kadyszewski, M. F. W. Festing, I. C. Cuthill, D. Fry, J. Hutton, and D. G. Altman. Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. *PLoS ONE*, 4(11), 2009.

[Knops et al. 2001] M. Knops, R. Altenburger, and H. Segner. Alterations of physiological energetics, growth and reproduction of Daphnia magna under toxicant stress. *Aquatic Toxicology*, 53:79–90, 2001.

[Kooijman et al. 2000] S. A. L. M. Kooijman J. S. Andersen, J. J. M. Bedaux and H. Holst. The influence of design characteristics on statistical inference in nonlinear estimation: A simulation study based on survival data and hazard modeling. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(3):323–341, 2000.

[Lamonica et al. 2016a] D. Lamonica, U. Herbach, F. Orias, B. Clément, S. Charles, and C. Lopes. Mechanistic modelling of daphnid-algae dynamics within a laboratory microcosm. *Ecological modelling*, 320:213-230, 2016.

[Lamonica et al. 2016b] D. Lamonica, B. Clément, S. Charles, and C. Lopes. Modelling algae?duckweed interaction under chemical pressure within a laboratory microcosm. *Ecotoxicology and Environmental Safety*, 128:252-265, 2016.

[Mance 1987] G. Mance. *Pollution threat of heavy metals in aquatic environments.* 1987.

[Nebeker et al. 1986] V. Nebeker, T. Onjukka, A. Cairns, F. Krawczyk, and E. T. Al. Survival of Daphnia Magna and Hyalella Azteca in Cadmium-Spiked Water and Sediment. *Environmental Toxicology and Chemistry*, 5:933–938, 1986.

[Plummer 2009] M. Plummer. rjags: Bayesian graphical models using mcmc. Rpackage version 1.0.3-12, 2009.

[R Core team 2013] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[Sitter and Torsney 1995] R. R. Sitter and B. Torsney. Optimal designs for binary response experiments with two design variables. *Statistica Sinica*, 5:405–419, 1995.

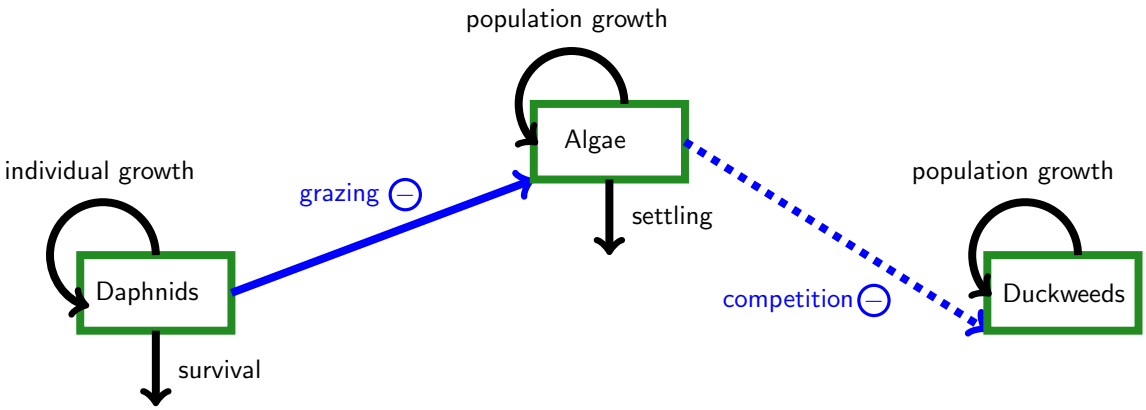
[Underwood 2000] A.J. Underwood. Importance of experimental design in detecting and measuring stresses in marine populations. *Journal of Aquatic Ecosystem Stress and Recovery*, 7(1):3–24, 2000.

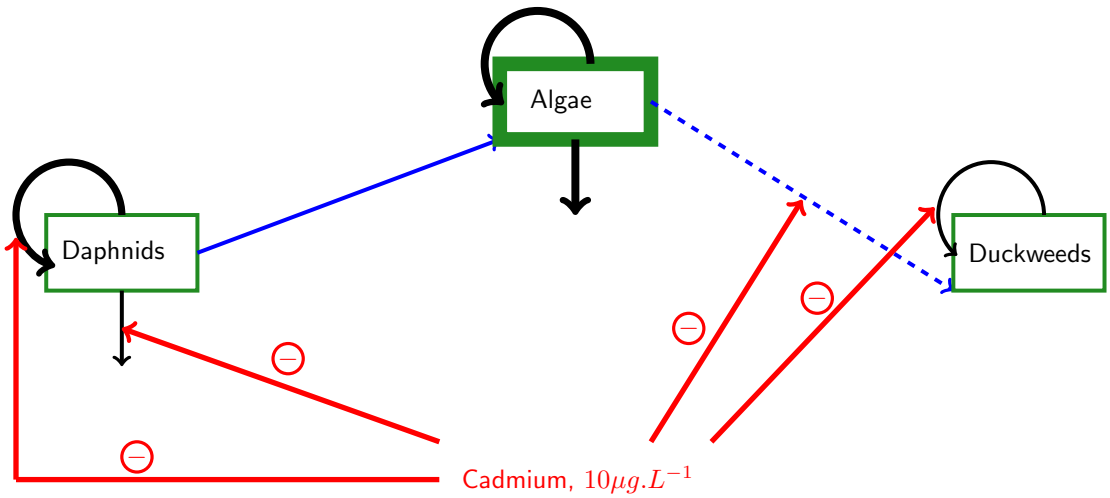
[Van Den Brink et al. 2016] P. J. Van Den Brink, C. B. Choung, W. Landis, M. Mayer-Pinto, V. Pettigrove, P. Scanes, R. Smith, and J. Stauber. New approaches to the ecological risk assessment of multiple stressors. *Marine and Freshwater Research*, 67(4):429–439, 2016.

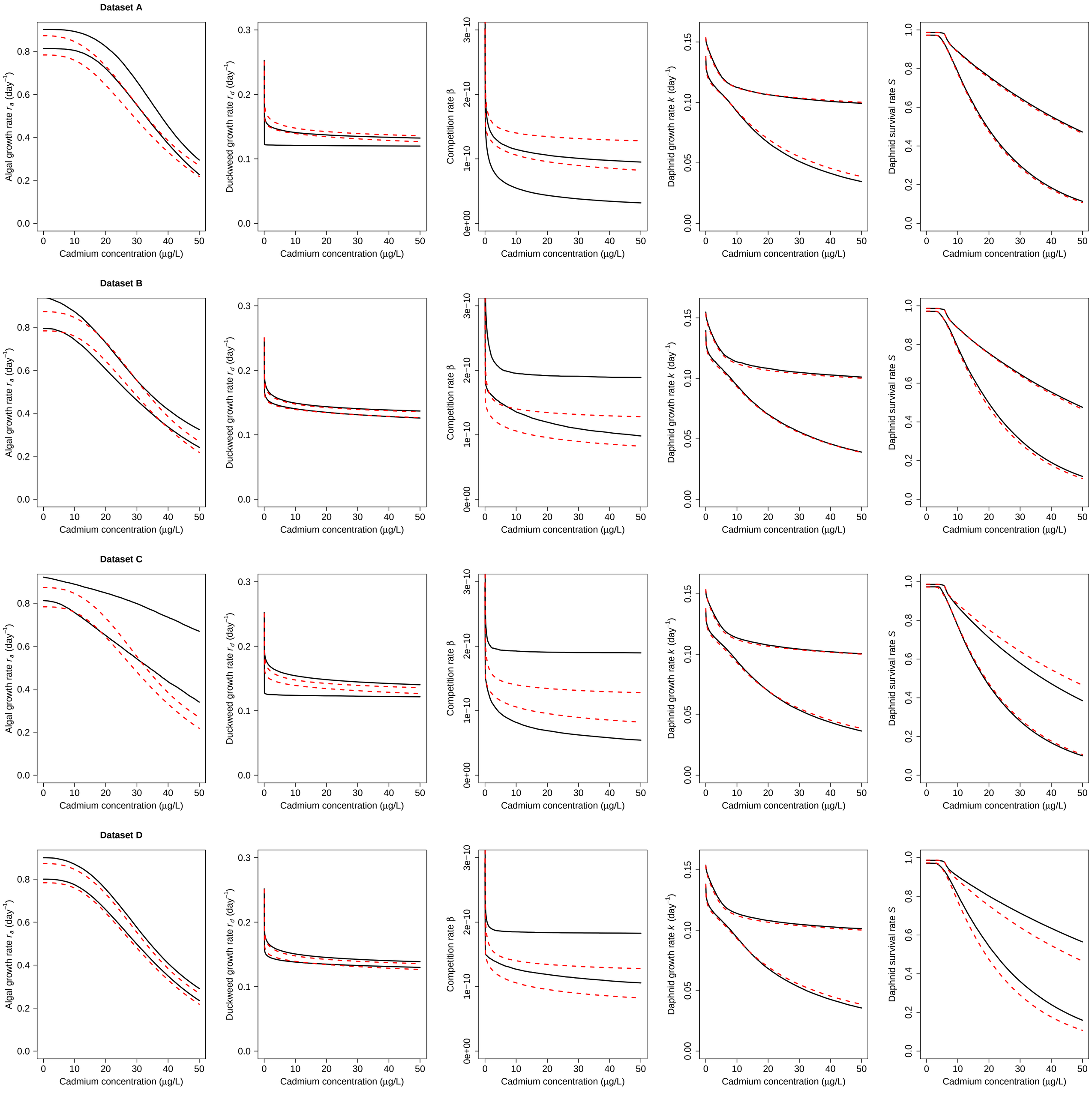
[Wang 2006] Y. Wang, R. H. Myers, E. P. Smith, and K. Ye. D-optimal designs for Poisson regression models. *Journal of Statistical Planning and Inference*, 136(8):2831–2845, 2006.

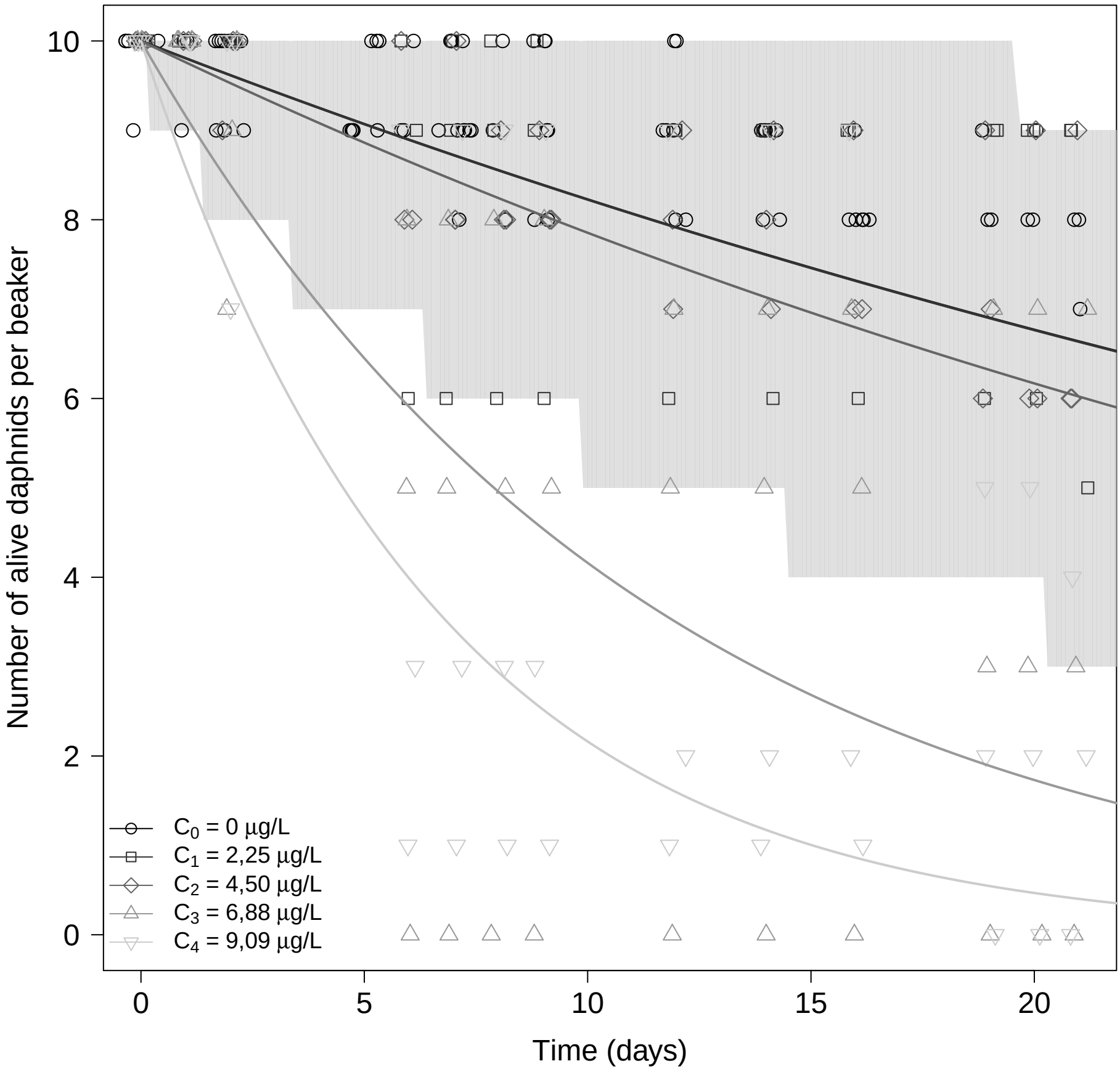
[Wright and Bailer 2006] S. E. Wright and A. J. Bailer. Optimal experimental design for a nonlinear response in environmental toxicology. *Biometrics*, (3):886, 2006.

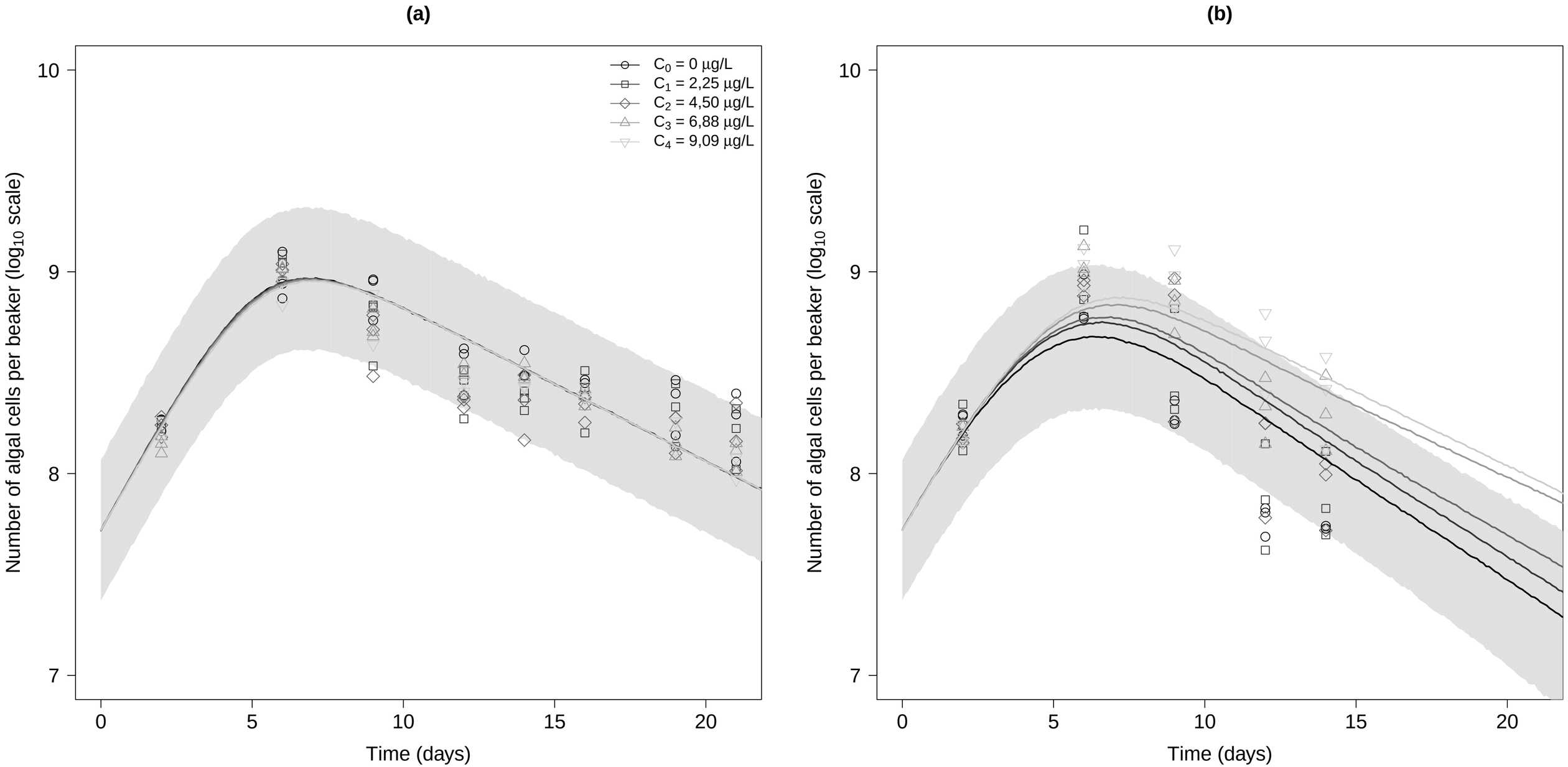
Appendix

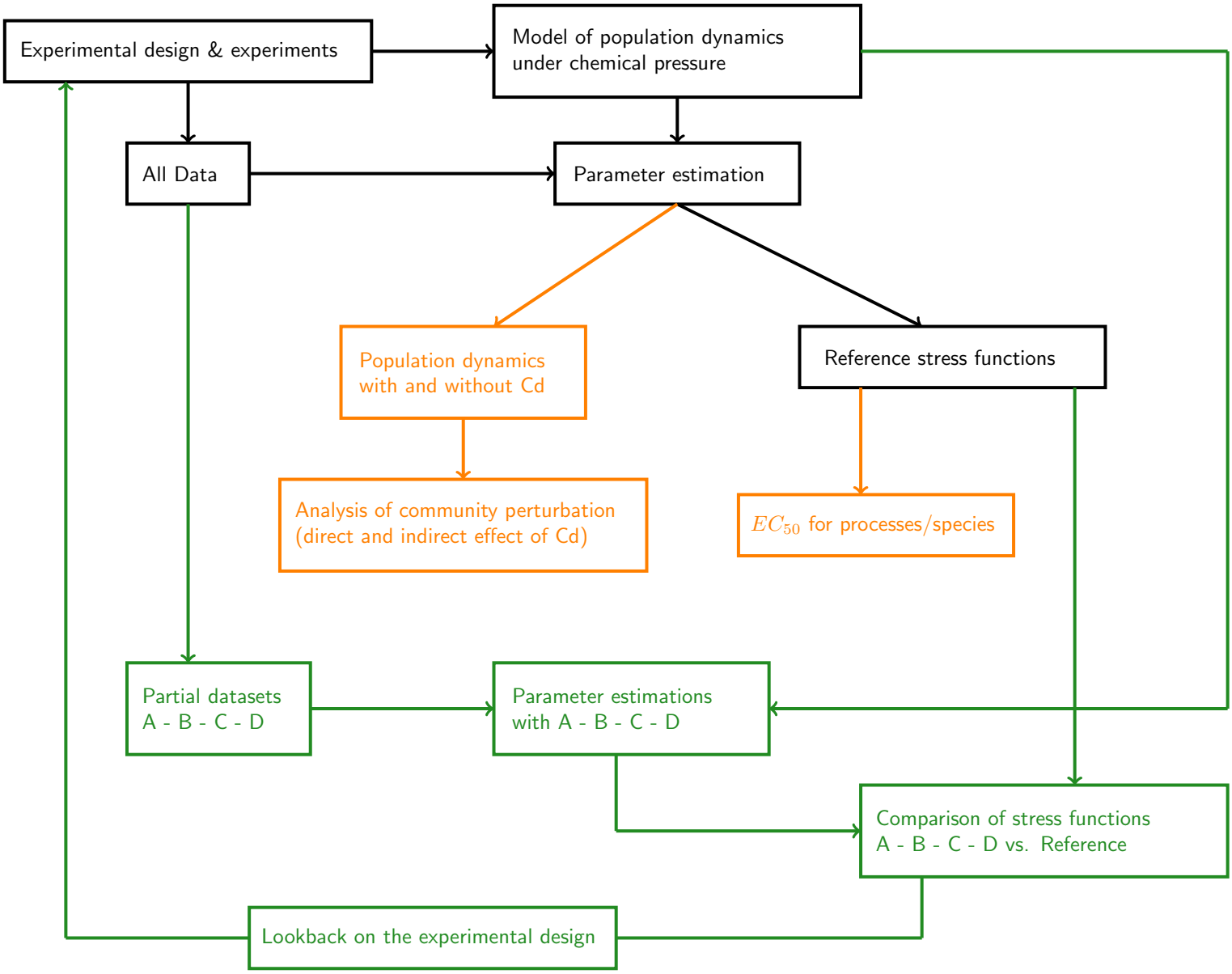

population growth
Algae
individual growth
grazing
settling
population growth
Daphnids
competition
Duckweeds
survival


Algae
Daphnids
Duckweeds
Cadmium, $10\mu g.L^{-1}$


Dataset A
Dataset B
Dataset C
Dataset D
Algal growth rate r_a (day^{-1})
Duckweed growth rate r_d (day^{-1})
Competition rate β
Daphnid growth rate k (day^{-1})
Daphnid survival rate S
Cadmium concentration (µg/L)


Number of alive daphnids per beaker
Time (days)
0
2
4
6
8
10
0
5
10
15
20
C0 = 0 µg/L
C1 = 2,25 µg/L
C2 = 4,50 µg/L
C3 = 6,88 µg/L
C4 = 9,09 µg/L


(a)
(b)
C_0 = 0 µg/L
C_1 = 2,25 µg/L
C_2 = 4,50 µg/L
C_3 = 6,88 µg/L
C_4 = 9,09 µg/L
Number of algal cells per beaker (log_{10} scale)
Time (days)
10
9
8
7
0
5
10
15
20


Experimental design & experiments
Model of population dynamics under chemical pressure
All Data
Parameter estimation
Population dynamics with and without Cd
Reference stress functions
Analysis of community perturbation (direct and indirect effect of Cd)
EC_{50} for processes/species
Partial datasets A - B - C - D
Parameter estimations with A - B - C - D
Comparison of stress functions A - B - C - D vs. Reference
Lookback on the experimental design