



HAL
open science

Vers une amélioration de l'analyse des données et une optimisation des plans d'expérience pour une analyse quantitative du risque en écotoxicologie

Carole Forfait-Dubuc

► To cite this version:

Carole Forfait-Dubuc. Vers une amélioration de l'analyse des données et une optimisation des plans d'expérience pour une analyse quantitative du risque en écotoxicologie. Ecotoxicologie. Université Claude Bernard - Lyon I, 2013. Français. NNT : 2013LYO10039 . tel-01174870

HAL Id: tel-01174870

<https://theses.hal.science/tel-01174870>

Submitted on 10 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre : 39-2013

Année 2013

THÈSE

Présentée

devant L'UNIVERSITÉ CLAUDE BERNARD - LYON 1

pour l'obtention

du DIPLÔME DE DOCTORAT
(arrêté du 7 août 2006)

et soutenue publiquement le
27 mars 2013

par Carole FORFAIT DUBUC

**Vers une amélioration de l'analyse
des données et une optimisation
des plans d'expérience pour une analyse
quantitative du risque en écotoxicologie**

Directrices de thèse : Sandrine CHARLES
Marie Laure DELIGNETTE-MULLER

JURY :	Sandrine CHARLES,	Directrice
	Bruno COMBOURIEU,	Examineur
	Marie Laure DELIGNETTE-MULLER,	Directrice
	Patrick FLAMMARION,	Rapporteur
	Laurent LAGADIC,	Rapporteur
	Marie SIMON-CORNU,	Examinatrice

UNIVERSITÉ CLAUDE BERNARD-LYON 1

Président de l'Université

Vice-Président du Conseil d'Administration

Vice-Président du Conseil des Etudes et
de la Vie Universitaire

Vice-Président du Conseil Scientifique

Secrétaire Général

M. François-Noël GILLY

M. le Professeur Hamda BEN HADID

M. le Professeur Philippe LALLE

M. le Professeur Germain GILLET

M. Alain HELLEU

COMPOSANTES SANTÉ

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique Lyon Sud

– Charles Mérieux

UFR d'Odontologie

Institut des Sciences Pharmaceutiques et Biolo-
giques

UFR d'Ontologie

Institut des Sciences et Techniques de la Réadap-
tation

Département de formation et Centre de Re-
cherche en Biologie Humaine

Directeur : M. le Professeur J. ETIENNE

Administrateur provisoire :

M. le Professeur G. KIRKORIAN

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C. VINCIGUERRA

Directeur : M. O. ROBIN

Directeur : M. le Professeur Y. MATILLON

Directeur : M. le Professeur P. FARGE

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

Département Sciences de la Terre

UFR Sciences et Techniques des Activités Phy-
siques et Sportives

Observatoire de Lyon

Polytech Lyon

Ecole Supérieure de Chimie Physique Electro-
nique

Institut Universitaire de Technologie de Lyon 1

Institut Universitaire de Formation des Maîtres

Institut de Science Financière et d'Assurances

Directeur : M. le Professeur F. De MARCHI

Directeur : M. le Professeur F. FLEURY

Directeur : Mme la Professeure H. PARROT

Directeur : M. N. SIAUVE

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur A. GOLDMAN

Directeur : M. le Professeur H. BEN HADID

Directeur : Mme S. FLECK

Directeur : Mme la Professeure I. DANIEL

Directeur : M. C. COLLIGNON

Directeur : M. B. GUIDERDONI

Directeur : M. P. FOURNIER

Directeur : M. G. PIGNAULT

Directeur : M. C. VITON

Directeur : M. R. BERNARD

Directeur :

Mme la Professeure V. MAUME-DESCHAMPS

Articles publiés pendant la thèse :

Forfait-Dubuc C, Charles S, Billoir E and Delignette-Muller ML (2012) Survival data analyses in ecotoxicology : critical effect concentrations, methods and models. What should we use? *Ecotoxicology* 21 : 1072–1083

Billoir E, Delhayé H, Forfait C, Clément B, Triffault-Bouchet G, Charles S and Delignette-Muller ML (2012) Comparison of bioassays with different exposure time patterns : The added value of dynamic modelling in predictive ecotoxicology. *Ecotoxicology and Environmental Safety* 75 : 80–86

Delignette-Muller ML, Forfait C, Billoir E and Charles S (2011) A new perspective on the Dunnett procedure : filling the gap between NOEC/LOEC and EC_x concepts. *Environmental Toxicology and Chemistry* 30 :2888–2891

Je tiens à remercier mes directrices de thèse Sandrine Charles et Marie Laure Delignette-Muller de m'avoir permis de réaliser cette thèse. Merci de votre suivi au quotidien, de votre intérêt et de m'avoir appris énormément de choses. Ce manuscrit ne serait pas ce qu'il est sans vos suggestions et vos corrections. Merci Marie Laure pour nos petites conversations très matinales.

Je tiens à remercier Patrick Flammarion et Laurent Lagadic d'avoir accepté d'être rapporteurs sur cette thèse ainsi que Bruno Combourieu et Marie Simon Cornu d'avoir accepté de faire partie du jury. Mes remerciements vont également à Elise Billoir (qui a guidé mes premiers pas dans cette thèse), Virginie Ducrot, Jeanne Garric et Eric Parent, pour avoir participé à mes comités de pilotage et m'avoir prodigué de précieux conseils. Je suis très reconnaissante à Jeanne Garric de m'avoir permis de participer à des expériences au laboratoire du CEMAGREF (IRSTEA) afin de faire plus ample connaissance avec les daphnies. Je tiens également à remercier Virginie Ducrot qui a permis d'orienter cette thèse vers les plans d'expérience ce qui m'a vraiment beaucoup intéressée.

Un grand MERCI à Cécile pour tous tes conseils (plus ou moins bons.....je plaisante) mais en tout cas je t'écoute toujours depuis plus de 4 ans! Tu m'as énormément soutenue et j'espère pouvoir t'aider autant que ce que tu m'as aidé surtout dans les moments critiques...je te souhaite pleins de bonnes choses pour cette dernière année de thèse et pour le reste!!!

Danke à Janice, je suis tellement contente de t'avoir rencontrée, merci pour tous tes conseils, ta bonne humeur et nos bons moments de rigolades...je sais qu'ils dureront encore longtemps et encore félicitations pour ton petit Célestin.

Merci à Ariane et Véronique mes copines de master, aux membres de l'équipe MEPS : Christelle, Guillaume et Dom, à l'équipe BAOBAB (ou ex) : Patricia, Florence, Cécilia, Marie France, Martin et aux autres personnes avec qui j'ai passé de bons moments comme Philippe (merci pour ton soutien psy) et Clothilde.

Merci à mes amis plus délocalisés comme Vava *et al.* et Marie *et al.*

Je tiens également à remercier ma soeur et sa tribu mais surtout mes parents qui ont tellement donné pour moi. Cette thèse est l'occasion de vous dire "Maman, papa, merci pour tout" je suis fière d'être votre fille, c'est vous qui m'avez appris à toujours me battre et à ne jamais laisser tomber et cela a été vraiment utile pour aller au bout de cette thèse.

Et enfin merci à Laurent de m'avoir suivie dans cette aventure, de m'avoir permis de relativiser tout au long de la thèse et de m'avoir soutenue.....surtout merci pour la super petite famille que nous avons construite ensemble, nos plus belles réussites Lola, Paul et Clément. Lola et Paul, merci pour vos encouragements et votre regard si juste....désolée de vous avoir imposé une maman stressée depuis septembre... d'ailleurs mon petit Clément, tu me l'as bien fait comprendre en arrêtant de faire tes nuits depuis cette période, tu peux dormir maintenant! Merci à tous les 4.

Table des matières

1	Introduction	1
1.1	Contexte et problématique	1
1.2	L'évaluation quantitative du risque	4
1.3	Les tests de toxicité et traits d'histoire de vie : exemple de <i>Daphnia magna</i>	6
1.4	Les concentrations critiques d'effet	8
1.4.1	Les NOEC/LOEC	8
1.4.2	Les EC_x	13
1.4.3	Les NEC	18
1.5	Problématique	21
1.6	Approches utilisées	21
1.7	Organisation du manuscrit	24
2	Les données de survie	25
2.1	Analyse des données de survie	25
2.1.1	Introduction	25
2.1.2	Les jeux de données de survie utilisés	26
2.1.3	Sélection des méthodes et modèles afin de déterminer les concentrations critiques d'effet	26
2.1.4	Comparaison des concentrations critiques d'effet	43
2.1.5	Discussion et conclusion sur l'analyse des données de survie	43
2.2	Optimisation des plans d'expérience pour des données de survie	48
2.2.1	Introduction	48
2.2.2	Etude théorique utilisant la D-optimalité localement	51
2.2.3	Etude globale basée sur des plans d'expérience réalistes	64
2.2.4	Discussion et conclusion sur l'optimisation des plans d'expérience dans le cadre des données de survie	81
3	Les données de reproduction	87
3.1	Analyse des données de reproduction	87
3.1.1	Introduction	87
3.1.2	Choix de la variable étudiée	89
3.1.3	Choix de la partie stochastique et inférence	94
3.1.4	Choix de la partie déterministe	113
3.1.5	Conclusion sur l'analyse des données de reproduction	118

3.2	Optimisation d'un plan d'expérience en vue d'une standardisation chez la limnée	121
3.2.1	Introduction	121
3.2.2	Réduction de la durée de l'expérience	123
3.2.3	Réduction du nombre d'individus utilisés dans l'expérience	130
3.2.4	Conclusion sur l'optimisation d'un plan d'expérience en vue d'une standardisation chez la limnée.	137
4	Conclusion et Perspectives	139
	Références	145

Chapitre 1

Introduction

Sommaire

1.1	Contexte et problématique	1
1.2	L'évaluation quantitative du risque	4
1.3	Les tests de toxicité et traits d'histoire de vie : exemple de <i>Daphnia magna</i>	6
1.4	Les concentrations critiques d'effet	8
1.4.1	Les NOEC/LOEC	8
1.4.2	Les EC _x	13
1.4.3	Les NEC	18
1.5	Problématique	21
1.6	Approches utilisées	21
1.7	Organisation du manuscrit	24

1.1 Contexte et problématique

De nos jours, nul ne peut contester l'importance de la pollution de l'environnement. Cette pollution peut être accidentelle comme l'explosion de la plateforme pétrolière Deepwater Horizon en avril 2010 qui a provoqué une gigantesque marée noire dans le Golfe du Mexique. Mais elle est aussi chronique et liée aussi bien à l'activité industrielle, à l'activité agricole qu'à nos actions quotidiennes. La mise en place du tri sélectif des déchets ménagers, la récupération des médicaments non utilisés et la réduction de la vitesse autorisée sur les routes en cas "d'alerte pollution" montrent à quel point l'opinion publique et les pouvoirs publics sont concernés par la pollution et les impacts qu'elle peut avoir sur nos vies. Les polluants (contaminants ayant un effet néfaste sur l'écosystème) sont considérés comme une menace, à la fois pour la santé publique notamment à travers leur impact sur l'environnement, mais aussi sur le développement économique et enfin sur la biosphère vu leur nombre et leur omniprésence (Ramade, 2010). Ainsi, c'est durant les années 1960 qu'on a pu observer une intensification des recherches concernant la pollution environnementale. C'est en 1969 que le Pr René Truhaut, célèbre toxicologue, a proposé le terme d'écotoxicologie pour désigner

"la branche de la toxicologie qui étudie les effets toxiques des polluants naturels ou synthétiques sur les organismes vivants des écosystèmes" (Molle, 1984). L'écotoxicologie est alors vue comme une branche de la toxicologie mais aussi comme une nouvelle subdivision des sciences écologiques (Ramade, 2010). L'écotoxicologie a été considérée longtemps comme une sous branche de ces deux sciences, la toxicologie et l'écologie. Elle est maintenant considérée comme une science à part entière (Newman, 1995). En effet, l'écotoxicologie pourrait être facilement subdivisée en plus petites sous-disciplines mais il est important de ne pas prendre en compte les différentes composantes individuellement puisque l'écotoxicologie est une science hiérarchique (Figure 1.1) (Newman and Clements, 2008). Ainsi, pour un processus observé à un niveau, on peut expliquer son mécanisme par le niveau inférieur et trouver ses conséquences au niveau supérieur. Cependant, comme toute structure hiérarchique, il est parfois difficile de passer d'un niveau à un autre.

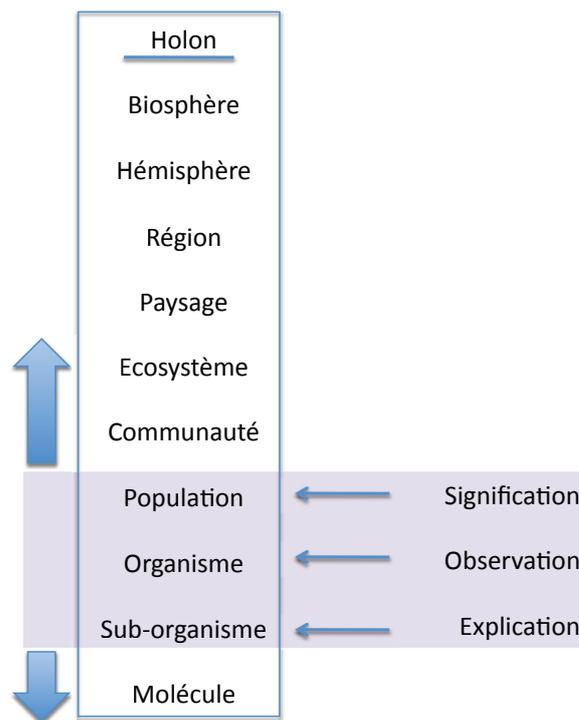


FIGURE 1.1 – Schéma représentant la succession "explication - observation - signification" (non fixée) pour les holons (sub-divisions) de la structure hiérarchique (Newman, 2008).

Au niveau moléculaire, la génomique, la transcriptomique, la protéomique, la métabolomique ainsi que la cancérogenèse peuvent être utilisées en écotoxicologie. De même, les transformations des toxiques organiques et inorganiques sont décrits au niveau mo-

léculaire avec les phénomènes d'activation, d'élimination et de séquestration. Au niveau sub-organisme, l'écotoxicologie fait appel à l'étude de la nécrose ou l'inflammation des tissus, des transferts d'énergie et de matériel entre l'organisme et son environnement avec notamment l'étude des phénomènes de bioaccumulation. C'est dans ces niveaux supra individuels que l'on retrouve l'étude des biomarqueurs, comme par exemple l'induction de l'activité enzymatique EROD (Ethoxy Résorufine-O-Dééthylase) chez les poissons en réponse à une exposition aux HAPs (Hydrocarbure Aromatiques Polycycliques) (Flammarion et al, 2000). Au niveau de l'organisme, ou encore appelé le niveau individuel, les effets létaux et sublétaux sont étudiés avec la possibilité de s'intéresser à différentes voies d'exposition (qui vont dépendre du milieu que l'on étudie, aquatique ou terrestre), différentes durées d'exposition (aiguë ou chronique), différentes espèces et différents contaminants seuls ou en mélange. Il y a donc une grande variété de tests de toxicité possibles au niveau individuel.

C'est à ce niveau individuel que nous allons nous intéresser dans cette thèse. Il s'agit du niveau qui semble le plus simple à observer, puisque des observations sont faites sur un individu afin de déterminer s'il est vivant, s'il s'est développé normalement, s'il a pu se reproduire, s'il a un comportement normal. Alors que les niveaux hiérarchiques inférieurs nécessitent généralement des équipements et des outils perfectionnés pour réaliser les observations, les niveaux supérieurs sont plus complexes et difficilement transposables en laboratoire. Pour autant, l'analyse de ces données réalisées au niveau individuel n'est pas triviale et suscite toujours des questionnements. Ainsi, rien qu'en 2012 dans la revue "Integrated Environmental Assessment and Management", sept communications débattant du traitement des données létales et sublétales ont été publiées (van Dam et al, 2012; Van der Vliet et al, 2012; van Dam et al, 2012; Sanchez-Bayo, 2012; Fox et al, 2012; Newman, 2012; Green et al, 2012). Ces publications débattaient de l'intérêt d'utiliser les modèles concentration-réponse plutôt que les tests d'hypothèse afin de déterminer des concentrations critiques d'effet ou seuils de toxicité à partir des données observées.

L'enjeu est d'autant plus important que depuis 2007, avec l'entrée en vigueur de REACH (Registration, Evaluation, Authorisation and restriction of CHemicals) mis en place par la communauté européenne, toutes les substances chimiques produites (ou importées) à plus d'une tonne par an et par fabricant doivent faire l'objet d'une analyse de risque (European parliament and the council of european union, 2007). Avec la mise en oeuvre de REACH, on assiste à une augmentation des connaissances sur les substances chimiques (nouvelles et existantes) afin de protéger la santé humaine et l'environnement. Ainsi, l'écotoxicologie est dans un schéma global d'évaluation du risque. A partir de données issues de tests de toxicité sur les différents traits d'histoire de vie (survie, croissance, reproduction), il faut fournir des concentrations critiques d'effet pour un contaminant. Il existe différentes concentrations critiques d'effet (NOEC, LOEC, EC_x, NEC) qui peuvent être utilisées mais elles sont issues de méthodes statistiques différentes. C'est le choix des méthodes statistiques utilisées pour obtenir les concentrations critiques d'effet qui est discuté dans la communauté des écotoxicologues. Et par extension, ce sont même les plans d'expérience permettant de fournir ces données (nombre d'observations par concentration, nombre de concentrations étudiées...) qui peuvent être discutés.

C'est pourquoi nous proposons dans cette thèse d'étudier les différentes concentrations critiques d'effet existantes et les méthodes ou modèles permettant de les obtenir, afin d'aller vers une amélioration de l'analyse des données individuelles en écotoxicologie. Nous nous intéresserons également à l'optimisation des plans d'expérience des tests de toxicité permettant de produire les données individuelles afin d'améliorer l'estimation des concentrations critiques d'effet les plus adaptées.

1.2 L'évaluation quantitative du risque

L'évaluation quantitative du risque liée aux substances chimiques dans les milieux aquatiques (ECB, 2003) est basée sur les mêmes principes que toute évaluation du risque. Il y a tout d'abord une phase d'identification du danger (un contaminant qui peut se retrouver dans l'environnement), puis une phase d'appréciation de l'exposition (occurrence et concentrations retrouvées de ce contaminant dans l'environnement), une phase de caractérisation du danger (estimation des effets en fonction de la concentration en contaminant sur une cible donnée) et enfin l'estimation du risque qui se base sur les trois premières phases. Ainsi, dans le cadre de l'évaluation du risque liée aux substances chimiques, deux concentrations sont déterminées :

- La PEC (Predictive Environmental Concentration) ou concentration prédite dans l'environnement qui permet d'estimer l'exposition environnementale. Cette concentration est déterminée à partir de calculs et de modèles de prédiction basés sur les quantités produites ou importées, les quantités utilisées et les propriétés physico-chimiques de la substance qui vont jouer sur sa rémanence dans les différents compartiments (sol, eau, sédiments).
- La PNEC (Predictive No Effect Concentration) ou concentration prédite ne produisant aucun effet, qui permet d'estimer la dangerosité d'une substance. Plus la PNEC sera faible plus la substance présentera un danger important. Elle est déterminée à partir de tests de toxicité dans les laboratoires (biotests, bioessais) sur des organismes de référence et selon des protocoles standardisés. Les résultats de ces tests sont fournis sous forme de concentrations critiques d'effet, encore appelées seuils de toxicité. Les concentrations critiques d'effet habituellement utilisées sont la NOEC, qui est la concentration pour laquelle aucun effet significatif n'a été observé, et la LC₅₀ ou EC₅₀, qui est la concentration provoquant 50% de létalité (L) ou d'effet (E). Des facteurs d'extrapolation sont appliqués sur les concentrations critiques d'effet les plus protectrices pour l'environnement en fonction des durées de test (test aiguë ou chronique), des concentrations critiques d'effet calculées (NOEC, EC₅₀) et des différences observées entre les espèces (algues, daphnies et poissons). Les facteurs d'extrapolation ont pour but de prendre en compte les incertitudes concernant les variations inter-spécifiques, le passage d'un test court à une durée longue et concernant le passage des données de laboratoire à celles du terrain. Le Tableau 1.1 reprend les concentrations critiques d'effet et les facteurs d'extrapolation qui interviennent dans la détermination d'une PNEC.

Données disponibles	Facteurs d'extrapolation
Au moins une L(E)C50 en exposition aiguë pour chacun des trois niveaux trophiques (poisson, daphnie et algue)	1000
Une NOEC en exposition chronique (soit pour le poisson ou la daphnie)	100
Deux NOECs en exposition chronique pour des espèces représentant deux niveaux trophiques (poisson et/ou daphnie et/ou algue)	50
Des NOECs en exposition chronique pour au moins trois espèces (normalement poisson, daphnie et algue) représentant les trois niveaux trophiques	10
Méthode basée sur la distribution de la sensibilité des espèces (SSD)	5-1

Tableau 1.1 – Facteurs d'extrapolation utilisés pour déterminer une PNEC (ECB, 2003).

Ainsi, une fois la PEC et la PNEC calculées, il est possible de déterminer le rapport PEC/PNEC. Si ce rapport est supérieur à un, alors on considère qu'il y a un risque pour l'environnement.

Quand, pour un contaminant donné, des résultats de tests de toxicité sont disponibles pour un grand nombre d'espèces (au minimum 10 mais préférentiellement plus de 15) représentant au minimum huit groupes taxonomiques différents, il est possible de caractériser la distribution de la sensibilité des espèces (SSD). C'est une méthode d'extrapolation statistique dont le principe est de considérer que les sensibilités des différentes espèces à un contaminant donné suivent une distribution paramétrique dont on peut déduire une concentration dangereuse pour moins de 5% des espèces. Dans le cas où une SSD peut être estimée, nous pouvons voir que le facteur d'extrapolation ne varie plus que de 1 à 5 (Tableau 1.1). Avec cette alternative, la variabilité inter-espèces est prise en compte directement et n'a plus besoin d'être prise en compte au niveau du facteur d'extrapolation.

Dans un premier temps, nous allons nous intéresser aux tests de toxicité à partir desquels les concentrations critiques d'effet sont déterminées ou estimées avant d'être utilisées en évaluation du risque. Dans un second temps, nous nous intéresserons aux méthodes ou modèles utilisés pour analyser les données issues des tests de toxicité et permettant l'estimation des concentrations critiques d'effet.

1.3 Les tests de toxicité et traits d'histoire de vie : exemple de *Daphnia magna*

Les principaux traits d'histoire de vie observés pendant un test de toxicité sont la survie, la croissance ou le développement et la reproduction. L'analyse des données issues de ces tests de toxicité va être liée à la spécificité des différents traits d'histoire de vie. Par exemple, l'analyse de la survie implique d'analyser des données de type "vivant ou mort", "oui ou non" : ce type de réponse binaire nécessite une approche spécifique au niveau statistique. La distribution de ces données suivant souvent une loi binomiale.

Les données de croissance, quant à elles, peuvent être des mesures de taille ou de poids en fonction de l'organisme étudié. Les données sont alors continues et non nulles. Les outils statistiques classiques développés pour les données distribuées selon une loi gaussienne peuvent alors souvent être utilisés.

Pour la reproduction, des comptages (de pontes, d'oeufs, de juvéniles...) sont effectués. Pour analyser ce type de données, il peut être recommandé de les transformer (en log ou en racine carrée) afin d'en normaliser la distribution. Pour les jeux de données comportant des zéros, ce qui peut être le cas de données de reproduction, la transformation en log nécessite de s'interroger sur la gestion des zéros et peut aboutir à des prédictions négatives après modélisation (O'Hara and Kotze, 2010). L'utilisation de distributions de type Poisson, dans ces cas de figures, semble plus adaptée à ce type de données.

Ainsi, le choix du trait d'histoire de vie et de la réponse étudiée va conditionner l'analyse des données. De plus, le choix de l'étude d'un de ces traits d'histoire de vie va

également conditionner la durée du test. Les effets d'un toxique sur la reproduction et la croissance s'observent généralement sur des durées plutôt longues avec des concentrations en toxique plutôt faibles afin de ne pas obtenir trop de mortalité. Des tests chroniques sont alors réalisés et on parle d'effets sublétaux. Les effets sur la survie, appelés effets létaux, peuvent être observés pendant des tests chroniques mais aussi dans le cas d'une exposition à une forte dose de toxique pendant une durée courte, lors des tests de toxicité aiguë. La notion de durée du test de toxicité va être relative à la durée du cycle de vie des espèces étudiées. En effet, les tests de toxicité peuvent être effectués sur de nombreuses espèces mais l'évaluation du risque est basée sur des espèces de référence pour lesquelles des tests standards ont été développés. Ainsi, comme nous avons pu le voir dans le Tableau 1.1, des espèces différentes ont été choisies en fonction à la fois de l'importance de leur présence dans le milieu naturel, de leur capacité à être élevées en laboratoire et de leur appartenance à différents niveaux trophiques (producteurs primaires, secondaires, consommateurs...).

Dans les tests de toxicité standard, un petit crustacé d'eau douce, *Daphnia magna*, est souvent utilisé (Figure 1.2) (ECB, 2003). Les daphnies sont des organismes du zooplancton qui sont très importants dans la chaîne alimentaire et sont omniprésents. En effet, on les trouve dans toutes les rivières où elles jouent un rôle important dans l'écosystème puisqu'elles se nourrissent de petites particules suspendues dans l'eau en la filtrant (Ebert, 2005). De plus, les daphnies sont facilement cultivables en laboratoire. Parmi toutes les espèces de daphnies, l'espèce *Daphnia magna* est la plus grande et donc la plus facile à manipuler et à observer (Adema, 1978). Les daphnies mesurent environ 0,8 mm à l'éclosion et peuvent atteindre 4 à 6 mm à l'âge adulte. Leur croissance se fait par mues successives. De plus, en conditions favorables, les daphnies se reproduisent par parthénogenèse (mode de reproduction monoparentale à partir d'un individu femelle) ce qui permet de maintenir une stabilité génétique dans les élevages en laboratoire (Figure 1.3). C'est seulement quand les conditions deviennent défavorables (changement de température, manque de nourriture) que des mâles apparaissent et qu'il y a accouplement. Les oeufs issus de l'accouplement, appelés éphippies, n'éclorent que lorsque les conditions redeviendront favorables. Cependant ce mode de reproduction est assez rare par rapport à la reproduction parthénogénétique. Concernant la reproduction parthénogénétique, des oeufs sont produits à chaque mue puisque *Daphnia magna* est munie d'une carapace bivalve avec fente ventrale. Une fois produits, les oeufs sont placés dans une chambre d'incubation dorsale et y restent environ 3 jours. Les mues sont généralement espacées de 2 à 3 jours. Les daphnies sont capables de se reproduire à partir de 5 à 10 jours selon la température et la taille des pontes peut atteindre 100 oeufs et en laboratoire, les femelles peuvent vivre jusqu'à 2 mois.

Les deux principaux tests de toxicité sur *Daphnia magna* sont le test d'immobilité et le test de reproduction. Concernant le test d'immobilité, les daphnies ne restent généralement pas longtemps immobiles et une immobilité prolongée (plusieurs dizaines de secondes) est un signe de mort ou de mort imminente. Ainsi ce test d'immobilité correspond à un test de survie. Le principe de l'essai est d'exposer de jeunes daphnies (âgées de moins de 24h) à la substance d'essai à différentes concentrations pendant une durée de 48h. L'immobilisation est enregistrée à 24 et 48h puis comparée à des valeurs de contrôle. Les résultats sont utilisés pour déterminer une concentration critique d'effet



FIGURE 1.2 – Femelle adulte *Daphnia magna* : (Ebert, 2005).

à 48h, celle à 24h étant facultative. Les recommandations concernant ce test sont les suivantes : "utiliser pour chaque concentration d'essai et chaque témoin au moins 20 organismes répartis de préférence en 4 groupes de 5. Il faut utiliser au moins cinq concentrations formant une série géométrique de concentrations successives séparées, de préférence, par un facteur inférieur ou égale à 2,2. L'utilisation de moins de cinq concentrations doit être justifiée. De préférence la concentration la plus élevée doit provoquer 100% d'immobilisation et la moins élevée ne doit donner lieu à aucun effet observable" (OCDE, 2004). L'autre principal test de toxicité est le test de reproduction, dont le principe est d'exposer de jeunes daphnies (âgées de moins de 24h) à la substance d'essai pendant 21 jours, pendant lesquels la survie doit également être observée. Dix individus doivent être répartis individuellement pour chaque concentration et pour le contrôle. Il faut utiliser au moins cinq concentrations formant une série géométrique de concentrations successives séparées, de préférence, d'un facteur inférieur ou égal à 3,2 (OCDE, 2012).

1.4 Les concentrations critiques d'effet

Généralement, ce que nous appelons "les concentrations critiques d'effet" peuvent être appelées "seuils de toxicité" ou "valeurs de référence". Nous avons choisi ce terme qui permet, à notre avis, de mieux décrire ce qu'elles sont. En effet, ce sont des concentrations pour lesquelles nous observons un certain niveau d'effet qui nous intéresse. Nous n'utilisons pas le terme "seuils" qui peut faire penser qu'en dessous du seuil, il n'y a pas de toxicité. Il existe plusieurs concentrations critiques d'effet que nous allons présenter. Afin d'essayer d'illustrer le principe de calcul de ces concentrations critiques d'effet, nous avons choisi un exemple de jeu de données représentant la taille mesurée en fonction des concentrations testées (2, 4, 8, 16, 32 en unité de concentration) d'un composé quelconque (Figure 1.4).

1.4.1 Les NOEC/LOEC

La NOEC (No Observed Effect Concentration) est la concentration en substance testée la plus élevée pour laquelle aucun effet significatif par rapport au témoin (ou

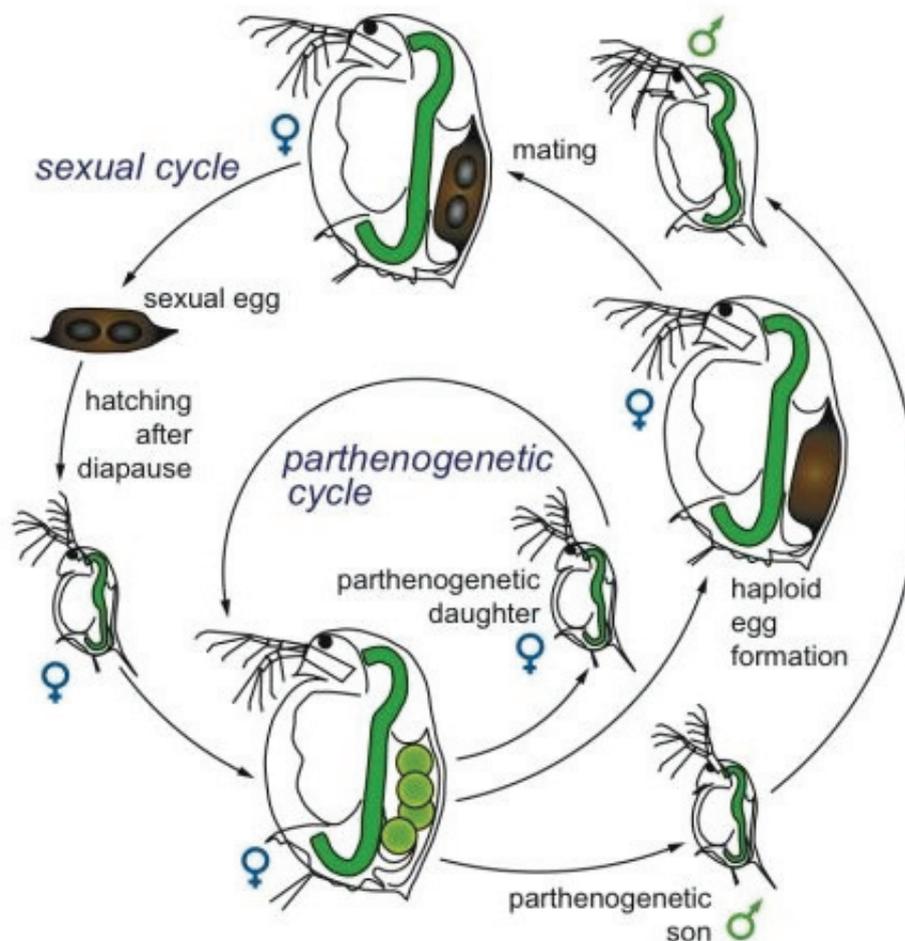


FIGURE 1.3 – Cycle de vie de *Daphnia magna* (Ebert, 2005). Le cycle parthénogénétique est le cycle prépondérant et en cas de conditions défavorables, des mâles sont produits et un cycle sexuel avec apparition d'oeufs issus d'accouplements apparaissent. Ces oeufs sont une forme de résistance et n'éclore que lorsque les conditions redeviendront favorables.

contrôle) n'a été observé. La LOEC (Lowest Observed Effect Concentration) correspond à la concentration la plus petite pour laquelle un effet significativement différent par rapport au témoin a été observé. La LOEC est donc la concentration qui suit immédiatement la NOEC dans la gamme des concentrations testées. Les valeurs des NOEC et LOEC sont, par définition, obligatoirement des concentrations testées et donc sont issues d'un choix préalable de l'écotoxicologue. C'est une des critiques faites et reprises depuis des années à l'encontre de cette concentration critique d'effet. De plus, les NOEC/LOEC ne peuvent être déterminées que pour un temps d'exposition donné. C'est généralement celui de la fin d'expérience qui est naturellement choisi. Des tests de comparaisons multiples sont utilisés pour déterminer ces valeurs. En effet, il s'agit de comparer les données de plusieurs concentrations par rapport au témoin. Il y a

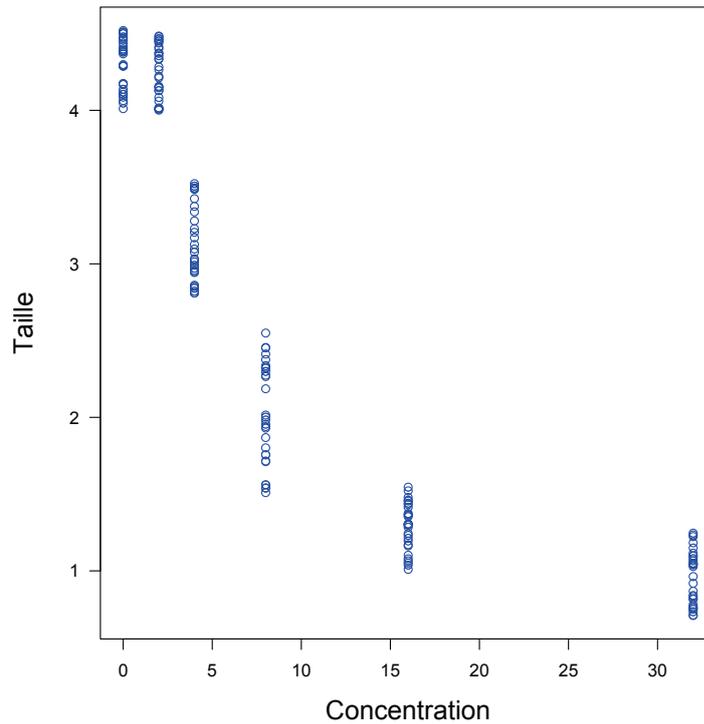


FIGURE 1.4 – Représentation de la taille mesurée des organismes en fonction de la concentration testée en contaminant (jeu de données fictif)

donc nécessité de faire plusieurs tests d'hypothèse. Le principe d'un test d'hypothèse est de formuler une hypothèse nulle H_0 qui présuppose une homogénéité entre deux échantillons. En l'occurrence pour l'écotoxicologie, les échantillons sont :

- les données observées pour une concentration testée
- les données observées pour le groupe témoin

et de vérifier si l'hypothèse H_0 est rejetée ou non. Pour cela, la distance entre deux échantillons est calculée et logiquement plus la distance sera grande plus la probabilité de rejeter H_0 sera grande. A chaque mise en oeuvre du test, il y a un risque de rejeter à tort H_0 sachant que H_0 est vraie, cette probabilité correspond au risque de première espèce appelé α . De même, on a un risque d'accepter H_0 à tort alors qu'elle est fautive, cette probabilité est appelée risque de deuxième espèce et est notée β . Cette probabilité est difficile à évaluer et donc elle n'est quasiment jamais utilisée, alors qu'elle a une importance cruciale dans notre contexte. En effet, elle permet d'évaluer le risque de conclure qu'il n'y a pas de différence d'effet entre une concentration et le témoin (de conclure que le produit n'est pas toxique pour une concentration), alors qu'il y en a une (le produit a bien un effet à cette concentration). Sous H_0 , la valeur de p (p-value en anglais) est calculée. Elle représente la probabilité d'observer, sous H_0 une différence au moins aussi grande que celle observée sur l'échantillon dont nous disposons. La conclusion du test se fait en fonction du risque α (généralement 5%), si p est inférieure à α alors H_0 est rejetée et nous en déduisons que les deux échantillons sont significativement différents. Les tests sont généralement utilisés pour rejeter H_0 , et la probabilité

de rejeter H_0 alors qu'elle est fautive constitue la puissance du test, qui est liée à la taille de l'échantillon. Ainsi, dans le cadre des tests pratiqués en écotoxicologie, il faut avoir un grand nombre de données par concentration testée pour avoir une puissance acceptable, d'où un nombre élevé d'individus par concentrations recommandé dans les tests de toxicité (par exemple, les tests d'immobilisation et de reproduction de *Daphnia magna*). Ainsi, il est imprudent d'accepter H_0 quand p est supérieure à α et c'est ce qui est fait lorsque la NOEC est déterminée puisque cette concentration est considérée comme non significativement différente du témoin. La NOEC est donc basée sur une interprétation abusive et peu prudente du résultat d'un test d'hypothèse.

Dans le guide OCDE traitant de l'analyse des données (OECD, 2006), la distinction est faite entre les données de type binaire (données de type survie), les données continues (données de type croissance) et les données discrètes (données de comptages comme pour la reproduction). Concernant ces dernières données, il est indiqué qu'elles sont souvent traitées comme étant des données continues. Les principaux tests recommandés par le guide OCDE (OECD, 2006) en fonction du type de données et de l'hypothèse sur la monotonie de la réponse sont présentés sur les (Figures 1.5 et 1.6). Il n'est pas forcément simple de choisir le bon test d'hypothèse (Chapman et al, 1996) en fonction des données que l'on veut analyser.

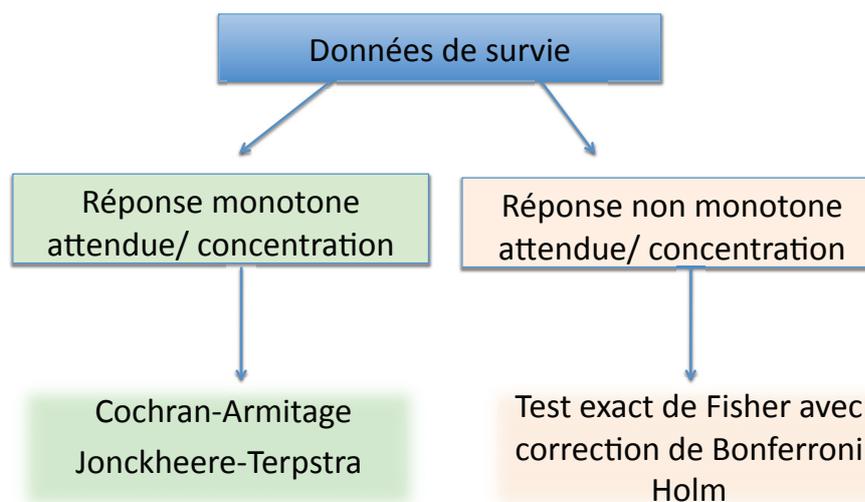


FIGURE 1.5 – Analyse des données de type binaire d'après OECD (2006)

Pour analyser les données de notre exemple, vu que nous attendons une réponse monotone par rapport aux concentrations testées, nous avons réalisé une ANOVA (analyse de variance), mais les conditions de son utilisation n'étaient pas respectées (pas d'homogénéité des résidus). Au final, le test de Jonckheere-Terpstra (Annexe 1) a été

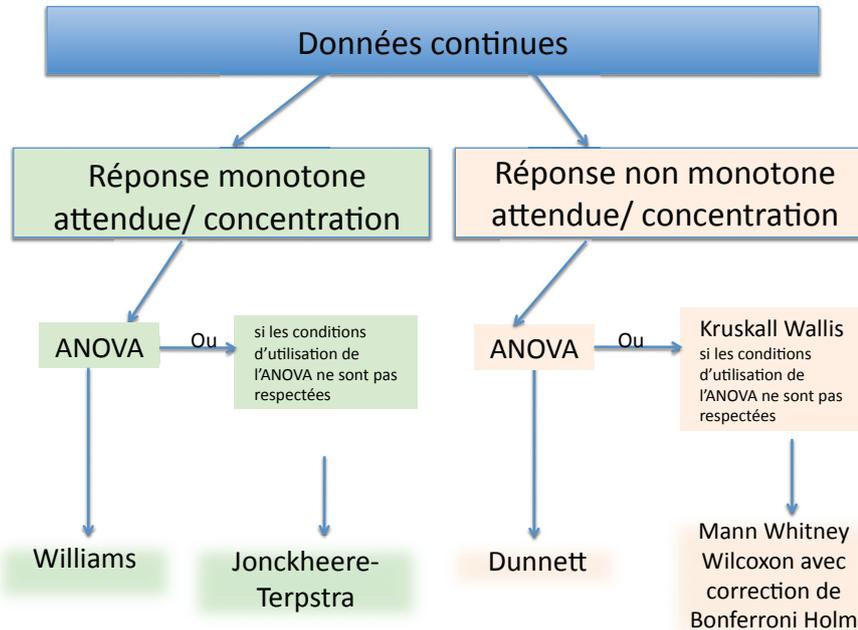


FIGURE 1.6 – Analyse des données de type continu d’après [OECD \(2006\)](#)

utilisé et a montré qu’il n’y avait pas de différence significative entre la première concentration testée (égale à 2 unités de concentration) et le témoin. Ainsi, par définition, la NOEC prend la valeur de la première concentration testée (2 unités de concentration) et la LOEC la valeur de la deuxième concentration testée (4 unités de concentration). Les valeurs de concentrations critiques d’effet obtenues sont alors des concentrations testées et choisies par l’écotoxicologue (Figure 1.7). De plus, sur cette figure, il est facile d’imaginer que si l’écotoxicologue n’avait pas choisi de tester la deuxième concentration (4 unités de concentration) alors la LOEC aurait été égale à la troisième concentration qui est égale à 8 unités de concentration.

Alors que les "défenseurs" de la NOEC arguent que les solutions alternatives ne sont pas satisfaisantes (utilisation des EC_x) ([Green et al, 2012](#)), les reproches faits sur l’utilisation des tests d’hypothèse pour calculer des NOEC sont repris par un grand nombre de scientifiques ([Fox et al, 2012](#); [Jager, 2012](#); [Delignette-Muller et al, 2011](#); [Newman, 2008](#); [Fox, 2008](#); [Isnard et al, 2001](#); [Crane et Newman, 2000](#); [Nelder, 1999](#); [Chapman et al, 1996](#)) et peuvent être résumés en trois parties :

- l’utilisation abusive des tests d’hypothèse pour estimer une valeur seuil. En effet, les tests d’hypothèse, comme nous l’avons vu, permettent de comparer des échantillons. Le résultat du test peut être "non, ils ne sont pas identiques" ou "il n’est pas possible de déceler une différence significative" mais ne peut pas être "oui, les échantillons sont identiques". Ce problème peut être résumé par cette citation d’[Altman and Bland \(1995\)](#) : "L’absence de preuve n’est pas la preuve de l’absence". Ainsi, ces tests ne peuvent pas avoir pour résultat une valeur (concen-

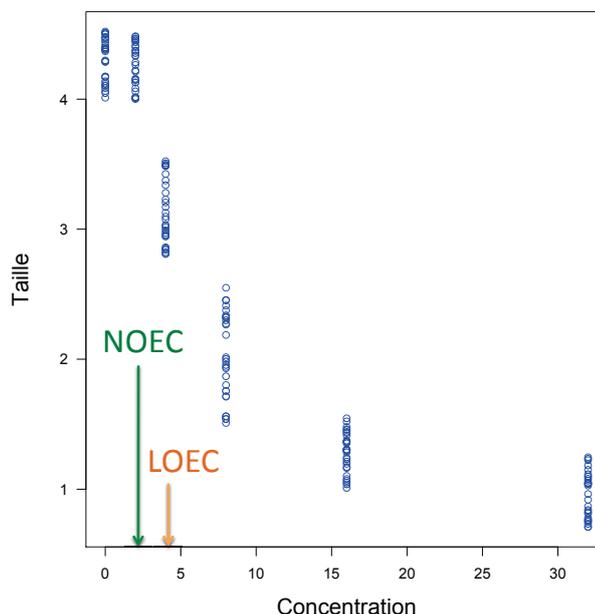


FIGURE 1.7 – Détermination des NOEC/LOEC sur notre jeu de données fictif

tration critique d'effet) (Fox et al, 2012; Delignette-Muller et al, 2011; Newman, 2008; Fox, 2008).

- la NOEC obtenue ne peut être qu'une concentration testée expérimentalement (comme nous l'avons vu dans l'exemple) et elle est donc clairement dépendante du choix initial de l'écotoxicologue.
- seul le risque α (probabilité de dire qu'une concentration a un effet alors qu'elle n'en a pas) est contrôlé alors qu'en écotoxicologie, c'est le risque β qui est important. C'est la probabilité de considérer qu'une concentration n'a pas d'effet alors qu'elle en a un, qui est intéressante. Ainsi, dans le cas d'un plan expérimental trop léger (*e.g.* taille d'échantillon trop petite, concentrations trop espacées) ces tests ne permettent pas de différencier les échantillons et peuvent amener à conclure qu'une substance est moins toxique que ce qu'elle est réellement (Moore et Caux, 1997).

1.4.2 Les EC_x

Nous allons voir que les EC_x peuvent être estimées en utilisant des modèles concentration-réponse mais aussi par les intervalles de confiance simultanés des ratios des moyennes.

Les EC_x obtenues par les modèles concentration-réponse

Les EC_x correspondent aux concentrations produisant $x\%$ d'effet sur la réponse mesurée. Dans le cas de la survie, on parle de concentration létale (LC_x). Les valeurs des EC_x sont estimées en ajustant des modèles concentration-réponse sur les données

pour une durée d'exposition donnée (21 jours d'exposition pour le test de reproduction chez la daphnie). Les valeurs de x sont choisies entre 5 et 50, mais c'est la valeur x égale à 50 qui fait souvent référence (ECB, 2003). En théorie, un grand nombre de modèles concentration-réponse peut être utilisé mais en pratique, les modèles concentration-réponse utilisés pour analyser ce type de données sont le modèle log-logistique (ou encore appelé modèle de Hill dans sa forme à trois paramètres), le modèle log-probit et le modèle de Weibull (OECD, 2006; Moore et Caux, 1997). Voici l'écriture que peuvent prendre les modèles concentration-réponse dans le cas de l'analyse de données de croissance. La taille $f(\text{conc}_i)$ est fonction de la concentration conc_i avec comme paramètres : d la taille en absence de contaminant, c la taille minimale pouvant être atteinte et b est un paramètre de pente.

Deux formes classiques équivalentes de l'écriture du modèle log-logistique :

$$f(\text{conc}_i) = c + \frac{d - c}{1 + \exp(b \times (\log(\text{conc}_i) - \log(EC_{50})))} \quad (1.1)$$

$$f(\text{conc}_i) = c + \frac{d - c}{\frac{\text{conc}_i^b + EC_{50}^b}{EC_{50}^b}} \quad (1.2)$$

Si $c = 0$, on retrouve la forme du modèle de Hill :

$$f(\text{conc}_i) = d \times \frac{EC_{50}^b}{EC_{50}^b + \text{conc}_i^b} \quad (1.3)$$

La fonction Log-probit s'écrit de la manière suivante :

$$f(\text{conc}_i) = c + (d - c) \times (1 - \phi(b \times \log_{10}(\frac{\text{conc}_i}{EC_{50}}))) \quad (1.4)$$

avec ϕ la fonction de répartition de la loi normale.

La fonction de Weibull s'écrit de la manière suivante :

$$f(\text{conc}_i) = c + (d - c) \times \exp(-\log(2) \times (\frac{\text{conc}_i}{EC_{50}})^b) \quad (1.5)$$

Pour modéliser la taille (y_{ij}) d'un individu (j) exposé à une i^e concentration, deux parties sont nécessaires dans le modèle (Figure 1.8) :

- une partie déterministe qui donne la valeur ou la courbe moyenne définie par la fonction $f : f(\text{conc}_i)$
- une partie stochastique qui décrit comment les valeurs observées y_{ij} se répartissent autour de cette moyenne.

Pour une partie stochastique suivant une loi normale :

$$y_{ij} \sim \text{Normale}(f(\text{conc}_i), \sigma) \quad (1.6)$$

Pour des variables non continues la partie stochastique peut s'écrire différemment en fonction du type de données :

- dans le cas de données de survie, le nombre de survivants N_i à la concentration $conc_i$ va suivre une loi binomiale avec $S(conc_i)$ le taux de survie et N_{i0} le nombre initial d'individus.

$$N_i \sim \text{Binomiale}(S(conc_i), N_{i0}) \quad (1.7)$$

- dans le cas de données de reproduction, le nombre d'oeufs N_{ij} pour un individu j à la concentration $conc_i$ va suivre une loi de Poisson avec $\lambda(conc_i)$ le nombre moyen d'oeufs par individu.

$$N_{ij} \sim \text{Poisson}(\lambda(conc_i)) \quad (1.8)$$

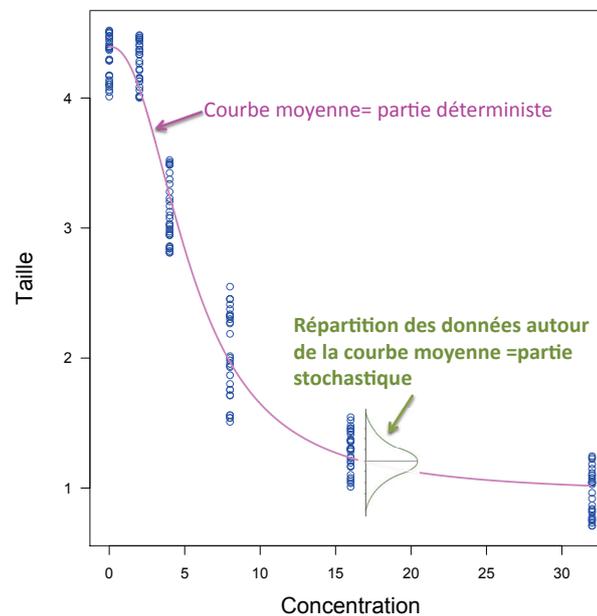
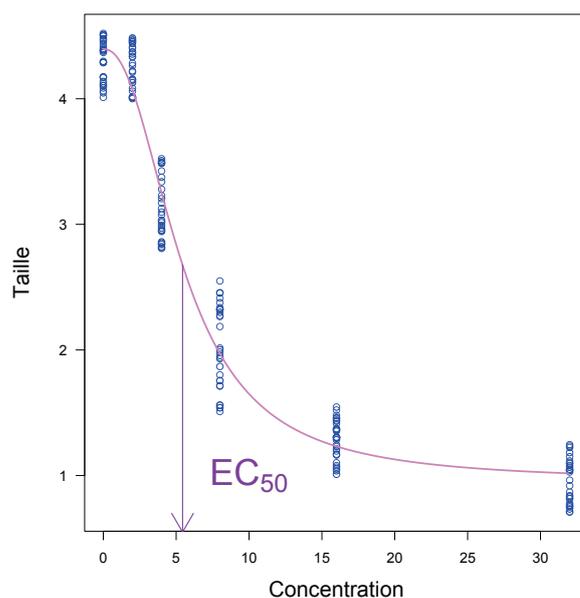


FIGURE 1.8 – Parties déterministe et stochastique d'un modèle.

Un modèle log-logistique a été ajusté sur les données de notre exemple, ce qui a permis d'estimer une valeur de EC_{50} à 5.44 en unité de concentration (Figure 1.9).

L'utilisation des EC_x est considérée comme la meilleure alternative (et la plus ancienne proposée) à l'utilisation des NOEC/LOEC (van der Hoeven, 1997). Cependant, les changements ne se font pas rapidement (Fox et al, 2012; van Dam et al, 2012). En effet, même si l'utilisation des LC_x et des EC_x résout beaucoup des problèmes rencontrés avec l'utilisation de la NOEC, elle peut être aussi controversée.

L'un des inconvénients souvent avancé contre l'utilisation des EC_x est que l'on doit choisir le x et que ce choix est très important (Green et al, 2012). En effet, sur la Figure 1.9, la EC_{50} correspond à 50% d'effet, ainsi la EC_{20} par exemple donnera un résultat bien différent. En ce qui concerne l'évaluation du risque l'effet intéressant à connaître

FIGURE 1.9 – Détermination de la EC_{50}

est le 0% d'effet. Cependant, en dessous de $x=10\%$, le risque est plus grand d'avoir des résultats dépendant du modèle utilisé (Moore et Caux, 1997) et la EC_0 est à zéro d'après le modèle. Les détracteurs des EC_x , arguent également que le choix du modèle concentration-réponse n'est pas facile à faire et que la prise en compte des réplicats n'est pas faite avec les modèles concentration-réponse (Green et al, 2012). Un autre point négatif qui est mis en avant contre l'utilisation des modèles concentration-réponse, est qu'ils ne prennent pas en compte le côté mécaniste et dynamique des processus de toxicité (Jager, 2012, 2011).

Les EC_x obtenues par les intervalles de confiance simultanés des rapports des moyennes

Nous avons essayé de faire le lien entre les méthodes statistiques utilisées pour déterminer des NOEC et la détermination des EC_x en utilisant une nouvelle approche basée sur les intervalles de confiance plutôt que sur le calcul de valeurs de p (p-value) (Gardner et Altman, 1986). Par exemple, pour comparer l'effet d'une concentration en contaminant par rapport au témoin (ou contrôle) et quantifier l'effet de cette concentration, on calcule l'intervalle de confiance du rapport de la moyenne de la réponse à cette concentration sur la moyenne de la réponse pour le témoin. Si l'ensemble des valeurs de cet intervalle est strictement en dessous de $x\%$, alors on peut dire qu'il y a moins de $x\%$ d'effet entre la concentration et le témoin et donc que la EC_x est supérieure à cette concentration.

Dans le cas de tests de comparaisons multiples, il est important d'estimer des intervalles de confiance simultanés pour les différents rapports des moyennes (chaque concentration comparée au témoin) (Dilba et al, 2006) afin de ne pas cumuler les risques d'erreur. Un outil a été récemment développé qui permet d'obtenir ces inter-

valles : le package *mratios* (Dilba et al, 2011) de R (R Development Core Team, 2010). Cet outil permet d'interpréter les résultats d'une procédure de Dunnett (Dunnett, 1964) qui compare chaque concentration testée avec le témoin, et d'éviter d'utiliser les tests statistiques pour déterminer des valeurs (NOEC et LOEC) (Delignette-Muller et al, 2011). Pour illustrer, cette méthode nous allons reprendre l'exemple des données de taille après exposition à différentes concentrations (Figure 1.10).

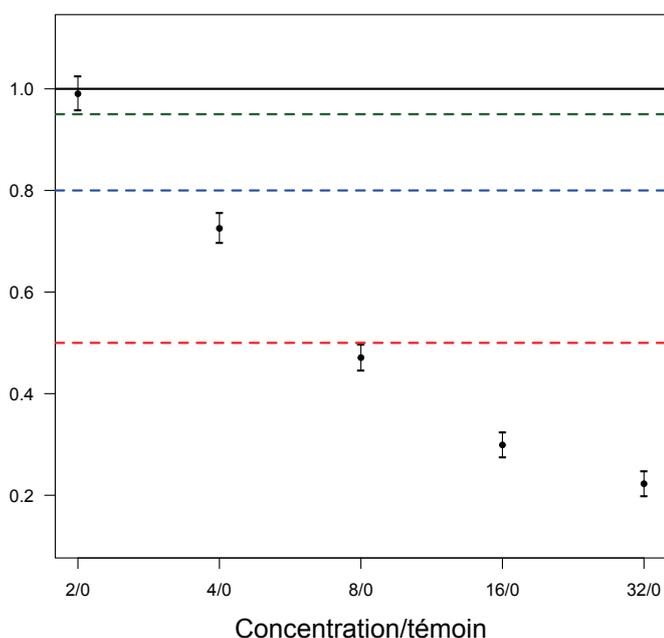


FIGURE 1.10 – Intervalles de confiance simultanés à 95% des rapports des moyennes de la taille à chaque concentration testée par rapport à la moyenne de la taille chez le témoin. Les lignes horizontales représentent le niveau d'effet, en noir aucun effet (EC_0 : rapport=1), en vert 5% d'effet (EC_5 : rapport= 0,95), en bleu 20% d'effet (EC_{20} : rapport= 0,80) et en rouge 50% d'effet (EC_{50} : rapport=0,50).

Pour un pourcentage d'effet égal à $x\%$ et sa ligne horizontale correspondante, si un intervalle de confiance simultané pour un rapport concentration/témoin est entièrement en dessous (ou en dessus) de cette ligne alors on peut dire que la EC_x est en-dessous (ou au-dessus) de cette concentration. Quand un intervalle de confiance coupe la ligne horizontale alors rien ne peut être conclu. Si l'intervalle est au-dessus de la ligne horizontale correspondant à $x=0\%$ alors on a un effet positif de la concentration par rapport au témoin. Il peut s'agir, par exemple, d'un effet hormésis pour les faibles concentrations en contaminant. Sur la Figure 1.10, l'intervalle de confiance du rapport "2/0" (de la moyenne de la taille à la concentration 2 sur la moyenne de la taille du témoin) coupe la ligne horizontale noire, on ne peut rien dire de cette concentration 2 par rapport à la EC_0 . Par contre, cet intervalle est au-dessus de la ligne verte et l'intervalle de confiance du rapport "4/0" (de la moyenne de la taille à la concentration

4 sur la moyenne de la taille du témoin) est bien en-dessous de la ligne verte. Nous pouvons alors conclure que la EC_0 est entre 0 et 4 et la EC_5 est entre 2 et 4 en unités de concentration. De même, nous pouvons dire que la EC_{20} est entre 2 et 4 et que la EC_{50} est entre 4 et 8 en unités de concentration. Dans l'article de [Delignette-Muller et al \(2011\)](#), le but de cette nouvelle approche était d'exploiter les informations fournies par les tests statistiques de manière correcte en tentant d'estimer des EC_x d'intérêt sans utiliser de modèle concentration-réponse. Dans cet article, les valeurs de EC_x obtenues avec cette méthode sont comparées avec celles obtenues avec ajustement d'une courbe concentration-réponse. Il en ressort que l'utilisation des modèles concentration-réponse reste toutefois plus efficace pour estimer des EC_x . L'utilisation de cette nouvelle approche permet de fournir des bornes autour des valeurs importantes de concentrations critiques, même celle ne menant à aucun effet (EC_0). Pour l' EC_0 , l'intervalle est comme attendu entre 0 et la LOEC.

1.4.3 Les NEC

L'utilisation d'un modèle à seuil permet d'estimer une concentration critique en-dessous de laquelle aucun effet n'est observé, c'est la NEC (No Effect Concentration). Au dessus de cette concentration, les effets vont augmenter en suivant une fonction de stress dépendante de la quantité en contaminant (Figure 1.11). Cette concentration critique d'effet a été proposée par [Kooijman et Bedaux \(1996\)](#) dans le cadre de l'approche DEBtox, elle-même s'appuyant sur la théorie DEB.

La théorie DEB (Dynamic Energy Budget) vise à décrire de manière générale l'acquisition de l'énergie par un organisme et son utilisation pour la croissance, la reproduction et la survie en liant tous les processus entre eux ([Nisbet et al, 2000](#)). La théorie DEB propose des lois générales qui peuvent être appliquées quel que soit l'organisme étudié et qui permettent de relier entre elles, les différentes étapes liées à l'allocation de l'énergie (Figure 1.12). L'utilisation de l'approche DEBtox permet d'étudier les effets d'un contaminant sur les différentes étapes liées à l'allocation de l'énergie. Les modèles DEBtox prennent en compte : la cinétique de contamination (modèle toxicocinétique liant la concentration interne à la concentration externe), les effets létaux et sub-létaux via une fonction de stress nulle quand la concentration en contaminant est inférieure à la NEC et qui devient proportionnelle à la concentration moins la NEC, quand la concentration en contaminant est supérieure à la NEC. Les avantages de cette approche sont de gagner en pertinence biologique et toxicologique par rapport aux méthodes permettant de calculer une NOEC ou une EC_x . La NEC peut être calculée pour l'ensemble d'une expérience, en prenant en compte tous les temps (elle n'est donc pas fonction du temps) et tous les traits d'histoire de vie affectés (croissance, reproduction et survie) ([Ducrot et al, 2010](#); [Billoir et al, 2008a](#); [Jager, 2004](#)). Cependant, les modèles DEBtox restent assez difficiles à utiliser, car les jeux de données standards ne sont souvent pas suffisants pour estimer l'ensemble des paramètres. Des travaux tendent à simplifier les modèles DEBtox ([Jager et Zimmer, 2012](#)) ou à utiliser la NEC comme une concentration seuil en utilisant d'autres modèles ([Pires et al, 2002](#)). Certains écotoxicologues sont contre l'utilisation d'un modèle à seuil en biologie car cela présuppose l'existence véritable de ce seuil, certains arguments sont résumés dans [van Straalen \(1997\)](#).

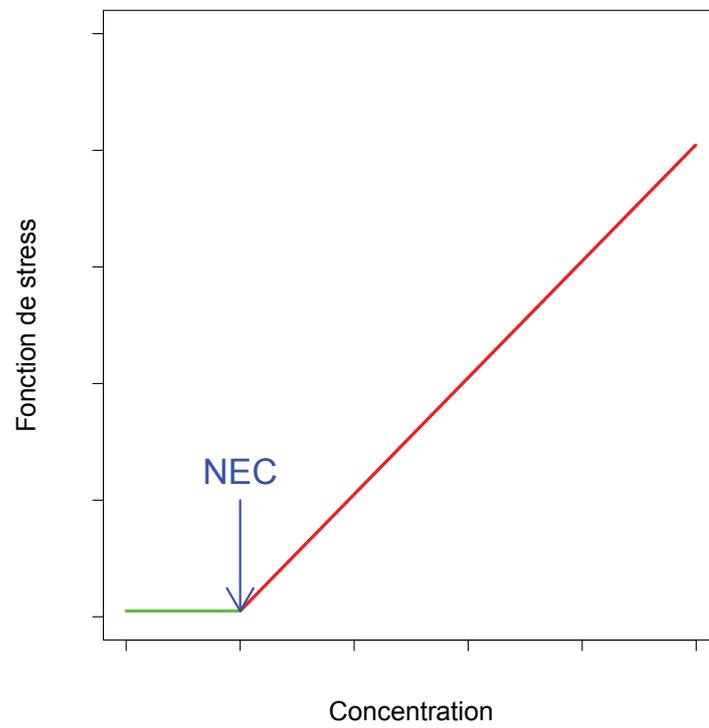


FIGURE 1.11 – Représentation de la fonction de stress par rapport à la NEC. Pour les concentrations en-dessous de la NEC, la fonction de stress est nulle puis elle va augmenter en fonction de la concentration.

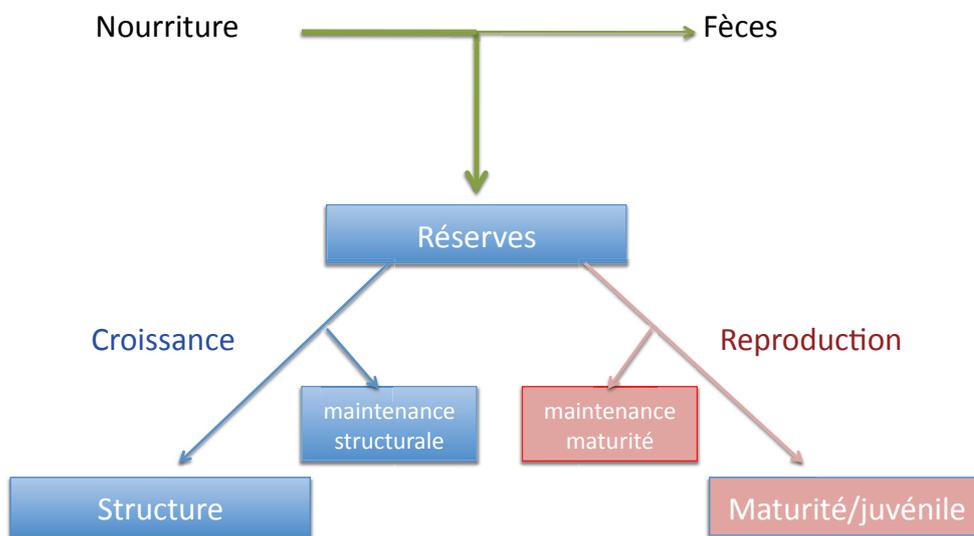


FIGURE 1.12 – Schéma de la théorie DEB (Dynamic Energy Budget) (Nisbet et al, 2000).

1.5 Problématique

L'évaluation quantitative du risque en écotoxicologie est basée sur la détermination ou l'estimation de concentrations critiques d'effet obtenues à partir de tests de toxicité en laboratoire. Ces bioessais sont standardisés mais l'analyse statistique de ces bioessais ne l'est pas vraiment et de nombreux choix (concentrations critiques d'effet, méthodes d'analyse, modèles) sont laissés à l'écotoxicologue sans qu'il n'ait forcément les informations pour faire ces choix. D'ailleurs, ces choix sont toujours débattus comme nous avons pu le voir précédemment. Le premier objectif de ce travail de thèse est de partir des données issues des bioessais et d'offrir aux écotoxicologues la possibilité de faire ces choix de manière éclairée afin d'estimer les concentrations critiques d'effet de la manière la plus pertinente possible. Le second objectif de cette thèse est de voir si on peut améliorer l'estimation des concentrations critiques d'effet en modifiant les plans d'expérience permettant la réalisation des bioessais. En effet, les plans d'expérience ayant été définis pour déterminer des NOEC et des LOEC, il est nécessaire d'explorer les pistes qui pourraient permettre de les adapter à l'estimation des autres concentrations critiques d'effet.

1.6 Approches utilisées

Choix des traits d'histoire de vie étudiés

Pour répondre à notre problématique, nous avons choisi de nous placer dans deux cas de figure : l'analyse de données de survie et l'analyse de données de reproduction. Ces deux types de données ont en commun qu'elles ne sont pas des données continues et qu'elles ne suivent pas une loi normale. Les données de survie sont de type binaire ou dichotomique (vivant ou mort) et les données de reproduction sont des comptages.

Choix des jeux de données étudiés

Pour l'analyse de la survie, nous nous sommes intéressés aux jeux de données de *Daphnia magna* issus de la littérature. Pour l'optimisation des plans d'expérience, à partir d'informations *a priori* issues de la littérature sur les jeux de données de survie, nous avons construit des jeux de données simulées.

En ce qui concerne la reproduction, nous avons pu travailler sur des données de *Lymnaea stagnalis* exposée au cadmium. *Lymnaea stagnalis* est un gastéropode pulmoné d'eau douce qui se nourrit d'algues (Figure 1.13). C'est un animal hermaphrodite simultané, c'est-à-dire qu'il a besoin d'un autre individu pour se reproduire. Il est utilisé en écotoxicologie depuis plusieurs années (Jumel et al, 2002; Russo and Lagadic, 2004; Russo et al, 2007) notamment pour sa sensibilité aux perturbateurs endocriniens (Lagadic et al, 2007; Ducrot et al, 2010). C'est d'ailleurs l'une des deux espèces retenues avec *Potamopyrgus antipodarum*, pour l'élaboration d'une ligne directrice de l'OCDE (Organisation de Coopération et de Développement Economique) concernant le test de reproduction chez les mollusques (Matthiessen, 2008; OCDE, 2010). Dans ce cadre, un groupe de travail s'est formé avec pour objectif de développer et de valider un test standard de reproduction chez les mollusques pour cette ligne directrice de l'OCDE.

Ce groupe de travail est composé de différentes structures comme des universités, des laboratoires et des organismes nationaux de différents pays européens dont le Royaume-Uni, l'Allemagne, le Danemark et la France (Annexe 2). La mise en place de ce type de test standard nécessite deux phases :

- Une phase de pré-validation pendant laquelle les experts mettent en commun leurs connaissances sur les espèces étudiées afin d'écrire un protocole standard d'opérations (POS) pour la culture et la reproduction de ces espèces. Des tests de pré-validation (basés sur le POS) sont alors réalisés par plusieurs laboratoires selon un plan d'expérience commun. Les résultats sont analysés et les tests sont optimisés (si nécessaire) avant de soumettre le POS et ses résultats à un comité de validation de l'OCDE pour achever cette première phase.
- Une phase de validation peut alors être enclenchée avec réalisation d'un ring-test à plus large échelle, basé sur le POS consolidé. Les résultats sont compilés et analysés avant d'être validés. Une nouvelle ligne directrice de l'OCDE concernant une version standardisée du test peut alors être soumise.

Nous avons eu l'opportunité en mai 2012 de rejoindre ce groupe de travail pour analyser les résultats de la pré-validation (analyse des données de reproduction) et optimiser le plan d'expérience. Dans cette thèse, nous présenterons les résultats obtenus sur *Lymnaea stagnalis*, les tests de toxicité sur *Potamopyrgus antipodarum* ayant été réalisés plus tardivement. Le bioessai (test) sur *Lymnaea stagnalis* a été réalisé sur 56 jours avec six réplicats de cinq organismes par concentration et pour le contrôle. La question du groupe de travail était de savoir s'il est possible de réduire la durée d'exposition et éventuellement le nombre d'organismes testés sans nuire à l'estimation des EC_x . Ainsi, que ce soit pour l'analyse de données ou l'optimisation du plan d'expérience, nous avons travaillé sur des jeux réels de données, de taille assez importante (un peu plus de 3200 données).



FIGURE 1.13 – *Lymnaea stagnalis* (Photo prise par Michel KUPFER (doris.ffesm.fr)).

Choix de l'inférence bayésienne comme méthode d'ajustement

Concernant la méthode utilisée pour ajuster les modèles concentration-réponse ou les modèles à seuil, nous avons choisi d'utiliser l'inférence bayésienne. L'intérêt de l'approche bayésienne est depuis longtemps reconnu mais les outils manquaient pour la mettre en pratique. Depuis les années 1990, avec le développement de l'informatique et

de sa puissance, l'inférence bayésienne peut être utilisée quelle que soit la complexité du cas envisagé (Parent et Bernier, 2007). D'ailleurs, elle a été utilisée en écotoxicologie dans Billoir et al (2008a,b); Fox (2010); Ducrot et al (2010); Billoir et al (2012); Biron et al (2012). L'inférence bayésienne a pour but, comme l'inférence fréquentiste, d'estimer des paramètres d'un modèle. Elle permet, contrairement à l'inférence fréquentiste, de fournir une distribution *a posteriori* des paramètres du modèle à partir d'une information *a priori* plus ou moins informative et des données. Elle est basée, comme son nom l'indique, sur le théorème de Bayes

$$P(\theta/Y) = \frac{P(\theta)P(Y/\theta)}{P(Y)}$$

avec

- Y : les données
- θ : le vecteur de paramètres du modèle
- $P(Y/\theta)$: la probabilité des données connaissant les paramètres, c'est à dire la vraisemblance (L).
- $P(\theta/Y)$: la probabilité des paramètres connaissant les données, c'est à dire la distribution *a posteriori* des paramètres
- $P(\theta)$: la distribution de probabilité des paramètres, non conditionnée aux données, c'est la distribution *a priori* des paramètres
- $P(Y)$: la distribution des données non conditionnée aux paramètres : elle ne dépend que de Y qui est constant, et est donc indépendante du paramètre θ que l'on souhaite estimer.

Nous pouvons donc écrire que la distribution *a posteriori* des paramètres est proportionnelle au produit de la distribution *a priori* des paramètres par la vraisemblance des données :

$$P(\theta/Y) \propto P(\theta)P(Y/\theta)$$

L'inférence bayésienne a été appliquée en utilisant JAGS et le package R nommé *rjags* (Plummer, 2010). Les méthodes Monte Carlo Markov Chain (MCMC) permettent de réaliser des tirages dans la distribution *a posteriori* des paramètres. Généralement la phase d'inférence est menée sur plusieurs chaînes MCMC indépendantes (souvent trois) et le diagnostic de convergence se fait en analysant visuellement la trace des chaînes MCMC mais aussi en calculant les statistiques de Gelman et Rubin (1992).

L'avantage majeur de l'utilisation de l'inférence bayésienne dans notre problématique est sa souplesse. Elle permet de choisir une partie stochastique non conventionnelle et ainsi de pouvoir comparer les parties stochastiques des modèles pour des jeux de données ne suivant pas une loi normale. La comparaison de modèle peut se faire en utilisant le DIC (Deviance Information Criterion) qui permet de calculer la déviance du modèle (D) en tenant compte de la complexité du modèles avec le calcul d'une déviance pénalisée (pD) (Spiegelhalter et al, 2002).

$$D(\theta) = -2\log P(Y/\theta) \tag{1.9}$$

et

$$pD = \bar{D} - D(\bar{\theta}) \tag{1.10}$$

$$DIC = \bar{D} + pD \quad (1.11)$$

Un autre avantage à ne pas négliger est l'apport de l'information *a priori*, qui est toujours présente dans le cas des tests de toxicité en écotoxicologie. En effet, l'écotoxicologue a une idée de la gamme de concentrations dans laquelle se situe la concentration critique d'effet qu'il veut estimer. De plus, avec l'inférence bayésienne, les lois *a posteriori* permettent de calculer l'incertitude sur n'importe quelle fonction des paramètres du modèle, ce qui n'est pas évident à réaliser en fréquentiste.

1.7 Organisation du manuscrit

Dans le chapitre 2, nous présenterons le travail réalisé sur les données de survie et dans le chapitre 3, nous aborderons le cas des données de reproduction dans le contexte du groupe de travail sur le test standardisé. Les deux chapitres s'organisent de la même manière : dans un premier temps, un travail sur l'analyse des données et dans un second temps un travail sur l'optimisation des plans d'expérience en vue d'obtenir la meilleure estimation possible de la concentration critique d'effet d'intérêt.

Chapitre 2

Les données de survie

Sommaire

2.1	Analyse des données de survie	25
2.1.1	Introduction	25
2.1.2	Les jeux de données de survie utilisés	26
2.1.3	Sélection des méthodes et modèles afin de déterminer les concentrations critiques d'effet	26
2.1.4	Comparaison des concentrations critiques d'effet	43
2.1.5	Discussion et conclusion sur l'analyse des données de survie	43
2.2	Optimisation des plans d'expérience pour des données de survie	48
2.2.1	Introduction	48
2.2.2	Etude théorique utilisant la D-optimalité localement	51
2.2.3	Etude globale basée sur des plans d'expérience réalistes	64
2.2.4	Discussion et conclusion sur l'optimisation des plans d'expérience dans le cadre des données de survie	81

2.1 Analyse des données de survie

2.1.1 Introduction

Dans cette première partie concernant la survie, nous nous intéressons aux méthodes et modèles utilisés pour analyser ce type de données afin d'estimer les concentrations critiques d'effet. Pour chaque concentration critique d'effet : NOEC, LC_x (puisque nous sommes dans le cas de la survie) et la NEC, plusieurs méthodes et modèles sont possibles. En ce qui concerne les NOEC et LC_x , nous avons choisi d'étudier et de comparer les méthodes et modèles décrits dans le guide OCDE (OECD, 2006). Dans le cas de l'estimation d'une NEC, il n'y a pas encore de référence et les modèles qui existent ne sont pas particulièrement simples (Billoir et al, 2008a). C'est pourquoi, nous proposons de développer un modèle à seuil simple et pragmatique. Dans une deuxième

partie, nous comparerons les valeurs obtenues pour les différentes concentrations critiques d'effet. Pour ces deux différentes étapes, nous utiliserons neuf jeux de données issus de la littérature.

2.1.2 Les jeux de données de survie utilisés

Les neuf jeux de données de survie concernent *Daphnia magna*, exposée à neuf substances toxiques :

- trois métaux : cuivre, zinc (Billoir et al, 2008b), cadmium (Santojanni et al, 1995),
- une substance oxydante : dichromate de potassium (Bedaux et Kooijman, 1994),
- un insecticide : le chlordane (Manar et al, 2009),
- un antiparasitaire : l'ivermectine (Lopes et al, 2009),
- trois antibiotiques : l'acide oxolinique, la streptomycine et la sulfadiazine (Wollenberger et al, 2000).

Les plans d'expérience ayant permis l'obtention des mesures de survie sont présentés dans le Tableau 2.1. Ces derniers sont réalisés sur une durée de 21 jours (test de toxicité chronique). Un exemple de données issues de ces tests figure dans le Tableau 2.2.

2.1.3 Sélection des méthodes et modèles afin de déterminer les concentrations critiques d'effet

Sélection des méthodes utilisées pour la détermination des NOEC et LOEC

La NOEC se calcule à un temps donné, généralement en fin de test de toxicité, soit 21 jours dans le cas d'un test chronique (OECD, 2006).

Trois méthodes basées sur les tests d'hypothèse sont recommandées par l'OCDE pour le traitement des données de type survie, que l'on choisit selon deux cas de figure :

1) La relation concentration-réponse attendue est monotone, les réponses sont strictement croissantes ou décroissantes, deux tests sont alors proposés :

- le test de Jonckheere-Terpstra basé sur les rangs avec la procédure « step-down » (Jonckheere, 1954) (Annexe 1)
- le test de Cochran-Armitage (Armitage and Colton, 2005) avec la procédure « step-down » (Armitage and Colton, 2005)

La procédure « step-down » est constituée de différentes étapes : une comparaison globale des données de survie est réalisée pour les différentes concentrations (dont le témoin). Si cette comparaison globale ne révèle pas de différence significative entre les traitements alors la NOEC prendra la valeur de la plus forte concentration testée. Si au contraire, le test d'hypothèse conclut à une différence significative entre les traitements, la concentration la plus forte est enlevée et un test global est refait sur l'ensemble des données restantes. L'opération est renouvelée jusqu'à ce que le test global de comparaison ne montre plus de différence significative et que la NOEC prenne la valeur de la plus forte concentration parmi les restantes.

Contaminant	Nombre de concentrations testées (k)	Gamme de concentrations (c_i) [nombre initial de <i>D.magna</i>] (N_i) avec $i=1, k$	Nombre de mesures de 0 à 21 jours
Cuivre ($\mu\text{g L}^{-1}$)	5	0, 2.5, 5, 10, 20 [60, 60, 60, 60, 60]	21
Zinc (mg L^{-1})	5	0, 0.074, 0.22, 0.66, 2 [57, 56, 58, 60, 60]	21
Cadmium ($\mu\text{g L}^{-1}$)	4	0, 2.5, 5, 7.5 [40, 40, 40, 40]	8
Dichromate (mg L^{-1}) de potassium	6	0, 0.1, 0.18, 0.32, 0.56, 1 [50, 50, 50, 50, 50, 50]	10
Chlordane ($\mu\text{g L}^{-1}$)	6	0, 0.18, 0.73, 1.82, 2.9, 7 [10, 10, 10, 10, 10, 10]	21
Ivermectine (pg L^{-1})	5	0, 0.3, 1, 10, 100 [10, 10, 10, 10, 10]	21
Acide oxolinique (mg L^{-1})	7	0, 0.094, 0.19, 0.38, 0.75, 1.5, 3 [10, 10, 10, 10, 10, 10, 10]	10
Streptomycine (mg L^{-1})	6	0, 4, 8, 16, 32, 64 [10, 10, 10, 10, 10, 10]	10
Sulfadiazine (mg L^{-1})	7	0, 3.1, 6.2, 12.5, 25, 50, 100 [10, 10, 10, 10, 10, 10, 10]	10

Tableau 2.1 – Plans d'expérience des différents jeux de données.

Jour	Témoin	Concentrations testées					
		3.1	6.2	12.5	25	50	100
0	10	10	10	10	10	10	10
3	10	10	10	10	10	10	10
5	10	10	10	9	10	10	3
7	10	10	10	9	10	10	1
10	10	10	10	9	9	9	0
12	10	10	10	9	8	9	0
14	10	10	10	9	8	9	0
17	9	10	10	9	8	8	0
19	9	10	10	9	8	8	0
21	9	10	10	9	8	7	0

Tableau 2.2 – Données de survie de *Daphnia magna* exposée à la sulfadiazine (mg L^{-1}).

2) Soit la relation concentration-réponse attendue est quelconque. C'est parfois le cas lorsqu'un composé a un effet hormesis (effet positif pour des petites concentrations d'exposition) avant d'avoir un effet négatif. Il est alors recommandé d'utiliser le test exact de Fisher avec correction de Bonferroni-Holm (Holm, 1979) (Annexe 3). Cette correction correspond à une variante de la méthode classique de Bonferroni.

Tous les tests ont été réalisés avec le logiciel R (R Development Core Team, 2010). Les tests de Cochran-Armitage et de Fisher ont été effectués à l'aide de fonctions R respectivement *prop.trend.test* et *fisher.test*. Des fonctions "maisons" ont également été développées pour le test de Jonckheere-Terpstra ainsi que la procédure "step-down" et la correction de Bonferroni-Holm.

Les valeurs des NOEC et LOEC ont été déterminées à l'aide des trois méthodes (Jonckheere-Terpstra et Cochran-Armitage avec procédure « step-down » ainsi que le test exact de Fisher avec correction de Bonferroni-Holm) pour chaque jeu de données après 21 jours d'exposition. Les résultats sont présentés dans le Tableau 2.3.

Contaminant	Méthodes		
	C-A et SDP <i>NOEC</i> ; <i>LOEC</i>	J-T et SDP <i>NOEC</i> ; <i>LOEC</i>	Fisher avec BH <i>NOEC</i> ; <i>LOEC</i>
Cuivre ($\mu\text{g L}^{-1}$)	10 ; 20	20 ; ND	10 ; 20
Zinc (mg L^{-1})	0.074 ; 0.22	0.66 ; 2	0.074 ; 0.22
Cadmium ($\mu\text{g L}^{-1}$)	0 ; 2.5	7.5 ; ND	0 ; 2.5
Dichromate de Potassium (mg L^{-1})	0.18 ; 0.32	1 ; ND	0.32 ; 0.56
Chlordane ($\mu\text{g L}^{-1}$)	1.82 ; 2.9	7 ; ND	1.82 ; 2.9
Ivermectine (pg L^{-1})	1 ; 10	10 ; 100	1 ; 10
Acide Oxolinique (mg L^{-1})	0.38 ; 0.75	3 ; ND	0.75 ; 1.5
Streptomycine (mg L^{-1})	16 ; 32	64 ; ND	32 ; 64
Sulfadiazine (mg L^{-1})	25 ; 50	100 ; ND	50 ; 100

Tableau 2.3 – Valeurs des *NOEC/LOEC* déterminées pour chaque contaminant au 21^e jour d'exposition avec les trois méthodes basées sur les tests d'hypothèse : le test de Cochran-Armitage (C-A) avec une procédure "step-down" (SDP), le test de Jonckheere-Terpstra (J-T) avec SDP et le test exact de Fisher avec correction de Bonferroni-Holm (BH). ND : valeur Non Déterminée.

On peut noter que les valeurs des NOEC sont dépendantes des méthodes utilisées (Tableau 2.3). Le test de Jonckheere-Terpstra avec procédure "step-down" est assez rarement capable de déceler une différence significative entre le témoin et les concentrations testées et donne donc des valeurs de NOEC égales à la plus forte concentration

testée à l'exception de deux cas (le zinc et l'ivermectine). La méthode basée sur le test de Cochran-Armitage semble être plus efficace que celle basée sur le test de Fisher pour détecter des différences significatives, ce qui s'explique par l'hypothèse faite, pour cette méthode, d'une relation monotone entre la concentration et la réponse. Ainsi, la méthode basée sur le test de Cochran-Armitage est la plus protectrice pour l'environnement, puisqu'elle détermine des valeurs de NOEC plus faibles que les autres méthodes. C'est pourquoi, nous avons décidé d'utiliser les valeurs des NOEC obtenues par cette méthode pour comparer les différentes concentrations critiques d'effet.

Sélection des modèles concentration-réponse pour estimer des LC_x

Afin d'estimer des LC_x , nous utilisons les trois modèles les plus couramment rencontrés : log-logistique, log-probit et Weibull. Dans le cas de données de survie, un modèle à deux paramètres est classiquement utilisé pour la survie en supposant $d = 1$ et $c = 0$ dans les équations 1.1, 1.4 et 1.5 décrites dans la partie 1.4.2. Le diagramme acyclique dirigé (DAG) est représenté sur la Figure 2.1 et permet de visualiser les liens (relation de type déterministe ou stochastique) entre les différents noeuds (variables et paramètres) d'un modèle concentration-réponse en inférence bayésienne. Ces différents liens sont également résumés et précisés pour chacun des trois modèles dans le Tableau 2.4.

La partie stochastique pour ce type de données binaires est binomiale, ce qui est une différence majeure par rapport aux modèles concentration-réponse utilisés sur des données continues qui ont comme partie stochastique une loi normale. Les paramètres de la loi binomiale sont : $S(\text{conc}_i)$ (la probabilité de succès) et le nombre de daphnies initialement introduites dans l'essai (N_i , Tableau 2.1).

Modèles	Noeud	Type	Définition
Commun aux 3 modèles	y_i	Stochastique	$y_i \sim \text{binomiale}(S(\text{conc}_i), N_i)$
log-logistique	$S(\text{conc}_i)$	Logique	$S(\text{conc}_i) = 1 - \frac{1}{1 + (\frac{LC_{50}}{\text{conc}_i})^b}$
log-probit	$S(\text{conc}_i)$	Logique	$S(\text{conc}_i) = 1 - F(b \times \log_{10}(\frac{\text{conc}_i}{LC_{50}}))$
Weibull	$S(\text{conc}_i)$	Logique	$S(\text{conc}_i) = \exp(-(\frac{\text{conc}_i}{LC_{50}})^b \times \ln(2))$

Tableau 2.4 – Définition des liens entre les différents noeuds des modèles concentration-réponse.

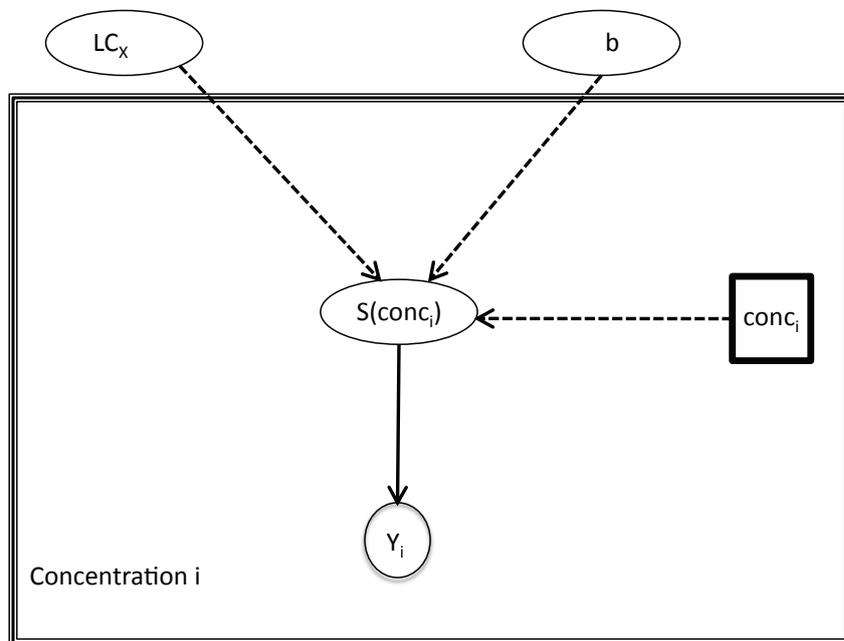


FIGURE 2.1 – Diagramme acyclique dirigé des modèles concentration-réponse. Les liens logiques entre chaque nœud sont représentés par des flèches en pointillés, les liens stochastiques par des flèches pleines. La covariable est représentée par un rectangle et les variables sont représentées par des ellipses.

Pour chaque jeu de données, les valeurs des LC_x ($x=5, 10, 20, 50$) ont été estimées et les déviations moyennes des trois modèles ont été comparées afin de choisir le modèle le plus adapté.

Pour les trois modèles concentration-réponse étudiés, une distribution *a priori* uniforme entre la plus faible et la plus forte concentration testée a été choisie pour le paramètre LC_{50} . En effet, il est présupposé que la LC_{50} se situe dans la gamme testée. En ce qui concerne le paramètre b des modèles log-logistique et log-probit, une distribution non informative uniforme entre 0 et 100 a été choisie. Pour le paramètre b du modèle de Weibull, c'est également une distribution non informative uniforme qui a été choisie mais entre 0 et 10. Même si le nom du paramètre b est commun aux trois modèles, il n'est pas équivalent pour les trois modèles.

Les neuf jeux de données à 21 jours ont été ajustés avec les modèles concentration-réponse log-logistique, log-probit et de Weibull en utilisant l'inférence bayésienne (Figure 2.2). On peut remarquer que les qualités d'ajustement sont assez proches et qu'il n'a pas été possible d'ajuster le modèle log-probit aux jeux de données chlordane et ivermectine ((e) et (f) dans la Figure 2.2). Hormis pour ces deux jeux de données, la représentation des ajustements ne permet pas de faire un choix de modèle. C'est pourquoi, les déviations moyennes des trois paramètres ont été calculées.

Toutes les valeurs des paramètres, ainsi que les déviations moyennes des trois modèles concentration-réponse étudiés, sont estimées pour chaque contaminant après 21 jours d'exposition et sont données dans le Tableau 2.5. Les trois modèles concentration-réponse estiment des valeurs de LC_{50} assez proches les unes des autres. Les déviations moyennes d'un modèle à l'autre sont également très proches. Cependant, le modèle log-probit ne permet pas d'estimer une LC_{50} pour le chlordane et l'ivermectine, ce qui fait que ce modèle ne sera pas retenu pour la suite. De plus, le modèle de Weibull donne dans tous les cas la déviance moyenne la plus faible (à une unité près) et donc le meilleur ajustement aux données. Nous choisissons ce modèle pour la suite de notre travail.

Pour le zinc, le dichromate de potassium, l'ivermectine, l'acide oxolinique et la streptomycine, le modèle de Weibull s'ajuste globalement assez bien aux données à 21 jours (Figure 2.2 (b), (d), (f), (g), (h)). Alors que pour le cuivre, le cadmium, le chlordane et la sulfadiazine la tendance des effets observés n'est pas tout à fait décrite par le modèle, des données se trouvent en dehors de l'intervalle de prédiction (Figure 2.2 (a), (c), (e), (i)). Ceci peut être dû à la mortalité naturelle observée dans ces jeux de données et qui n'est pas prise en compte par le modèle à deux paramètres.

En plus des valeurs de LC_{50} , nous avons estimé les valeurs d'autres LC_x ($x=5, 10, 20$) et leurs intervalles de crédibilité. Ces valeurs sont représentées dans le Tableau 2.6. Comme attendu, nous observons un effet important du choix du x sur les valeurs des LC_x . Toutefois les valeurs estimées par les trois modèles concentration-réponse sont assez proches. Nous pouvions nous attendre à avoir des intervalles de crédibilité plus larges pour les valeurs de x plus petites, ce qui n'est pas le cas.

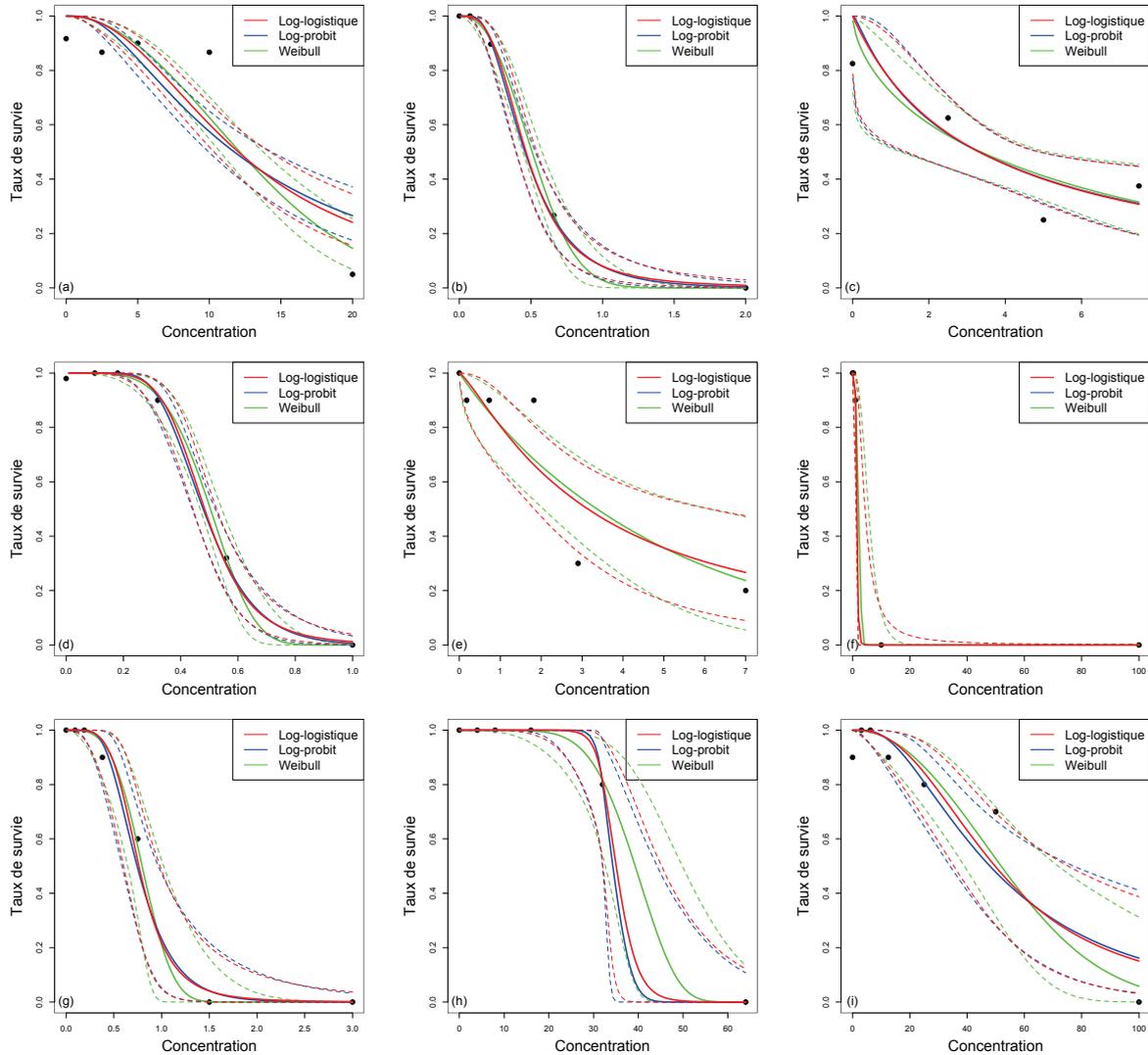


FIGURE 2.2 – Ajustements des trois modèles concentration-réponse (log-logistique, log-probit et Weibull) aux neuf jeux de données (taux de survie en fonction de la concentration) à 21 jours ainsi que les intervalles de crédibilité à 95%. Les points noirs représentent les données. (a) cuivre ($\mu\text{g/L}$), (b) zinc (mg/L), (c) cadmium (mg/L), (d) dichromate de potassium (mg/L), (e) chlordane ($\mu\text{g/L}$), (f) ivermectine (pg/L), (g) acide oxolinique (mg/L), (h) streptomycine ($\mu\text{g/L}$), (i) sulfadiazine ($\mu\text{g/L}$)

Contaminant	Paramètre	modèle log-logistique	modèle log-probit	modèle Weibull
Cuivre	LC_{50} ($\mu\text{g/L}$)	12.0 [10.3 - 14.2]	11.7 [9.94 - 14.4]	12.1 [10.7 - 13.7]
	b	5.18 [3.92 - 6.60]	2.70 [2.08 - 3.36]	2.04 [1.58 - 2.54]
	DM	65.8	71.6	52.5
Zinc	LC_{50} (mg/L)	0.465 [0.398 - 0.541]	0.458 [0.395 - 0.532]	0.508 [0.444 - 0.578]
	b	7.33 [5.63 - 9.48]	4.15 [3.24 - 5.25]	2.40 [1.76 - 3.26]
	DM	11.7	10.5	10.7
Cadmium	LC_{50} ($\mu\text{g/L}$)	3.32 [1.40 - 4.92]	3.32 [1.37 - 4.93]	3.36 [1.28 - 5.12]
	b	2.28 [0.561 - 4.27]	1.39 [0.345 - 2.60]	0.627 [0.148 - 1.19]
	DM	19.2	19.3	19.7
Dichromate de potassium	LC_{50} (mg/L)	0.483 [0.443 - 0.526]	0.479 [0.440 - 0.522]	0.504 [0.467 - 0.543]
	b	13.8 [10.4 - 18.4]	7.77 [5.97 - 10.0]	4.56 [3.27 - 6.40]
	DM	11.6	10.5	10.4
Chlordane	LC_{50} ($\mu\text{g/L}$)	3.13 [1.80 - 6.11]	pas de convergence	3.35 [2.02 - 6.06]
	b	2.88 [1.31 - 5.12]	pas de convergence	0.985 [0.459 - 1.64]
	DM	19.3	-	18.7
Ivermectine	LC_{50} (pg/L)	1.44 [1.04 - 4.28]	pas de convergence	2.09 [1.14 - 5.45]
	b	16.00 [4.27 - 89.72]	pas de convergence	3.36 [1.27 - 9.40]
	DM	3.41	-	3.38
Acide oxolinique	LC_{50} (mg/L)	0.761 [0.600 - 0.976]	0.745 [0.586 - 0.968]	0.794 [0.643 - 1.02]
	b	10.7 [5.85 - 20.6]	5.71 [3.29 - 9.48]	3.59 [1.91 - 7.36]
	DM	8.53	8.41	7.18
Streptomycine	LC_{50} (mg/L)	35.0 [32.1 - 45.1]	34.2 [32.0 - 44.1]	39.2 [32.7 - 49.8]
	b	36.3 [11.0 - 94.9]	26.6 [6.62 - 93.9]	6.29 [3.00 - 9.75]
	DM	3.72	3.58	3.71
Sulfadiazine	LC_{50} (mg/L)	49.8 [35.7 - 75.9]	48.0 [34.2 - 76.5]	51.8 [39.0 - 71.8]
	b	5.66 [3.08 - 9.48]	3.08 [1.73 - 4.87]	2.14 [1.21 - 3.51]
	DM	15.2	15.1	13.0

Tableau 2.5 – Estimation des paramètres LC_{50} et b des trois modèles concentration-réponse pour chaque contaminant après 21 jours d'exposition. Les valeurs correspondent aux valeurs médianes des distributions *a posteriori* des paramètres et aux intervalles de crédibilité à 95%. La valeur de la déviance moyenne (DM) pour chaque modèle est donnée afin d'évaluer et de comparer la qualité d'ajustement de chaque modèle pour chaque contaminant. Le modèle avec la déviance minimale (en bleu) donne le meilleur ajustement.

Contaminant	Paramètre	Modèle log-logistique	Modèle log-probit	Modèle Weibull
Cuivre	LC_5 ($\mu\text{g/L}$)	3.25 [2.20 - 4.33]	2.89 [2.02 - 3.75]	3.39 [2.31 - 4.49]
	LC_{10} ($\mu\text{g/L}$)	4.53 [3.34 - 5.69]	3.94 [2.97 - 4.88]	4.83 [3.60 - 6.01]
	LC_{20} ($\mu\text{g/L}$)	6.49 [5.25 - 7.74]	5.74 [4.67 - 6.83]	6.96 [5.68 - 8.20]
Zinc	LC_5 (mg/L)	0.184 [0.133 - 0.243]	0.184 [0.136 - 0.234]	0.172 [0.110 - 0.241]
	LC_{10} (mg/L)	0.233 [0.178 - 0.293]	0.225 [0.174 - 0.277]	0.231 [0.165 - 0.303]
	LC_{20} (mg/L)	0.301 [0.242 - 0.364]	0.287 [0.234 - 0.341]	0.317 [0.249 - 0.386]
Cadmium	LC_5 ($\mu\text{g/L}$)	0.170 [1.46e-05 - 0.778]	0.216 [2.75e-05 - 0.908]	0.0520 [6.88e-08 - 0.466]
	LC_{10} ($\mu\text{g/L}$)	0.360 [0.000291 - 1.18]	0.394 [0.000314 - 1.26]	0.1645 [7.19e-06 - 0.856]
	LC_{20} ($\mu\text{g/L}$)	0.812 [0.00772 - 1.86]	0.816 [0.00560 - 1.90]	0.543 [0.000954 - 1.63]
Dichromate de potassium	LC_5 (mg/L)	0.296 [0.246 - 0.345]	0.294 [0.248 - 0.336]	0.284 [0.223 - 0.344]
	LC_{10} (mg/L)	0.335 [0.288 - 0.381]	0.327 [0.284 - 0.367]	0.333 [0.276 - 0.386]
	LC_{20} (mg/L)	0.383 [0.341 - 0.427]	0.373 [0.332 - 0.411]	0.393 [0.344 - 0.439]
Chlordane	LC_5 ($\mu\text{g/L}$)	0.296 [0.0223 - 0.819]	-	0.236 [0.0137 - 0.733]
	LC_{10} ($\mu\text{g/L}$)	0.537 [0.0787 - 1.19]	-	0.490 [0.0632 - 1.17]
	LC_{20} ($\mu\text{g/L}$)	1.03 [0.303 - 1.86]	-	1.05 [0.299 - 1.96]
Ivermectine	LC_5 (pg/L)	0.970 [0.480 - 1.661]	-	0.934 [0.351 - 2.05]
	LC_{10} (pg/L)	1.04 [0.672 - 2.02]	-	1.14 [0.572 - 2.59]
	LC_{20} (pg/L)	1.16 [0.898 - 2.55]	-	1.43 [0.873 - 3.38]
Acide oxolinique	LC_5 (mg/L)	0.406 [0.224 - 0.574]	0.385 [0.223 - 0.523]	0.390 [0.193 - 0.569]
	LC_{10} (mg/L)	0.477 [0.295 - 0.637]	0.445 [0.285 - 0.583]	0.476 [0.279 - 0.642]
	LC_{20} (mg/L)	0.569 [0.393 - 0.723]	0.531 [0.376 - 0.676]	0.585 [0.401 - 0.747]
Streptomycine	LC_5 (mg/L)	29.5 [20.7 - 32.9]	30.3 [21.4 - 32.3]	25.8 [16.5 - 33.2]
	LC_{10} (mg/L)	30.8 [23.8 - 34.7]	31.1 [24.0 - 33.7]	28.8 [20.7 - 36.6]
	LC_{20} (mg/L)	32.2 [27.5 - 37.6]	32.0 [27.3 - 36.2]	32.4 [25.7 - 40.8]
Sulfadiazine	LC_5 (mg/L)	15.0 [5.94 - 25.6]	14.1 [6.04 - 22.8]	15.5 [6.04 - 26.8]
	LC_{10} (mg/L)	20.3 [10.1 - 31.7]	18.5 [9.35 - 27.9]	21.6 [10.7 - 33.5]
	LC_{20} (mg/L)	28.3 [17.3 - 40.8]	25.6 [15.76 - 36.5]	30.7 [18.8 - 43.1]

Tableau 2.6 – Estimations des paramètres LC_5 , LC_{10} et LC_{20} et intervalles de crédibilité à 95% obtenus avec les trois modèles concentration-réponse pour chaque contaminant après 21 jours d'exposition.

Sélection d'un modèle pour estimer une NEC

Comme pour les LC_x , plusieurs modèles à seuil sont possibles pour estimer la NEC. Le paramètre d'intérêt (la NEC) a été introduit en tant que concentration seuil dans un modèle simple. L'ajout de ce paramètre ne présuppose pas l'existence d'un seuil biologique (Fox, 2010), cependant il permet d'estimer une concentration limite en dessous de laquelle il n'y a pas d'effet. Une caractéristique particulière de ce modèle à seuil, appelé par la suite modèle NEC, est qu'il dépend à la fois de la concentration ($conc_i$) et du temps d'exposition (t). Ce modèle NEC peut donc être ajusté sur un plus grand nombre de données (toutes les mesures à tous les temps) contrairement à un modèle concentration-réponse classique. Ainsi, un plus grand nombre de paramètres peut raisonnablement être estimé. La probabilité de survie $S(conc_i, t)$ décrit à la fois la concentration $conc_i$ et le temps d'exposition (t).

En effet, prenons Tm la variable aléatoire continue qui représente le moment où survient la mort. Notons f la fonction de densité de Tm et F sa fonction de répartition :

$$F(t) = P(Tm \leq t) = \int_0^t f(\tau) d\tau \quad (2.1)$$

Soit S la fonction de survie qui donne la probabilité pour un individu de survivre jusqu'à un certain temps t :

$$S(t) = 1 - F(t) = P(Tm > t) = \int_t^\infty f(\tau) d\tau \quad (2.2)$$

Si on dérive cette fonction de survie par rapport au temps on obtient :

$$\frac{dS}{dt} = \frac{d(1 - F)}{dt} = -\frac{dF}{dt} \quad (2.3)$$

Or d'après la première équation, la dérivée de F est f donc :

$$\frac{dS}{dt} = -f(t) \quad (2.4)$$

Le taux de mortalité instantanée $h(t)$ caractérise le risque instantané de mourir à un certain temps t étant donné que l'individu est encore vivant au temps t :

$$h(t) = \frac{-\frac{dS}{dt}}{S(t)} \quad (2.5)$$

D'où :

$$\int \frac{dS}{S} = \ln S(t) = - \int_0^t h(\tau) d\tau \quad (2.6)$$

D'où :

$$S(t) = e^{-\int_0^t h(\tau) d\tau} \quad (2.7)$$

Si h ne dépend pas du temps :

$$\Leftrightarrow S(t) = e^{-h \times t} \quad (2.8)$$

Ainsi, si l'on introduit la dépendance du taux instantané à la concentration :

$$S(\text{conc}_i, t) = e^{-h(\text{conc}_i) \times t} \quad (2.9)$$

Dans le modèle choisi, on écrit :

$$h(\text{conc}_i) = m_0 + \sigma(\text{conc}_i) \quad (2.10)$$

où m_0 (en jour⁻¹) correspond à la mortalité naturelle instantanée et $\sigma(\text{conc}_i)$ est la fonction de stress, positive à partir de la NEC (en unité de concentration). Cette fonction augmente linéairement avec une pente k_S (en 1/ unité de concentration jour⁻¹) quand la concentration augmente (Figure 2.3) :

$$\sigma(\text{conc}_i) = \min(k_S \times \max(0, \text{conc}_i - NEC), m_{max} - m_0) \quad (2.11)$$

Le paramètre m_{max} (en jour⁻¹) correspond à la saturation potentielle de la mortalité instantanée $h(\text{conc}_i)$ pour des concentrations élevées. Le nombre de survivants à la concentration conc_i et au temps t suit une distribution binomiale conditionnelle (BC) :

$$BC\left(\frac{1 - (S(\text{conc}_i, t - 1) - S(\text{conc}_i, t))}{S(\text{conc}_i, t - 1)}, N_{\text{conc}_i, t}\right) \quad (2.12)$$

Le premier paramètre est égal à 1-probabilité de mourir entre $t - 1$ et t , sachant qu'on est vivant à $t - 1$ et le second paramètre est égal au nombre d'individus vivants à la concentration conc_i et au temps $t - 1$. Ce modèle stochastique est équivalent au modèle multinomial utilisé par (Jager et al, 2011).

Le modèle NEC que nous utilisons a quatre paramètres : NEC , k_S , m_{max} et m_0 . Tous les paramètres ne sont pas nécessaires pour décrire tous les jeux de données. Dans certains cas, le modèle complet avec quatre paramètres peut être simplifié par un des trois sous-modèles possibles présentés sur la Figure 2.4 : un modèle à deux paramètres (NEC , k_S) et deux modèles à trois paramètres (NEC , k_S , m_0) et (NEC , k_S , m_{max}).

Le paramètre m_0 est estimé seulement si on observe une mortalité dans le groupe témoin ou si une réponse non monotone est observée, sinon m_0 est fixé à zéro. Le paramètre m_{max} est estimé seulement quand le nombre de survivants à la plus forte concentration est supérieur ou égal au nombre de survivants à une concentration inférieure, pour au moins une durée d'exposition.

Les valeurs de NEC sont estimées pour chaque jeu de données en incluant toutes les mesures intermédiaires jusqu'à 21 jours, contrairement aux autres concentrations critiques d'effets pour lesquelles seules les données à 21 jours sont utilisées. Afin de choisir le modèle NEC le plus approprié pour chaque jeu de données, les paramètres nécessaires ont été retenus, dans un premier temps, en fonction des critères énoncés précédemment ; dans un second temps, la précision d'estimation des paramètres m_0 et m_{max} a été quantifiée et les paramètres ont été retenus uniquement s'il n'y avait

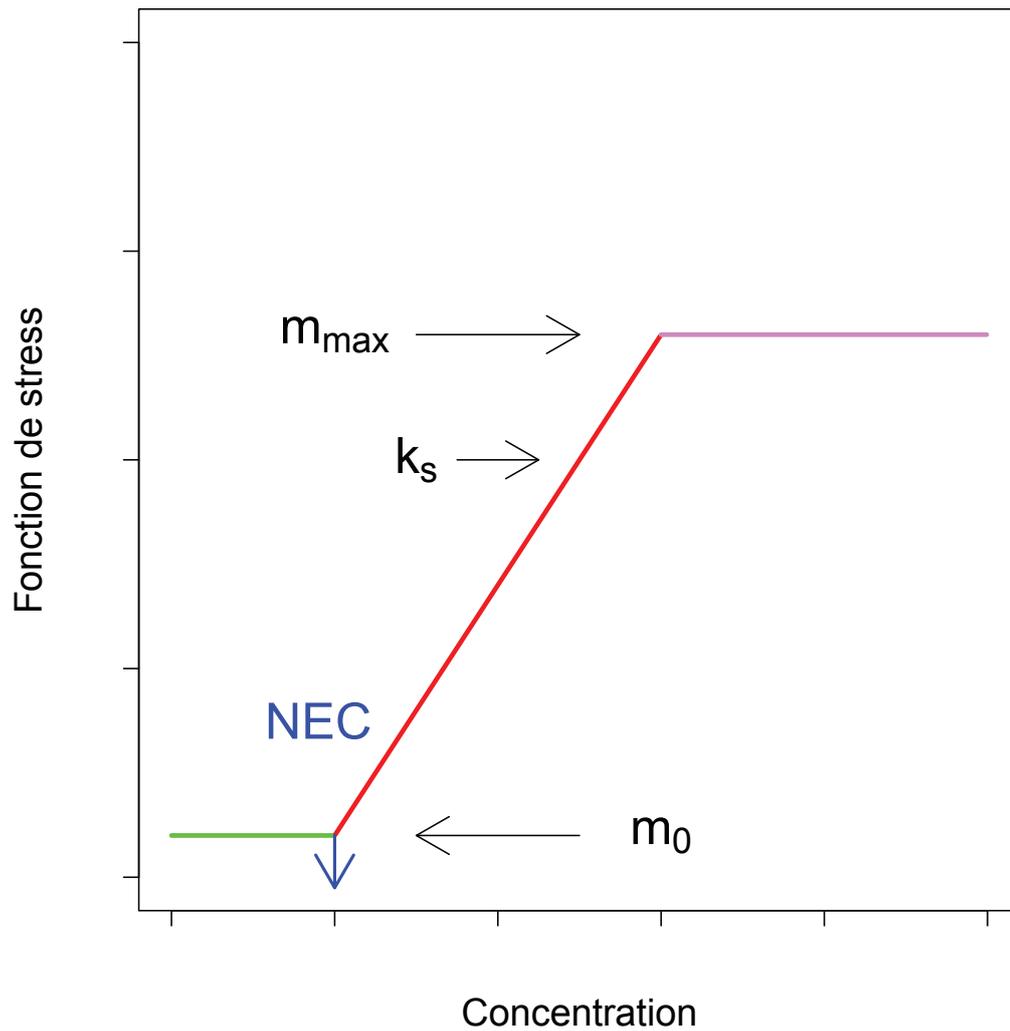


FIGURE 2.3 – Schéma représentant l'évolution de la fonction de stress par rapport à la concentration.

pas plus de trois \log_{10} entre les deux bornes de l'intervalle de crédibilité à 95%. Les estimations obtenues avec le modèle le plus approprié seront comparées avec les autres concentrations critiques d'effet.

Le diagramme acyclique dirigé a été représenté dans la Figure 2.5. Les différents liens entre les noeuds du modèle sont résumés dans le Tableau 2.7.

Noeud	Type	Definition
$S(\text{conc}_i, t)$	Logique	$S(\text{conc}_i, t) = e^{-m_0 + \min(k_s \times \max(0, \text{conc}_i - \text{NEC}), m_{\max} - m_0) \times t}$
$y_{i,t}$	Stochastique	$y_{i,t} \sim \text{binomial}(1 - (S(\text{conc}_i, t - 1) - S(\text{conc}_i, t)), N_{i,t-1})$

Tableau 2.7 – Définition des liens entre les différents noeuds du modèle NEC.

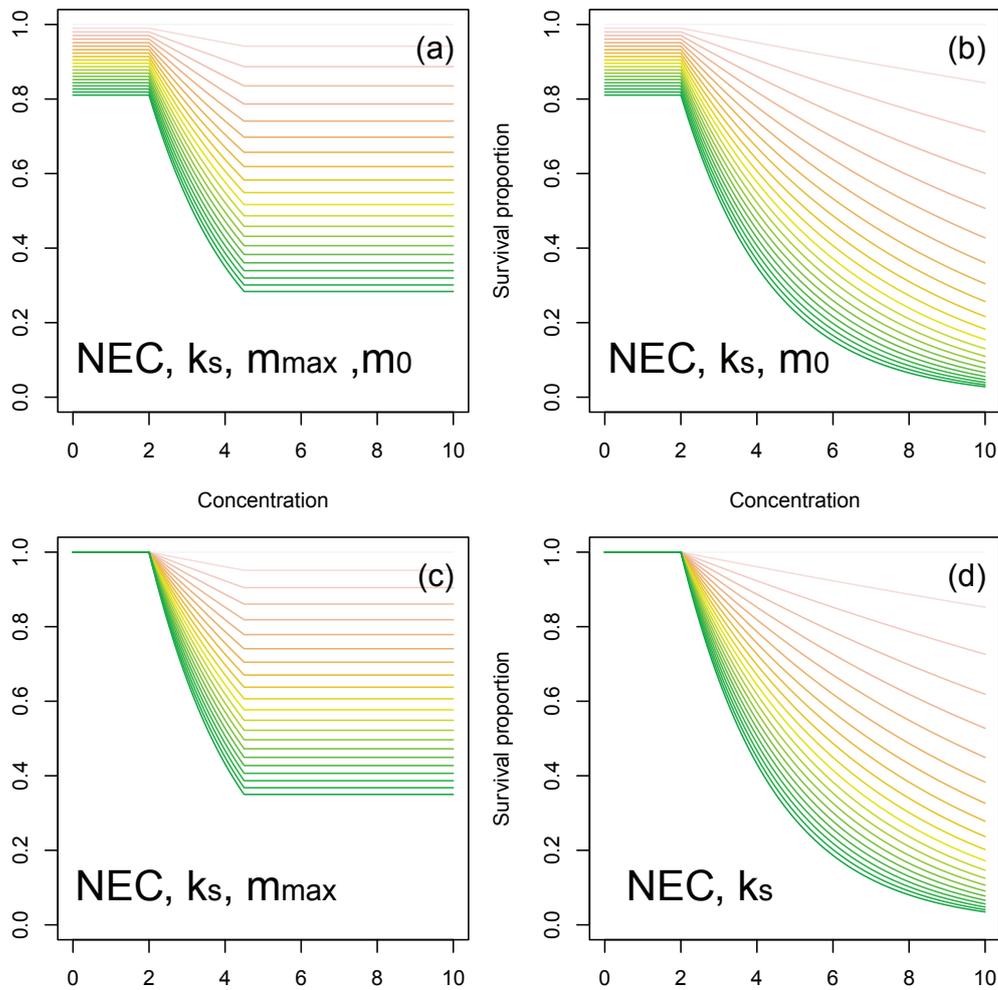


FIGURE 2.4 – Les modèles NEC avec quatre paramètres (a), trois paramètres (b, c), et deux paramètres (d) sont représentés par le taux de survie en fonction de la concentration en contaminant. Les lignes de couleurs de plus en plus foncées (du beige vers le vert foncé) représentent 22 durées d'exposition de 0 à 21 jours. Des valeurs de paramètres ont été arbitrairement choisies : $NEC = 2$ mg/L ; $k_S = 0.02$ L/mg/jour ; $m_{max} = 0.05$ /jour et $m_0 = 0.01$ /jour.

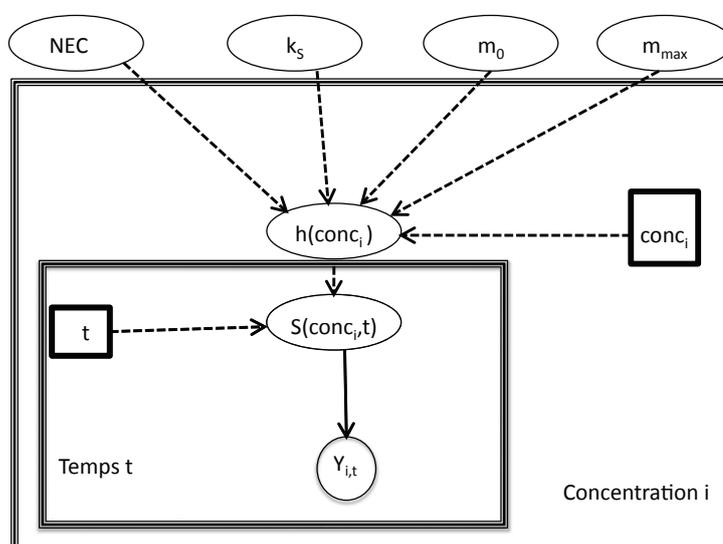


FIGURE 2.5 – Diagramme acyclique dirigé du modèle NEC. Les liens logiques entre chaque nœuds sont représentés par des flèches en pointillés, les liens stochastiques par des flèches pleines. La covariable est représentée par un rectangle et les variables sont représentées par des ellipses.

En ce qui concerne le paramètre NEC , la même distribution *a priori* informative que celle utilisée pour la LC_{50} a été utilisée : une distribution uniforme entre 0 et la plus forte concentration testée.

D'après le guide de l'OCDE (OCDE, 1998), une expérimentation menée sur *Daphnia magna* ne doit pas aboutir à plus de 20% de mortalité dans le témoin après 21 jours d'exposition. Ainsi, la distribution *a priori* du $\log_{10}(m_0)$ a été bornée au maximum par le \log_{10} de la mortalité maximale acceptée dans le témoin : si $S(\text{conc}_i, 21) = 1 - 0.2 = e^{-\text{max}_{m_0} \times 21}$ alors $\text{max}_{m_0} = -\ln(0.8)/21 \simeq 0.01 \text{ jour}^{-1}$. Une distribution *a priori* uniforme de $\log_{10}(m_0)$ entre $[-5; -2]$ a donc été choisie pour tous les jeux de données pour lesquels le paramètre m_0 était nécessaire.

Concernant le paramètre k_S , la borne supérieure de la distribution *a priori* du $\log_{10}(k_S)$ a été en premier lieu considérée. Considérons la plus petite différence entre deux concentrations successives c_j et c_{j+1} . Nous supposons k_S maximal dans le cas expérimental extrême pour lequel la concentration c_j ne mène à aucun effet tandis que c_{j+1} mène à 99% de mortalité. La valeur de 99% a été choisie à la place de 100% pour éviter des problèmes mathématiques dus à une non définition des fonctions. Cette situation peut être considérée comme extrême puisque habituellement les concentrations sont choisies afin d'observer des effets entre 0% et 100%. Ainsi, si $S(c_{j+1} - c_j, 21) = 0.01 = e^{-\text{max}_{k_S} \times (c_{j+1} - c_j) \times 21}$ alors $\text{max}_{k_S} = \frac{-\ln(0.01)}{(c_{j+1} - c_j) \times 21}$. Pour la borne inférieure de la distribution *a priori* uniforme de $\log_{10}(k_S)$, il a été choisi arbitrairement : $\log_{10}(\text{min}_{k_S}) = \log_{10}(\text{max}_{k_S}) - 2$.

Pour chaque jeu de données et chaque modèle, après une période initiale de "chauffe" d'environ 20 000 itérations, entre 20 000 et 50 000 itérations ont été utilisées pour faire converger l'algorithme bayésien. Généralement, la convergence a été vérifiée à la fois en observant les traces des chaînes MCMC et en analysant les statistiques de Gelman et Rubin (Gelman et Rubin (1992)).

Les ajustements des modèles NEC les plus appropriés aux données au cours du temps sont présentés sur la Figure 2.6, où le parallèle peut être fait entre la tendance observée dans les données (graphes (a1) à (i1)) et celle décrite par le modèle NEC choisi (graphes ((a2) à (i2))). Nous pouvons remarquer que les modèles semblent bien décrire les principales tendances observées dans les données. La mortalité naturelle observée dans les jeux de données du cuivre, du cadmium, du chlordan et de la sulfadiazine ((a1), (c1), (e1) et (i1)), est assez bien décrite par le modèle NEC sélectionné pour chacun des jeux de données ((a2), (c2), (e2) et (i2)). Le Tableau 2.8 résume les informations principales concernant les paramètres des modèles NEC sélectionnés, avec les paramètres nécessaires pour chaque jeu de données, leurs estimations médianes et les intervalles de crédibilité à 95% autour de ces valeurs médianes. La transition entre les concentrations où nous n'observons aucun effet et celles où nous observons un effet est assez bien décrite par les modèles NEC (par exemple (d2) et (h2) dans la Figure 2.6). De plus, les intervalles de crédibilité autour des valeurs estimées des paramètres sont assez précis pour la plupart des jeux de données, ce qui reflète une bonne précision des estimations.

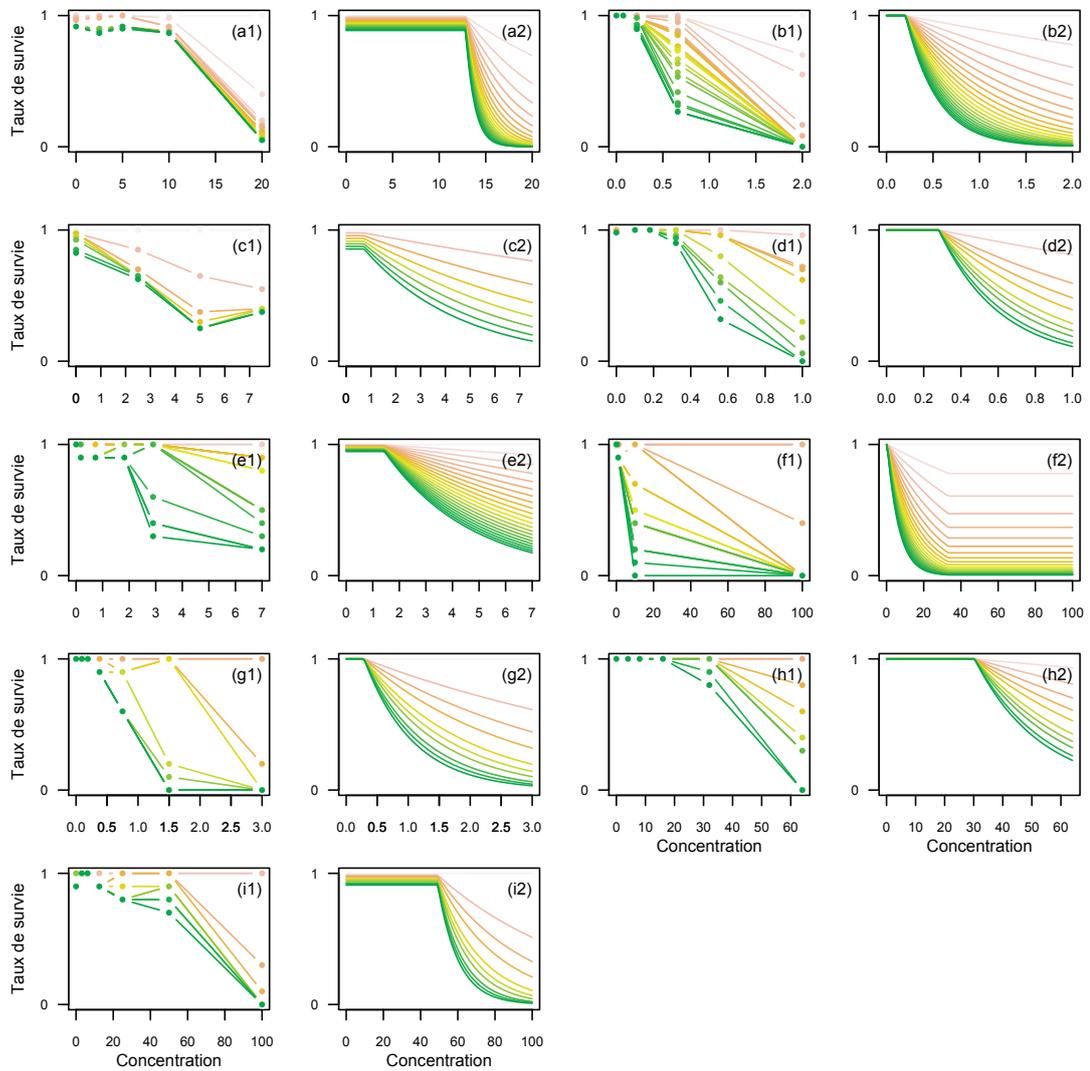


FIGURE 2.6 – Les couples de graphiques avec la même lettre correspondent au même contaminant. Chaque graphique représente le taux de survie en fonction de la concentration en contaminant. Les points sur les graphiques de gauche (indiqués 1) représentent les données avec des couleurs allant du plus clair au plus foncé correspondant à des temps d'exposition allant de 0 à 21 jours. Chaque ligne reliant les points représente une durée d'exposition. Les graphiques de droite (indiqués 2) décrivent les ajustements obtenus avec les modèles NEC adaptés pour chaque contaminant pour chaque durée d'exposition (lignes de couleurs). (a) cuivre ($\mu\text{g/L}$), (b) zinc (mg/L), (c) cadmium (mg/L), (d) dichromate de potassium (mg/L), (e) chlordane ($\mu\text{g/L}$), (f) ivermectine (pg/L), (g) acide oxolinique (mg/L), (h) streptomycine ($\mu\text{g/L}$), (i) sulfadiazine ($\mu\text{g/L}$).

Contaminant	Paramètres	Valeurs médianes et intervalle de crédibilité à 95%
Cuivre	NEC, k_S, m_0	$NEC = 12.84$ [9.94 - 16.01] $\mu\text{g L}^{-1}$ $k_S = 0.0501$ [0.0309 - 0.0842] $\text{L}/\mu\text{g jour}^{-1}$ $m_0 = 0.00574$ [0.00379 - 0.00824] jour^{-1}
Zinc	NEC, k_S	$NEC = 0.184$ [0.139 - 0.206] mg L^{-1} $k_S = 0.143$ [0.112 - 0.180] $\text{L}/\text{mg jour}^{-1}$
Cadmium	NEC, k_S, m_0	$NEC = 0.720$ [0.0427 - 1.78] $\mu\text{g L}^{-1}$ $k_S = 0.0121$ [0.00840 - 0.0176] $\text{L}/\mu\text{g jour}^{-1}$ $m_0 = 0.00744$ [0.0049 - 0.00985] jour^{-1}
Dichromate de potassium	NEC, k_S	$NEC = 0.281$ [0.228 - 0.306] mg L^{-1} $k_S = 0.145$ [0.112 - 0.184] $\text{L}/\text{mg jour}^{-1}$
Chlordane	NEC, k_S, m_0	$NEC = 1.44$ [0.113 - 2.41] $\mu\text{g L}^{-1}$ $k_S = 0.0145$ [0.0123 - 0.0240] $\text{L}/\mu\text{g jour}^{-1}$ $m_0 = 0.00261$ [0.000144 - 0.00816] jour^{-1}
Ivermectine	NEC, k_S, m_{max}	$NEC = 0.531$ [0.0513 - 0.952] pg L^{-1} $k_S = 0.00990$ [0.00744 - 0.00865] $\text{L}/\text{pg jour}^{-1}$ $m_{max} = 0.185$ [0.0833 - 0.342] jour^{-1}
Acide oxolinique	NEC, k_S	$NEC = 0.216$ [0.0393 - 0.352] mg L^{-1} $k_S = 0.0379$ [0.0251 - 0.0587] $\text{L}/\text{mg jour}^{-1}$
Streptomycine	NEC, k_S	$NEC = 22.2$ [8.60 - 30.2] mg L^{-1} $k_S = 0.00126$ [0.000617 - 0.00261] $\text{L}/\text{mg jour}^{-1}$
Sulfadiazine	NEC, k_S, m_0	$NEC = 49.2$ [16.3 - 94.5] mg L^{-1} $k_S = 0.00431$ [0.00106 - 0.0370] $\text{L}/\text{mg jour}^{-1}$ $m_0 = 0.00447$ [0.00106 - 0.00900] jour^{-1}

Tableau 2.8 – Estimations des paramètres et intervalles de crédibilité à 95% des modèles NEC les plus appropriés aux contaminants.

2.1.4 Comparaison des concentrations critiques d'effet

A partir du travail fait précédemment, nous avons pu sélectionner une méthode pour la détermination de la NOEC (Cochran-Armitage avec procédure "step-down"), un modèle pour l'estimation des LC_x d'intérêt (Weibull) et un modèle NEC à quatre paramètres (pouvant être adapté en fonction du jeu de données). Nous avons reporté les valeurs obtenues sur la Figure 2.7 afin de pouvoir les comparer.

Nous pouvons remarquer d'après cette Figure 2.7, qu'excepté pour le cadmium, les valeurs de NOEC, LC_5 et NEC sont toutes supérieures ou égales à la plus petite concentration en contaminant testée. Ceci montre que les plans d'expérience sont plutôt bien adaptés. Pour 6 contaminants sur 9, les incertitudes obtenues autour des valeurs de la LC_5 sont plus petites que celles obtenues autour de la NEC et de la NOEC. Pour les trois autres contaminants ((b), (d) et (g) sur la Figure 2.7), ce sont les incertitudes autour de la NEC qui sont les plus petites, pour ces contaminants un modèle NEC à deux paramètres avait été utilisé. Dans les cas où la mortalité naturelle est négligeable comme c'est le cas du zinc, du dichromate de potassium, de l'ivermectine, de l'acide oxolinique et de la streptomycine, les estimations ponctuelles de la NEC sont proches voire en dessous de celles de la LC_5 (Figure 2.7 (b), (d), (f), (g) et (h)). Par contre, quand on observe de la mortalité naturelle (cuivre, cadmium, chlordane et sulfadiazine) et qu'on utilise un modèle NEC pour l'estimer, alors les valeurs des estimations des NEC sont supérieures à celles des LC_5 (Figure 2.7 (a), (c), (e) et (i)). Nous avons d'ailleurs observé (Figure 2.5 (a), (c), (e), (i)) que le modèle de Weibull n'arrivait pas à bien décrire les observations pour ces contaminants. Ainsi, nous pouvons dire que les valeurs de LC_5 pour ces contaminants sont sous estimées et que la mortalité naturelle quand elle n'est pas prise en compte dans le modèle est imputée au contaminant. Nous pouvons donc considérer que le modèle NEC semble donner une estimation plus réaliste de la NEC. De plus, dans tous les cas, les estimations ponctuelles de la NEC se situent à l'intérieur de l'intervalle $[0 ; LOEC]$ ou de l'intervalle de crédibilité autour de la LC_5 .

2.1.5 Discussion et conclusion sur l'analyse des données de survie

Dans cette partie, nous avons comparé les méthodes et les modèles habituellement utilisés pour déterminer ou estimer des NOEC, des LC_x et des NEC pour un test de toxicité de 21 jours chez la daphnie. Contrairement aux LC_x et aux NEC, il n'est pas nécessaire de choisir un modèle pour déterminer une NOEC, ce qui pourrait être considéré comme un avantage (van der Hoeven et al, 1997). Cependant, en accord avec la littérature (Chapman et al, 1996), les valeurs de NOEC sont très dépendantes du test d'hypothèse utilisé. Ainsi, nous avons pu voir que le test de Cochran-Armitage avec procédure "step-down" donnait les valeurs de NOEC les plus protectrices pour l'environnement. Mais l'un des inconvénients majeurs de la NOEC est qu'elle est inévitablement l'une des concentrations testées (de Bruijn and Hof, 1997). La véritable valeur de concentration pour laquelle il n'y a aucun effet est comprise entre 0 et la LOEC, ce qui n'est pas très informatif en terme de gestion du risque.

Nous avons pu constater que les valeurs obtenues de LC_{50} pour les trois modèles

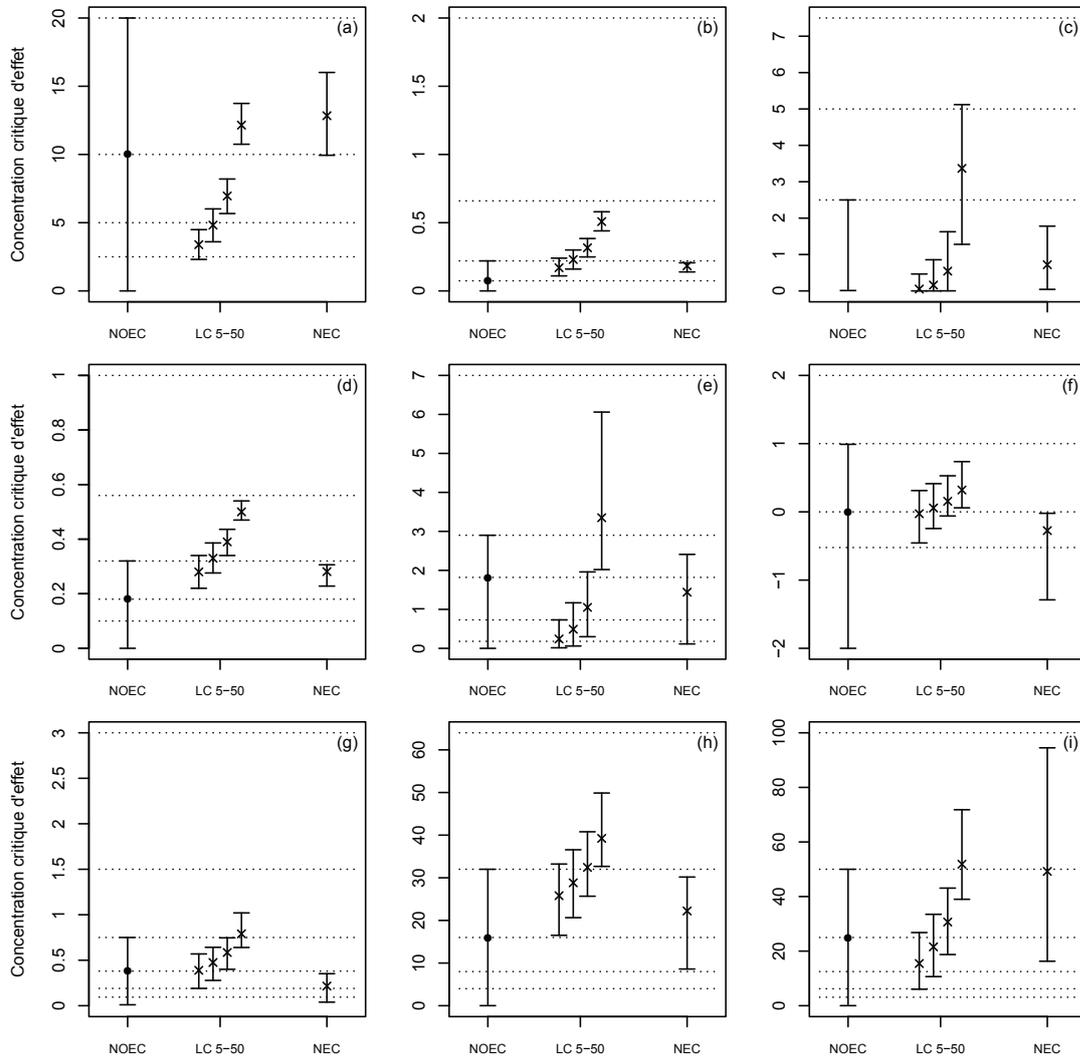


FIGURE 2.7 – Dans chaque graphique sont représentées les concentrations critiques d'effet obtenues pour chaque contaminant. Les NOEC et les intervalles $[0 ; \text{LOEC}]$ ont été déterminés en utilisant le test de Cochran-Armitage avec une procédure "step-down"; les LC_5 , LC_{10} , LC_{20} et LC_{50} et leurs intervalles de crédibilité à 95% ont été estimés en utilisant le modèle de Weibull; les NEC et leurs intervalles de crédibilité à 95%, ont été estimés en utilisant le modèle NEC le plus approprié. Les lignes horizontales en pointillés représentent les concentrations testées. Pour l'ivermectine, les résultats sont présentés en \log_{10} pour gagner en visibilité. (a) cuivre ($\mu\text{g/L}$), (b) zinc (mg/L), (c) cadmium (mg/L), (d) dichromate de potassium (mg/L), (e) chlordane ($\mu\text{g/L}$), (f) ivermectine ($\log_{10} \text{pg/L}$), (g) acide oxolinique (mg/L), (h) streptomycine ($\mu\text{g/L}$), (i) sulfadiazine ($\mu\text{g/L}$).

testés étaient très proches. Ces résultats sont en accord avec ceux de la littérature (van der Hoeven et al, 1997; Ritz, 2010) et montrent que le choix du modèle n'a pas une très forte influence sur l'estimation de cette concentration critique d'effet. A l'issue de notre comparaison des trois modèles concentration-réponse, nous avons retenu le modèle de Weibull comme étant le plus approprié. De plus ce modèle ne nécessite pas une conversion des concentrations en log et les données du témoin peuvent être prises en compte (Ritz, 2010). Le modèle log-logistique peut également être adapté, contrairement au modèle log-probit pour lequel les paramètres ne peuvent pas toujours être estimés. Ces résultats sont en accord avec les travaux de Moore et Caux (1997) pour des modèles à deux paramètres et ceux d'Isnard et al (2001), qui soulignaient également les problèmes rencontrés avec l'utilisation d'un modèle log-probit comparé aux mêmes modèles mais dans le cas d'analyse de données de reproduction et de croissance en utilisant une approche fréquentiste. Le fort impact prévisible du x sur la valeur de LC_x est un inconvénient majeur pour les gestionnaires du risque qui ont à choisir une valeur de x pour déterminer un niveau d'effet acceptable (van der Hoeven, 1997, 2004; Green et al, 2012). Cependant, nous avons pu voir précédemment qu'une utilisation correcte des tests statistiques avec les intervalles de confiance des ratios des moyennes implique également le choix d'un x . Contrairement à ce qui était attendu, les estimations des LC_5 étaient globalement aussi précises que celles des LC_{50} et même plus précises que les estimations des LC_{50} quand on n'observait pas une mortalité importante aux fortes concentrations à 21 jours comme pour le cadmium et le chlordane (Figure 2.6 (c1) et (e1)). Ainsi, la plus petite LC_x (c'est-à-dire LC_5) parmi celles recommandées semble être estimée de manière suffisamment précise pour l'ensemble des jeux de données que nous avons étudiés.

Nous avons estimé des valeurs de NEC en utilisant un modèle à seuil où la mortalité instantanée est fonction à la fois de la concentration et du temps, tout en ayant le souci de garder un modèle simple. Le nombre important de données disponibles entre 0 et 21 jours nous a permis d'ajuster un modèle à seuil à deux, trois ou quatre paramètres. Un modèle flexible de la sorte semble être adapté et efficace pour décrire nos neuf jeux de données dans leur ensemble. Le modèle NEC est par sa définition plus flexible que les modèles concentration-réponse. Ainsi, quand c'est nécessaire, la mortalité instantanée naturelle peut être estimée facilement et précisément. Même s'il est possible d'ajouter une mortalité naturelle dans le modèle de Weibull, le nombre de données à 21 jours seulement ne permet en général pas d'estimer proprement trois paramètres (Ritz, 2010). Dans les cas où la mortalité naturelle n'est pas négligeable, les estimations des LC_5 (à 21 jours) étaient en dessous de celles des NEC, puisque le modèle concentration-réponse représente à tort la mortalité "naturelle" comme étant due au contaminant. Dans ces cas, les intervalles de crédibilité autour des LC_5 étaient plus petits que ceux autour de la NEC, ce qui est lié à la plus grande complexité du modèle NEC (un paramètre en plus) pour estimer la mortalité naturelle.

La comparaison des concentrations critiques d'effet montre que :

- aucune des trois concentrations critiques d'effet (valeurs ponctuelles) n'est systématiquement plus protectrice pour l'environnement pour les neuf contaminants étudiés ;
- les intervalles de crédibilité autour des valeurs de la NEC sont dans la plupart

des cas relativement précis ;

- et les valeurs ponctuelles des NEC appartiennent aux intervalles autour de la NOEC et des LC_x .

Ainsi, il est raisonnable de proposer l'utilisation d'un modèle NEC, qui serait une meilleure option en terme de précision, de protection de l'environnement et d'aide à la décision pour les gestionnaires du risque. Ceci rejoint le travail de Pires et al (2002). Le modèle NEC permet en plus d'utiliser les informations disponibles au cours du temps qui ne sont pas valorisées avec les autres estimations de concentrations critiques d'effet. Cependant, l'utilisation d'une NEC, et donc d'un seuil, peut être contestée pour des raisons biologiques comme van Straalen (1997) l'a résumé. En effet, certaines réactions physiologiques mènent à une réponse graduée et non pas à un seuil. Nous pouvons ainsi argumenter en faveur de l'utilisation de la NEC que :

- un modèle à seuil, comme un modèle NEC, peut être simplement considéré comme une approximation d'un modèle sans seuil (par exemple comme le modèle de Weibull) et que la valeur du seuil estimée ne doit pas être vue comme une valeur absolue (Wennig, 2000; Cairns, 1992). Elle n'est pas la preuve de l'existence d'un vrai seuil biologique mais permet d'avoir une information sur la quantité de contaminant à partir de laquelle on peut commencer à observer des effets ;
- de plus, l'utilisation d'une concentration ne produisant "pas d'effet" en gestion du risque est toujours une réalité et un besoin pour les décisionnaires qui utilisent toujours les NOEC/LOEC. Seuls quelques pays comme le Canada, l'Australie et la Nouvelle-Zélande commencent à exclure la production de NOEC dans les récentes révisions de leur guide sur la qualité des eaux au profit des EC_x et des NEC (van Dam et al, 2012).

Ce travail sur la survie a fait l'objet d'une publication :

Forfait-Dubuc C, Charles S, Billoir E and Delignette-Muller ML (2012) Survival data analyses in ecotoxicology : critical effect concentrations, methods and models. What should we use ? Ecotoxicology 21 : 1072–1083

Conclusion :

Dans cette première partie sur l'analyse des données de survie, nous avons vu que :

- les valeurs de NOEC et de LOEC à 21 jours étaient dépendantes de la méthode utilisée et que le test de Cochran-Armitage associé à une procédure "step-down" semblait être la méthode donnant les valeurs de NOEC et de LOEC les plus protectrices pour l'environnement ;
- les trois modèles concentration-réponse que nous avons comparés (log-logistique, log-probit et Weibull) donnent des valeurs très proches de LC_{50} à 21 jours, néanmoins le modèle de Weibull semble avoir la déviance moyenne la plus faible globalement pour les neuf contaminants testés ;
- un modèle à seuil, simple et flexible, appelé modèle NEC, permet de décrire l'ensemble du jeu de données au cours du temps (qui est sous exploité quand on utilise seulement un modèle concentration-réponse classique à un temps donné) et permet d'estimer de manière assez précise une NEC.

Après avoir comparé ces trois concentrations critiques d'effet, nous pensons qu'il pourrait être intéressant que des outils appropriés et faciles à utiliser soient développés pour mettre les modèles NEC à disposition des écotoxicologues pour analyser les données de survie au cours du temps et ainsi optimiser l'analyse de ces données.

Publication associée : *Forfait-Dubuc C, Charles S, Billoir E and Delignette-Muller ML (2012) Survival data analyses in ecotoxicology : critical effect concentrations, methods and models. What should we use ? Ecotoxicology 21 : 1072–1083*

2.2 Optimisation des plans d'expérience pour des données de survie

2.2.1 Introduction

Dans ce chapitre, nous nous intéressons au deuxième aspect qui peut permettre d'améliorer les estimations des concentrations critiques d'effet d'intérêt : l'optimisation des plans d'expérience. En effet, les jeux de données sont issus de plans d'expérience qui, validés par les guides OCDE, ont pour vocation à déterminer à la fois des NOEC ou des LOEC mais aussi des EC_x ou LC_x . Ainsi, historiquement, des plans d'expérience "type" basés sur l'estimation des NOEC et des LOEC se sont imposés et sont repris dans les nouveaux guides OCDE.

L'optimisation des plans d'expérience (ou designs) est étudiée depuis de nombreuses années dans des domaines d'application tels que la toxicologie, la pharmacologie, la physique... avec, d'une part, l'utilisation de méthodes locales de détermination d'un plan d'expérience optimal (local optimal design) (Chernoff, 1953), et d'autre part l'utilisation d'une approche bayésienne globale (optimal Bayesian design) (Chaloner et al, 1989).

La première approche consiste à choisir un plan d'expérience performant pour estimer au mieux les paramètres du modèle fixés au préalable. Il faut donc connaître les paramètres du modèle avant de pouvoir les estimer, ce qui constitue un inconvénient majeur (Chaloner et al, 1989; King et Wong, 2000) surtout pour une utilisation pratique. Si les paramètres sont mal fixés, le "plan d'expérience optimal" obtenu ne sera peut être pas du tout optimal pour les vraies valeurs des paramètres. Malgré cet inconvénient, cette approche reste très intéressante pour répondre à des questions générales et théoriques sur les plans d'expérience.

L'approche globale, quant à elle, permet de prendre en compte toutes les valeurs possibles des paramètres du modèle que l'on veut estimer, en tenant compte de l'information disponible sur ces paramètres sous forme de distributions *a priori*. Une approche intermédiaire aux deux précédentes consiste à déterminer un plan d'expérience optimal quand seule l'étendue des valeurs des paramètres est connue et pas leurs distributions (Minimax D-optimal design) (King et Wong, 2000).

En écotoxicologie, l'intérêt pour les plans d'expérience optimaux est récent (Andersen et al, 2000; Chèvre et Brazzale, 2008; Wright et Bailer, 2006; Albert et al, 2012). Cet intérêt peut être relié à l'utilisation de plus en plus fréquente de modèles concentration-réponse (plus ou moins mécanistes) qui sont voués à remplacer les tests d'hypothèse, très critiqués dans leur utilisation pour déterminer des concentrations critiques d'effet (Fox, 2008; Warne et van Dam, 2008). En effet, historiquement, les plans d'expérience ont plutôt été conçus pour déterminer des NOEC et LOEC par tests d'hypothèse comme décrits dans le guide OCDE pour l'analyse des données (OECD, 2006). Cependant, idéalement la NOEC et les LC_x devraient être issues de plans d'expérience différents (Green et al, 2012).

Parmi les trois modèles concentration-réponse recommandés par le guide OCDE (OECD, 2006) (modèles de Weibull, log-probit et log-logistique), nous avons choisi de travailler dans un premier temps avec le modèle log-logistique à deux paramètres.

Nous avons choisi ce modèle car, avec le modèle de Weibull, il donnait de meilleurs ajustements dans notre étude comparant ces trois modèles (Forfait-Dubuc et al, 2012) et c'est celui qui est le plus utilisé et connu des deux.

Notons $S(\text{conc}_i)$ la probabilité de survie en fin d'essai à la concentration conc_i . Pour le modèle log-logistique à deux paramètres, elle s'écrit :

$$S(\text{conc}_i) = 1/(1 + (\frac{\text{conc}_i}{LC_{50}})^{b_1}) \tag{2.13}$$

Comme le paramètre b_1 n'est pas très explicite en terme de toxicité nous avons décidé de le réécrire en fonction de LC_{50} et de LC_5 (concentration conduisant à 5 % de mortalité soit 95 % de survie) et plus particulièrement du $\log_2(\frac{LC_{50}}{LC_5})$ à partir de la relation :

$$0.95 = 1/(1 + (\frac{LC_5}{LC_{50}})^{b_1}) \tag{2.14}$$

On obtient ainsi :

$$b_1 = \frac{\log_2(0.95/0.05)}{\log_2(LC_{50}/LC_5)} = \frac{4.25}{\log_2(LC_{50}/LC_5)} \tag{2.15}$$

Rappelons que la partie stochastique associée à ce type de données est de type binomial. Le nombre de survivants N_i à la concentration conc_i va suivre une loi binomiale de paramètres $S(\text{conc}_i)$ et N_{i0} le nombre d'individus initialement introduits.

$$N_i \sim \text{Binomiale}(S(\text{conc}_i), N_{i0}) \tag{2.16}$$

L'objectif est ici d'optimiser les plans d'expérience, et pour cela, nous cherchons à quantifier l'impact que peut avoir le choix de l'écotoxicologue sur les composantes du plan d'expérience comme : 1) la position des concentrations testées par rapport à une valeur attendue de la concentration d'intérêt, 2) le nombre de concentrations testées ainsi que la raison de la progression géométrique entre chaque concentration, 3) le poids de chaque concentration, ce qui correspond au nombre d'individus testés par concentration et 4) le nombre total d'individus testés.

Un plan d'expérience optimal est intimement dépendant des paramètres de la courbe concentration-réponse et donc du contaminant étudié. Nous avons décidé de raisonner en relatif par rapport à la LC_{50} , sachant que ces valeurs en tant que telles ne nous intéressent pas, mais c'est l'incertitude autour de cette valeur qui est intéressante. Le paramètre b_1 nous intéresse également et donc le rapport $\log_2(LC_{50}/LC_5)$ est important pour caractériser la courbe (il est inversement proportionnel à b_1). C'est donc pour ce rapport que nous avons besoin d'information *a priori*.

Il nous a paru intéressant de raisonner sur le \log_2 des concentrations du fait que les gammes de concentrations dans les plans d'expérience sont souvent obtenues par des dilutions de raison deux. Nous avons donc cherché des informations sur ce rapport $\log_2(LC_{50}/LC_5)$ à partir de la littérature :

- soit à partir de jeux de données complets pour lesquels nous avons estimé la LC_{50} et la LC_5 en utilisant le package *drc* (Ritz et Streibig, 2005) sous R (R Development Core Team, 2010)
- soit à partir d'une valeur de LC_x , de la pente et du modèle
- soit à partir de deux valeurs de LC_x différentes et du modèle

Ainsi, nous avons pu tirer de l'information de 22 études issues de la littérature pour des tests de toxicité différents :

- 15 tests de survie à 21 jours : cuivre et zinc (Billoir et al, 2008b), cadmium (Santojanni et al, 1995), bichromate de potassium (Bedaux et Kooijman, 1994), chlordanes (Manar et al, 2009), ivermectine (Lopes et al, 2009), acide oxolinique, streptomycine et sulfadiazine (Wollenberger et al, 2000), composé à base de nanoparticules de TiO_2 (Wiensch et al, 2009), acides déhydroabiétique et abiétique (Kamaya et al, 2012), uranium (Zeman et al, 2008), sulfonate de perfluorooctane et perfluorooctanoate d'ammonium (Li, 2010)
- 7 tests de survie en toxicité aiguë : acide oxolinique, streptomycine, sulfadiazine, tiamuline et tylosine (Wollenberger et al, 2000), nonylphenol (Hirano et al, 2004) et le dinosèbe (Chèvre, 2000).

Sur la Figure 2.8, les valeurs des $\log_2(LC_{50})$ en fonction du rapport $\log_2(LC_{50}/LC_5)$ ont été représentées pour chaque étude. Il est possible de constater que les valeurs des $\log_2(LC_{50})$ sont réparties de manière assez aléatoire, ce qui est normal puisque ces valeurs sont caractéristiques d'un contaminant donné pour un test de toxicité donné. Le $\log_2(LC_{50}/LC_5)$, qui est inversement proportionnel à la pente de la courbe concentration-réponse, n'est pas dépendant du type de test de toxicité, ni de la valeur du $\log_2(LC_{50})$. Ainsi, nous pouvons considérer que les informations acquises sur ces 22 jeux de données peuvent constituer une information *a priori* valable sur le rapport $\log_2(LC_{50}/LC_5)$, quel que soit le contaminant étudié et le test de toxicité réalisé. Nous avons donc caractérisé cette distribution par une loi log-normale de moyenne 0,0524 et d'écart type 0,929 en utilisant le package *fitdistrplus* sous R (Delignette-Muller et al, 2010).

Concernant la LC_{50} , c'est l'incertitude autour de la valeur de la LC_{50} qui est importante et donc le fait que la valeur de la LC_{50} n'est pas forcément celle qui était attendue. C'est pourquoi nous étudions deux cas de figures : dans le premier cas de figure, nous estimons que l'écotoxicologue a une idée relativement précise de la valeur de LC_{50} (que l'on nommera par la suite L) et la situe entre $L/2$ et $L \times 2$, ce qui donne en log base 2 : $\log_2(L) \in [\log_2(L) - 1; \log_2(L) + 1]$; dans le second cas de figure, nous supposons que l'incertitude autour de L est plus importante, entre $L/10$ et $L \times 10$, ce qui donne en log base 10 : $\log_{10}(L) \in [\log_{10}(L) - 1; \log_{10}(L) + 1]$. Ainsi, la distribution du $\log_2(LC_{50})$ a été caractérisée par une distribution normale de moyenne 0 et d'écart type 0,5 dans le cas d'une information *a priori* plutôt précise, donnée par l'écotoxicologue et dans le cas d'une information moins précise, c'est la distribution du $\log_{10}(LC_{50})$ qui suit une loi normale de moyenne 0 et d'écart type 0,5.

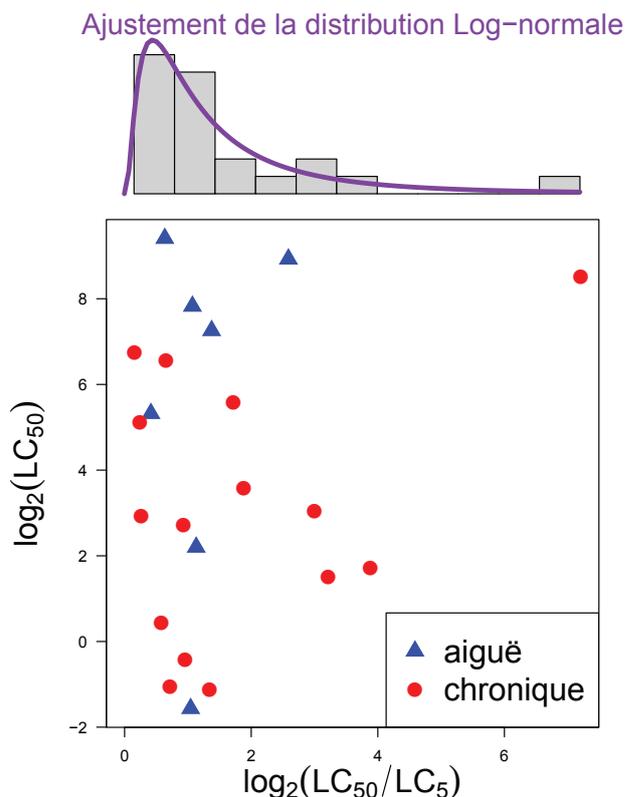


FIGURE 2.8 – Valeurs du $\log_2(LC_{50})$ par rapport au $\log_2(\frac{LC_{50}}{LC_5})$ pour les 7 jeux de données en toxicité aiguë (en bleu) et pour les 15 jeux de données en toxicité chronique (en rouge) pour les tests de survie de *Daphnia magna*. En haut, en gris est représentée la distribution des valeurs du $\log_2(\frac{LC_{50}}{LC_5})$ et en violet la distribution log-normale ajustée.

Dans un premier temps, c'est une étude théorique sur certaines composantes du plan d'expérience qui est présentée. Elle est basée sur la D-optimalité en local, c'est à dire pour un jeu de paramètres (LC_{50} et $\log_2(\frac{LC_{50}}{LC_5})$) correspondant aux valeurs les plus probables. D'autres analyses locales ont tout de même été envisagées en choisissant plusieurs valeurs du rapport $\log_2(\frac{LC_{50}}{LC_5})$. Puis dans un second temps, c'est une étude pratique tenant compte des contraintes techniques qui est réalisée sur des plans d'expérience réalistes. Elle est réalisée globalement et est basée sur la D-optimalité et sur le calcul du gain de connaissance (Albert et al, 2012) sur les concentrations critiques d'intérêt. Cette approche globale permet de tester un grand nombre de paires de paramètres possibles à partir de leurs distributions *a priori* en passant par la construction de jeux de données simulées réalistes pour le calcul du gain de connaissance.

2.2.2 Etude théorique utilisant la D-optimalité localement

La D-optimalité, comme utilisée dans le calcul du "minimax D-optimal design" pour le modèle logistique dans le cas de la survie King et Wong (2000), est la maximisation du déterminant de la matrice d'information de Fisher, équivalente à la minimisation des

variances et covariances des paramètres. En effet, il y a optimalité, quand on minimise :

- les incertitudes sur les estimations des paramètres du modèle = les variances
- les dépendances entre les paramètres = les covariances.

L'Annexe 4 présente la démonstration permettant d'obtenir l'écriture de la matrice de Fisher (notée \mathbf{F}) pour le modèle logistique dans le cas de données de survie :

$$\mathbf{F} = \begin{pmatrix} b^2 \sum_{i=1}^K n_i p_i (1 - p_i) & - \sum_{i=1}^K n_i b (x_i - a) p_i (1 - p_i) \\ - \sum_{i=1}^K n_i b (x_i - a) p_i (1 - p_i) & \sum_{i=1}^K n_i (x_i - a)^2 p_i (1 - p_i) \end{pmatrix} \quad (2.17)$$

Avec pour le modèle log-logistique qui nous intéresse, p_i correspondant à la probabilité de survie pour une concentration x_i en log, $a = \log(\text{LC}_{50})$, $b = -b_1$ la pente du modèle log-logistique et n_i le nombre d'individus à la concentration x_i . Nous pouvons écrire $n_i = w_i \times N$ avec w_i le poids donné à la concentration x_i et N le nombre total d'individus testés. La matrice \mathbf{F} peut alors s'écrire sous la forme :

$$\mathbf{F} = \begin{pmatrix} A & -B \\ -B & C \end{pmatrix} \quad (2.18)$$

et

$$\begin{aligned} \mathbf{A} &= b^2 \sum_{i=1}^K \omega_i N p_i (1 - p_i) \\ \mathbf{B} &= \sum_{i=1}^K b \omega_i N (x_i - a) p_i (1 - p_i) \\ \mathbf{C} &= \sum_{i=1}^K \omega_i N (x_i - a)^2 p_i (1 - p_i) \end{aligned} \quad (2.19)$$

Le déterminant à maximiser est :

$$|\mathbf{F}| = AC - (-B)(-B) = AC - B^2 \quad (2.20)$$

$$|\mathbf{F}| = N^2 \times \left(\sum_{i=1}^k w_i p_i (1 - p_i) \times \sum_{i=1}^k w_i (b(x_i - a))^2 p_i (1 - p_i) - \left(\sum_{i=1}^k b w_i (x_i - a) p_i (1 - p_i) \right)^2 \right) \quad (2.21)$$

$$= N^2 \times (\text{Somme}_1 \times \text{Somme}_2 - \text{Somme}_3^2)$$

L'étude locale se fait en prenant les valeurs les plus probables des paramètres du modèle, établies à partir des informations *a priori*. Ainsi, les impacts de plusieurs choix concernant les composantes des plans d'expérience sont présentés en étudiant le déterminant de la matrice de Fisher localement.

Impact de la position de la valeur attendue de la LC_{50} dans la gamme de concentrations (étape 1)

Nous avons exprimé mathématiquement le déterminant de la matrice de Fisher dans le cas où on aurait un plan d'expérience avec cinq concentrations testées. Ces concentrations sont séparées suivant une suite géométrique de raison égale à 2 avec c , une concentration quelconque en \log_2 . Le plan d'expérience est donc constitué des concentrations $c - 2$; $c - 1$; c ; $c + 1$; $c + 2$. Les w_i sont considérés comme étant tous identiques. Les composantes $Somme_1$, $Somme_2$ et $Somme_3$ peuvent s'écrire en fonction de c de la manière suivante :

$$Somme_1 = w \times [f(c - 2) \times (1 - f(c - 2)) + f(c - 1) \times (1 - f(c - 1)) + f(c) \times (1 - f(c)) + f(c + 1) \times (1 - f(c + 1)) + f(c + 2) \times (1 - f(c + 2))] \quad (2.22)$$

$$Somme_2 = w \times b^2 \times [(c - 2 - a)^2 \times f(c - 2) \times (1 - f(c - 2)) + [(c - 1 - a)^2 \times f(c - 1) \times (1 - f(c - 1))] + [(c - a)^2 \times f(c) \times (1 - f(c)) + (c + 1 - a)^2 \times f(c + 1) \times (1 - f(c + 1)) + (c + 2 - a)^2 \times f(c + 2) \times (1 - f(c + 2))] \quad (2.23)$$

$$Somme_3 = w \times b \times [(c - 2 - a) \times f(c - 2) \times (1 - f(c - 2)) + [(c - 1 - a)^2 \times f(c - 1) \times (1 - f(c - 1))] + [(c - a) \times f(c) \times (1 - f(c)) + (c + 1 - a)^2 \times f(c + 1) \times (1 - f(c + 1)) + (c + 2 - a) \times f(c + 2) \times (1 - f(c + 2))] \quad (2.24)$$

Nous pouvons écrire $|\mathbf{F}|$ de la manière suivante :

$$|\mathbf{F}| = N^2 \times w^2 \times b^2 \times H(c) \quad (2.25)$$

avec

$$\begin{aligned}
H(c) = & -(((-2 - a + c) \times (1 - 1/(1 + \exp(b \times (-2 - a + c)))))/(1 + \exp(b \times (-2 - a + c))) + \\
& ((-1 - a + c) \times (1 - 1/(1 + \exp(b \times (-1 - a + c)))))/(1 + \exp(b \times (-1 - a + c))) + \\
& ((-a + c) \times (1 - 1/(1 + \exp(b \times (-a + c)))))/(1 + \exp(b \times (-a + c))) + \\
& ((1 - a + c) \times (1 - 1/(1 + \exp(b \times (1 - a + c)))))/(1 + \exp(b \times (1 - a + c))) + \\
& ((2 - a + c) \times (1 - 1/(1 + \exp(b \times (2 - a + c)))))/(1 + \exp(b \times (2 - a + c)))^2 + \\
& ((1 - 1/(1 + \exp(b \times (-2 - a + c)))))/(1 + \exp(b \times (-2 - a + c))) + \\
& (1 - 1/(1 + \exp(b \times (-1 - a + c))))/(1 + \exp(b \times (-1 - a + c))) + \\
& (1 - 1/(1 + \exp(b \times (-a + c))))/(1 + \exp(b \times (-a + c))) + \\
& (1 - 1/(1 + \exp(b \times (1 - a + c))))/(1 + \exp(b \times (1 - a + c))) + \\
& (1 - 1/(1 + \exp(b \times (2 - a + c))))/(1 + \exp(b \times (2 - a + c))) \times \\
& (((-2 - a + c)^2 \times (1 - 1/(1 + \exp(b \times (-2 - a + c)))))/(1 + \exp(b \times (-2 - a + c))) + \\
& (((-1 - a + c)^2 \times (1 - 1/(1 + \exp(b \times (-1 - a + c)))))/(1 + \exp(b \times (-1 - a + c))) + \\
& (((-a + c)^2 \times (1 - 1/(1 + \exp(b \times (-a + c)))))/(1 + \exp(b \times (-a + c))) + \\
& (((1 - a + c)^2 \times (1 - 1/(1 + \exp(b \times (1 - a + c)))))/(1 + \exp(b \times (1 - a + c))) + \\
& (((2 - a + c)^2 \times (1 - 1/(1 + \exp(b \times (2 - a + c)))))/(1 + \exp(b \times (2 - a + c)))) \quad (2.26)
\end{aligned}$$

Ainsi pour maximiser le déterminant $|\mathbf{F}|$, il faut trouver pour quelles valeurs de c la fonction $H(c)$ trouve son maximum. Pour cela, nous avons choisi plusieurs couples de paramètres a et b_1 , avec deux valeurs de a testées et quatre valeurs de b_1 (1, 2, 3, 4) correspondant à certaines valeurs de $\log_2(\frac{LC_{50}}{LC_5})$ prises dans les informations *a priori* de la Figure 2.8 (respectivement 4.25, 2.12, 1.42 et 1.05). La valeur $b_1=4$, correspond à la valeur la plus probable. Les courbes concentration-réponse correspondant aux différents paramètres testés sont présentés dans la Figure 2.9.

Ainsi, pour chaque couple de paramètres a et b_1 , nous avons tracé les courbes $H(c)$ pour voir pour quelles valeurs de c elles sont maximales (Figure 2.10). Il est possible de constater que pour $b_1=1$ et 2, le maximum de la courbe est atteint pour un $c = a$. Pour des valeurs de b_1 plus élevées (3 et 4) et donc des pentes plus raides, il y a deux concentrations pour lesquelles $H(c)$ est maximale. Pour $a=4$, ces concentrations sont $c=3.5$ et $c=4.5$ et pour $a=2$, ces concentrations sont $c=1.5$ et 2.5 . Une pente raide implique que la zone importante où se situe la LC_{50} se situe sur une gamme de concentrations plus restreinte. Ainsi, si on ne diminue pas la raison de la série de concentrations testées, cette zone sera couverte avec un maximum de concentrations si la gamme n'est pas centrée sur la LC_{50} , mais centrée sur des concentrations proches. Ainsi, dans le cas $a=4$ et $b_1=4$, nous pouvons voir sur la Figure 2.11, que centrer la gamme de concentrations sur la valeur de a ne permet d'avoir qu'une seule concentration dans l'intervalle où le taux de survie est influencé par les concentrations (intervalle d'intérêt). Tandis que sur la Figure 2.12, nous pouvons voir que décaler légèrement la gamme de concentrations (centrée dans cet exemple sur 3.5) permet d'avoir deux concentrations dans l'intervalle d'intérêt. Ainsi, dans le cas des pentes raides, une raison plus petite permettrait d'obtenir la maximum de la courbe $H(c)$ pour $c = a$.

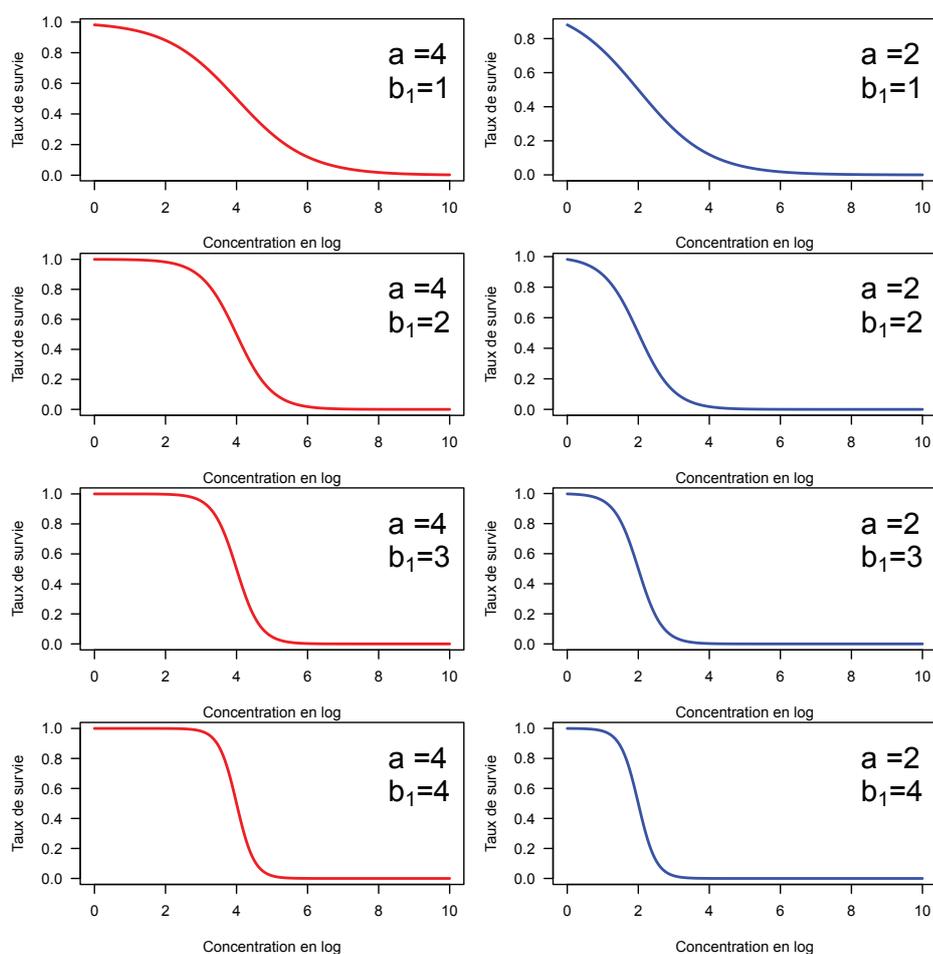


FIGURE 2.9 – Courbes concentration-réponse obtenues pour différentes valeurs de $a = \log(\text{LC}_{50})$ et de b_1 . Les courbes en rouge sont celles obtenues pour un $a=4$ et en bleu pour un $a=2$

Globalement, il semble favorable de centrer la gamme de concentrations sur la valeur de $a = \log(\text{LC}_{50})$ pour le modèle log-logistique. De plus, le paramètre b a de l'importance dans le choix de la gamme de concentrations surtout sur la raison de la série des concentrations. Il est donc intéressant de prendre en compte toutes les valeurs possibles de b en faisant une étude globale, et comme attendu, la valeur de a ne joue pas sur les conclusions.

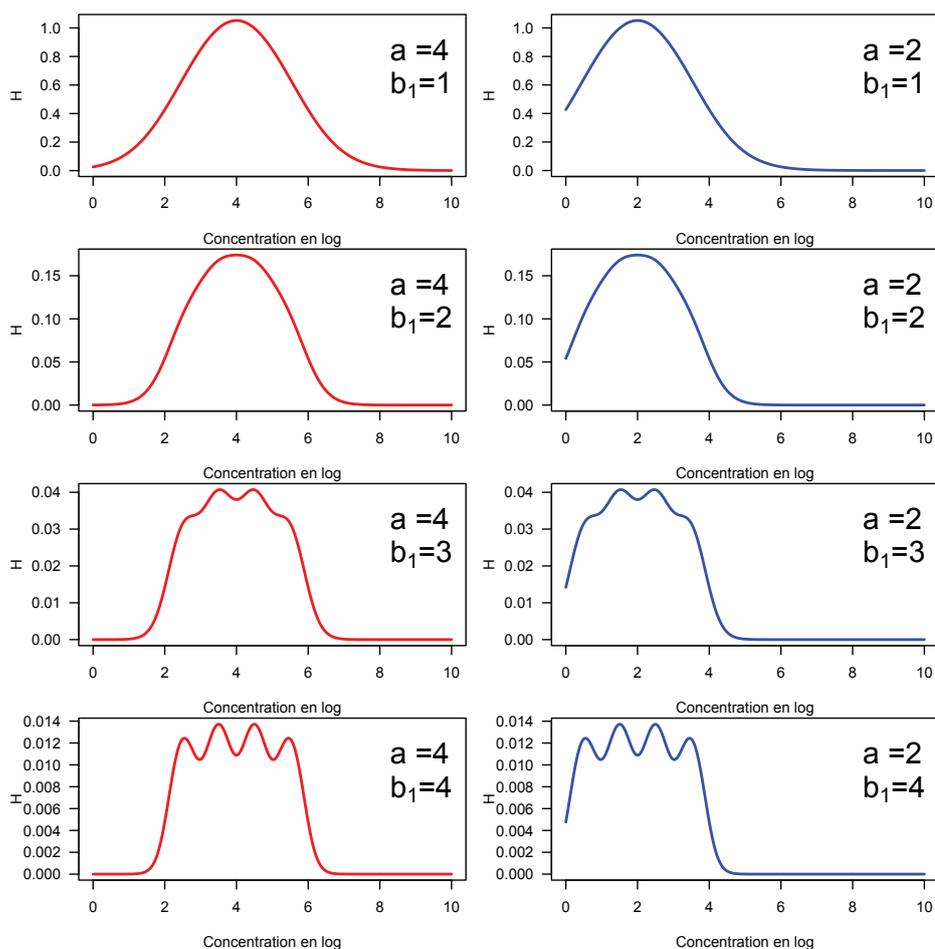


FIGURE 2.10 – Valeurs de la fonction $H(c)$, qui est proportionnelle au déterminant de la matrice de Fisher pour différents couples de paramètres a et b_1 . Les courbes en rouge sont celles obtenues pour un $a=4$ et en bleu pour un $a=2$.

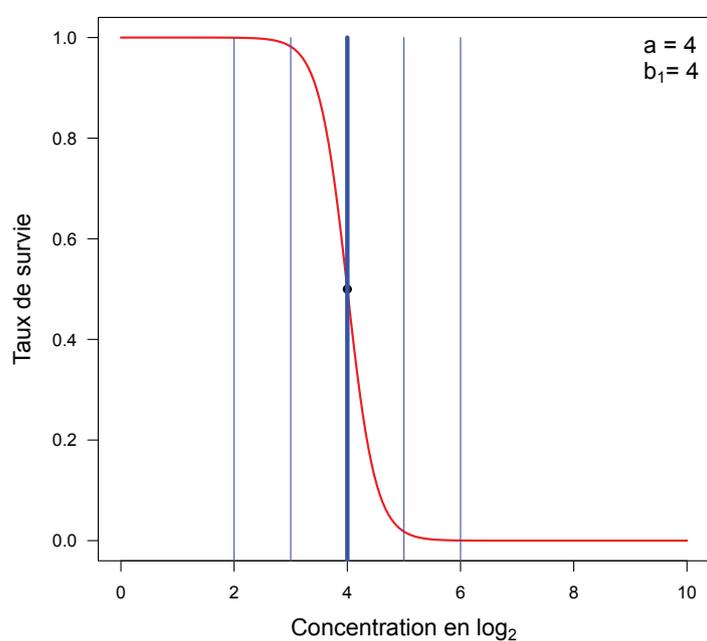


FIGURE 2.11 – Courbe concentration-réponse pour $a= 4$ et $b_1= 4$ (en rouge) et concentrations testées (lignes bleues) pour un plan d'expérience centré sur a . La concentration se situant dans l'intervalle où le taux de survie est influencé par les concentrations est représentée par une ligne épaisse.

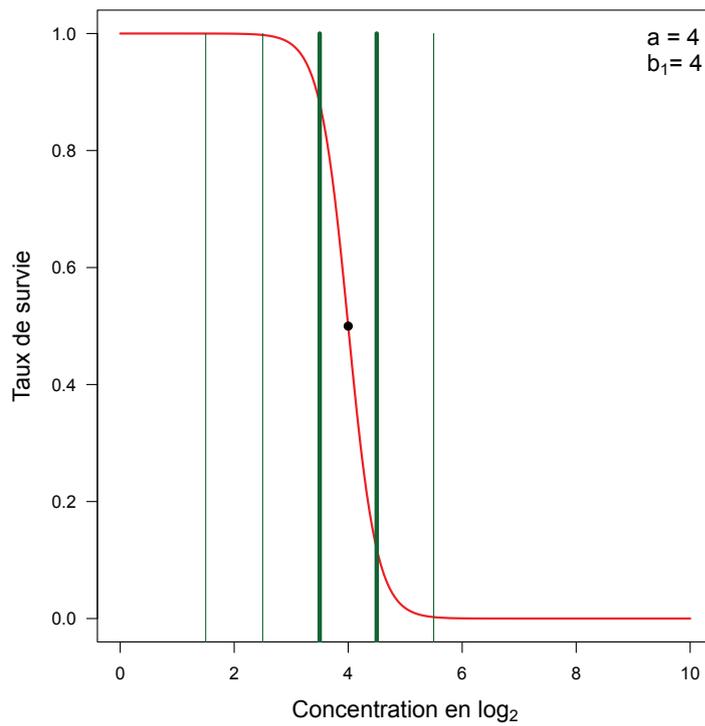


FIGURE 2.12 – Courbe concentration-réponse pour $a=4$ et $b_1=4$ (en rouge) et concentrations testées (lignes vertes) pour un plan d'expérience centré sur 3.5 (pour laquelle $H(c)$ est maximale). Les concentrations se situant dans l'intervalle où le taux de survie est influencé par les concentrations sont représentées par des lignes épaisses.

Impact du nombre de concentrations testées et de la raison de la progression géométrique (étape 2)

Nous étudions maintenant l'impact du choix du nombre de concentrations et donc de la raison de la progression géométrique séparant chaque concentration, sur les valeurs de déterminant. Cette étude se fait sur une étendue de gamme donnée, c'est-à-dire avec une concentration minimale et une concentration maximale fixées. Nous avons choisi une étendue de $4 \times \log_2$. La Figure 2.13 représente les résultats obtenus localement dans un cadre théorique où pour un nombre d'organismes fixé (120), nous avons comparé plusieurs plans d'expérience qui vont de seulement 2 concentrations testées à 120. La raison la plus petite est obtenue dans le cas où 120 concentrations sont testées et donc seulement un organisme par concentration. Il est possible de remarquer un effet de palier à partir d'environ 20 concentrations testées (en bleu). Entre 10 (en orange) à 20 concentrations, les valeurs de déterminants obtenues sont très proches. Il est possible également de remarquer qu'en dessous de 10 concentrations les valeurs des déterminants sont notablement plus faibles dès qu'on enlève une concentration. De plus, il ne semble pas raisonnable d'avoir moins de 6 concentrations testées (en jaune).

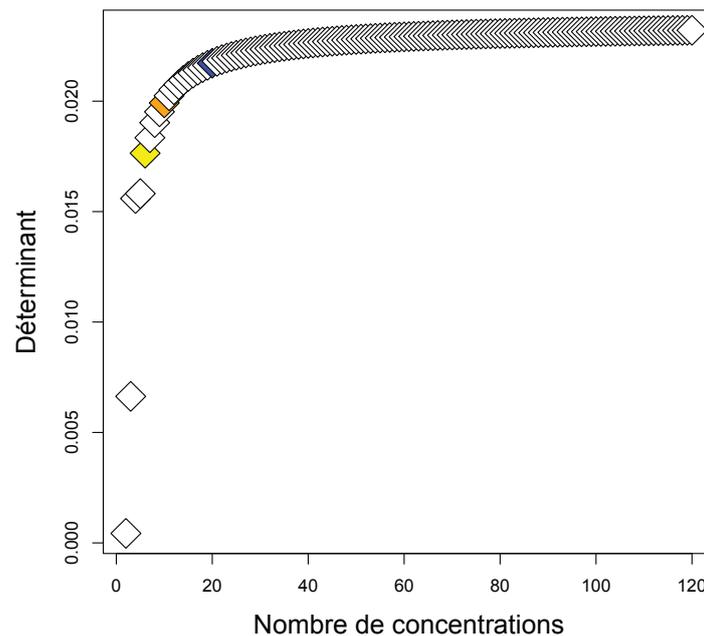


FIGURE 2.13 – Valeurs de déterminant de la matrice de Fisher déterminées localement ($a = \log(1)$ et $b_1 = 4$) en fonction du nombre de concentrations testées avec un nombre d'organismes fixés à 120. La valeur du déterminant pour 6 concentrations est représentée en jaune, celle pour 10 concentrations en orange et pour 20 concentrations en bleu.

Impact du poids des concentrations c'est-à-dire le nombre d'individus testés par concentration (étape 3)

Dans cette étape 3, il s'agit d'étudier l'impact que peut avoir le poids des concentrations sur les valeurs de déterminant. Pour cela, nous avons cherché à évaluer les poids optimaux permettant d'obtenir un déterminant maximum. Dans cette partie d'optimisation, en plus de la paire de valeurs les plus probables, nous avons testé également d'autres valeurs du paramètre de pente avec la même valeur la plus probable de LC_{50} . Nous avons considéré que les poids étaient symétriques autour de la concentration centrale afin de limiter le nombre de poids à optimiser. Ainsi, il a été possible d'optimiser les poids en utilisant la fonction *optim* de R mais uniquement dans le cas d'une information *a priori* précise (5 concentrations testées). En effet, dans le cas d'une information *a priori* moins précise (7 concentrations testées), les paramètres à optimiser étaient trop nombreux.

L'hypothèse forte faite sur la symétrie des poids, afin de pouvoir résoudre le problème de l'optimisation, permet de considérer que les concentrations 1 et 5 ont des poids identiques ($w_1 = w_5$), tout comme les concentrations 2 et 4 ($w_2 = w_4$). Le poids de la troisième concentration L (w_3) résulte de la différence entre 1 et la somme des autres poids. En effet, la somme totale de tous les poids doit être égale à 1. Ainsi, nous avons tracé les contours du déterminant en fixant dans un premier temps la valeur de $a = \log(LC_{50})$ et de $b = -b_1 = \frac{-4.25}{\log_2(LC_{50}/LC_5)}$ à leurs valeurs les plus probables soit $a = 0$ et $b = -4$ (troisième graphique de la Figure 2.14). Nous obtenons ainsi un optimal local représenté par une croix rouge sur le troisième graphique de la Figure 2.14. Cet optimal local correspond au plan d'expérience pour lequel $w_1 = w_5 = 0$, $w_2 = w_4 = 0.32$ et $w_3 = 0.36$.

Afin d'avoir une vision un peu plus globale, nous avons décidé de tester d'autres valeurs de b à partir de l'information *a priori* obtenue sur $\log_2(LC_{50}/LC_5)$. Nous avons choisi de faire l'analyse locale pour différentes valeurs de quantiles de la distribution log normale (meanlog=0.0524, sdlog=0.929). Les résultats sont présentés dans le Tableau 2.9 et sur la Figure 2.14 dans l'ordre croissant des valeurs de $b = -b_1$. Pour s'assurer de la convergence vers une solution unique (représentée par une croix rouge dans les graphiques), nous avons réalisé deux optimisations avec des valeurs initiales de paramètres différentes. Nous pouvons constater que tous les optima locaux donnent des poids nuls pour la première (w_1) et la cinquième concentration (w_5) et des poids $w_2=w_4$ qui varient selon les valeurs de b . L'optimisation d'un plan d'expérience à cinq concentrations nous donne en fait un plan avec trois concentrations. Nous pouvons retenir donc retenir deux plans d'expérience :

- Type 1 : Davantage de poids est donné à la concentration centrale (0 ; 0.25 ; 0.50 ; 0.25 ; 0)
- Type 2 : Quasiment le même poids est donné aux trois concentrations en favorisant légèrement le poids de la concentration centrale (0 ; 0.32 ; 0.36 ; 0.32 ; 0)

Ce résultat pourrait paraître en contradiction avec ceux de l'étape 2 concernant le nombre de concentrations (Figure 2.13) mais dans ce cas là, la raison était diminuée en même temps que le nombre de concentrations augmentaient et une même gamme

était testée. Par contre, le fait que l'optimum soit obtenu quand un poids nul est mis sur les deux concentrations extrêmes semblent rejoindre les résultats de l'étape 1. En effet, ce qui ressort c'est que la gamme de concentrations est trop étendue par rapport l'intervalle de concentrations où le taux de survie est le plus dépendant des concentrations testées.

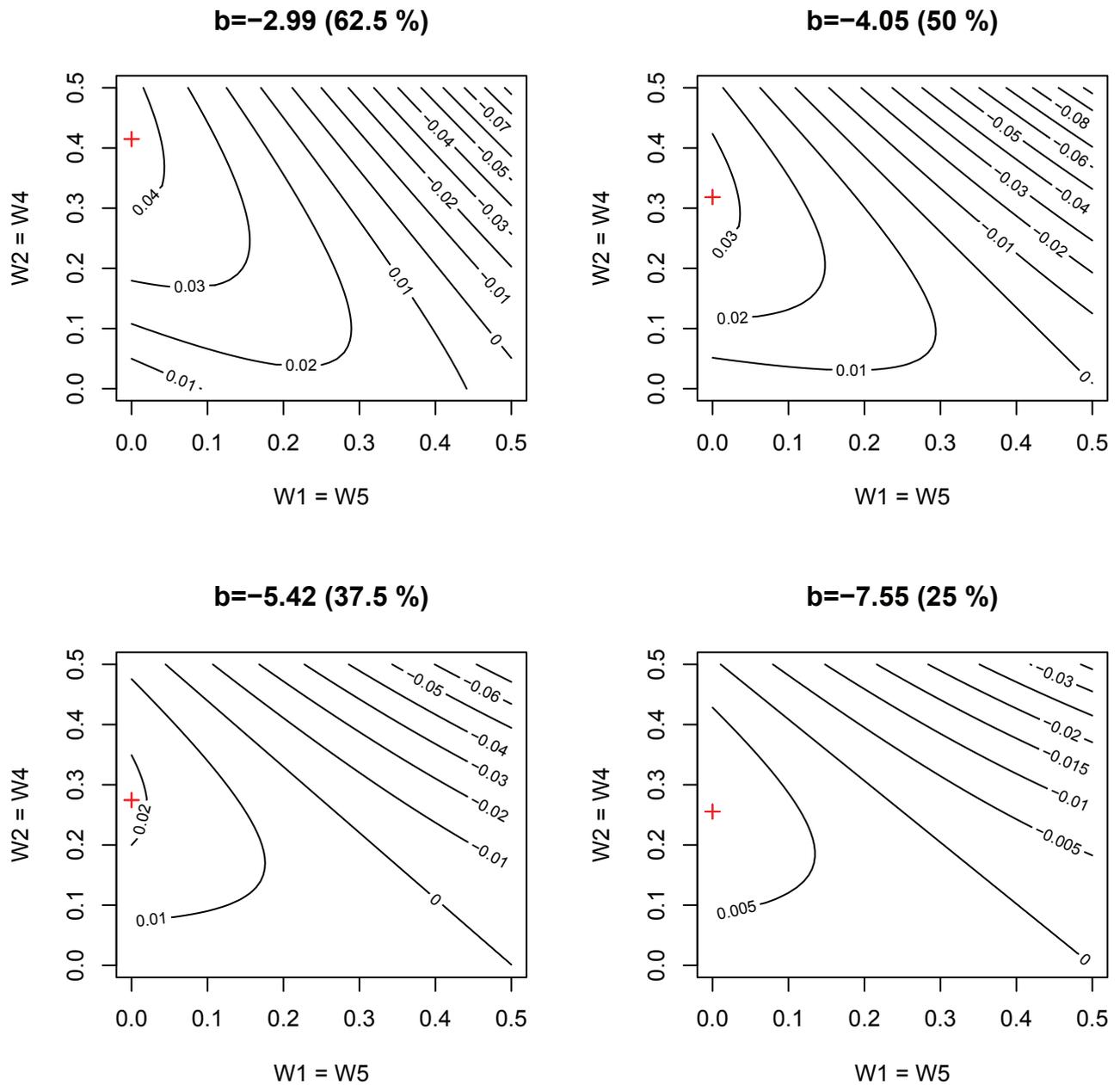


FIGURE 2.14 – Valeurs du déterminant en fonction des poids $w_2 = w_4$ et $w_1 = w_5$ pour différentes valeurs de b correspondant aux différents quantiles de la distribution $\log_2(LC_{50}/LC_5)$ précisés entre parenthèses. Les optima locaux sont représentés par une croix rouge.

Quantiles de la distribution	$\log_2(\text{LC}_{50}/\text{LC}_5)$	b	$w1=w5$	$w2 = w4$	$w3$
25%	0.563	-7.55	0	0.255	0.49
37.5%	0.784	-5.42	0	0.274	0.452
50%	1.05	-4.05	0	0.318	0.364
62.5%	1.42	-2.99	0	0.415	0.17

Tableau 2.9 – Quantiles de la distribution log normale du $\log_2(\text{LC}_{50}/\text{LC}_5)$ et valeurs associées de b et des différents poids obtenus par optimisation. $w1 = w5$ et $w2 = w4$ sont les poids respectifs des concentrations 1, 5, 2 et 4. De même, $w3$ correspond au poids de la troisième concentration.

L'étude faite localement ne permet d'avoir des résultats que pour un jeu de paramètres donnés. Il est donc nécessaire d'avoir une approche globale afin de visualiser les résultats pour l'ensemble des couples possibles de paramètres, basés sur les informations *a priori* dont nous disposons. De plus, il est nécessaire de se rapprocher des conditions de laboratoire pour avoir une idée des contraintes auxquelles il n'est pas possible d'échapper. En effet, on ne peut pas conseiller de faire une expérience avec 120 concentrations testées, ni même de tester une étendue de gamme de concentrations restreintes. Ainsi, nous avons défini des plans d'expérience à tester issus des recommandations des guides OCDE, qui serviront de base de comparaison pour notre étude globale.

2.2.3 Etude globale basée sur des plans d'expérience réalistes

Méthodes utilisées pour comparer les plans d'expérience

Pour comparer les plans d'expérience dans cette étude globale, nous avons utilisé la D-optimalité comme pour la partie locale mais aussi le calcul du gain de connaissance sur les concentrations critiques d'effet d'intérêt.

Nous avons utilisé la définition du gain de connaissance donnée par [Albert et al \(2012\)](#) et qui correspond au rapport de la variance de l'information *a priori* sur celle de l'*a posteriori* (Figure 2.15) afin d'estimer la qualité de l'estimation des paramètres d'intérêt. Pour chaque jeu de données, un gain de connaissance sur un ou plusieurs paramètres peut être calculé. Nous avons choisi de nous intéresser aux gains de connaissance sur les concentrations d'intérêt : la LC_{50} et la LC_5 .

La Figure 2.16 présente les différentes étapes permettant le calcul de 1000 valeurs de déterminants (D-optimalité) et de 1000 gains de connaissance sur la LC_{50} et la LC_5 . Ainsi, 1000 tirages ont été réalisés dans les distributions *a priori* de chacun des paramètres, obtenant ainsi 1000 paires de paramètres du modèle log-logistique, ce qui équivaut à 1000 courbes concentration-réponse théoriques. A partir de ces 1000 courbes concentration-réponse théoriques, 1000 valeurs du déterminant sont calculées pour un plan d'expérience à étudier. L'association de ces 1000 courbes concentration-réponse et d'un plan d'expérience à étudier permet également de construire 1000 jeux de données simulées. En effet, pour une courbe concentration-réponse donnée, seuls les taux de survie correspondant aux concentrations du plan d'expérience étudié, sont retenus. Ce taux de survie multiplié par le nombre d'individus par concentration permet alors d'obtenir un nombre de survivants pour chaque concentration testée et donc un jeu de données simulées. Pour chaque plan d'expérience, nous obtenons donc 1000 jeux de données simulées. L'inférence bayésienne est alors utilisée pour ajuster chaque jeu de données simulées avec le modèle log-logistique et ainsi estimer les paramètres LC_{50} et $\log_2(LC_{50}/LC_5)$ pour lesquels une distributions *a posteriori* des paramètres sera alors disponible. Le gain de connaissance sur les paramètres LC_{50} et $\log_2(LC_{50}/LC_5)$ peut alors être calculé pour chaque jeu de données. Pour chaque plan d'expérience étudié, une distribution de 1000 valeurs de gain de connaissance est alors obtenue.

Protocole d'étude des plans d'expérience

Nous avons choisi d'étudier les plans d'expérience selon un protocole constitué de quatre étapes (Figure 2.17) :

- Etape 1- à partir d'un plan d'expérience basé sur les recommandations des guides OCDE ([OCDE, 2004, 1998](#)), l'impact de la position de L (valeur de LC_{50} attendue par l'écotoxicologue) dans la gamme de concentrations a été étudié. Un plan d'expérience de "référence" a ainsi été retenu en fonction des résultats obtenus avec le calcul des 1000 valeurs de déterminant et des 1000 valeurs de gain de connaissance sur la LC_{50} et la LC_5 . Puis dans les autres étapes, à partir du plan d'expérience de "référence", nous avons étudié les différents impacts des choix faits par l'écotoxicologue, lors de la mise en place d'un plan d'expérience, sur

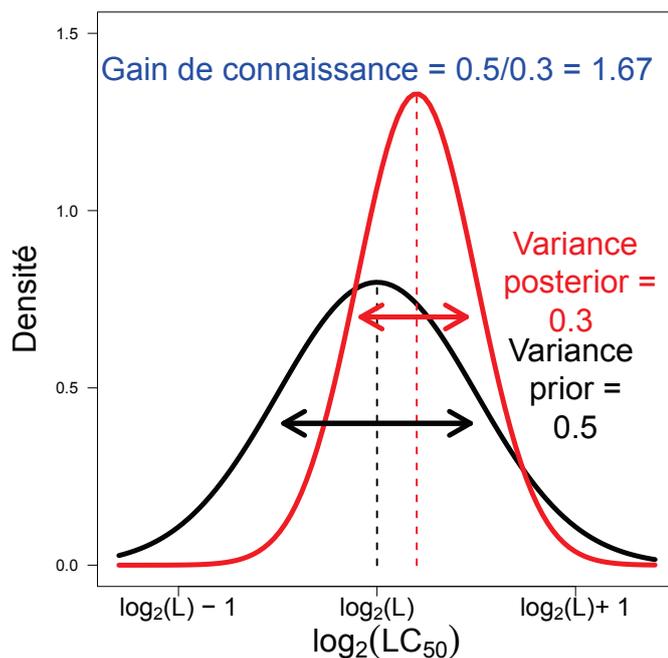


FIGURE 2.15 – Représentation d'une distribution *a priori* (en noir) et d'une distribution *a posteriori* (en rouge) et calcul du gain de connaissance sur le $\log_2(LC_{50})$ sur un cas fictif. Plus le rapport de la variance de l'*a priori* sur la variance de l'*a posteriori* est élevé, plus le gain de connaissance est important.

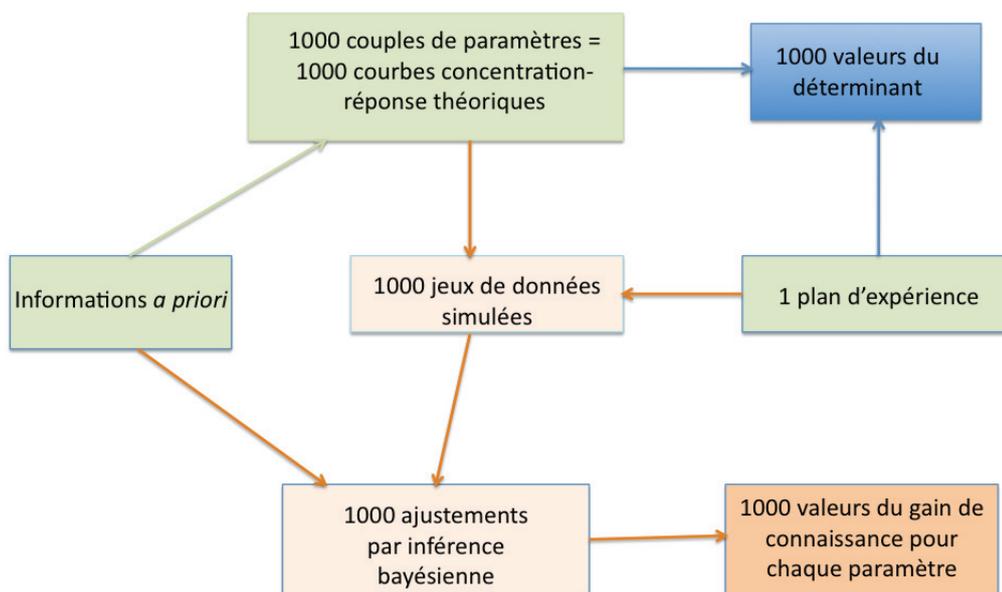


FIGURE 2.16 – Représentation des étapes permettant le calcul de 1000 déterminants et 1000 gains de connaissance pour un plan d'expérience donné.

l'estimation des concentrations critiques d'effet.

- Etape 2- l'impact du choix de la raison de la progression géométrique et du nombre de concentrations testées a été étudié. Ces deux composantes du plan d'expérience sont très liées et ont donc été étudiées conjointement.
- Etape 3- l'impact du choix du poids donné à chaque concentration c'est-à-dire l'impact d'un nombre différent d'individus par concentration
- Etape 4- l'impact du nombre total d'individus exposés dans l'expérience.

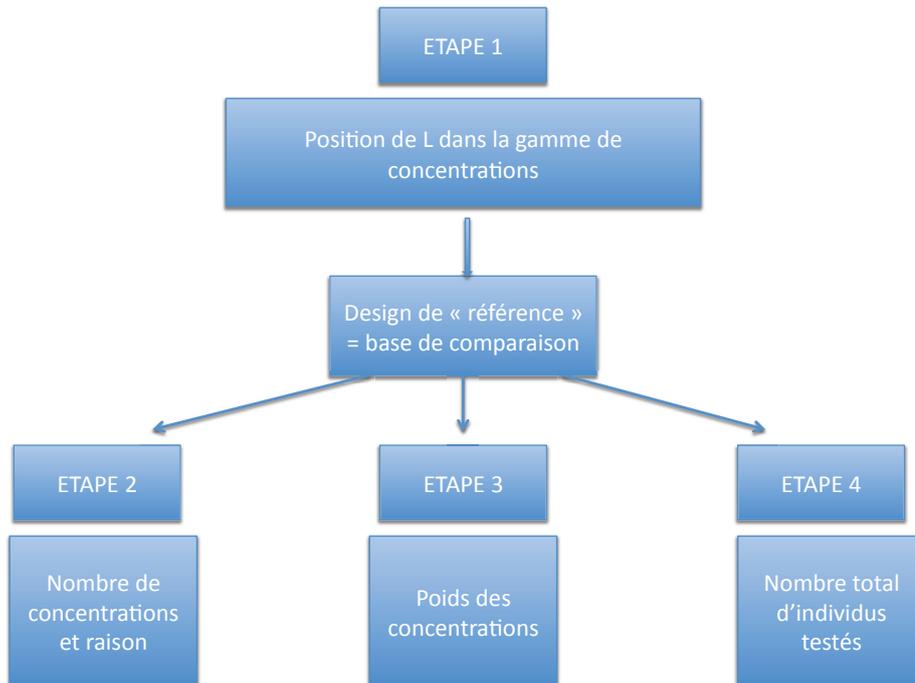


FIGURE 2.17 – Représentation des différentes étapes du protocole de choix des plans d'expérience à tester.

Pour chaque étape, nous allons présenter les résultats globaux obtenus à la fois en utilisant la D-optimalité avec le calcul des déterminants ainsi que ceux obtenus sur les gains de connaissance.

Impact de la position de la valeur attendue de la LC_{50} dans la gamme de concentrations (étape 1)

Au regard des résultats obtenus en 2.2.2, il est nécessaire de tester de manière globale l'impact de la position de la valeur attendue de LC_{50} dans la gamme des concentrations testées.

Les guides OCDE recommandent l'utilisation d'au moins 5 concentrations, espacées d'une raison inférieure ou égale à 2,2 dans le cas des tests aigus, et d'une raison

inférieure ou égale à 3,2 dans le cas des tests chroniques, avec 20 individus au minimum par concentration. Nous avons donc choisi de tester 5 concentrations en plus du témoin, espacées d'une raison égale à 2 et avec 20 individus par concentration, soit 120 individus au total dans le cas d'un *a priori* assez précis sur la valeur de la LC_{50} . Dans le cas d'un *a priori* moins précis, nous avons choisi de tester 7 concentrations en plus du témoin, espacées d'une raison égale à 3 avec 20 individus par concentration, soit 160 individus au total.

Dans les deux cas, pour la première étape, nous testons la place de L (valeur la plus probable de LC_{50}) dans la gamme des concentrations. Ainsi, nous comparons cinq plans d'expérience dans le cas d'une information *a priori* précise et sept plans d'expérience dans le cas d'une information *a priori* peu précise (Figure 2.18).

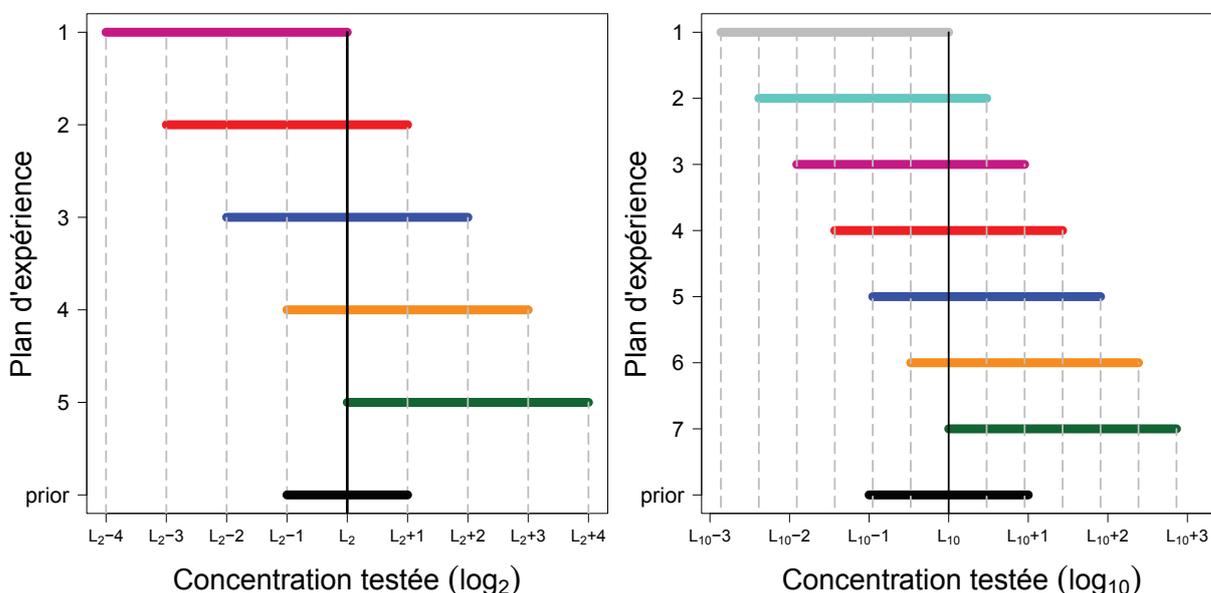


FIGURE 2.18 – Plans d'expérience testés pour l'étape 1 dans le cas 1) d'une information *a priori* précise sur la valeur attendue de LC_{50} (nommée L) : $[\frac{L}{2}; L \times 2]$ (graphique de gauche) et 2) d'une information *a priori* peu précise sur de L : $[\frac{L}{10}; L \times 10]$ (graphique de droite). L_2 correspond au $\log_2(L)$ et L_{10} correspond au $\log_{10}(L)$. L'information *a priori* est représentée par les segments noirs et les plans d'expérience testés par des segments de couleur. Les lignes pointillées représentent les concentrations testées pour chaque plan d'expérience.

Les résultats des distributions des 1000 valeurs de déterminant obtenues globalement pour chacun des plans d'expérience à comparer dans l'étape 1, sont présentés dans la Figure 2.19. Nous avons choisi de représenter ces distributions à l'aide de diagrammes en boîte (ou boîtes à moustaches) dont le rectangle représente les résultats allant du premier quartile (Q1) au troisième quartile (Q3) et est coupé par la médiane (Q2). Les segments de part et d'autre de la boîte représentent les minimum et maximum. Ainsi, 25% des valeurs de déterminant se situent en dessous de Q1, de même 50% et 75% de ces valeurs se situent respectivement en dessous de Q2 et Q3. Ces distributions obtenues pour chacun des plans d'expérience de l'étape 1 (Figure 2.19) permettent de conclure que les meilleurs plans d'expérience semblent être ceux qui sont centrés sur la valeur la plus probable de la LC_{50} (L). Les deux plans d'expérience extrêmes pour les deux *a priori* testés donnent des valeurs de déterminant nettement plus inférieures aux autres plans d'expérience. D'où la nécessité d'avoir des informations *a priori* (pré-tests et/ou bibliographie) sur L afin de ne pas placer L en extrémité de la gamme de concentrations testées. Ces résultats confirment ceux que nous avons obtenus lors de l'analyse locale (2.2.2).

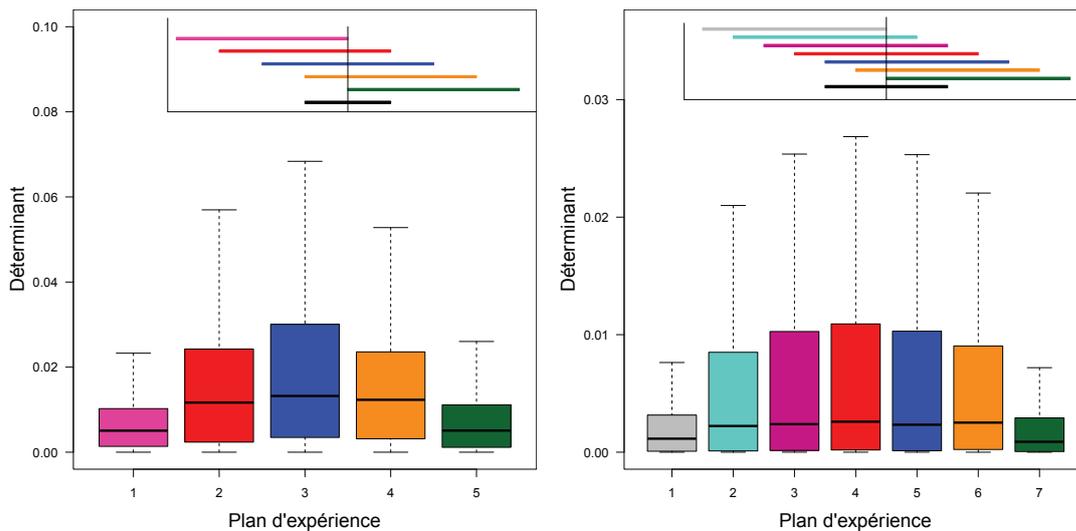


FIGURE 2.19 – Diagrammes en boîte des distributions des 1000 valeurs du déterminant de la matrice \mathbf{F} obtenues pour chacun des cinq plans d'expérience testés dans le cadre d'une information *a priori* précise (à gauche) et pour les sept plans d'expérience testés dans le cadre d'une information *a priori* peu précise (à droite).

Concernant les gains de connaissance, les résultats des distributions des 1000 valeurs de gain de connaissance obtenues pour chaque plan d'expérience sont présentés par les graphiques de la Figure 2.20. Les plans d'expérience 2, 3 et 4 (dans le cas d'un *a priori* précis) et les plans d'expérience 3, 4, 5 (dans le cas d'un *a priori* plus moins précis) donnent des résultats assez similaires en terme de gain de connaissance sur la LC_{50} , avec des résultats plutôt symétriques comme nous avons pu le remarquer avec le calcul du déterminant. Par contre, concernant le gain de connaissance sur la LC_5 , les résultats sont dissymétriques, et nous obtenons une information supplémentaire par rapport à ce qui a été fait avec le calcul du déterminant. En effet, contrairement à l'estimation de la LC_{50} , pour laquelle il n'y a pas de différence si on décale le plan d'expérience par rapport L entièrement d'un côté ou de l'autre (plan d'expérience 1 et 5 dans le cas d'un *a priori* précis et plan d'expérience 1 et 7 dans le cas d'un *a priori* peu précis), le gain de connaissance sur l'estimation de la LC_5 sera moins bon dans la cas d'un plan d'expérience 5 ou 7 (selon l'*a priori*) que dans le cas d'un plan d'expérience 1.

Ainsi, il faut centrer le plan d'expérience sur L pour avoir un bon gain de connaissance sur la LC_{50} et la LC_5 . En cas de doute, il est préférable de choisir plutôt des concentrations inférieures à L, afin d'éviter d'avoir un plan d'expérience (comme le 5) qui serait nettement moins avantageux concernant le gain de connaissance sur la LC_5 .

Pour la suite du travail, nous avons choisi de garder les plans d'expérience centrés sur la valeur de L, c'est-à-dire le plan d'expérience 3 dans le cas d'un *a priori* précis et le plan d'expérience 4 dans le cas d'un *a priori* moins précis.

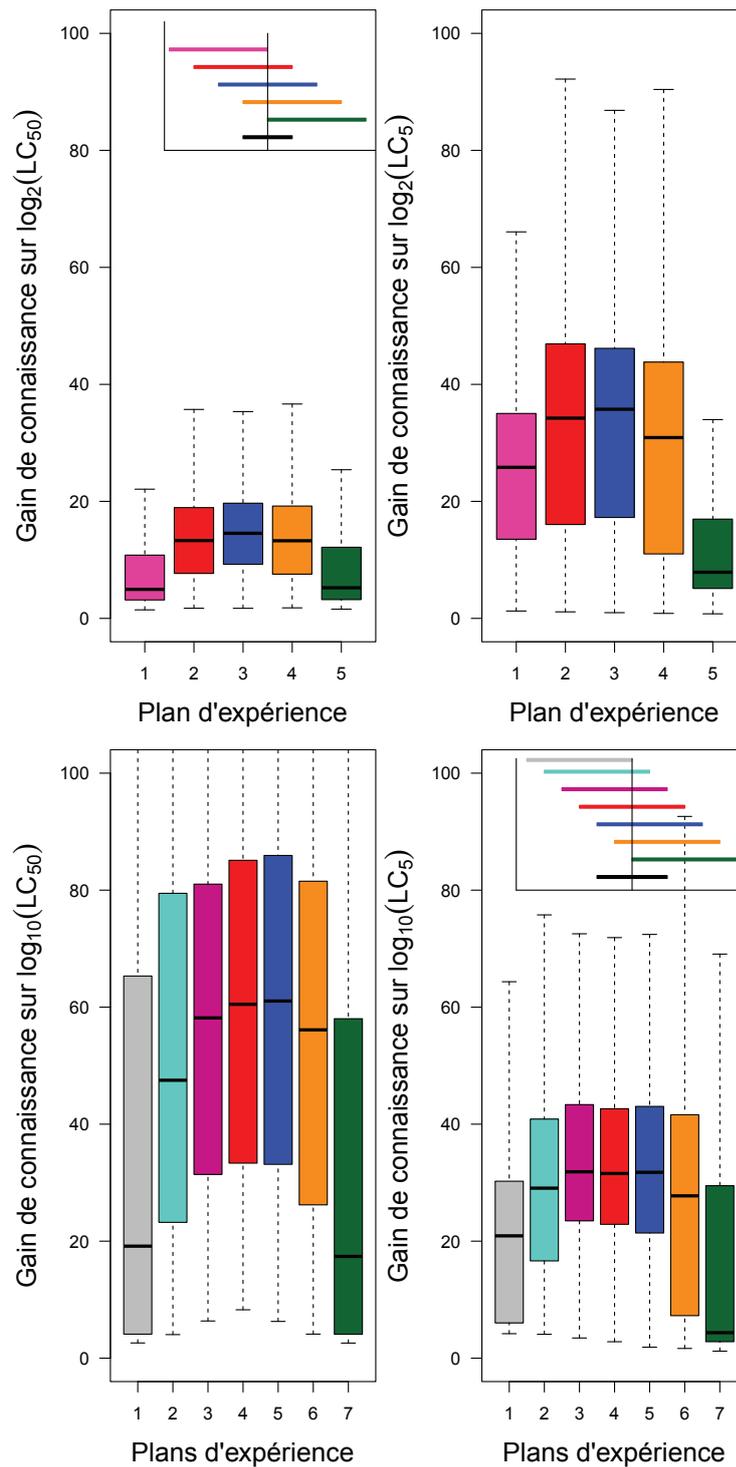


FIGURE 2.20 – Diagrammes en boîte de la distribution de 1000 gains de connaissance pour chacun des plans d'expérience testés dans l'étape 1 sur le $\log_2(LC_{50})$ (graphique de gauche) et le $\log_2(LC_5)$ (graphique de droite) dans le cas d'une information *a priori* précise sur L (graphes du haut) et dans le cas d'une information *a priori* peu précise sur L (graphes du bas).

Impact du nombre de concentrations testées et de la raison de la progression géométrique (étape 2)

Nous avons choisi de tester l'impact du choix de la raison de la progression géométrique suivie par les concentrations et du nombre de concentrations testées, puisque ces deux composantes du plan d'expérience sont intimement liées. En effet, pour comparer des plans d'expérience sur la même base, nous avons essayé de faire en sorte que les plans d'expérience testés couvrent à peu près la même étendue de gamme de concentrations (écart entre la plus faible concentration testée et la plus forte). Les plans d'expérience testés sont présentés sur la Figure 2.21, sachant que nous avons décidé de tous les centrer sur la valeur de L (sur la base des résultats obtenus à l'étape 1) et avec un nombre total d'individus de 120 dans le cas d'une information *a priori* précise et de 160 dans le cas d'une information *a priori* moins précise.

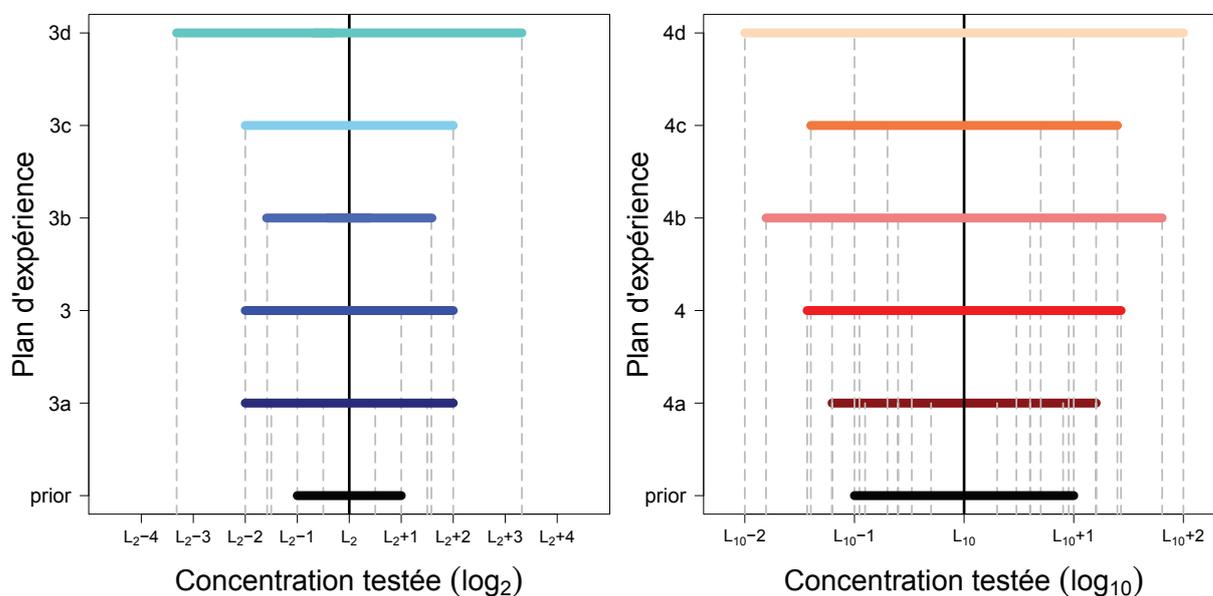


FIGURE 2.21 – Plans d'expérience testés pour l'étape 2 dans le cas 1) d'une information *a priori* précise sur L (la valeur attendu de LC_{50}) : $[\frac{L}{2}; L \times 2]$ (graphique de gauche) et 2) d'une information *a priori* peu précise sur L : $[\frac{L}{10}; L \times 10]$. L_2 correspond au $\log_2(L)$ et L_{10} correspond au $\log_{10}(L)$. L'information *a priori* est représentée par les segments noirs et les plans d'expérience testés par des segments de couleur. Les lignes pointillées représentent les concentrations testées pour chaque plan d'expérience. Par exemple, le plan d'expérience 3a est composé de 9 concentrations, en plus du témoin, qui sont séparées par une raison $\sqrt{2}$.

Dans le cas d'un *a priori* précis, les plans d'expérience testés sont (graphique de gauche de la Figure 2.21) :

- Plan d'expérience 3 : celui de l'étape 1 avec 5 concentrations testées espacées d'une raison 2
- Plan d'expérience 3a : 9 concentrations testées espacées d'une raison $\sqrt{2}$
- Plan d'expérience 3b : 3 concentrations testées espacées d'une raison 3

- Plan d'expérience 3c : 3 concentrations testées espacées d'une raison 4
- Plan d'expérience 3d : 3 concentrations testées espacées d'une raison 10

Dans le cas d'un *a priori* peu précis, les plans d'expérience testés sont (graphique de droite de la Figure 2.21) :

- Plan d'expérience 4 : celui de de l'étape 1 avec 7 concentrations testées avec une raison 3
- Plan d'expérience 4a : 9 concentrations testées espacées d'une raison 2
- Plan d'expérience 4b : 7 concentrations testées espacées d'une raison 4
- Plan d'expérience 4c : 5 concentrations testées espacées d'une raison 5
- Plan d'expérience 4d : 5 concentrations testées espacées d'une raison 10

Afin de rester dans des conditions expérimentales courantes, nous n'avons pas testé la raison $\sqrt{2}$ dans le cas d'un *a priori* plus large afin de ne pas avoir plus de 10 concentrations à tester dans la gamme.

Concernant l'étude globale des déterminants, les résultats sont présentés par la Figure 2.22 avec des plans d'expérience qui sont présentés de gauche à droite dans l'ordre croissant des valeurs des raisons. Les plans d'expérience 3a et 4a semblent donner des résultats légèrement meilleurs respectivement dans le cas d'une information *a priori* précise et peu précise. Ainsi, pour des étendues de la gamme de concentrations quasiment identiques, les plans d'expérience avec plus de concentrations donnent des valeurs de déterminants globalement plus élevées. Ce qui confirme les résultats obtenus localement dans la partie théorique (2.2.2). Cependant, dans le cas où une légère différence entre l'étendue des gammes est observée (le plan d'expérience 4b a une gamme de concentrations plus étendue que le plan d'expérience 4c), même si la raison de la progression géométrique de 4b est plus faible que celle de 4c, les résultats obtenus pour le plan d'expérience 4b sont un peu moins bons que ceux du plan d'expérience 4c (Figure 2.22). Ceci montre l'importance de l'étendue de la gamme : plus la gamme est étendue moins le déterminant est élevé.

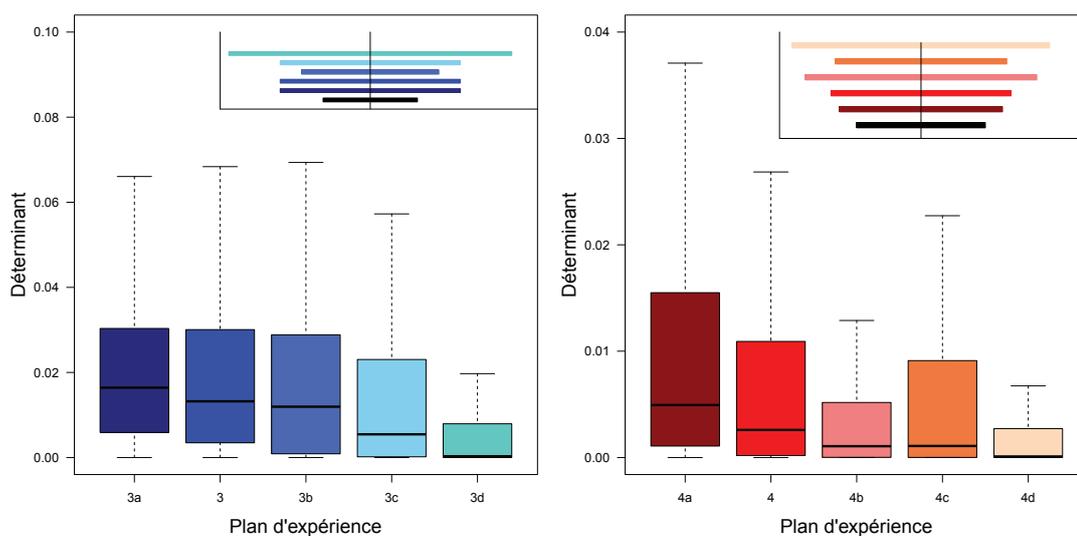


FIGURE 2.22 – Diagrammes en boîte des distributions des 1000 valeurs de déterminant de la matrice de Fisher déterminées globalement en fonction du nombre de concentrations testées pour les différents plans d'expérience à comparer dans le cadre d'un *a priori* précis (à gauche) et dans le cadre d'un *a priori* peu précis (à droite).

Concernant la distribution des gains de connaissance, elle est meilleure pour les plans d'expérience avec une raison faible (3a et 4a) pour les deux informations *a priori* (Figure 2.23). Ce qui va dans le sens des résultats obtenus grâce au calcul du déterminant (D-optimalité) mais avec des différences entre les résultats qui sont un peu plus marquées, et ceci aussi bien pour les gains de connaissance sur la LC_{50} que pour ceux sur la LC_5 . Nous pouvons constater l'importance du gain d'information que l'on peut obtenir en passant d'un plan d'expérience de 5 concentrations (3) à un plan à 9 concentrations (3a) dans le cas d'un *a priori* précis et d'un plan d'expérience de 7 concentrations (4) à un plan d'expérience à 9 concentrations (4a) dans le cas d'un *a priori* peu précis. Dans ce dernier cas, l'effort expérimental à fournir est moins important que dans le premier cas, puisque seulement deux concentrations sont ajoutées à la gamme initiale.

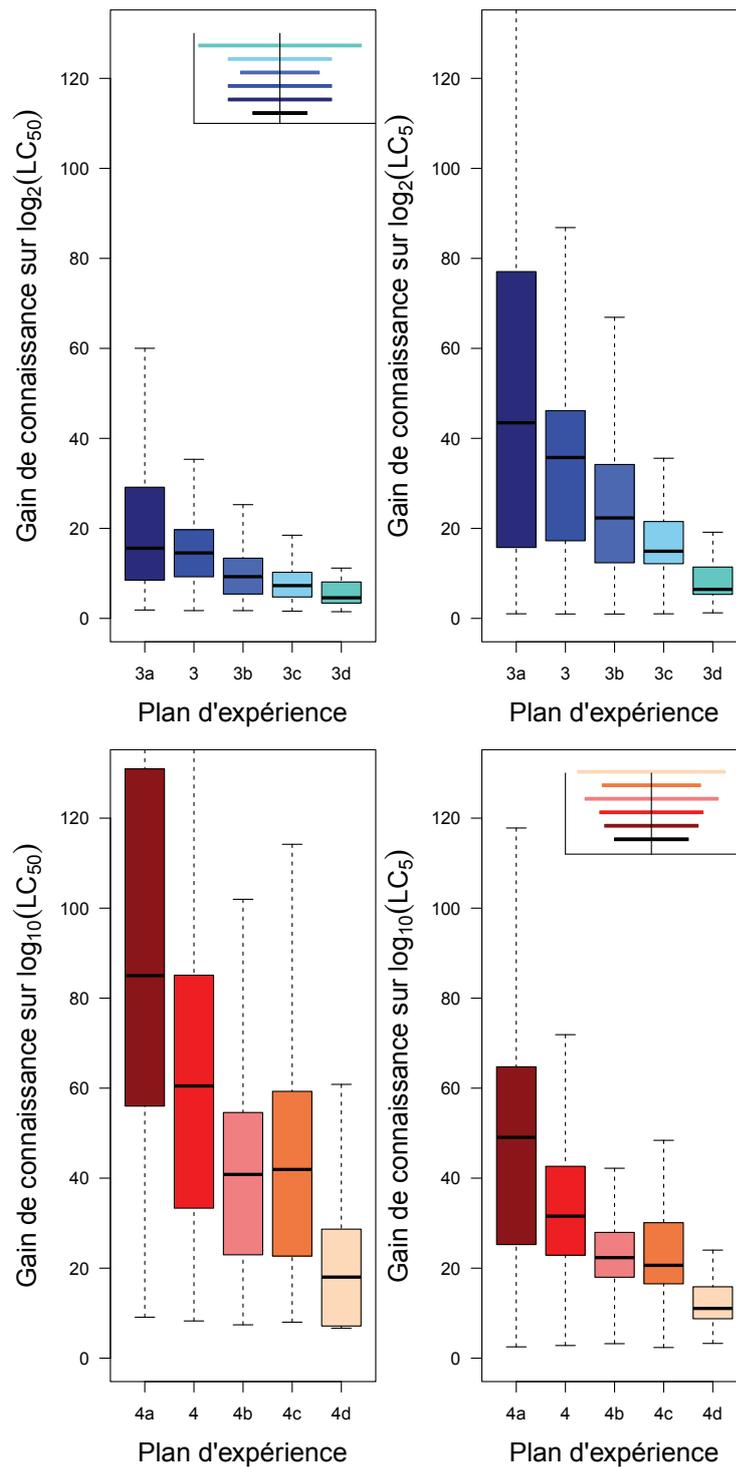


FIGURE 2.23 – Diagrammes en boîte des distributions de 1000 gains de connaissance pour chacun des plans d'expérience testés dans l'étape 2 sur le $\log_2(LC_{50})$ et le $\log_2(LC_5)$ dans le cas d'une information *a priori* précise sur L (graphes du haut) et sur le $\log_{10}(LC_{50})$ et le $\log_{10}(LC_5)$ dans le cas d'une information *a priori* peu précise sur L (graphes du bas).

Impact du poids des différentes concentrations c'est-à-dire le nombre d'individus testés par concentration (étape 3)

A l'issue de l'analyse locale, nous avons décidé de tester deux types de plan d'expérience que nous appellerons P1 et P2. Dans le cas d'une information *a priori* précise sur L, nous avons décidé de tester le plan d'expérience de référence "3" en mettant un poids minimum aux concentrations 1 et 5 et un poids plus important sur la concentration centrale (P1) ou des poids quasiment identiques sur les trois concentrations centrales (P2) (Tableau 2.10).

Dans le cas d'une information *a priori* peu précise sur L, nous avons étudié deux plans d'expérience correspondant à P1, P2 mais avec 7 concentrations testées, nous les avons appelé respectivement PL1 et PL2. Les concentrations centrales de ces plans d'expérience ont quasiment les mêmes proportions que celles des plans P1 et P2. Ce sont les poids des concentrations extrêmes initiales qui sont réparties sur 2 concentrations au lieu d'une.

Plans d'expérience	w_1	w_2	w_3	w_4	w_5	w_6	w_7
P1	0.10	0.20	0.40	0.20	0.10		
P2	0.10	0.26	0.28	0.26	0.10		
PL1	0.05	0.05	0.20	0.40	0.20	0.05	0.05
PL2	0.057	0.057	0.26	0.26	0.26	0.057	0.057

Tableau 2.10 – Poids (w) pour chaque concentration des plans d'expérience à comparer dans l'étape 3.

Nous avons représenté les valeurs des 1000 déterminants obtenus pour les plans d'expérience P1, P2 et PL1, PL2 (Figure 2.24). Nous pouvons remarquer, dans un premier temps, que pour un même nombre d'organismes total, le fait de mettre des poids différents aux concentrations en minimisant les poids sur les concentrations extrêmes, permet d'augmenter globalement les valeurs des déterminants par rapport aux plans d'expérience de "référence". Ces résultats sont beaucoup plus nets dans le cas d'une information *a priori* peu précise. Il est difficile de départager les deux plans d'expérience P1 et P2. Par contre, la différence entre les plans d'expérience PL1 et PL2 étudiés dans le cas d'une information *a priori* plus précise est un peu plus importante. PL2 semble permettre d'obtenir des valeurs de déterminant un peu plus élevées que PL1.

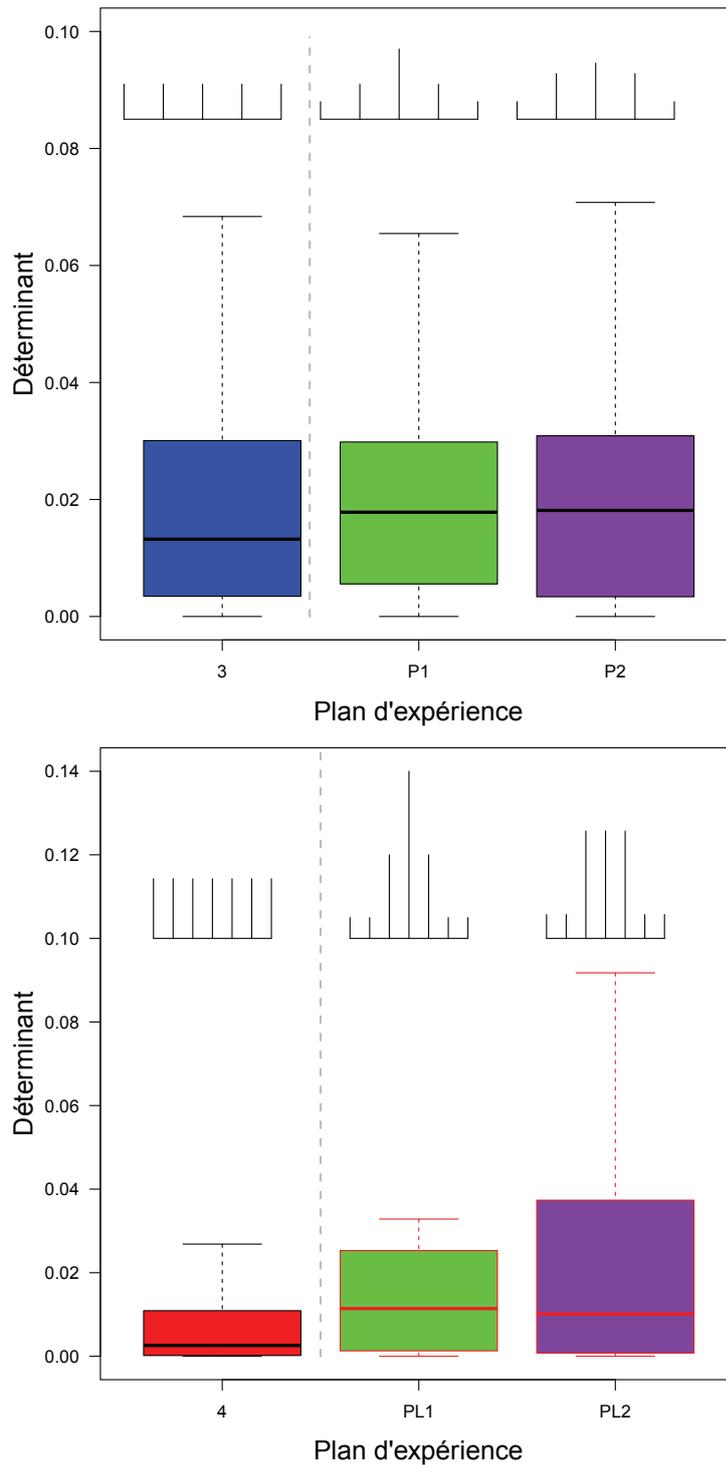


FIGURE 2.24 – Diagrammes en boîte des distributions des 1000 valeurs de déterminant de la matrice de Fisher déterminées globalement en fonction des plans d'expérience avec des poids différents donnés aux concentrations testées dans le cadre d'un *a priori* précis autour de L (graphe du haut) et peu précis autour de L (graphe du bas). En haut du graphique, des petits diagrammes schématisent l'importance relative des poids des différentes concentrations dans le plan d'expérience.

La Figure 2.25 représente les distributions des 1000 gains de connaissance pour chaque plan d'expérience à comparer dans le cadre d'un *a priori* précis. Nous pouvons constater que les plans d'expérience testés avec des poids différents (nombre d'individus différents) sur les concentrations semblent donner de meilleurs gains de connaissance sur le $\log_2(LC_{50})$ que le plan de référence 3 avec le même poids sur toutes les concentrations. Ces résultats vont dans le même sens que ce qui avait été observé avec le calcul du déterminant mais sont plus nets.

De même, dans le cas d'une information *a priori* peu précise, nous pouvons remarquer que les deux plans d'expérience PL1 et PL2 donnent globalement des meilleurs gains de connaissance que le plan d'expérience 4 (Figure 2.25) que ce soit pour le $\log_{10}(LC_{50})$ ou pour le $\log_{10}(LC_5)$. Ainsi, nous pouvons dire qu'à défaut de ne pouvoir réduire l'étendue de la gamme de concentrations testées, il peut être intéressant de mettre des poids minimum sur les concentrations extrêmes. Il est difficile de départager de manière nette les plans P1, P2 et PL1, PL2. La répartition des poids sur les concentrations centrales ne semble pas avoir un impact fort sur les gains de connaissance.

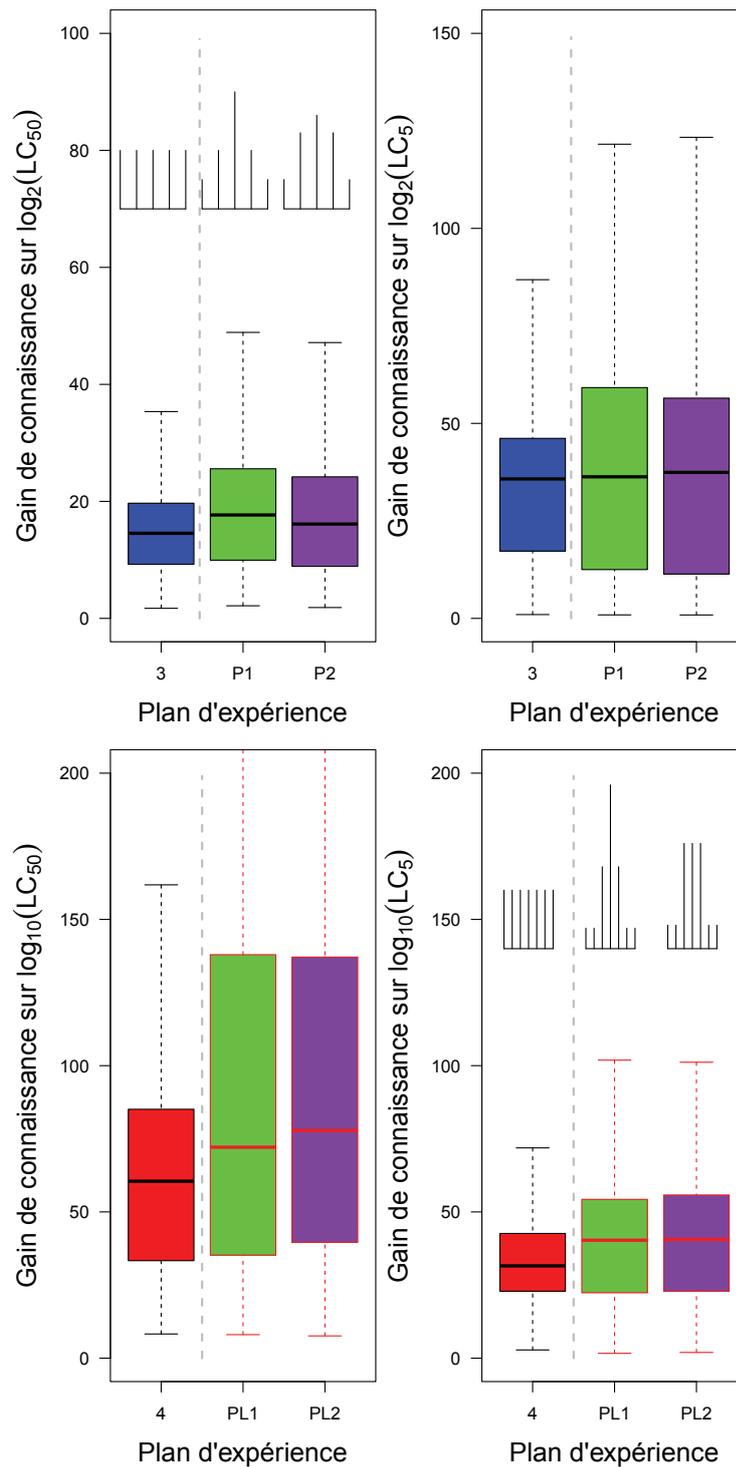


FIGURE 2.25 – Diagrammes en boîte des distributions de 1000 gains de connaissance pour chacun des plans d'expérience testés dans l'étape 3 sur le $\log_2(LC_{50})$ (à gauche) et sur le $\log_2(LC_5)$ (à droite), dans le cas de l'*a priori* précis sur de L (graphes du haut) et sur le $\log_{10}(LC_{50})$ (gauche) et sur le $\log_{10}(LC_5)$ (droite) dans le cas de l'*a priori* peu précis sur L (graphes du bas).

Impact du nombre d'individus exposés total (étape 4)

Dans cette partie, nous avons comparé des plans d'expérience dont les concentrations ont des poids identiques et nous avons regardé l'impact que pouvait avoir le nombre d'organismes (ou individus) exposés total sur le gain de connaissance. Nous avons choisi les plans d'expérience 3 et 4 pour lesquels le nombre d'individus par concentration est égale à 20 auxquels nous avons enlevé ou ajouté 10 individus par concentration. Nous avons ainsi testé en plus du plan d'expérience 3 et 4 les plans appelés :10, 30, 40, 50, 60 et 70 correspondant au nombre d'individus testés par concentration. Nous avons étudié l'impact de nombre d'individus par concentration uniquement sur le gain de connaissance car le déterminant est directement proportionnel à N^2 (Equation 2.21), avec N le nombre total d'organismes étudiés.

Les résultats obtenus sont présentés sur la Figure 2.26 et les plans de référence quelle que soit l'information *a priori* sur la LC_{50} sont ceux avec 20 organismes par concentration. Ces résultats montrent l'impact important que peut avoir le nombre d'individus par concentration sur le gain de connaissance. Plus le nombre d'individus par concentration est important plus le gain de connaissance sera élevé. De plus, l'augmentation du gain de connaissance en fonction du nombre d'individus par concentration semble linéaire et ne semble pas atteindre de "palier" du moins pour les plans d'expérience testés.

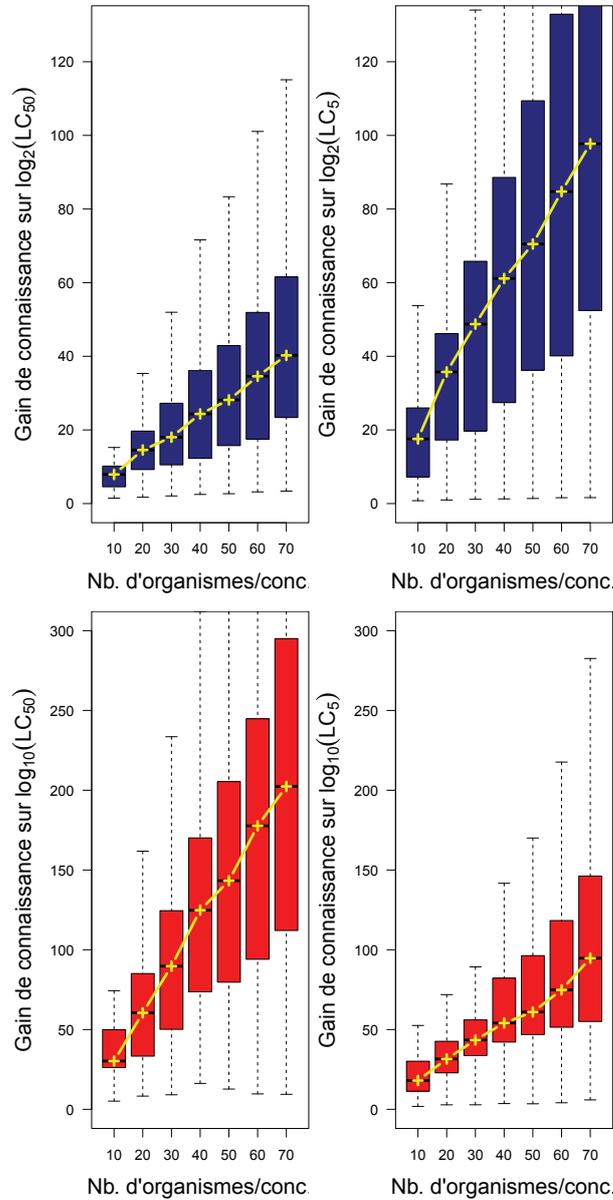


FIGURE 2.26 – Diagrammes en boîte des distributions de 1000 gains de connaissance pour chacun des plans d'expérience testés dans l'étape 4, sur le $\log_2(LC_{50})$ et le $\log_2(LC_5)$ dans le cas d'une information *a priori* précise sur L (graphes du haut) et sur le $\log_{10}(LC_{50})$ et le $\log_{10}(LC_5)$ dans le cas d'une information *a priori* peu précise sur L (graphes du bas). En jaune, sont représentées les valeurs médianes des distributions des 1000 valeurs de gain de connaissance.

2.2.4 Discussion et conclusion sur l'optimisation des plans d'expérience dans le cadre des données de survie

Conclusion sur les méthodes utilisées pour optimiser les plans d'expérience

Dans ce travail, nous avons utilisé deux méthodes de comparaison des plans d'expérience, la première est la D-optimalité et la seconde le gain de connaissance. La D-optimalité, avec calcul du déterminant, nous a permis d'avoir une référence concernant les résultats que nous pouvions obtenir. En effet, cette méthode est largement utilisée en optimisation. Toutefois, nous voulions tester une autre méthode qui est le calcul des gains de connaissance qui, nous pensons, présente certains avantages.

En effet, le calcul du gain de connaissance permet à la fois de :

- s'affranchir du calcul mathématique de la matrice de Fisher du modèle, celle-ci n'étant pas triviale à trouver surtout pour des modèles non classiques. A titre indicatif, nous avons pu réaliser la même étude sur les plans d'expérience dans le cas où le modèle concentration-réponse utilisé est un modèle de Weibull grâce aux gains de connaissance. Les résultats obtenus avec ce modèle de Weibull, concernant l'impact de la position de L dans le cadre d'un *a priori* précis, sont présentés dans la Figure 2.27. Les résultats obtenus nous permettent de tirer les mêmes conclusions que celles avec le modèle log-logistique (Figures 2.20 et 2.25).
- d'avoir une réponse directement et de manière spécifique sur la concentration critique d'effet d'intérêt (LC_5 , LC_{50} ...). Les résultats concernant la position de L dans la gamme de concentrations sont différents pour le gain de connaissance sur la LC_5 par rapport à la LC_{50} . Nous obtenons une information supplémentaire par rapport au calcul du déterminant, qui donne les mêmes informations que pour le gain de connaissance sur la LC_{50} .

Ainsi, nous avons pu vérifier que les résultats donnés par les deux méthodes (D-optimalité et gain de connaissance) concordent. Nous avons également pu obtenir des informations directement quantifiables en terme d'impact et ceci pour la LC_x d'intérêt. En effet, il est plus facile d'appréhender une différence entre deux gains de connaissance, puisqu'un gain de connaissance est un rapport de variances, qu'entre deux valeurs de déterminants.

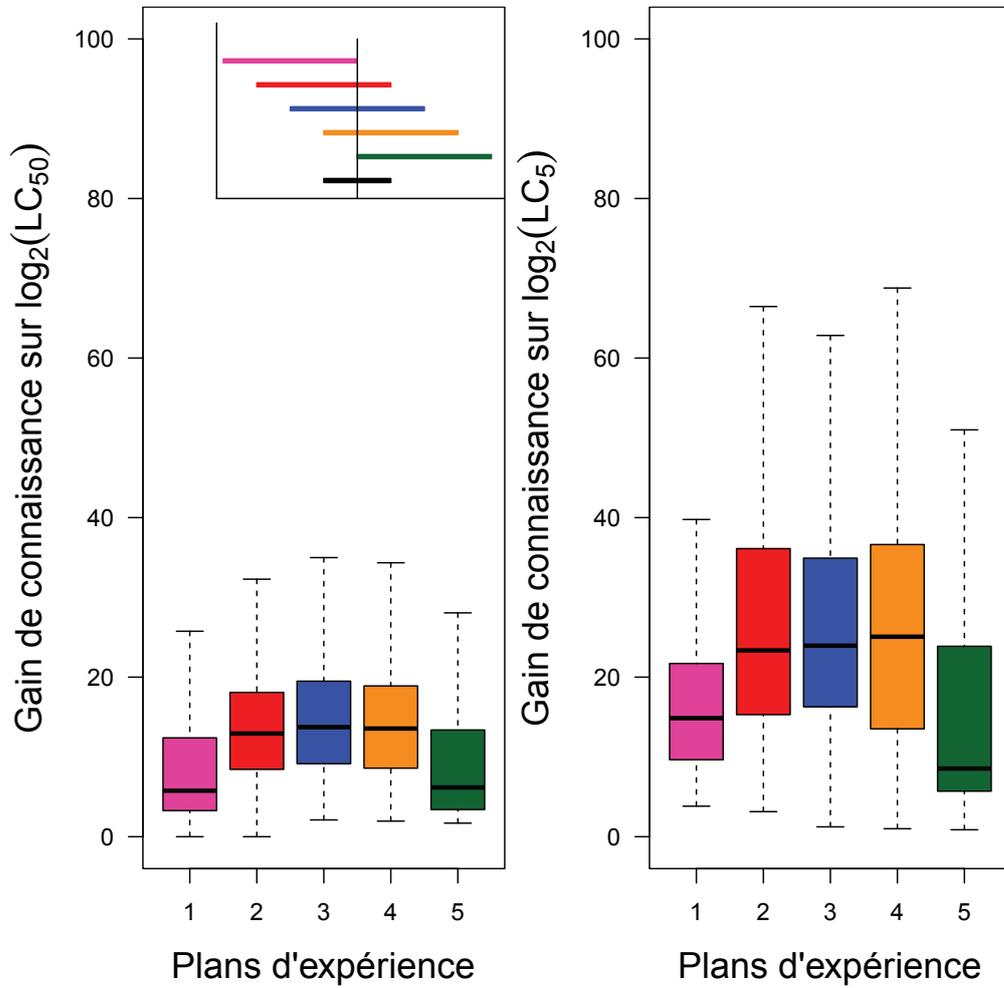


FIGURE 2.27 – Diagrammes en boîte des distributions de 1000 gains de connaissance pour chacun des plans d'expérience testés dans l'étape 1 (position de L dans la gamme) sur le $\log_2(LC_{50})$ et le $\log_2(LC_5)$ dans le cas d'une information *a priori* précise sur L avec comme modèle concentration-réponse le modèle de Weibull.

Impact des composantes des plans d'expérience sur les LC_x

Nous sommes partis des recommandations du guide OCDE concernant les tests en toxicité aiguë et chronique chez *Daphnie magna* (OCDE, 2004, 1998, 2012) et nous avons décliné plusieurs plans en fonction des composantes du plan d'expérience que nous voulions tester. La première conclusion globale de ce travail est qu'à partir des plans d'expérience recommandés, il est possible d'améliorer le gain de connaissance sur les LC_x d'intérêt en modifiant légèrement les protocoles existants. La mise en oeuvre de ces "petites améliorations" ne nécessitent pas d'attendre l'évolution des guides, puisque les plans d'expérience suivent les grandes lignes des recommandations.

Nous avons pu voir dans cette partie, l'intérêt de centrer les plans d'expérience sur les valeurs attendues de LC_{50} et pour cela, l'intérêt d'essayer d'avoir des informations *a priori*. Nous avons également pu constater que plus il y a de concentrations sur une même étendue de gamme (et donc plus la raison entre les concentrations est faible) plus le gain de connaissance sera important. Ceci rejoint les travaux de Moore et Caux (1997) qui recommande d'inclure plus de concentrations peu espacées dans les plans d'expérience pour les modèles de type log-logistique et de Weibull.

Nous n'avons pas choisi les concentrations testées uniquement parmi les concentrations probables pour la LC_{50} , car nous nous sommes placés du point de vue des écotoxicologues. Il est légitime de penser que les écotoxicologues auraient peut être des difficultés à baser leur gamme de concentrations uniquement sur l'information *a priori*. En effet, cette pratique pourrait être difficile à mettre en place d'une part pour des raisons de marge de sécurité par rapport à l'information *a priori* mais également par habitude puisque dans les guides, il est recommandé d'étendre la gamme des concentrations testées afin d'avoir des effets observés allant de 0% à 100%. Ainsi, nous sommes restés dans une démarche assez proche de ce qui se fait en pratique. Toutefois, nos résultats semblent montrer que plus la gamme de concentrations testée est étendue, moins le gain de connaissance sur les concentrations critiques d'effet est bon. Cependant, nous avons vu que des plans d'expérience P1, P2 ou PL1, PL2 (avec des poids minimum sur les concentrations extrêmes) pour une même étendue de gamme sont plus favorables que les plans d'expérience respectifs avec des poids identiques sur chaque concentration. Ainsi, dans le cas où il n'est pas possible de réduire l'étendue de la gamme de concentrations, il semble intéressant de mettre moins de poids (moins d'individus) sur les concentrations extrêmes de la gamme. Il paraît tout de même important d'étudier par la suite, l'impact des composantes des plans d'expérience en réduisant l'étendue des gammes testées à l'étendue des valeurs probables de LC_{50} issues de l'information *a priori*.

Enfin, concernant le nombre total d'individus testés dans l'expérience (à poids égal), plus le nombre d'individus est élevé, meilleur est le gain de connaissance. Mais étant donné que les tests de toxicité vont aller dans le sens d'une réduction du nombre d'organismes exposés, les résultats que nous avons obtenus nous permettent seulement de quantifier l'importance de cette composante du plan d'expérience par rapport aux autres.

D'ailleurs, pour illustrer l'importance relative de ces différentes composantes, nous avons résumé dans le Tableau 2.11 certains résultats de gain de connaissance concernant

des plans d'expérience étudiés. La position de la valeur attendue dans la gamme des concentrations est une composante "gratuite" en terme de coût expérimental et peut augmenter ou diminuer de manière importante le gain de connaissance, d'où l'importance de disposer d'une bonne information *a priori* sur les paramètres. Ainsi, contrairement à ce qui est indiqué dans le guide OCDE (OECD, 2006), l'adaptation du plan d'expérience à l'ajustement d'un modèle concentration-réponse ne nécessite pas obligatoirement un coût élevé, ce qui rejoint les travaux de Moore et Caux (1997). Le choix de la raison associée au nombre de concentrations testées permet aussi d'avoir une influence importante sur le gain de connaissance, mais son coût expérimental peut ne pas être gratuit. En effet, une augmentation du nombre de concentrations (sur une même étendue de gamme) implique de faire des dilutions supplémentaires du composé étudié, ce qui peut être un problème dans certains laboratoires. Le nombre d'organismes total de l'expérience est par définition bien plus coûteux en terme expérimental mais permet d'augmenter le gain de connaissance sans difficulté. Il n'y a donc pas un plan d'expérience optimal mais plusieurs qui vont être dépendants d'une part du laboratoire dans lequel se fait le test de toxicité et d'autre part de la concentration critique d'effet d'intérêt qu'il faut estimer. Ainsi, nous pouvons imaginer, à partir de ce travail, un outil disponible en ligne qui permettrait, en fonction des informations *a priori* fournies par l'écotoxicologue, d'optimiser un plan d'expérience en fonction des contraintes expérimentales (nombre d'organismes maximal, nombre de concentrations maximal), de la concentration critique d'effet à estimer et du modèle envisagé pour ajuster les données.

Enfin, il pourrait être intéressant de tester les plans d'expérience pour un modèle NEC comme développé dans la première partie sur les données de survie. Avec la particularité du modèle à seuil et de la prise en compte des données au cours du temps, il pourrait être intéressant de voir quels types de plans d'expérience seraient les plus adaptés.

Composantes						
Plans d'expérience	Position de L dans la gamme	Raison et Nb. de concentrations	Nb. organismes au total	GC sur $\log_2(LC_{50})$	% par rapport à la référence	
3	3	2 ; 5	120	12.65		
5	1	2 ; 5	120	7.89	- 38%	
3a	5	$\sqrt{2}$; 9	120	15.59	+ 23%	
10	3	2 ; 5	60	7.91	- 37%	
				GC sur $\log_{10}(LC_{50})$		
4	4	3 ; 7	160	60.5		
7	1	3 ; 7	160	16.9	- 72%	
4a	5	2 ; 9	160	79.7	+ 32%	
10	4	3 ; 7	80	30.2	- 50%	

Tableau 2.11 – Quantification de l'impact que peuvent avoir les différentes composantes d'un plan d'expérience sur le gain connaissance (GC) dans le cas d'une information *a priori* précise (première partie du tableau) et dans le cas d'une information *a priori* peu précise (deuxième partie du tableau). Dans chaque partie du tableau, les valeurs des gains de connaissance obtenus pour des plans d'expérience testés ont été rapportées et mises en pourcentage par rapport aux valeurs obtenues pour le plan d'expérience de référence 3 dans le cas d'une information *a priori* précise et pour le plan d'expérience de référence 4 dans le cas d'une information *a priori* peu précise.

Conclusion :

Dans cette deuxième partie sur l'optimisation des plans d'expérience dans le cadre des données de survie, nous avons pu voir :

- que l'utilisation des gains de connaissance sur les concentrations critiques d'effet (LC_x) est une bonne méthode pour comparer les plans d'expérience ;
- qu'il est important de bien récolter les informations *a priori* que l'on peut avoir avant de réaliser le plan d'expérience ;
- qu'il vaut mieux centrer la gamme de concentrations sur la valeur attendue de la LC_{50} même si l'on veut estimer une LC_5 ;
- qu'un nombre plus importants de concentrations pour une même étendue de gamme est préférable (raison de la progression géométrique plus faible) ;
- que lorsque l'étendue de la gamme ne peut être réduite, pour un même nombre total d'individus, un nombre d'individus moins importants sur les concentrations extrêmes est aussi favorable au gain de connaissance sur les concentrations critiques d'effet.

De plus, la mise à disposition d'un outil basé sur les gains de connaissance afin de mettre en place un plan d'expérience pourrait être envisagé. L'écotoxicologue pourrait fournir les informations plus ou moins précises qu'il a sur les concentrations critiques d'effet qu'il veut estimer, ainsi que sa capacité en terme de nombre d'organismes ou de nombre maximal de concentrations qu'il peut tester. Il aurait en réponse un ou plusieurs plans d'expérience permettant de maximiser les gains de connaissance en fonction des capacités indiquées.

Chapitre 3

Les données de reproduction

Sommaire

3.1	Analyse des données de reproduction	87
3.1.1	Introduction	87
3.1.2	Choix de la variable étudiée	89
3.1.3	Choix de la partie stochastique et inférence	94
3.1.4	Choix de la partie déterministe	113
3.1.5	Conclusion sur l'analyse des données de reproduction	118
3.2	Optimisation d'un plan d'expérience en vue d'une standardisation chez la limnée	121
3.2.1	Introduction	121
3.2.2	Réduction de la durée de l'expérience	123
3.2.3	Réduction du nombre d'individus utilisés dans l'expérience	130
3.2.4	Conclusion sur l'optimisation d'un plan d'expérience en vue d'une standardisation chez la limnée.	137

3.1 Analyse des données de reproduction

3.1.1 Introduction

Dans cette partie sur l'étude des données de reproduction, nous avons travaillé sur les jeux de données issus de bioessais de *Lymnaea stagnalis* exposées à du cadmium dans le cadre de la mise en place d'un test standard (voir 1.6). Ainsi, après avoir homogénéisé le plus possible les conditions expérimentales, sept laboratoires ont réalisé des tests basés sur un plan d'expérience dans lequel un nombre d'organismes important a été utilisé. Il s'agit donc d'analyser les données issues de ces tests afin d'estimer de la manière la plus correcte possible les concentrations critiques d'effet.

Quelles que soient les espèces étudiées, les données de reproduction sont toujours du même type : nombre de pontes, nombre d'oeufs, nombre de juvéniles. Ce nombre pouvant être ramené (ou non) au nombre d'individus parents initial (ou à la fin du test). Sauf dans le cas où aucune mortalité n'est observée au cours du test, ce qui s'avère

assez rare pour les tests chroniques, il n'est pas trivial de choisir la variable à prendre en compte. Comme le soulignent Wang et Smith (2000), le fait qu'il n'y ait pas de reproduction ("zéro" reproduction) peut être soit dû à un effet du contaminant sur la reproduction, soit dû à la mort du parent. Dans le guide OCDE sur la reproduction de *Daphnia magna* (OCDE, 2012), il est recommandé de ne pas choisir de concentrations pour lesquelles il y aurait une différence de survie statistiquement significative avec le témoin. De plus, il est indiqué que le but du test est de déterminer l'effet de la substance sur le nombre total de juvéniles vivants produits par animal parent vivant à la fin du test. Si une mortalité chez les parents est observée dans un réplicat alors ce dernier doit donc être exclu de l'analyse. Pour les guides concernant la reproduction du ver Enchytrée (OCDE, 2004) ou des oiseaux (OCDE, 1984) la variable d'intérêt est le nombre de juvéniles produits, la mortalité des parents n'étant pas prise en compte. Néanmoins, il est précisé que les autres signes de nocivité (comme la mortalité) doivent être consignés dans le rapport final. Le test de reproduction sur un acarien prédateur *Hypoaspis (Geolaelaps) aculeifer* (OCDE, 2008) ne prend pas non plus en compte la mortalité des parents et la variable étudiée est le nombre de juvéniles vivants produits pendant l'essai, divisé par le nombre de parents femelles introduits en début d'essai. Souvent, dans les publications concernant la reproduction, il n'est pas précisé si les nombres cumulés de juvéniles (oeufs ou pontes) correspondent à ceux issus des mères encore vivantes à la fin de l'essai ou non (Chèvre et al, 2005; Li, 2010; Kim et al, 2005). Le guide concernant la reproduction des poissons (OCDE, 2009) se distingue un peu des autres du fait qu'il recommande de relever le nombre d'oeufs par femelle survivante et par jour. Ceci se rapproche de ce qui est fait en épidémiologie avec le calcul d'un nombre de personne-temps permettant de prendre en compte la contribution temporelle de chaque personne dans le calcul d'un taux d'incidence d'une maladie par exemple.

Dans ce travail sur la reproduction, il paraît important d'explorer l'impact de différents choix de variables tout en choisissant des outils adéquats dans l'analyse de ces données. Notre étude sur la survie nous a montré l'intérêt d'utiliser la modélisation plutôt que les tests statistiques pour estimer les concentrations critiques d'effet. De plus, s'agissant d'un test standard, ce sont les modèles concentration-réponse qui sont préconisés et donc sur lesquels nous devons travailler. Comme nous l'avons déjà vu dans le chapitre 1, les données de reproduction sont des données de comptage et ne suivent pas une loi normale. Ainsi, dans un premier temps, en nous basant sur les jeux de données issus des pré-tests en vue d'une standardisation, nous étudierons le choix de la variable à analyser, puis nous nous intéresserons à l'ajustement de modèles concentration-réponse sur la variable retenue.

3.1.2 Choix de la variable étudiée

Les jeux de données utilisés sont ceux produits lors de la mise en place d'un test standard sur *Lymnaea stagnalis*. Sept laboratoires ont réalisé des tests de toxicité du cadmium et ont étudié la reproduction tous les deux ou trois jours pendant 56 jours dans des conditions les plus homogènes possibles. Pour chaque concentration et le témoin, six réplicats de cinq individus ont été exposés sauf pour le témoin du laboratoire 3 où on avait uniquement quatre réplicats de cinq individus. Pour cinq jeux de données (laboratoires 1, 2, 3, 6 et 7), cinq concentrations en plus du témoin ont été testées et pour deux jeux de données (laboratoires 4 et 5), 7 concentrations en plus du témoin ont été testées (Tableau 3.1). Dans ce test reproduction, le nombre de morts, le nombre de pontes par réplikat ainsi que le nombre d'oeufs par réplikat ont été relevés tous les deux à trois jours.

Concentrations nominales ($\mu\text{g L}^{-1}$)	53	80	120	180	270	405	608
Moyenne de 3 concentrations mesurées ($\mu\text{g L}^{-1}$)							
Laboratoire 1		53	78	124	232	284	
Laboratoire 2		46	67	95	137	244	
Laboratoire 3		55	88	122	168	307	
Laboratoire 4	23	38	58	85	123	220	319
Laboratoire 5	41	59	88	131	197	291	486
Laboratoire 6		50	109	164	229	297	
Laboratoire 7		58	93	126	211	284	

Tableau 3.1 – Concentrations testées en cadmium en valeurs nominales et mesurées pour chaque laboratoire.

En ce qui concerne les bioessais réalisés sur la limnée, deux mesures concernant la reproduction sont effectuées : le nombre de pontes et le nombre d'oeufs. Cependant différentes variables peuvent être étudiées et comparées en fonction de la méthode utilisée pour prendre en compte ou non la mortalité des parents.

Les différentes méthodes sont :

- Méthode 1 : Cette méthode consiste à calculer le nombre de pontes ou d'oeufs cumulés à la fin du test divisé par le nombre de parents au début du test. Il s'agit uniquement du nombre de pontes ou d'oeufs cumulés qui est comparé entre le témoin et les différentes concentrations testées. L'effet observé correspond à la fois à l'effet sur la reproduction et sur la survie dans le cas où il y aurait de la mortalité. Il n'est pas possible de différencier les deux effets avec cette variable.
- Méthode 2 : Cette méthode, parfois pratiquée mais préconisée par aucun guide, consiste à diviser le nombre de pontes ou d'oeufs cumulés à la fin du test par le nombre de parents en vie à la fin du test. Ceci n'est raisonnable que si la mortalité est faible ou observée dès le début du test. En effet, le problème avec le calcul de cette variable, c'est que l'on attribue des produits de la reproduction de parents morts à des parents vivants à la fin du test. Il y a donc une surestimation du

nombre de pontes ou d'oeufs par parent vivant à la fin du test et donc une sous estimation de l'effet réel de la substance sur la reproduction.

- Méthode 3 : Cette méthode consiste à calculer le nombre de pontes ou d'oeufs cumulés à la fin du test provenant uniquement de parents en vie à la fin du test. Seuls les produits de la reproduction des parents en vie jusqu'à la fin du test, sont pris en compte. L'effet de la substance étudiée sur la reproduction seule est ainsi observée, puisque les réplicats pour lesquels on a observé de la mortalité sont retirés. Le problème est que le nombre de réplicats peut être ainsi fortement diminué.
- Méthode 4 : Cette méthode consiste à calculer le nombre de pontes ou d'oeufs cumulés à la fin du test divisé par le nombre de parent-jours ou individu-jours. Par analogie avec ce qui est fait en épidémiologie, le calcul d'un nombre d'individu-jours pourrait permettre de prendre en compte la contribution de chaque individu dans le test en fonction de sa durée de vie. L'effet observé est donc lié uniquement à la reproduction tout en conservant un maximum d'information sur les réplicats. Exemple du calcul d'un nombre de pontes par individu-jour : A $t=3$ jours on a 10 mères vivantes et à $t=7$ jours on a 6 mères vivantes. A $t=7$ jours, le nombre d'individu-jours est : $6 \times 7 + (10 - 6) \times (\frac{7+3}{2}) = 62$. Nous supposons que les individus morts entre 3 et 7 jours sont en moyenne morts à $\frac{7+3}{2}$ jours.

La survie des limnées en fonction de la concentration mesurée en cadmium aux 28^e et 56^e jours d'exposition et pour les différents jeux de données est représentée sur la Figure 3.1.

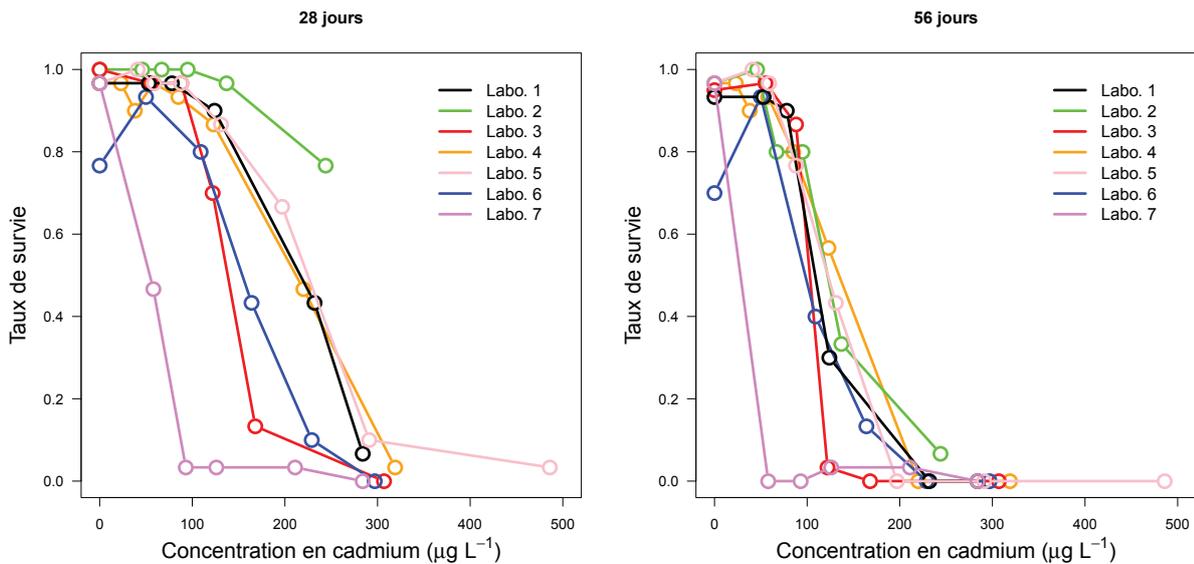


FIGURE 3.1 – Taux de survie en fonction des différentes concentrations en cadmium pour les différents jeux de données à 28 jours et à 56 jours (temps final).

Une mortalité est observée pour toutes les concentrations et tous les jeux de données et pas seulement à la durée finale de l'expérience, mais également à 28 jours et ceci à partir de la deuxième concentration testée. Il est fréquent de trouver des jeux de données de reproduction dans lesquels une mortalité est observée, il est parfois difficile d'étudier les effets sub-létaux sans avoir d'effet létaux. Par définition, les méthodes 1 et 2 ne peuvent pas convenir dans le cas où une mortalité assez importante est observée durant le bioessai. Pour illustrer les inconvénients des méthodes 1 et 2 et comparer les différentes méthodes, les données issues du laboratoire 1 (témoin et concentration égale à 124 $\mu\text{g/L}$) ont été utilisées (Tableau 3.2). Pour chaque méthode, les variables associées et le pourcentage de réduction de cette variable entre le témoin et la concentration testée ont été calculés (Tableau 3.3).

- La méthode 1 ne prend pas en compte du tout la mortalité et donne une réduction de 69% du nombre d'oeufs cumulés entre le témoin et la concentration testée. Ce résultat additionne l'effet sur la mortalité et l'effet sur la reproduction.
- La méthode 2 quant à elle, ne donne une réduction que de 3.6% de la variable étudiée puisqu'elle attribue le nombre cumulé d'oeufs issus des individus morts au cours du test, aux individus vivants à la fin du test.
- La méthode 3 ne permet pas de calculer un pourcentage de réduction de la variable étudiée entre le témoin et la concentration testée, puisqu'aucun réplicat de la concentration testée ne peut être retenu (mortalité dans chacun des réplicats). Avec la méthode 3, les 10 680 oeufs comptés au cours du temps, l'ont été pour rien car ils ne sont pas pris en compte dans l'analyse finale. De plus, nous ne pouvons tirer aucune information de l'effet du cadmium à cette concentration sur la reproduction.
- La méthode 4 quant à elle, prend en compte uniquement l'effet sur la reproduction sans gaspiller de l'information et donne une réduction de 63% de la variable étudiée entre le témoin et la concentration.

	Témoin	124 $\mu\text{g/L}$
Nombre d'oeufs cumulés à la fin du test (N)	34 461	10 680
Nombre d'individus au début du test (nd)	30	30
Nombre d'individus vivants à la fin du test (nf)	28	9
Nombre d'individu-jours à la fin du test (NID)	1 609.5	1 347.5
Nombre de réplicats avec mortalité sur un total de 6	0	6

Tableau 3.2 – Données du laboratoire 1 pour le témoin et la concentration en cadmium égale à 124 $\mu\text{g L}^{-1}$

Suite à la comparaison de ces différentes méthodes, la variable "nombre de pontes ou d'oeufs par individu-jour" (méthode 4) est retenue. Ainsi, cette variable est représentée en fonction de la concentration en cadmium pour les jeux de données des sept laboratoires (Figure 3.2). La même échelle a été utilisée pour représenter le nombre de pontes par individu-jour pour chaque laboratoire, de même pour le nombre d'oeufs par

Méthode	Variable étudiée	Témoin (t56)	124 $\mu\text{g/L}$ (mesure56)	% Réduction $(1-\text{mesure56}/t56)*100$
1	N ou $\frac{N}{nd}$	$\frac{34461}{30} = 1148.7$	$\frac{10680}{30} = 356$	69%
2	$\frac{N}{nf}$	$\frac{34461}{28} = 1230.75$	$\frac{10680}{9} = 1186.7$	3.6 %
3	N	$\frac{34461}{30} = 1148.7$	Pas de données	Non défini
4	$\frac{N}{NID}$	$\frac{34461}{1609.5} = 21.41$	$\frac{10680}{1347.5} = 7.93$	63%

Tableau 3.3 – Comparaison des différentes variables liées aux méthodes de prise en compte (ou non) de la mortalité.

individu-jour. Pour tous les témoins, les nombres de pontes ou d'oeufs par individu-jour varient entre les laboratoires. Des critères comme la dureté de l'eau peuvent expliquer en partie ces différences, mais nous n'aborderons pas le problème ici, notre objectif étant de déterminer la meilleure méthode d'analyse des données de reproduction.

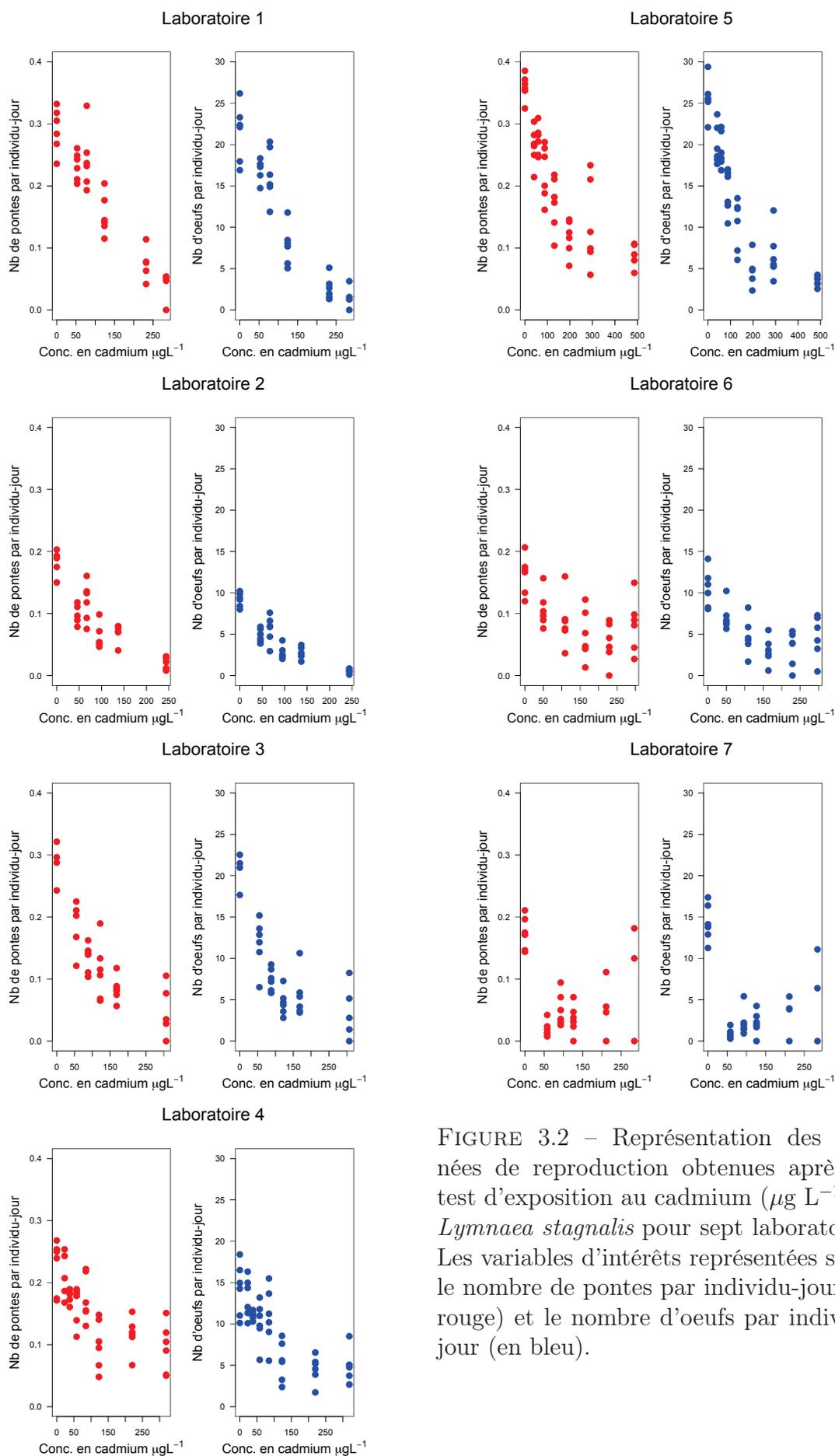


FIGURE 3.2 – Représentation des données de reproduction obtenues après un test d'exposition au cadmium ($\mu\text{g L}^{-1}$) de *Lymnaea stagnalis* pour sept laboratoires. Les variables d'intérêts représentées sont : le nombre de pontes par individu-jour (en rouge) et le nombre d'oeufs par individu-jour (en bleu).

Pour analyser les données nous avons choisi d'ajuster un modèle concentration-réponse. Rappelons qu'un modèle est constitué de deux parties : une partie déterministe et une partie stochastique (voir 1.4.2). Comme les données de reproduction sont des données de comptage, il est opportun de s'intéresser en premier lieu à la partie stochastique du modèle. C'est pourquoi, dans un premier temps nous étudierons la partie stochastique pour une partie déterministe donnée, puis dans un second temps, nous étudierons l'impact du choix de la partie déterministe pour une partie stochastique fixée.

3.1.3 Choix de la partie stochastique et inférence

Afin de comparer les plusieurs parties stochastiques, il est nécessaire de fixer au préalable la partie déterministe. Le modèle log-logistique à trois paramètres a été choisi dans un premier temps pour comparer les différentes parties stochastiques, puisqu'il est très classiquement utilisé comme modèle concentration-réponse.

Les différentes parties stochastiques que nous proposons d'étudier sont :

– **Loi de Poisson**

Les données observées sont des comptages que l'on divise par la suite par un nombre d'individu-jours (NIJ). Les écarts entre les données observées sont d'autant plus grands que la moyenne est grande. Il est classique, pour des données de comptages d'utiliser une loi de Poisson de paramètre λ (de moyenne égale à la variance égale à λ) pour décrire la distribution des données observées par rapport à la courbe théorique (Bailer and Oris, 1997; Zhang et al, 2012) :

$$f(\text{conc}_i) = d \times \frac{EC_{50}^b}{EC_{50}^b + \text{conc}_i^b} \quad (3.1)$$

et

$$N_{i,j} \sim \text{Poisson}(\lambda_{i,j}) \quad (3.2)$$

avec

$$\lambda_{i,j} = f(\text{conc}_i) \times NIJ_{i,j} \quad (3.3)$$

Notons $f(\text{conc}_i)$, la fonction définissant le nombre de pontes ou d'oeufs par individu-jour en fonction de la concentration conc_i . $N_{i,j}$ est le nombre total de pontes ou d'oeufs observés et $NIJ_{i,j}$ le nombre d'individu-jour pour une concentration conc_i et un répliquat j . Cependant, dans certains cas, il est possible d'observer une sur-dispersion des données observées par rapport aux données prédites par le modèles de Poisson.

– **Loi binomiale négative**

Une telle sur-dispersion est souvent décrite par une loi binomiale négative à deux paramètres π et r :

La moyenne d'une loi binomiale négative est égale à $\frac{r(1-\pi)}{\pi}$;

et sa variance à : $\frac{r(1-\pi)}{\pi^2}$

Nous considérons que la moyenne de cette loi, définie via la fonction f , est égale, comme dans le cas de la loi de Poisson, à $f(\text{conc}_i) \times NIJ_{ij}$ et nous prenons π comme paramètre fixe décrivant la dispersion autour de la moyenne. Nous pouvons alors définir r_{ij} par :

$$NIJ_{i,j} \times f(\text{conc}_i) = \frac{r_{ij}(1 - \pi)}{\pi}$$

$$r_{ij} = NIJ_{i,j} \times f(\text{conc}_i) \times \frac{\pi}{1 - \pi} \quad (3.4)$$

et nous pouvons décrire N_{ij} , le nombre total de pontes ou d'oeufs observés par :

$$N_{ij} \sim BN(\pi, r_{ij}) \quad (3.5)$$

- **Loi gamma-Poisson** La loi binomiale négative peut être considérée comme le mélange d'une loi gamma et d'une loi de Poisson. Elle est ainsi parfois appelée la loi gamma-Poisson. En effet, on peut obtenir cette loi en considérant que la variable étudiée suit une loi de Poisson dont la moyenne est elle-même variable suivant une loi gamma. Mais dans la cas qui nous intéresse, il est possible de définir une loi gamma-Poisson un peu différemment, en attribuant cette variabilité de la moyenne de la loi de Poisson directement à la variabilité entre les réplicats. Nous écrivons ainsi :

$$N_{ij} \sim P(\lambda_{ij}) \quad (3.6)$$

avec

$$\lambda_{ij} = F_{i,j} \times (NIJ_{i,j}) \quad (3.7)$$

En supposant F_{ij} variable entre les réplicats, suivant une loi gamma de moyenne $f(\text{conc}_i)$ et de dispersion caractérisée par le paramètre β . Sachant que la loi gamma(α, β) a pour moyenne ($\frac{\alpha}{\beta}$) et pour variance ($\frac{\alpha}{\beta^2}$), on peut écrire :

$$F_{i,j} \sim \text{Gamma}(\alpha_i, \beta) \quad (3.8)$$

$$\alpha_i = f(\text{conc}_i) \times \beta \quad (3.9)$$

Pour contourner le problème, certaines pratiques comme les transformations en racine carrée ou en logarithme sont parfois aussi rencontrées. Toutefois, elles ne sont pas conseillées pour les données de comptages (O'Hara and Kotze, 2010). En effet, au voisinage de zéro, elles peuvent conduire à des prédictions irréalistes comme des nombres d'oeufs négatifs. Nous ne les envisagerons donc pas.

Nous avons choisi d'utiliser l'inférence bayésienne car sa souplesse permet d'utiliser des modèles avec des parties stochastiques non traditionnelles (voir 1.6). Pour mettre en

oeuvre l'inférence bayésienne et ajuster les modèles, nous avons récolté des informations *a priori* sur les paramètres (d , EC_{50}) et sur le nombre de pontes et le nombre d'oeufs de *Lymnaea stagnalis* en conditions normales. A dire d'experts, il semblerait que la ponte ait lieu environ tous les deux à trois jours et que dans chaque ponte, il y ait entre 40 à 120 oeufs. A partir de ces informations, il a été possible de définir un prior sur le paramètre d du modèle log-logistique dans le cas où c'est le nombre de pontes par individu-jour qui est la variable, en considérant que le nombre de pontes varie de zéro à une ponte par individu-jour : $d = ponte0 \sim Uniforme(0, 1)$. Dans le cas où c'est le nombre d'oeufs par individu-jour qui est étudié, en considérant que le nombre d'oeufs par individu-jour varie de zéro à 120 : $d = oeuf0 \sim Uniforme(0, 120)$. Une loi uniforme a également été choisie pour décrire l'information *a priori* sur la EC_{50} en supposant qu'elle se situait dans la gamme des concentrations testées.

Le Tableau 3.4 résume les liens entre les noeuds du modèle. Les diagrammes acycliques dirigés des modèles avec comme partie stochastique une loi de Poisson et une loi de gamma-Poisson sont représentés sur la Figure 3.3.

Pour chaque jeu de données et chaque modèle, après une période initiale de "chauffe" d'environ 20 000 itérations, entre 20 000 et 50 000 itérations ont été utilisées pour faire converger l'algorithme bayésien. Généralement, la convergence a été vérifiée à la fois en observant les traces des chaînes MCMC et en analysant les statistiques de Gelman et Rubin [Gelman et Rubin \(1992\)](#). Pour chaque ajustement effectué en utilisant l'inférence bayésienne, nous avons vérifié que la distribution des informations *a posteriori* n'était pas trop contrainte par celle des *a priori* en superposant graphiquement les deux distributions.

Modèles	Noeud	Definition
Partie déterministe		
log-logistique	$f(conc_i)$	$f(conc_i) = \frac{d}{1 + (\frac{LC_{50}}{conc_i})^b}$
Parties stochastiques		
Poisson	N_{ij}	$N_{ij} \sim P(f(conc_i) \times NIJ_{i,j})$
Binomiale négative	N_{ij}	$N_{ij} \sim NB(\pi, r_{ij})$ $\log_{10}\pi \sim Uniforme(-5, 1)$ $r_{ij} = NIJ_{i,j} \times f(conc_i) \times \frac{\pi}{1-\pi}$
Gamma-Poisson	N_{ij}	$N_{ij} \sim P(F_{i,j}, NIJ_{i,j})$ $F_{i,j} \sim Gamma(f(conc_i), \beta)$ $\log_{10}\beta \sim uniforme(-3, 3)$

Tableau 3.4 – Définition des liens du modèle log-logistique avec différentes parties stochastiques.

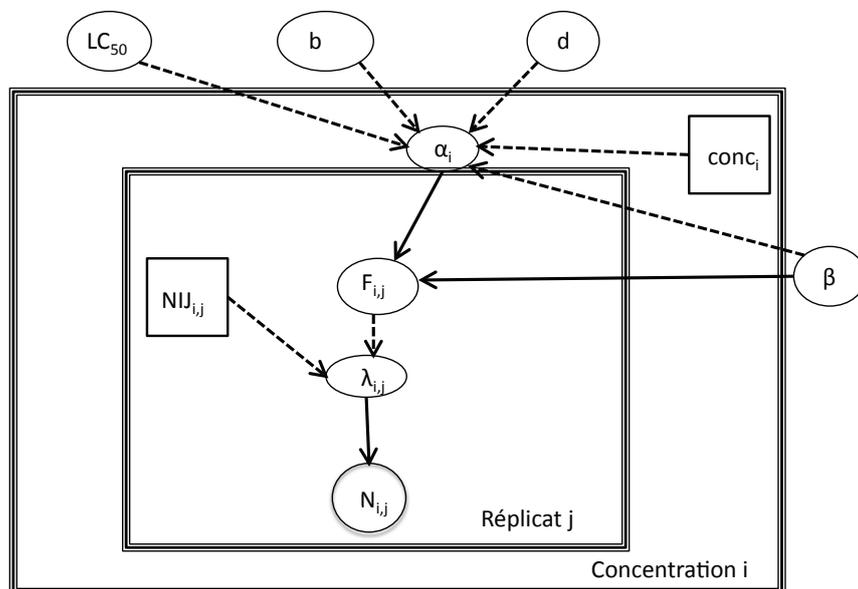
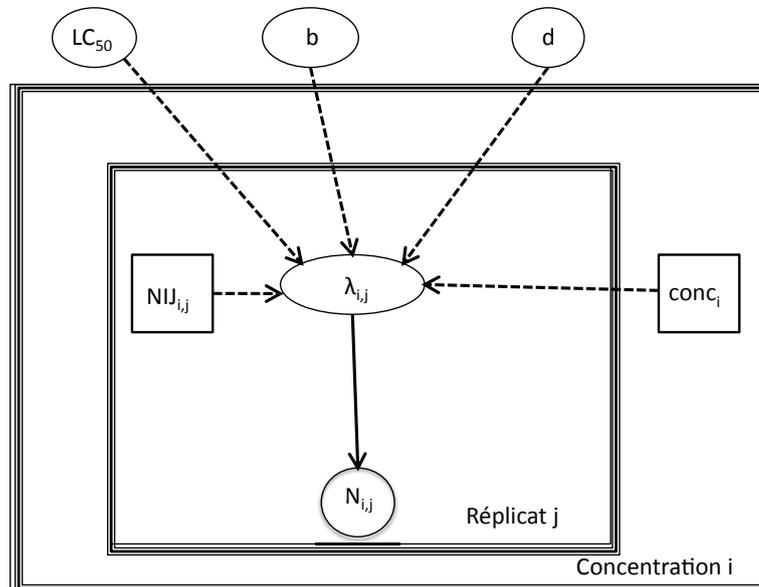


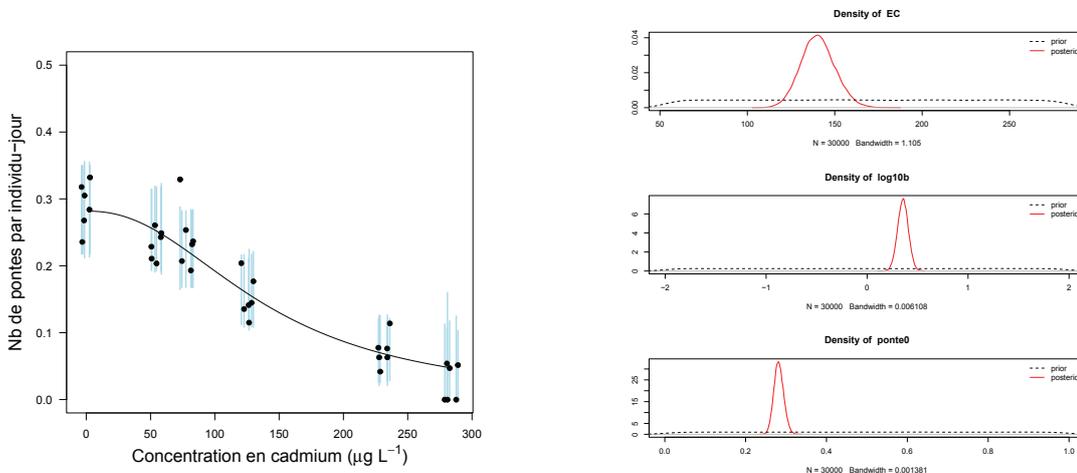
FIGURE 3.3 – Diagrammes Acycliques Dirigés (DAG) du modèle concentration-réponse avec comme partie stochastique une loi de Poisson (DAG du haut) et une loi gamma-Poisson (DAG du bas). Les rectangles indiquent les covariables et les ellipses indiquent les paramètres ou variables latentes ou variables observées. Les flèches en pointillés représentent les liens logiques entre les différents noeuds et les flèches pleines représentent les liens stochastiques. $N_{i,j}$ représente le nombre total de pontes ou d’oeufs et $NIJ_{i,j}$ le nombre d’individu-jours pour une exposition à une concentration $conc_i$ pour un réplicat j .

Les résultats des ajustements du modèle log-logistique avec les différentes parties stochastiques pour les cinq laboratoires sont présentés sur les Figures 3.4, 3.5, 3.6, 3.7 et 3.8 pour le nombre de pontes par individu-jour. En effet, seules les données de cinq laboratoires ont été ajustées, les données des laboratoires 6 et 7 n'ont pas pu être ajustées avec le modèle concentration-réponse. Ces deux laboratoires avaient comme particularité d'avoir eu une mortalité importante chez les témoins (laboratoire 6) et une mortalité quasiment totale dès la première concentration testée pour le laboratoire 7 (Figure 3.1).

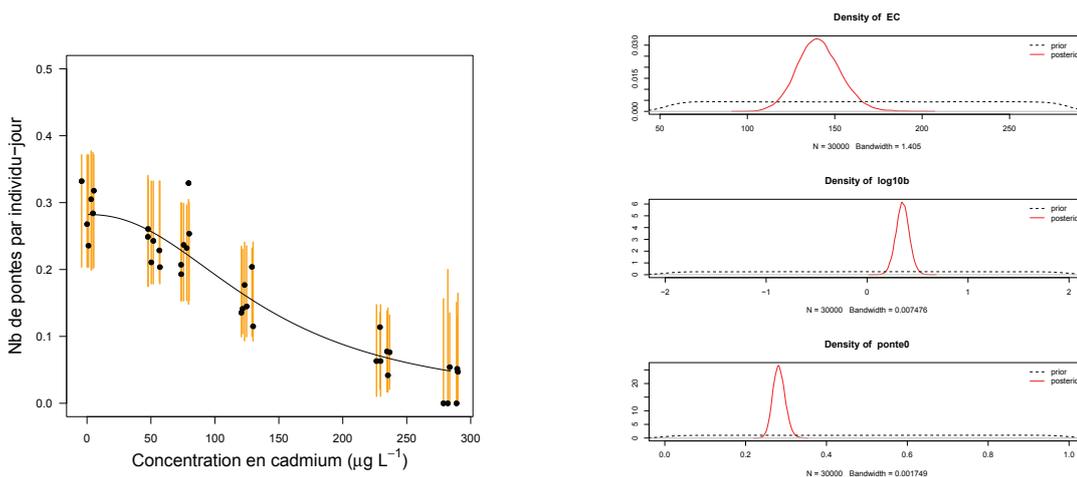
Sur les graphiques de gauche, les données sont représentées ainsi que l'ajustement du modèle log-logistique avec comme partie stochastique : une loi de Poisson (graphique du haut), une loi binomiale négative (graphique du milieu) et une loi gamma-Poisson (graphique du bas). Les données sont présentées pour des concentrations légèrement décalées pour des raisons visuelles, afin de pouvoir mieux apprécier les différents intervalles de crédibilité à 95 % (notés IC) des prédictions associées à chaque concentration testée. En effet, la partie stochastique du modèle porte uniquement sur le nombre de pontes ou d'oeufs et donc les prédictions doivent être calculées à l'aide du nombre d'individu-jours différents pour chaque réplicat. Sur les graphiques de droite sont présentés les distributions *a posteriori* (en rouge) par rapport aux distributions *a priori* (en pointillés noirs) des principaux paramètres d'intérêt du modèle déterministe : la EC_{50} (notée EC), le logarithme décimal de b (noté $\log_{10}b$) et $d = ponte0$ ou $d = oeuf0$ (en fonction de la variable étudiée). Les résultats obtenus pour les différents laboratoires ne sont pas présentés avec la même échelle, l'essentiel ici étant de voir le mieux possible les données et les prédictions.

Globalement, pour le nombre de pontes par individu-jour quelle que soit la partie stochastique, le modèle log-logistique donne d'assez bonnes prédictions pour l'ensemble des laboratoires. Les distributions *a priori* ne semblent pas contraindre l'estimation des paramètres, c'est-à-dire que les distributions *a posteriori* se distinguent nettement de celles *a priori*. Pour le laboratoire 4 (lois binomiale négative et gamma-Poisson), les estimations obtenues sur la EC_{50} ont une précision moins bonne que celles obtenues pour les autres laboratoires. En effet, les nombres de pontes par individu-jour à la cinquième concentration ne semblent pas différent de ceux de la septième concentration d'exposition. Nous avons calculé des intervalles de crédibilité à 95% sur les valeurs prédites, nous nous attendons donc à ce qu'ils contiennent 95% des observations. Même si les trois modèles donnent des résultats qui semblent globalement satisfaisants, le modèle log-logistique avec comme partie stochastique une loi gamma-Poisson donne des prédictions plus larges que les deux autres modèles. Ainsi, l'ensemble des données des laboratoires 1, 2 et 3 sont comprises dans les prédictions obtenues avec ce modèle, ce qui n'est pas le cas pour les autres modèles (loi de Poisson et loi binomiale négative).

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

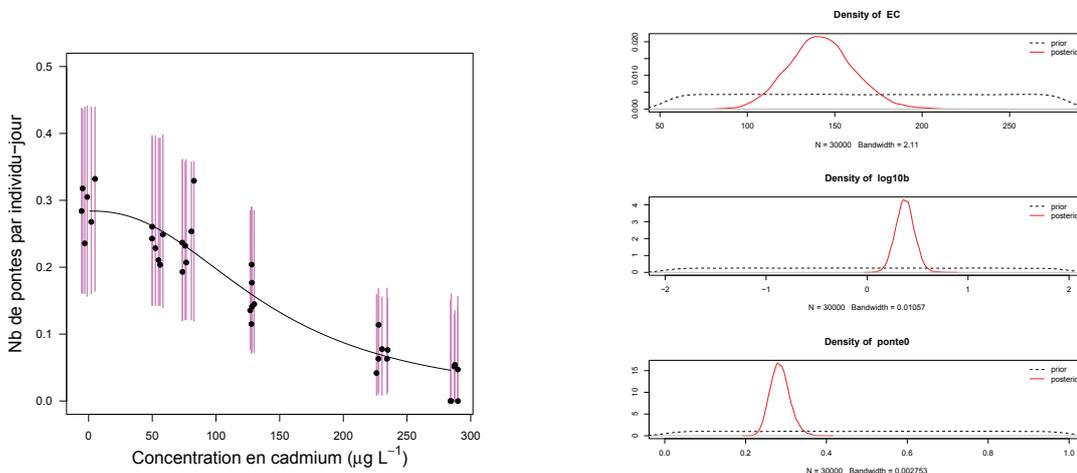
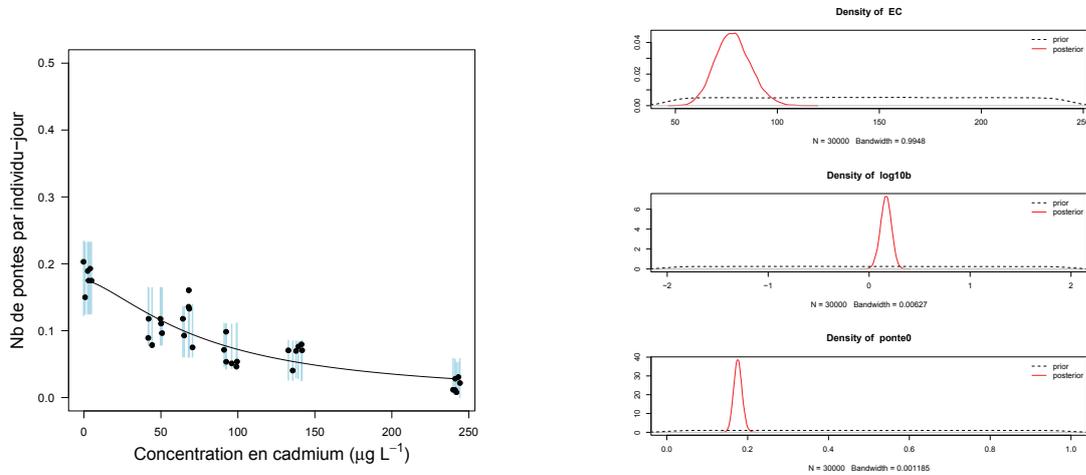
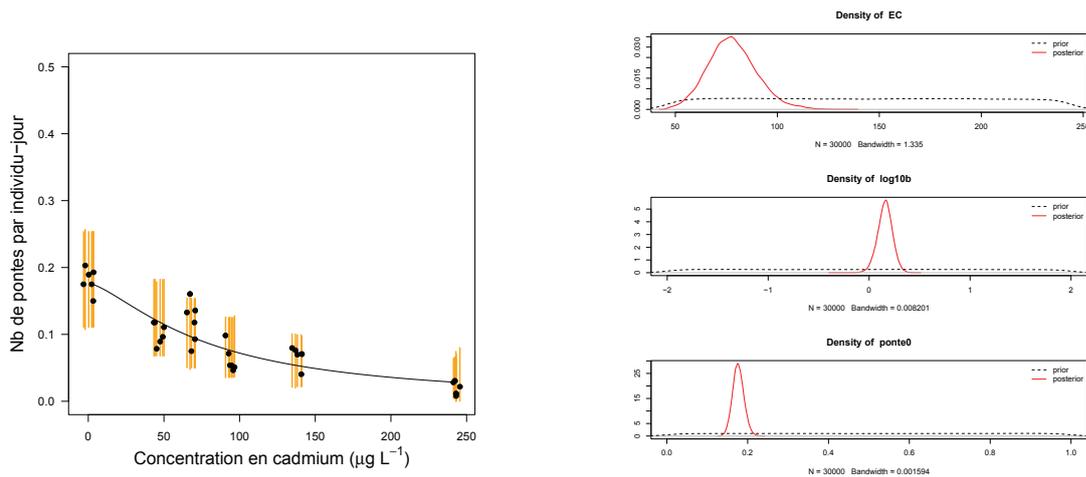


FIGURE 3.4 – Laboratoire 1 (pontes) : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

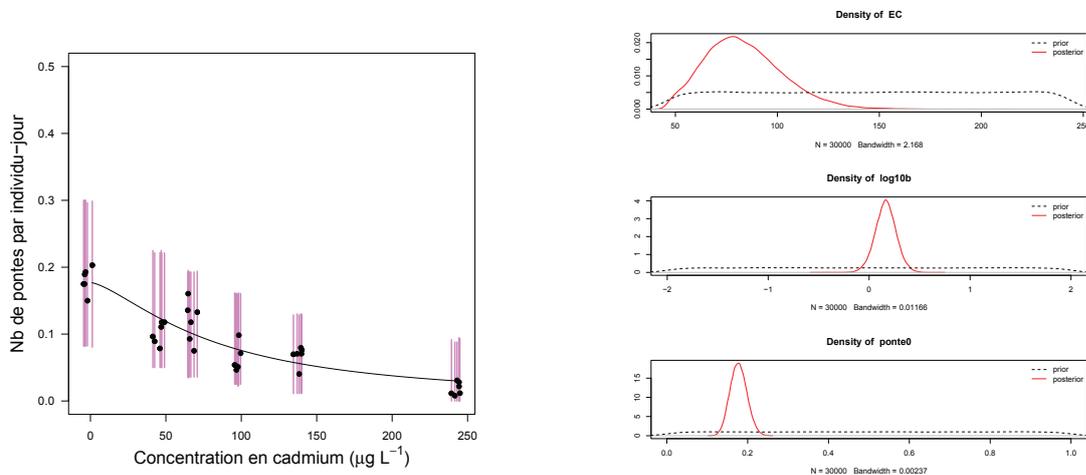
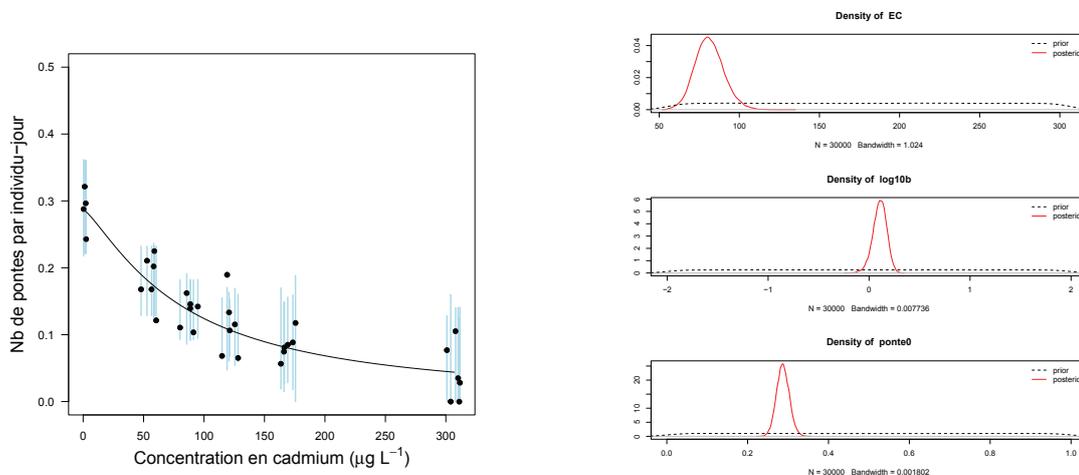
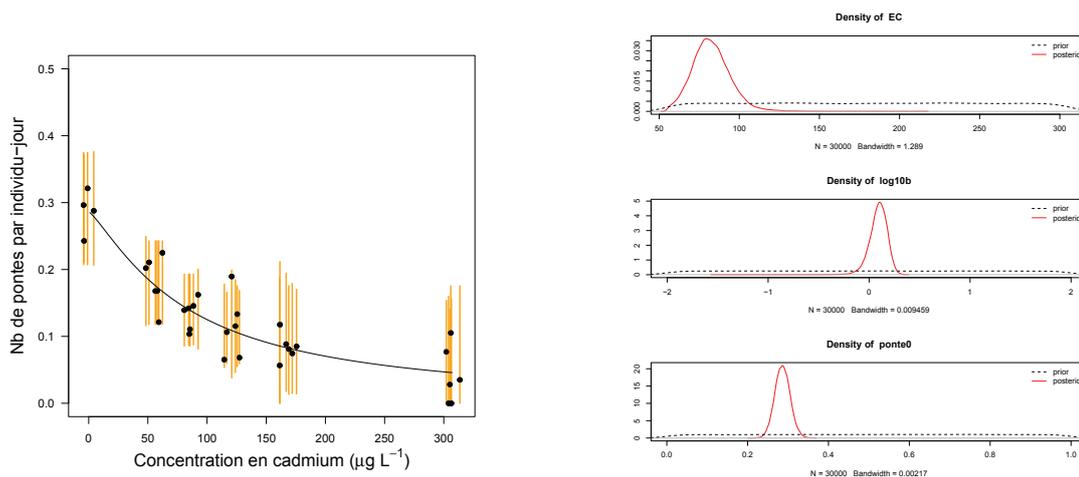


FIGURE 3.5 – Laboratoire 2 (pontes) : à gauche, courbe moyenne et IC à 95% après ajustement d’un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

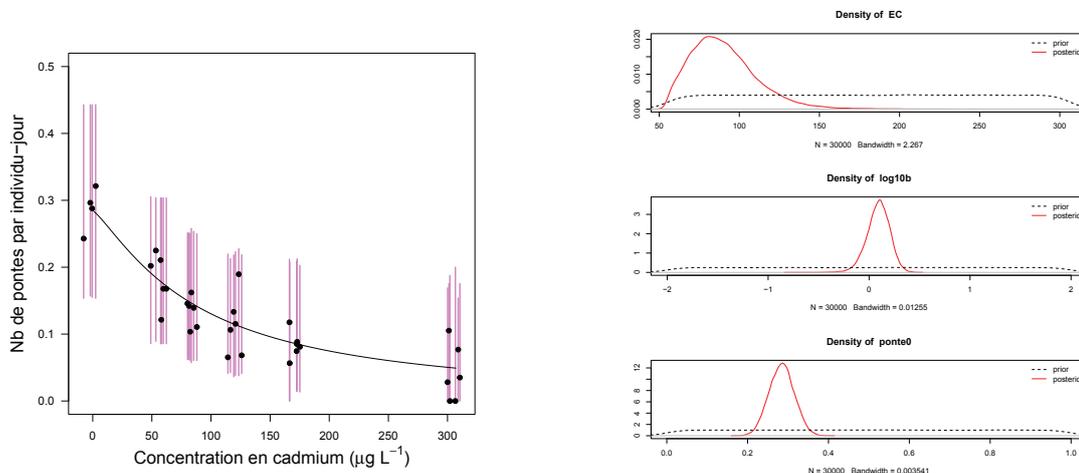
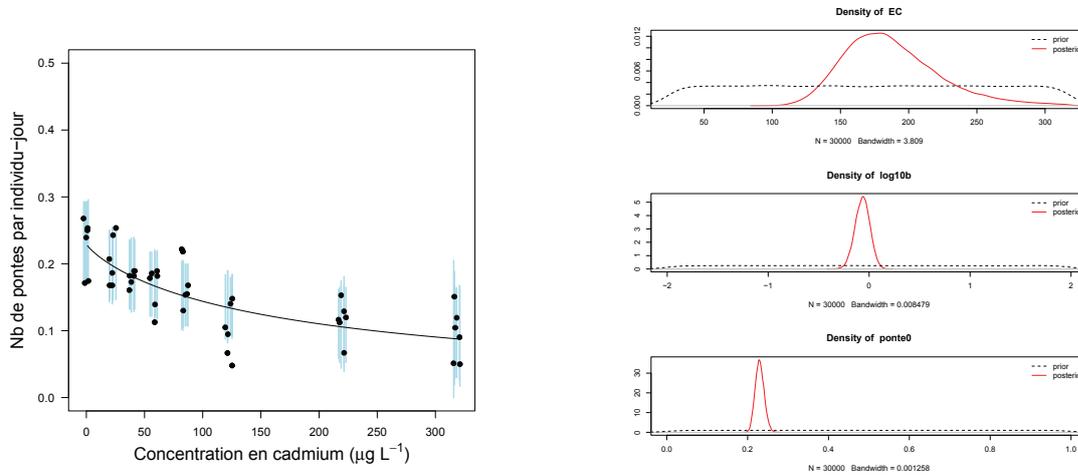
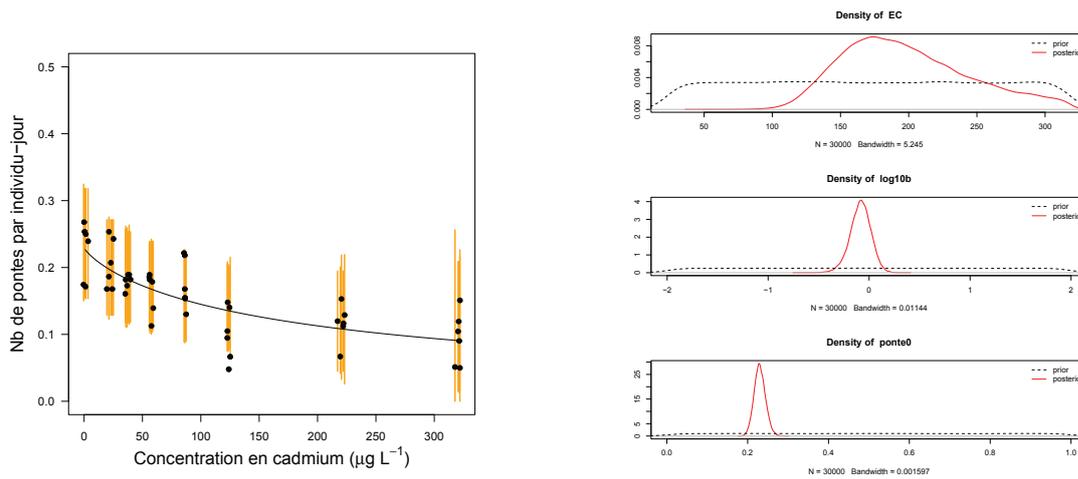


FIGURE 3.6 – **Laboratoire 3 (pontes)** : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

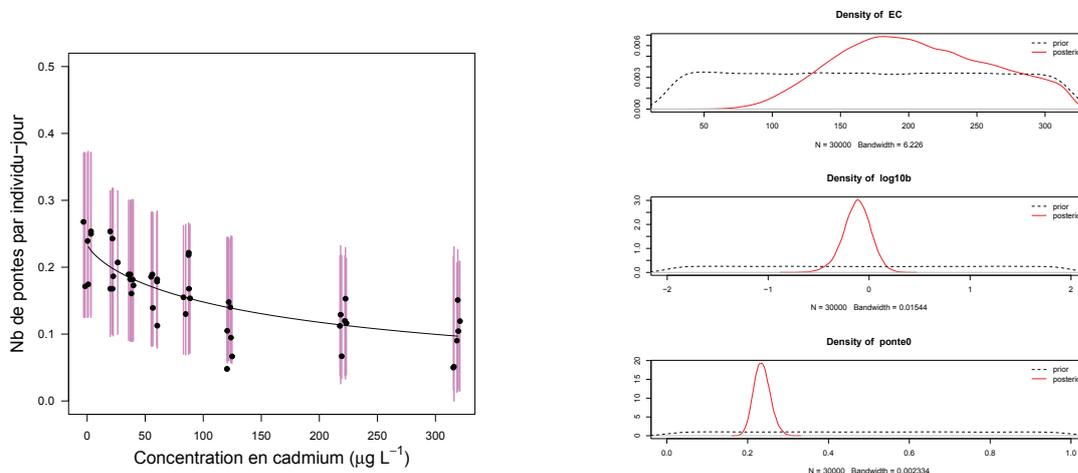
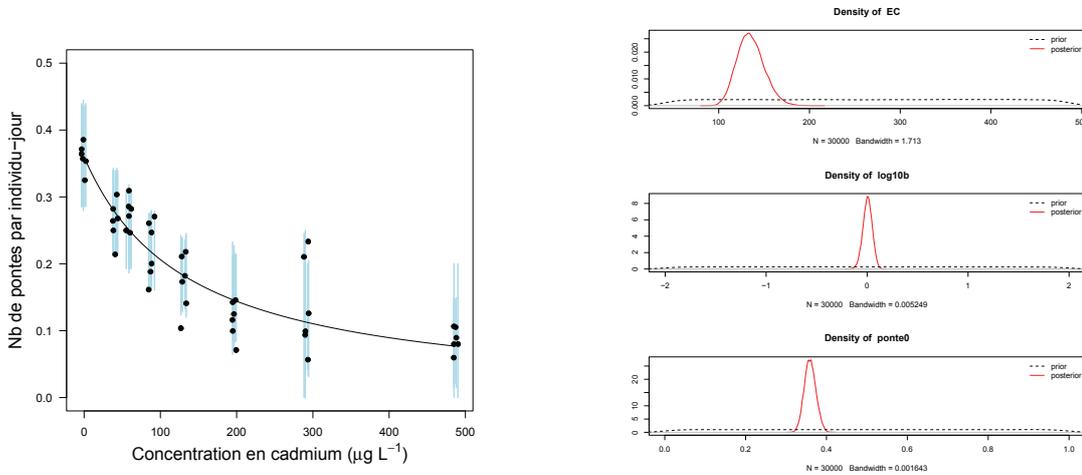
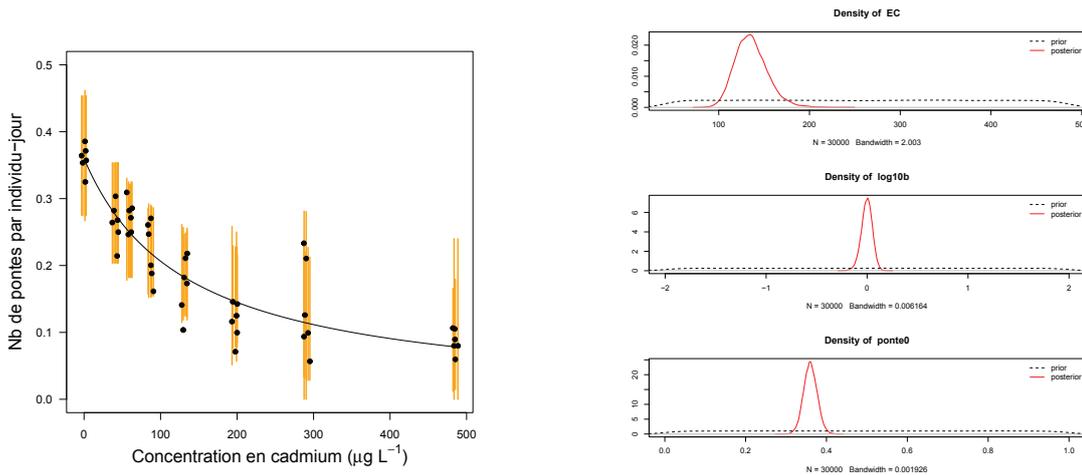


FIGURE 3.7 – Laboratoire 4 (pontes) : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

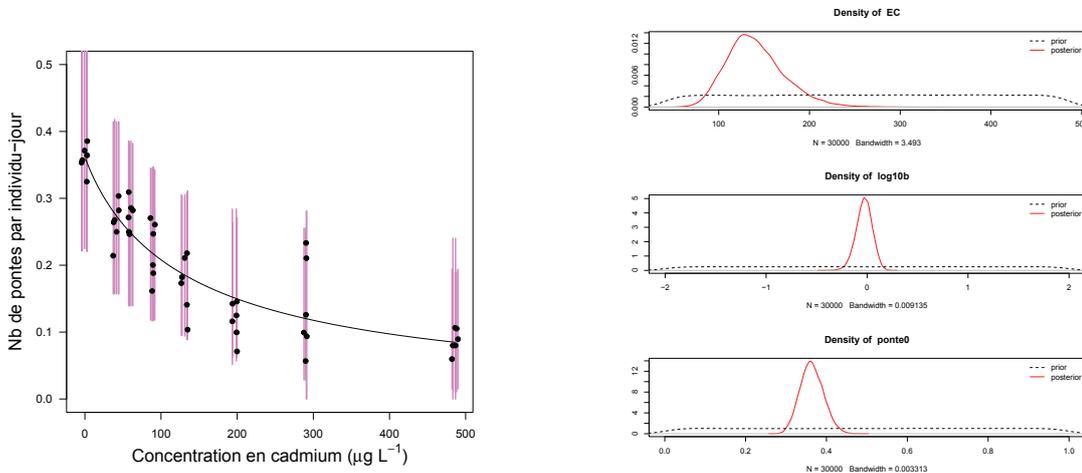
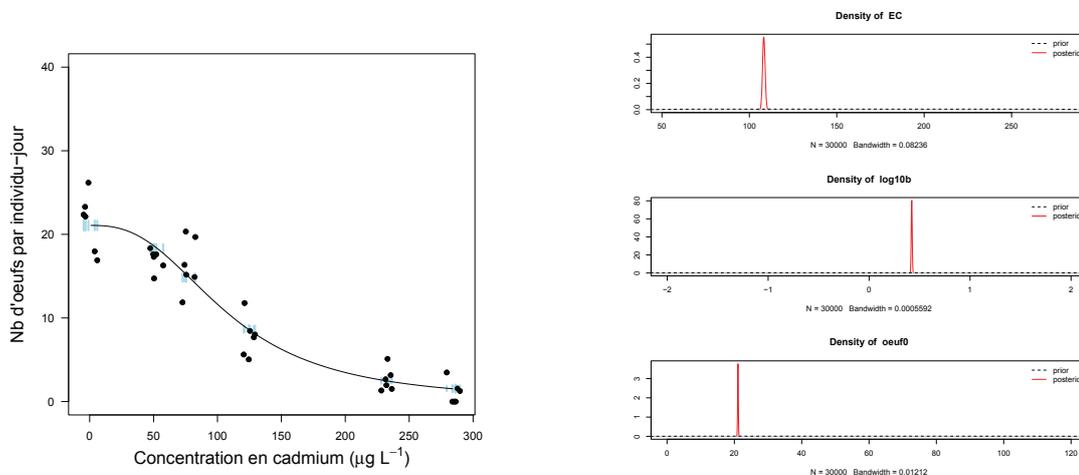


FIGURE 3.8 – **Laboratoire 5 (pontes)** : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

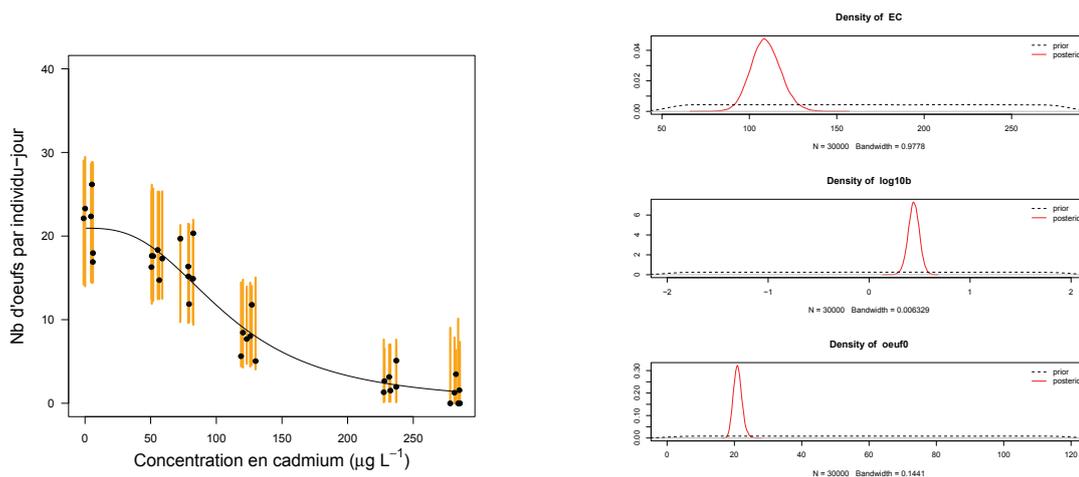
Pour le nombre d'oeufs par individu-jour, les ajustements, les intervalles de crédibilité et les distributions *a posteriori* par rapport aux *a priori* des paramètres des modèles sont présentés sur les Figures 3.9, 3.10, 3.11, 3.12 et 3.13. Contrairement à ce qui a été observé pour le nombre de pontes par individu-jour, pour les oeufs le choix de la partie stochastique du modèle a une importance majeure. En effet, pour la partie stochastique suivant une loi de Poisson, une sur-dispersion des données par rapport aux prédictions est observée. Ainsi, pour l'étude du nombre d'oeufs par individu-jour, il est déconseillé d'utiliser un modèle avec comme partie stochastique la loi de Poisson. Pour les deux autres modèles, les prédictions semblent bien décrire les données. De plus, les distributions *a posteriori* des paramètres par rapport à celles des *a priori* semblent correctes.

Afin d'affiner l'analyse de ces ajustements, nous nous intéresserons aux valeurs des DIC (voir 1.6) des modèles et aux estimations des EC_{50} et EC_{10} puisque ce sont ces valeurs qui nous intéressent.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

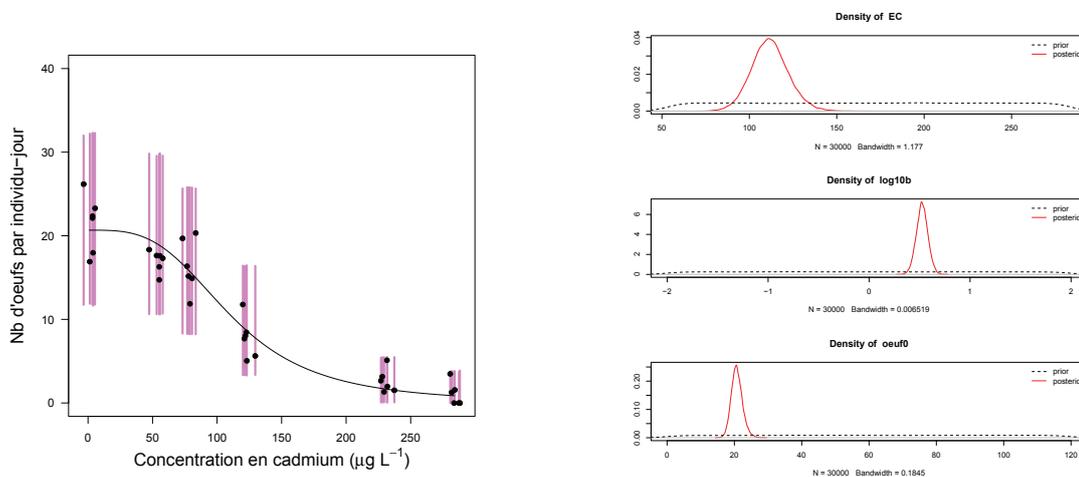
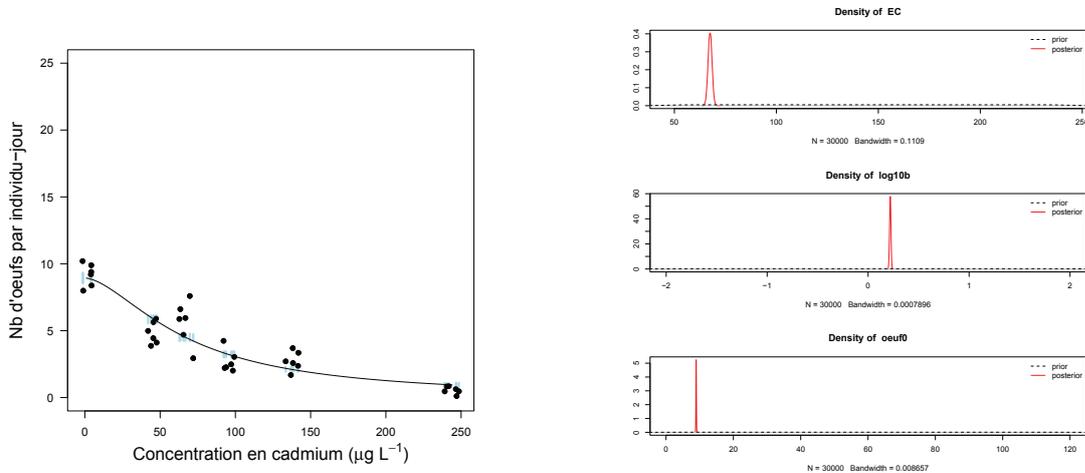
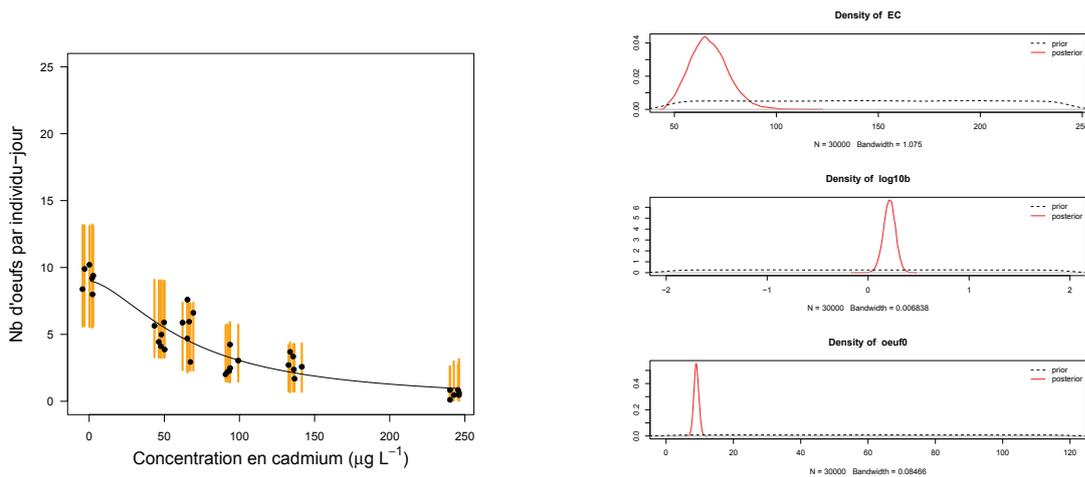


FIGURE 3.9 – Laboratoire 1 (oeufs) : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

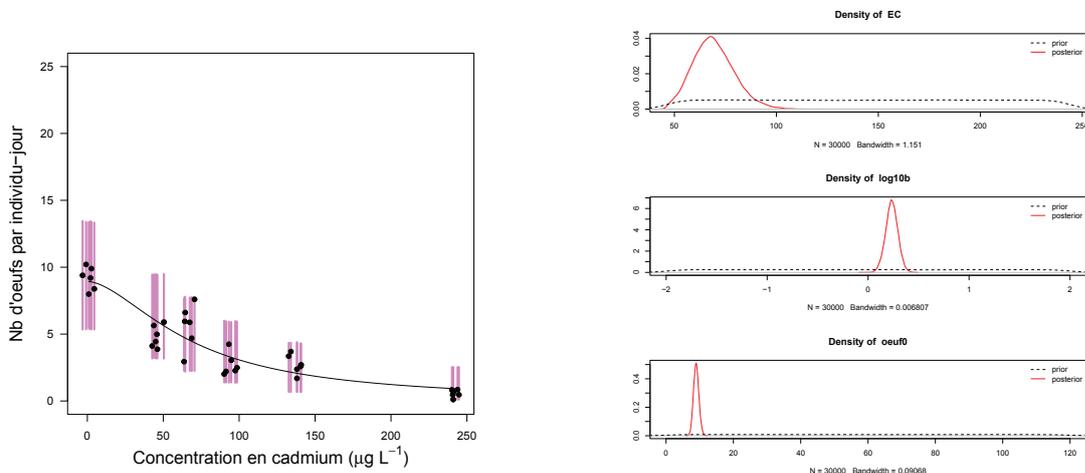
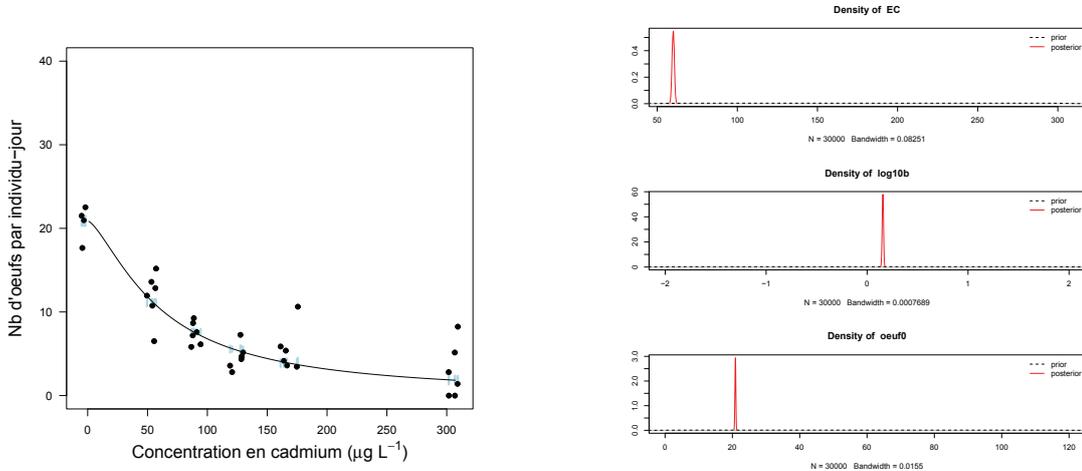
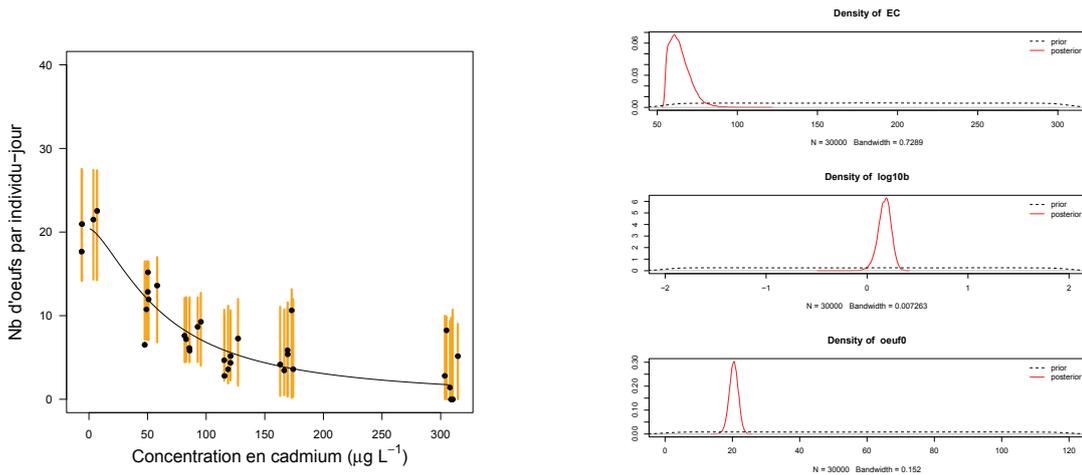


FIGURE 3.10 – Laboratoire 2 (oeufs) : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

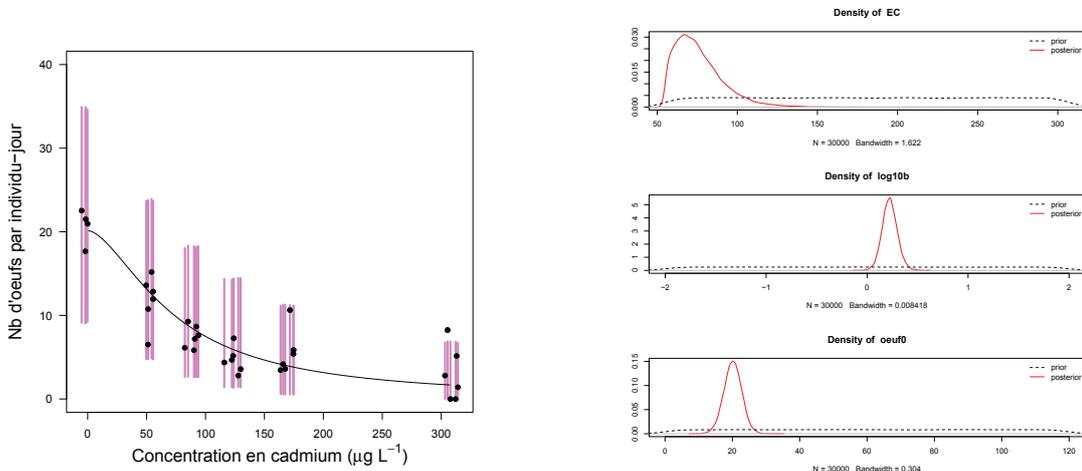
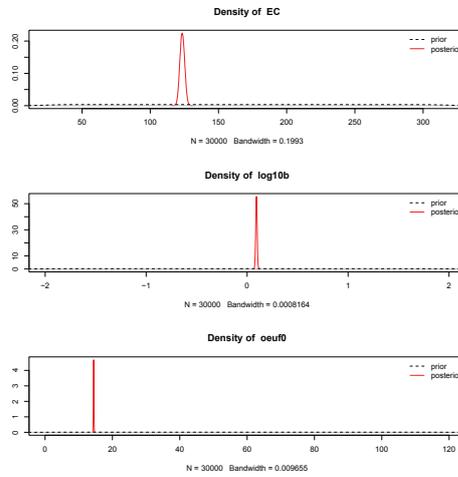
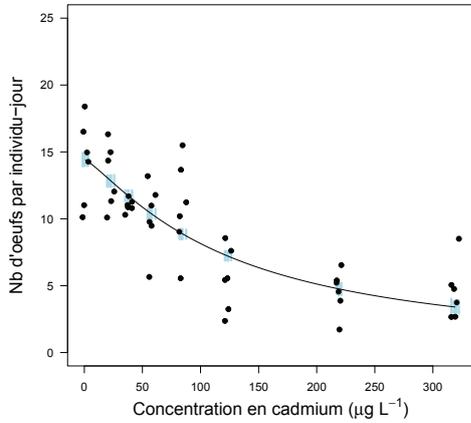
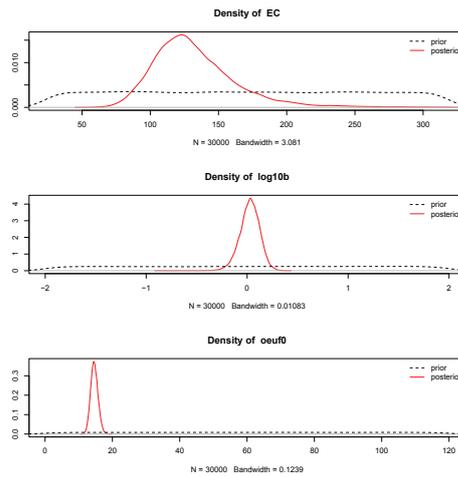
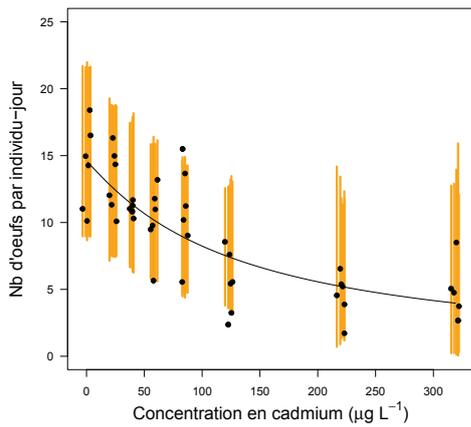


FIGURE 3.11 – Laboratoire 3 (oeufs) : à gauche, courbe moyenne et IC à 95% après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

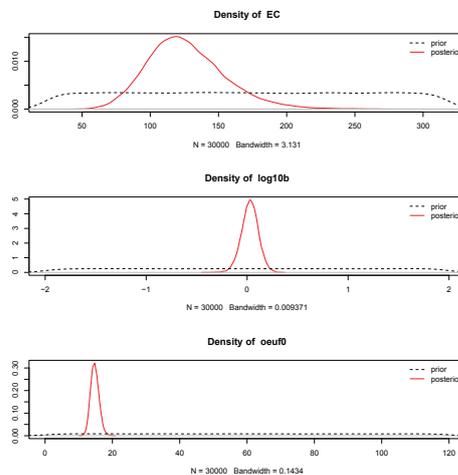
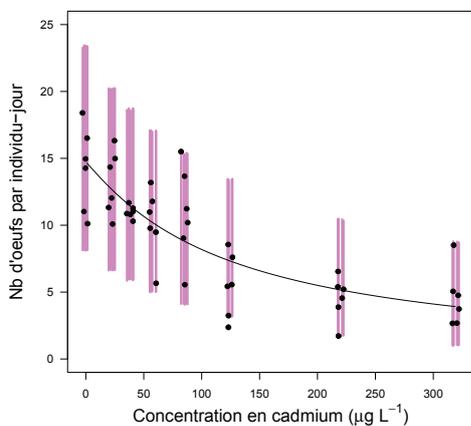
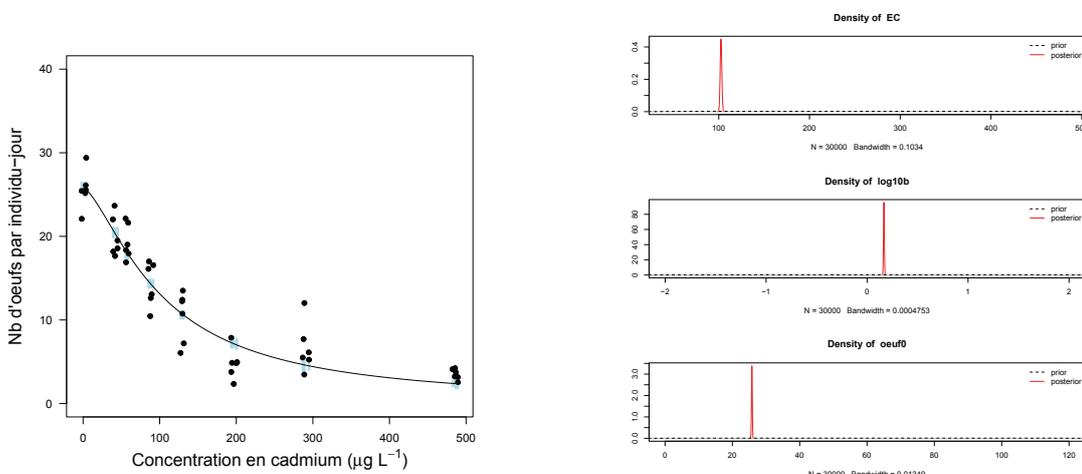
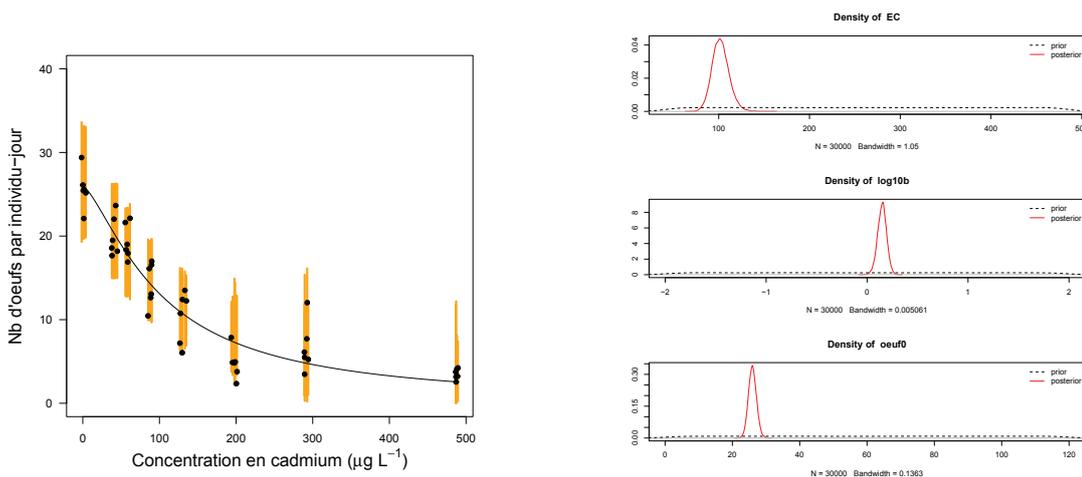


FIGURE 3.12 – Laboratoire 4 (oeufs) : à gauche, courbe moyenne et IC après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Loi de Poisson



Loi binomiale négative



Loi gamma-Poisson

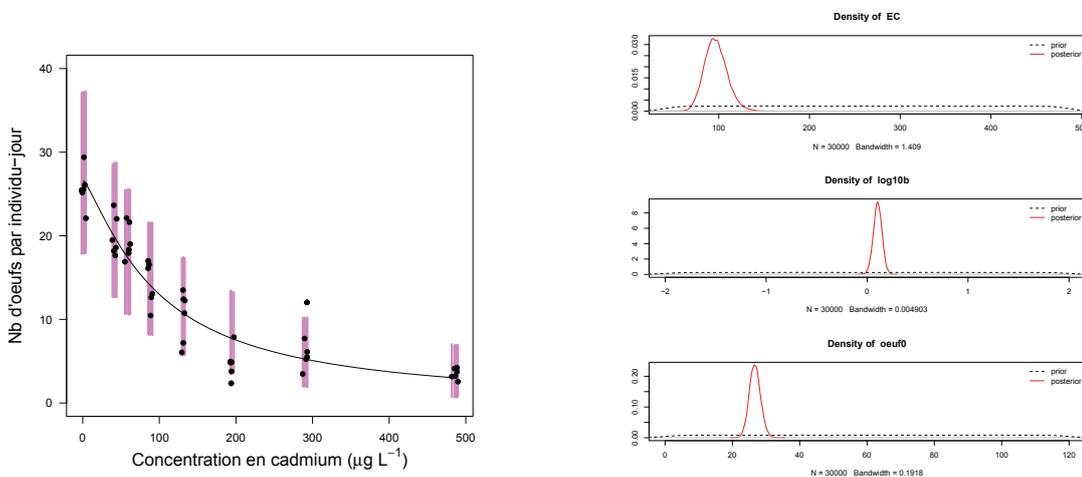


FIGURE 3.13 – Laboratoire 5 (oeufs) : à gauche, courbe moyenne et IC après ajustement d'un modèle log-logistique avec différentes parties stochastiques (Poisson, binomiale négative, gamma-Poisson) et à droite les distributions des *a posteriori* (rouges) et *a priori*.

Les DIC obtenus pour chaque modèle et chaque jeu de données sont présentés dans le Tableau 3.5. Les modèles avec les DIC les plus faibles (en bleu dans le tableau 3.5) sont ceux qui s'ajustent mieux aux données. Les valeurs des DIC des trois modèles étudiés sont beaucoup plus proches pour la variable "nombre de pontes par individu-jour" que pour la variable "nombre d'oeufs par individu-jour". En ce qui concerne la première variable, les trois modèles semblent équivalents (DIC les plus faibles dans deux cas sur cinq pour chacun des modèles). Pour la seconde variable, c'est le modèle avec comme partie stochastique la loi gamma-Poisson qui donne très nettement des valeurs de DIC plus faibles (dans cinq cas sur cinq). La loi de Poisson ne convient clairement pas, comme nous avons pu le voir en comparant les données et les prédictions (Figures 3.9, 3.10, 3.11, 3.12 et 3.13). Ainsi, la comparaison de ces trois modèles en utilisant les DIC, permet de retenir celui avec comme partie stochastique la loi gamma-Poisson.

	Poisson	Binomial négative	gamma-Poisson
Pontes par individu-jour			
Labo. 1	230	230	232
Labo. 2	238	234	231
Labo. 3	199	200	202
Labo. 4	339	334	325
Labo. 5	316	317	322
Oeufs par individu-jour			
Labo. 1	4285	516	380
Labo. 2	3387	510	378
Labo. 3	3776	481	350
Labo. 4	ND	743	537
Labo. 5	5043	725	540

Tableau 3.5 – Valeurs des DIC (Deviance Information Criterion) obtenues pour les six laboratoires pour chaque variable (nombre de pontes par individu-jour et nombre d'oeufs par individu-jour). Les valeurs de DIC les plus faibles sont écrites en bleu. ND : Non Déterminé pour des raisons techniques.

Les estimations des intervalles de crédibilité à 95% des concentrations critiques d'effet (EC_{50} et EC_{10}) sont représentées sur la Figure 3.14 pour les deux variables étudiées, les 5 laboratoires et les trois modèles (parties stochastiques différentes). Les résultats du modèle avec comme partie stochastique une loi de Poisson pour la variable "nombre d'oeufs par individu-jour", ont été représentés dans un but illustratif sachant que ce modèle ne convient pas. En effet, un mauvais modèle (celui avec la loi de Poisson) peut donner des résultats qui semblent idéaux en terme de courbe moyenne et de précision, d'où la nécessité de représenter les prédictions du modèle. Dans un premier temps, il est possible de constater l'importance de la différence d'estimations des EC_{50} entre les laboratoires. Cette différence est moins importante quand on compare les EC_{10} . Mais globalement, les différences d'estimations entre les laboratoires sont plus importantes que celles entre les modèles qui ont des parties stochastiques différentes. Les intervalles de crédibilité à 95% autour des EC_{50} et des EC_{10} sont plus restreints pour la variable "nombre d'oeufs par individu-jour". Globalement, les intervalles de crédibilité donnés par les différents modèles retenus pour un même jeu de données sont cohérents et proches. Il nous semble intéressant de nous préoccuper des bornes inférieures des intervalles de crédibilité à 95% autour des EC_{10} . En effet, plus la valeur de la borne inférieure sera faible plus ce sera protecteur pour l'environnement, puisque la vraie valeur de la EC_{10} estimée par le modèle a 95% de chance de se situer dans cet intervalle. Dans les cas où les trois modèles diffèrent de manière nette, le modèle avec partie stochastique de type gamma-Poisson donne les valeurs les plus protectrices.

Ainsi, la visualisation des ajustements et des distributions des paramètres *a posteriori* par rapport aux *a priori*, la comparaison des valeurs des DIC et la qualité des estimations et leur caractère protecteur pour l'environnement pour les deux types de variables, nous font choisir le modèle avec comme partie stochastique une loi gamma-Poisson. Ce modèle sera donc utilisé pour comparer les parties déterministes.

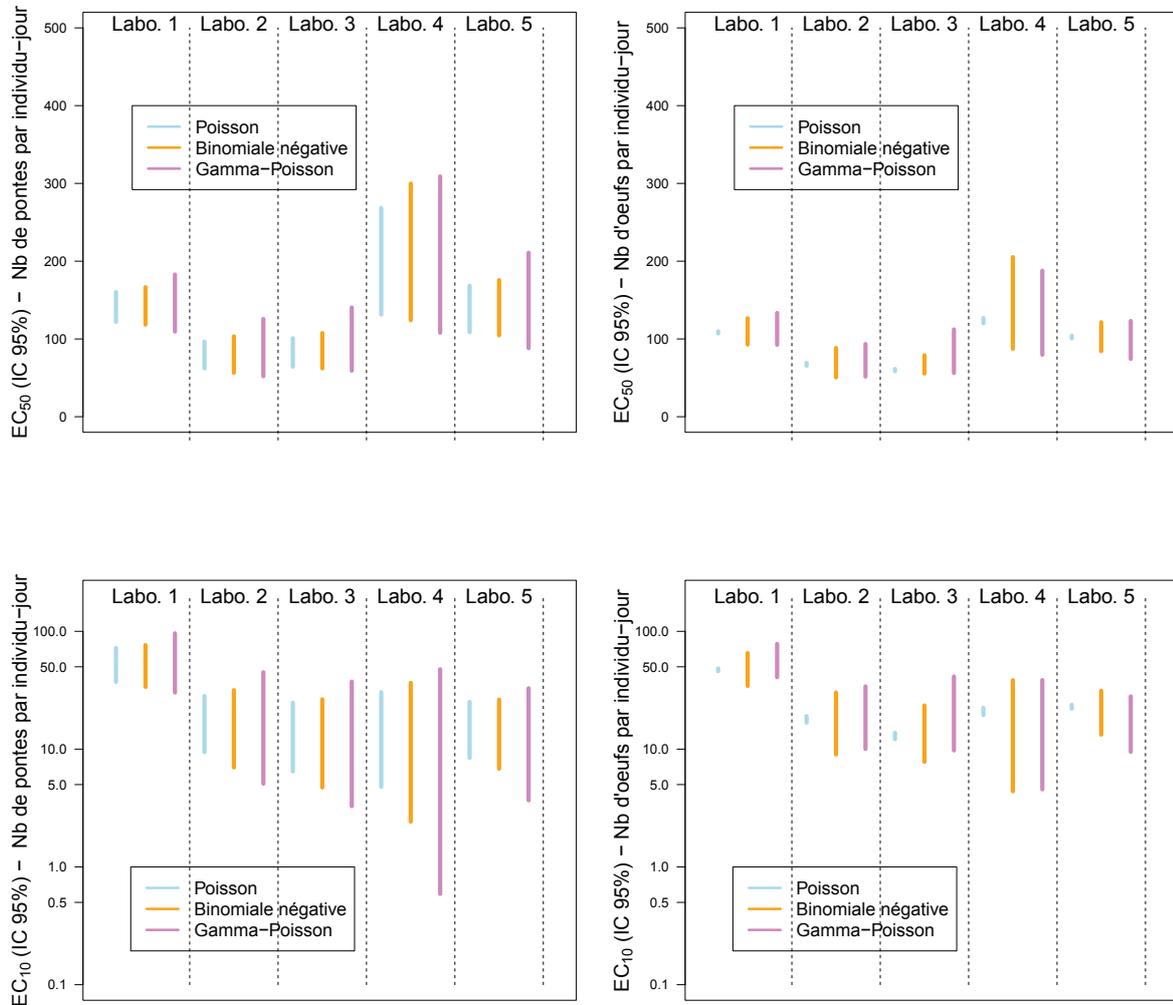


FIGURE 3.14 – Intervalles de crédibilité à 95% autour des EC_{50} (graphes du haut) et autour des EC_{10} (graphes du bas) pour le nombre de pontes par individu-jour (graphes de gauche) et pour le nombre d'oeufs par individu-jour (graphes de droite) obtenus pour chaque laboratoire, pour les modèles avec comme partie stochastique une loi de Poisson, une loi binomiale négative et une loi gamma-Poisson.

3.1.4 Choix de la partie déterministe

Pour choisir la partie déterministe du modèle, nous allons comparer trois modèles dont la partie stochastique est de type gamma-Poisson et la partie déterministe peut être une fonction log-logistique (Equation 3.1), une fonction log-probit (Equation 1.4) ou une fonction de Weibull (Equation 1.5).

Les résultats des ajustements des modèles avec des parties déterministes différentes pour les cinq laboratoires sont présentés sur la Figure 3.15 pour le nombre de pontes par individu-jour et sur la Figure 3.16 pour le nombre d'oeufs par individu-jour. Nous n'avons pas représenté les intervalles de crédibilité sur ces figures car cela nous aurait empêché de mettre les trois modèles sur la même figure et parce que seule la partie déterministe (courbe moyenne) nous intéresse ici. Les résultats obtenus pour les différents laboratoires ne sont pas présentés avec la même échelle, l'essentiel ici étant de voir au mieux les données. Il n'est pas possible de différencier les résultats des trois modèles après visualisation des ajustements. Nous nous intéresserons donc aux valeurs des DIC obtenues.

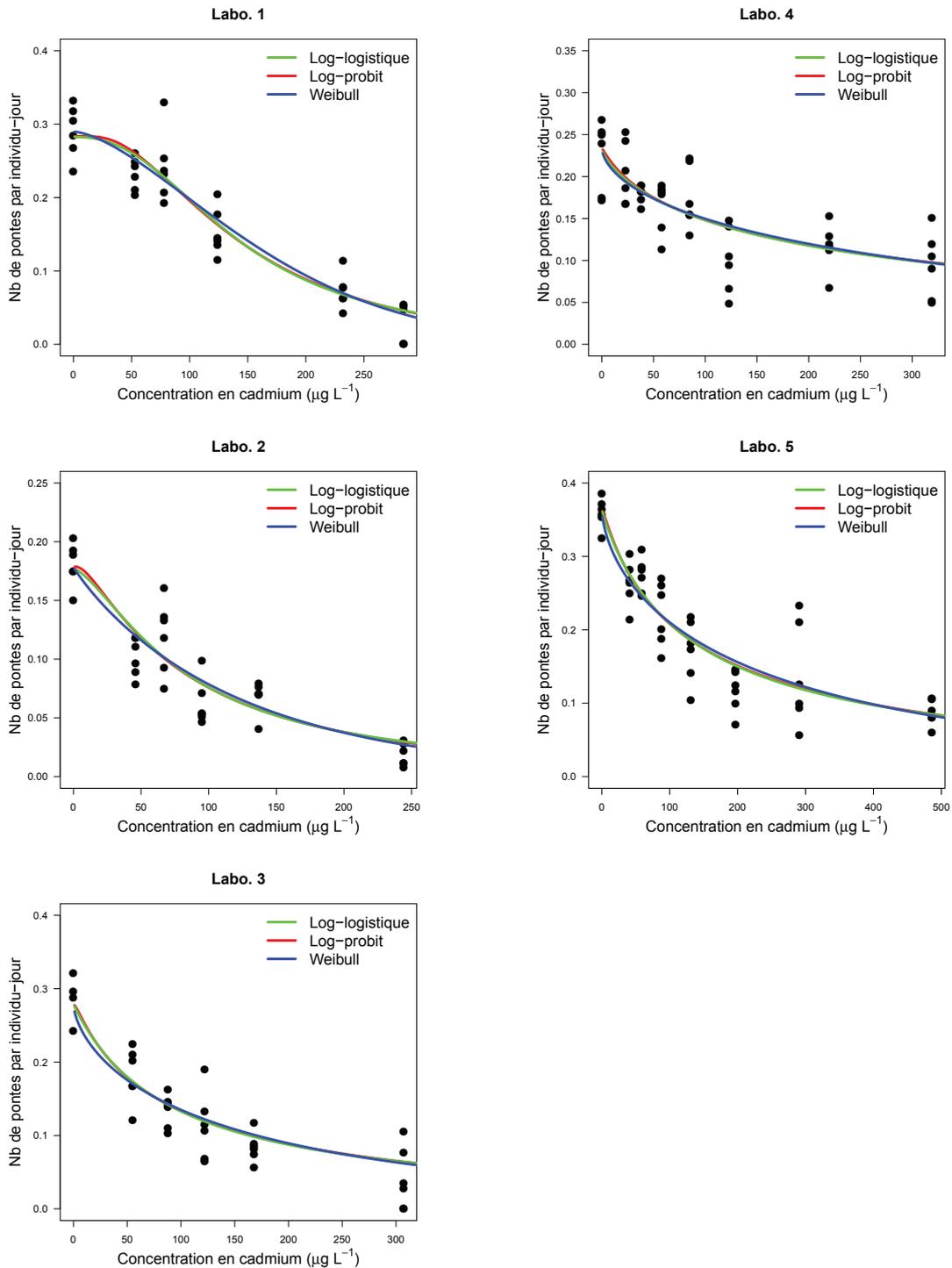


FIGURE 3.15 – Nombre de pontes par individu-jour (points noirs) et courbes moyennes (lignes de couleur) obtenues après ajustement d'un modèle avec comme partie stochastique une loi gamma-Poisson et comme partie déterministe une fonction log-logistique (en vert), une fonction log-probit (en rouge) et une fonction de Weibull (en bleu).

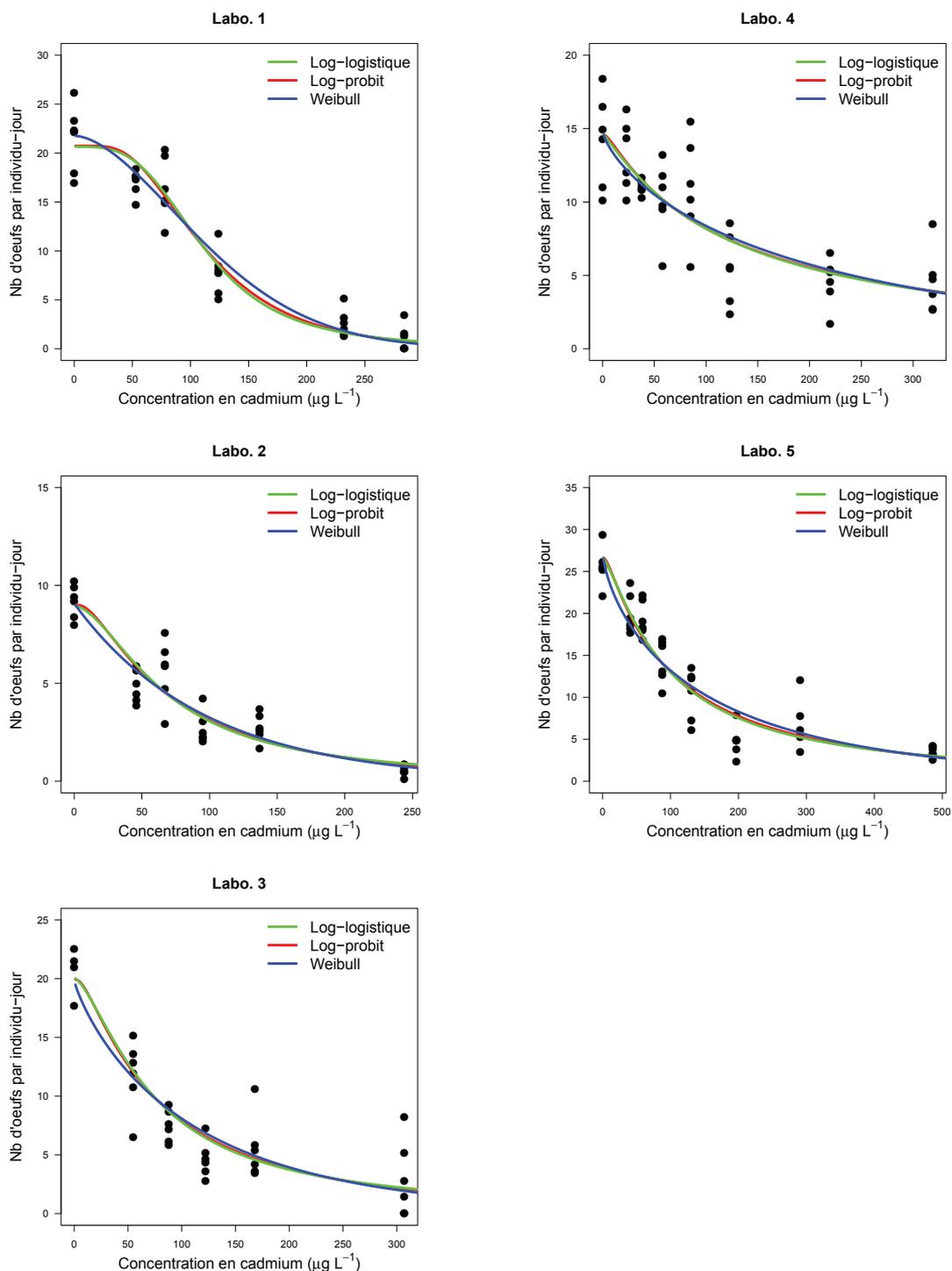


FIGURE 3.16 – Nombre d’œufs par individu-jour (points noirs) et courbes moyennes (lignes de couleur) obtenues après ajustement d’un modèle avec comme partie stochastique une loi gamma-Poisson et comme partie déterministe une fonction log-logistique (en vert), une fonction log-probit (en rouge) et une fonction de Weibull (en bleu).

Les DIC des trois modèles avec des parties déterministes différentes pour chaque jeu de données sont présentés dans le Tableau 3.6. Les trois modèles ont exactement les mêmes DIC dans huit cas sur douze. Les quatre fois où il y a des différences, il ne s'agit que d'un point de DIC, ce qui ne peut pas être considéré comme une différence importante. Ainsi, l'étude des valeurs des DIC ne nous permet pas de choisir entre les trois modèles.

	Log-logistique	Log-probit	Weibull
Pontes par individu-jour			
Labo. 1	232	232	232
Labo. 2	231	231	231
Labo. 3	202	202	202
Labo. 4	325	325	325
Labo. 5	322	322	323
Oeufs par individu-jour			
Labo. 1	380	380	379
Labo. 2	378	378	378
Labo. 3	350	350	350
Labo. 4	537	537	537
Labo. 5	540	540	540

Tableau 3.6 – Valeurs des DIC (Deviance Information Criterion) obtenues pour les cinq laboratoires pour chaque variable (nombre de pontes par individu-jour et nombre d'oeufs par individu-jour). Les valeurs de DIC les plus faibles sont en bleu.

Voyons maintenant quelles informations nous pouvons déduire des estimations des concentrations critiques d'effet. Les estimations des intervalles de crédibilité à 95% des EC_{50} et des EC_{10} pour les deux variables étudiées, les cinq laboratoires et les trois modèles avec des parties déterministes différentes sont représentées sur la Figure 3.17. Les intervalles de crédibilité à 95% autour des EC_{50} pour un même jeu de données (un même laboratoire) sont quasiment identiques quel que soit le modèle. Les différences entre les trois modèles avec parties déterministes différentes sont plus marquées quand on compare les intervalles de crédibilité autour des valeurs de EC_{10} . Comme nous l'avons fait pour comparer les parties stochastiques, nous comparons les bornes inférieures des intervalles de crédibilité à 95% autour des EC_{10} . Le modèle avec comme partie déterministe une fonction de Weibull (en bleu sur la Figure 3.17) donne les valeurs les plus protectrices pour l'environnement. C'est donc le modèle avec comme partie déterministe la fonction de Weibull et comme partie stochastique la loi gamma-Poisson qui semble être le plus protecteur vis à vis de l'environnement parmi tous les modèles testés.

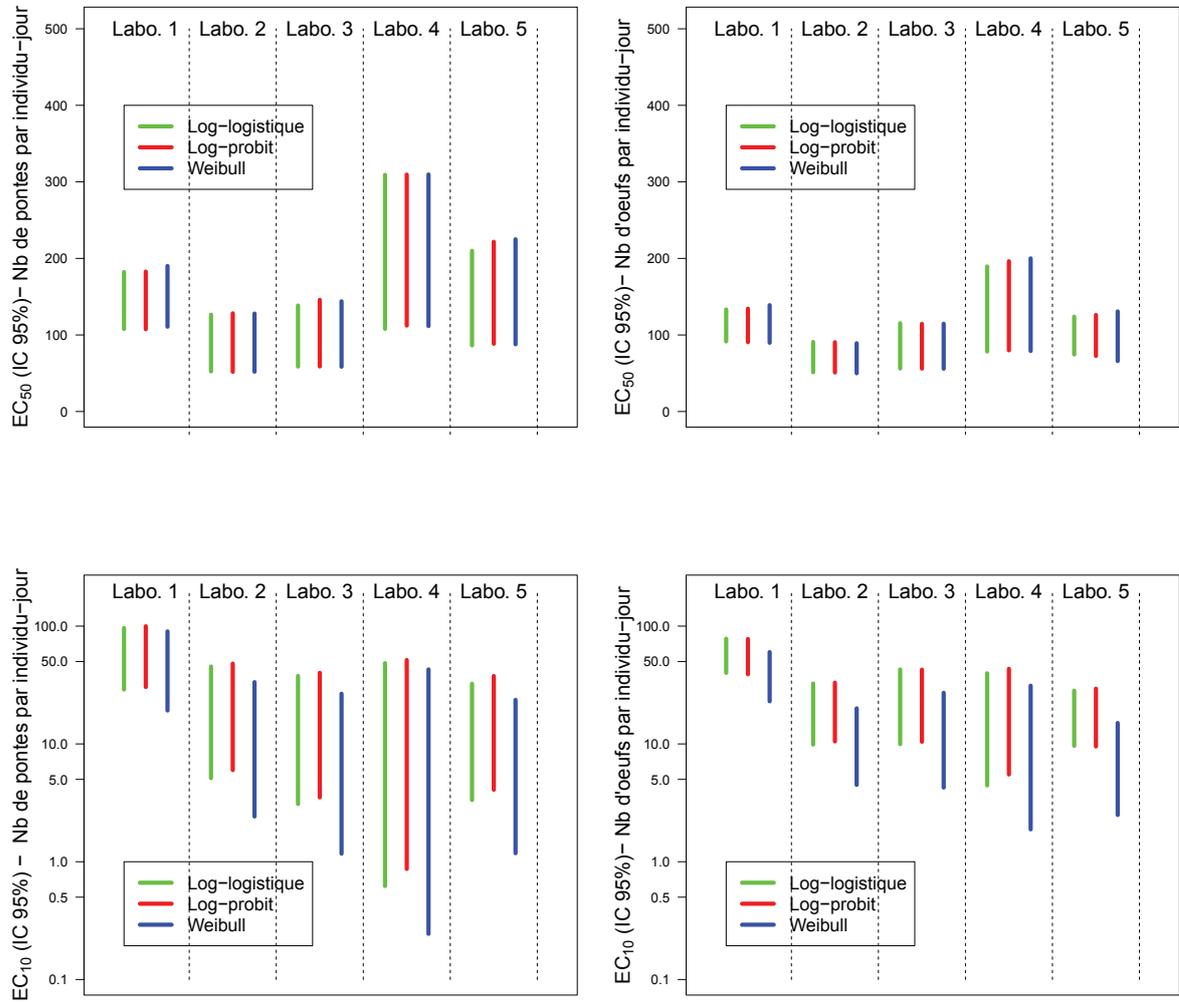


FIGURE 3.17 – Intervalles de crédibilités à 95% autour des EC_{50} (graphes du haut) et autour des EC_{10} (graphes du bas) pour le nombre de pontes par individu-jour (graphes de gauche) et pour le nombre d'oeufs par individu-jour (graphes de droite) obtenus pour chaque laboratoire, pour les modèles avec une partie stochastique de type gamma-Poisson et comme partie déterministe une fonction log-logistique (en vert), une fonction log-probit (en rouge) et une fonction de Weibull (en bleu).

Les valeurs des estimations obtenues après ajustement du modèle de Weibull (partie déterministe) associé à une loi gamma-Poisson (partie stochastique) sont présentées dans le tableau 3.7. Généralement, que ce soit pour les estimations des EC_{50} ou des EC_{10} (hormis pour le laboratoire 3), les valeurs obtenues pour la variable "nombre d'oeufs par individu-jour" sont plus faibles que celles pour la variable "nombre de pontes par individu-jour". Il est donc permis de penser que le nombre d'oeufs est plus sensible à une exposition de 56 jours au cadmium que le nombre de pontes. Les valeurs estimées des EC_{50} (pontes) semblent proches pour les laboratoires 2 et 3 (environ 6% de différence) et pour les laboratoires 1 et 5 (environ 3% de différence). De même, les valeurs des estimations des EC_{10} (oeufs) semblent être proches pour les laboratoires 2, 3 et 4, elles diffèrent de moins de 7%.

	EC_{50}	EC_{10}
Labo. 1 (pontes)	147,1 [110,8 - 190,1]	44,21 [19,20 - 90,28]
Labo. 1 (oeufs)	112,0 [89,78 - 138,9]	37,97 [22,99 - 60,25]
Labo. 2 (pontes)	83,49 [51,98 - 128,0]	10,99 [2,419 - 33,47]
Labo. 2 (oeufs)	66,85 [49,92 - 89,21]	9,937 [4,488 - 20,03]
Labo. 3 (pontes)	88,79 [58,50 - 144,1]	8,011 [1,170 - 26,70]
Labo. 3 (oeufs)	71,93 [55,92 - 114,70]	9,438 [4,261 - 27,15]
Labo. 4 (pontes)	208,1 [111,6 - 309,7]	7,740 [0,2439 - 42,95]
Labo. 4 (oeufs)	128,3 [79,07 - 200,0]	9,499 [1,881 - 31,22]
Labo. 5 (pontes)	143,3 [87,83 - 225,0]	6,836 [1,181 - 23,65]
Labo. 5 (oeufs)	93,35 [65,87 - 130,8]	6,676 [2,492 - 15,07]

Tableau 3.7 – Valeurs des estimations (médianes des distributions *a posteriori*) et des intervalles de crédibilité à 95% des EC_{50} et EC_{10} obtenues pour les deux variables et les cinq laboratoires après ajustement d'un modèle de Weibull associé à une loi gamma-Poisson.

3.1.5 Conclusion sur l'analyse des données de reproduction

Dans cette partie sur l'analyse des données de reproduction, nous avons montré que la mortalité doit être prise en compte dans la variable d'intérêt pour estimer l'effet d'une substance sur la reproduction de *Lymnaea stagnalis*. Comme en épidémiologie, l'utilisation du nombre d'individu-jours permet de prendre en compte la durée de participation de chaque individu au cours du test.

Du fait de la nature des données, à savoir des données de comptage, différentes parties stochastiques du modèle peuvent être envisagées. La comparaison de ces différentes parties stochastiques à partie déterministe fixée pour cinq jeux de données a montré qu'une partie stochastique de type gamma-Poisson semble être la plus appropriée pour la modélisation du nombre d'oeufs par individu-jour. La comparaison de différentes parties déterministes fréquemment utilisées avec comme partie stochastique la loi gamma-Poisson a montré que les résultats des estimations étaient vraiment très proches. Toutefois, en choisissant une fonction de Weibull comme partie déterministe, les bornes inférieures des intervalles de crédibilité autour de la EC_{10} semblent légèrement plus protectrices pour l'environnement pour les différents jeux de données étudiés. Toutefois, il faudrait valider ces résultats en utilisant d'autres jeux de données issus d'autres tests de toxicité. Les résultats ont également montré que les différences d'estimations entre les laboratoires étaient relativement importantes.

Conclusion :

Dans cette première partie sur l'analyse des données de reproduction, nous avons pu voir que :

- la prise en compte correcte de la mortalité dans l'analyse des données de reproduction peut se faire en calculant le nombre d'individu-jours
- la partie stochastique du modèle concentration-réponse la plus adaptée à l'analyse de ce type de donnée semble être de type gamma-Poisson au moins pour la modélisation du nombre d'oeufs par individu-jour.
- le choix de la partie déterministe du modèle concentration-réponse a moins d'impact que la partie stochastique, la comparaison des trois parties déterministes ne permet pas de les différencier de manière nette. Toutefois, le modèle de Weibull semble être plus protecteur en ce qui concerne l'estimation des EC_{10} .

3.2 Optimisation d'un plan d'expérience en vue d'une standardisation chez la limnée

3.2.1 Introduction

Après avoir étudié l'analyse des données de reproduction, nous allons nous intéresser à l'optimisation des plans d'expérience pour ce type de données. Dans le cadre des jeux de données sur la limnée exposée au cadmium, la demande du groupe de travail préparant le test standard pour le guide OCDE était dans un premier temps de réduire la durée du test qui représente un coût important en terme de temps de travail et de monopolisation du laboratoire. Une durée minimale de 21 jours, correspondant à la durée du test de reproduction chez *Daphnia magna*, a tout de même été considérée comme raisonnable. Dans un deuxième temps, l'objectif du groupe de travail était de réduire le nombre d'individus testés. Sachant qu'initialement, il a été choisi de travailler avec un nombre assez élevé de réplicats par concentration testée (six réplicats). Ainsi, nous allons étudier l'impact que peut avoir la réduction de la durée du test ainsi que celle du nombre de réplicats par concentration sur l'estimation des concentrations critiques d'effet afin d'optimiser le plan d'expérience du test standard initial. Les jeux de données sur lesquels nous nous baserons pour la réduction de la durée du test sont ceux présentés dans la partie reproduction. Seuls les jeux de données des laboratoires 1 à 5 sont utilisés, les laboratoires 6 et 7 n'ont pas été retenus en raison de la forte mortalité observée. Nous avons choisi de ne pas prendre en compte, dans l'optimisation du plan d'expérience, les deux concentrations supplémentaires ajoutées à chaque extrémité de la gamme initiale des laboratoires 4 et 5. En effet, l'optimisation du plan d'expérience standard doit se faire en travaillant sur des jeux de données homogènes (même nombre de concentrations initiales). D'ailleurs, les estimations, que ce soit pour les EC_{50} ou les EC_{10} pour les deux variables, sont très proches, que l'on prenne en compte ou non les concentrations extrêmes (Tableau 3.8). Ainsi, les deux concentrations extrêmes n'apportent pas d'information supplémentaire aux estimations pour les laboratoires 4 et 5. Au regard de ce que nous avons remarqué pour l'optimisation des plans d'expérience en survie, nous pouvons penser que l'ajout de ces concentrations supplémentaires à l'intérieur de la gamme plutôt qu'aux extrémités aurait pu être plus intéressant.

	EC ₅₀ (7 conc.)	EC ₅₀ (5 conc)
Labo. 4 (pontes)	208,1 [111,6 - 309,7]	209,8 [112,2 - 309,4]
Labo. 4 (oeufs)	128,3 [79,07 - 200,0]	129,2 [79,83 - 202,1]
Labo. 5 (pontes)	143,3 [87,83 - 225,0]	141,0 [85,30 - 217,7]
Labo. 5 (oeufs)	93,35 [65,87 - 130,8]	92,79 [66,40 - 128,1]
	EC ₁₀ (7 conc)	EC ₁₀ (5 conc)
Labo. 4 (pontes)	7,740 [0,2439 - 42,95]	7,869 [0,2481 - 42,56]
Labo. 4 (oeufs)	9,499 [1,881 - 31,22]	9,695 [1,879 - 31,37]
Labo. 5 (pontes)	6,836 [1,181 - 23,65]	6,711 [1,101 - 22,20]
Labo. 5 (oeufs)	6,676 [2,492 - 15,07]	6,558 [2,555 - 14,49]

Tableau 3.8 – Estimations (valeurs médianes des distributions *a posteriori*) et intervalles de crédibilité à 95% des EC₅₀ et des EC₁₀ pour les deux variables (pontes et oeufs par individu-jour) pour les laboratoires 4 et 5 obtenus en prenant en compte toutes les données (7 concentrations) ou en enlevant les deux concentrations extrêmes (5 concentrations).

3.2.2 Réduction de la durée de l'expérience

Analyse des données au cours du temps

Il s'agit dans un premier temps d'observer l'évolution des estimations des EC_{50} et des EC_{10} pour les deux variables (nombre de pontes par individu-jour et nombre d'oeufs par individu-jour) au cours du temps entre le 21^e et le 56^e jour, en utilisant le modèle de Weibull associé à une loi gamma-Poisson. Les résultats sont présentés sur la Figure 3.18 pour la variable "nombre d'oeufs par individu-jour". Il est possible de remarquer que les différents laboratoires présentent des évolutions de concentrations critiques d'effet différentes au cours du temps. Certains laboratoires présentent des profils d'estimation des concentrations critiques d'effet plutôt stables dans le temps avec des intervalles de crédibilité plutôt restreints (laboratoires 1 et 3). Par exemple, pour le laboratoire 1, les valeurs des estimations sont sensiblement les mêmes à partir de 38 jours. Le laboratoire 2, quant à lui, présente un profil temporel particulièrement "instable".

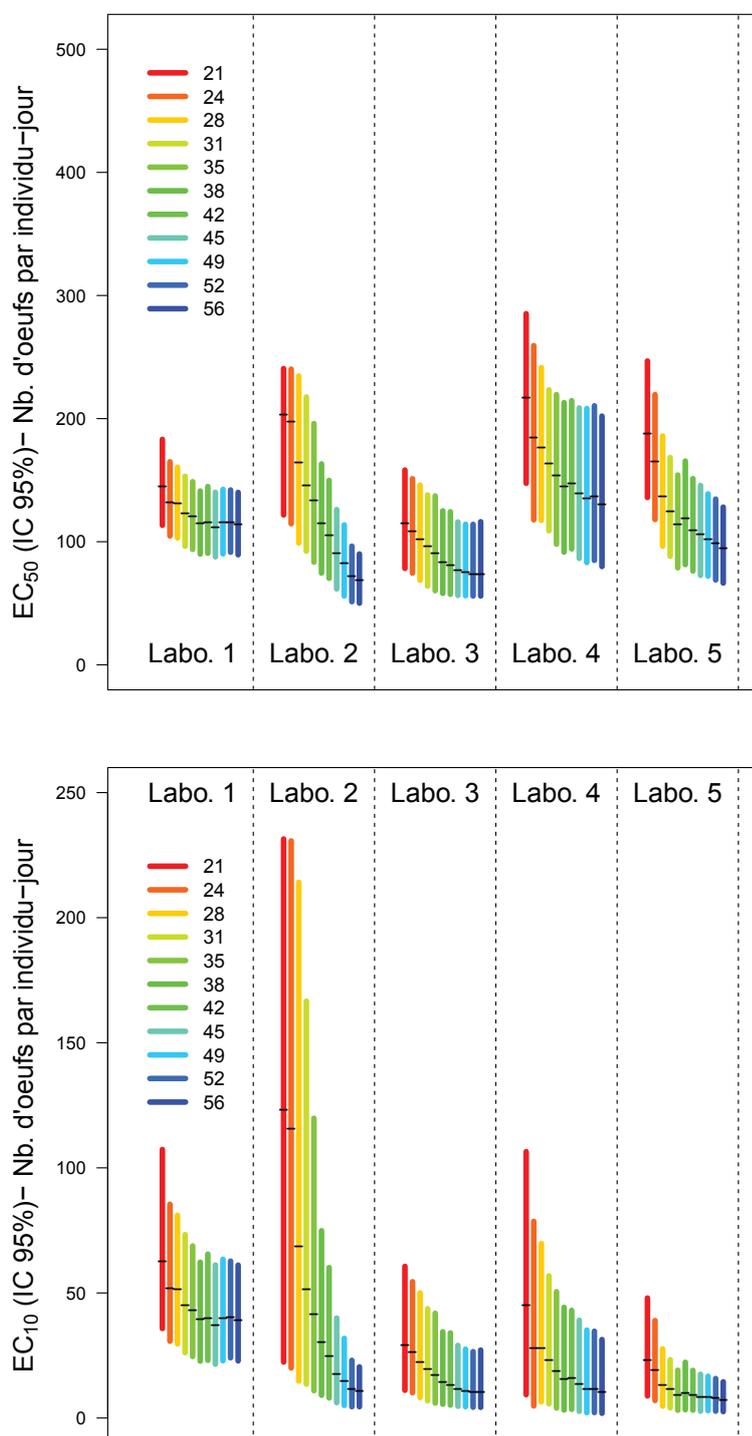


FIGURE 3.18 – Estimations (tirets) et intervalles de crédibilité autour des estimations des EC_{50} (graphe du haut) et des EC_{10} (graphe du bas) pour le nombre d'oeufs par individu-jour, pour chaque laboratoire et pour chaque durée d'exposition allant de 21 jours à 56 jours codés par différentes couleurs. Pour améliorer la lisibilité, les échelles des deux graphiques ne sont pas les mêmes.

Stratégies de réduction de la durée de l'expérience

Quelques petits rappels quant aux résultats obtenus sont utiles pour la suite de cette partie concernant les stratégies de réduction de la durée d'exposition. D'après les jeux de données des cinq laboratoires, la variable "nombre d'oeufs par individu-jour" est la plus sensible au cadmium. Concernant cette variable, les valeurs des EC_{10} des laboratoires 2, 3 et 4 donnaient les résultats les plus proches, avec des estimations qui différaient d'environ 7% à 56 jours (Tableau 3.7). De plus, l'évolution des EC_x au cours du temps est différente selon les laboratoires (Figure 3.18). Pour certains laboratoires, un palier semble être atteint ce qui n'est pas le cas pour d'autres. Deux stratégies ont été mises en oeuvre afin de réduire la durée de l'expérience : la stratégie du "palier" et la stratégie du "minimum".

La première est basée sur les paliers observés lors de l'évolution des EC_x au cours du temps, et la deuxième basée sur la valeur minimale des EC_x qui n'est pas obligatoirement la valeur à 56 jours.

- **Stratégie du "palier"** : basée sur l'étude de la variable "nombre d'oeufs par individu-jour", qui est la plus sensible au cadmium, la différence des EC_{50} (respectivement des EC_{10}) entre chaque pas de temps a été calculée. Les résultats des laboratoires 2 et 3 sont relativement proches à 56 jours, une différence d'environ 6% est observée. Il est raisonnable de considérer qu'une différence inférieure à 10% entre deux valeurs, séparées d'un seul pas de temps, peut être une différence acceptable. De plus, nous avons considéré que d'avoir trois fois de suite moins de 10% de différence entre les valeurs consécutives peut être considéré comme un palier. Ainsi, quand à la fois pour la EC_{50} et la EC_{10} un palier est atteint, alors la dernière durée d'exposition sera retenue comme nouvelle durée de l'expérience.
- **Stratégie du "minimum"** : comme précédemment, cette stratégie est basée sur le "nombre d'oeufs par individu-jour". Pour chaque laboratoire, les estimations minimales des EC_{50} et des EC_{10} au cours du temps sont retenues comme base de comparaison. A chaque durée d'exposition, à partir du 21^e jour, chaque estimation (EC_{50} et EC_{10}) est comparée respectivement à la valeur minimale retenue. Quand pour un même temps d'exposition, les différences à la fois pour la EC_{50} et la EC_{10} , entre les estimations et les valeurs minimales de base sont inférieures à 10%, alors cette durée d'exposition sera retenue comme nouvelle durée de l'expérience.

Les deux stratégies sont appliquées à l'ensemble des laboratoires pour les EC_{50} et les EC_{10} . Ainsi, pour chaque laboratoire et chaque concentration critique d'effet, nous estimerons la durée à laquelle le test peut être raisonnablement réduit.

Résultats de la réduction du temps d'exposition

1) Durée d'expérience retenue après application de la stratégie du "palier"

Afin de bien comprendre la stratégie de réduction de la durée d'exposition dite du "palier", les résultats obtenus pour le laboratoire 1 sont présentés dans le Tableau 3.9. A partir de 24 jours d'exposition au cadmium, les estimations des EC_{50} varient de moins de 10% et cela pour trois durées d'exposition consécutives (24, 28 et 31 jours). Ainsi, d'après la stratégie du "palier", la durée de 31 jours semble suffire comme nouvelle durée de l'expérience si l'on souhaite estimer une EC_{50} . En ce qui concerne la EC_{10} , à partir de 35 jours la différence entre deux estimations consécutives (dans le temps)

est inférieure à 10%. Ainsi pour estimer une EC_{10} , la durée minimale que l'on pourrait conseiller avec cette stratégie "palier" serait de 42 jours.

Les nouvelles durées obtenues pour l'ensemble des laboratoires sont présentées dans le Tableau 3.11.

Durée d'exposition D_v avec v allant de 1 à 11	EC_{50}	% de différence entre EC_{50} à D_v et à D_{v+1}	EC_{10}	% de différence entre EC_{10} à D_v et à D_{v+1}
$D_1 = 21$	143,4		61,75	
$D_2 = 24$	130,6	8,92	51,17	17,12
$D_3 = 28$	129,7	0,670	50,65	1,03
$D_4 = 31$	121,5	6,33	44,33	12,5
$D_5 = 35$	118,9	2,12	42,39	4,36
$D_6 = 38$	113,2	4,79	38,66	8,80
$D_7 = 42$	114,04	-0,717	38,99	-0,844
$D_8 = 45$	110,3	3,24	36,46	6,49
$D_9 = 49$	114,0	-3,31	39,22	-7,58
$D_{10} = 52$	114,1	-0,0589	39,57	-0,878
$D_{11} = 56$	112,4	1,43	38,42	2,91

Tableau 3.9 – Estimations des EC_{50} et des EC_{10} pour chaque durée d'exposition après la durée minimale de 21 jours pour le laboratoire 1. A partir de 24 jours d'exposition, le pourcentage de différence entre l'estimation à la durée D_v et l'estimation à la durée D_{v+1} est calculé respectivement pour les EC_{50} et les EC_{10} . Quand le pourcentage de différence est inférieur à 10%, il est en vert sinon il est en rouge. La durée d'exposition retenue d'après la stratégie du "palier" est écrite en orange.

2) Durée d'expérience retenue après application de la stratégie du "minimum"

Afin de bien comprendre la stratégie de réduction de la durée d'exposition dite du "minimum", les résultats obtenus pour le laboratoire 1 sont présentés dans le Tableau 3.10. Les valeurs minimales des EC_{50} et des EC_{10} pour la variable nombre d'oeufs par individu-jour sont atteintes après 45 jours d'exposition au cadmium. Les valeurs (en bleu dans le tableau) constituent les valeurs minimales de référence. Ainsi, pour chaque durée d'exposition, le pourcentage de la différence entre l'estimation à la durée d'exposition étudiée et la valeur minimale de référence est calculée. Après 35 jours d'exposition, l'estimation de la EC_{50} est différente de moins de 10% de la valeur minimale de référence. Par contre, pour cette durée d'exposition, la différence entre l'estimation de la EC_{10} et la valeur minimale de référence de la EC_{10} est supérieure à 10% (16,3%). Cette différence est inférieure à 10% après 38 jours d'exposition. Ainsi, la durée de 35 jours peut-être retenue comme nouvelle durée de l'expérience après application de la stratégie du "minimum" si on veut estimer une EC_{50} , par contre si on veut estimer une EC_{10} la durée doit être au moins égale à 38 jours.

Le tableau 3.11, reprend les résultats obtenus pour les deux stratégies, pour les deux concentrations critiques d'intérêt et pour l'ensemble des laboratoires. On retrouve ce qui avait été observé auparavant, c'est à dire que le laboratoire 2 ne semble pas atteindre

Durée d'exposition	EC ₅₀	% de différence entre EC ₅₀ et "Min50"	EC ₁₀	% de différence entre EC ₁₀ et "Min10"
21	143,4	29,9	61,75	69,4
24	130,6	18,3	51,17	40,4
28	129,7	17,5	50,65	38,9
31	121,5	10,1	44,33	21,6
35	118,9	7,77	42,39	16,3
38	113,2	2,61	38,66	6,04
42	114,04	3,35	38,99	6,93
45	110,3 (Min50)	0	36,46 (Min10)	0
49	114,0	3,31	39,22	7,58
52	114,1	3,37	39,57	8,53
56	112,4	1,88	38,42	5,37

Tableau 3.10 – Estimations des EC₅₀ et des EC₁₀ pour chaque durée d'exposition après la durée minimale de 21 jours pour le laboratoire 1. Pour chaque durée d'exposition, le pourcentage de différence entre l'estimation de cette durée et l'estimation de référence (en bleu notées Min50 et Min10) est calculé respectivement pour les EC₅₀ et les EC₁₀. Quand le pourcentage de différence est inférieur à 10%, il est en vert, sinon il est en rouge. La durée d'exposition retenue d'après la stratégie du "minimum" est écrite en orange.

de palier ni pour l'estimation de la EC₅₀, ni pour celle de la EC₁₀. Ce laboratoire a un profil d'évolution des EC_x dans le temps bien particulier par rapport aux autres laboratoires. En ce qui concerne le laboratoire 4, la stratégie "palier" ne permet pas non plus de dégager une nouvelle durée d'exposition pour l'estimation de la EC₁₀. Dans le cas où l'on souhaite estimer une EC₅₀, la stratégie du "palier" permet de réduire la durée d'exposition entre 31 et 38 jours. Ainsi, si on réduit la durée de l'expérience à 38 jours pour l'ensemble des laboratoires (hormis le laboratoire 2) la valeur de l'estimation de la EC₅₀ obtenue sera sur un palier où les différences entre deux valeurs consécutives diffèrent de moins de 10%.

En ce qui concerne la stratégie du "minimum", elle ne permet pas de réduire autant la durée d'exposition. Les nouvelles durées varient de 35 à 49 jours (sans considérer le laboratoire 2). Ainsi, si on se place à 49 jours pour l'ensemble des laboratoires (hormis le laboratoire 2), on obtient une estimation de la EC₅₀ qui diffère de moins de 10% de la valeur minimale obtenue au cours du test 56 jours. En ce qui concerne la EC₁₀, les nouvelles durées varient de 42 à 56 jours pour la stratégie du "palier" et de 38 à 56 jours pour la stratégie du "minimum". On peut donc en conclure que la durée d'exposition doit être maintenue à 56 jours quand on se place uniquement dans le cas d'une estimation de la EC₁₀.

Dans un souci de fournir une durée réduite qui puisse convenir pour l'ensemble des laboratoires et pour les deux concentrations critiques d'effet, nous choisissons de sélectionner la durée la plus courte permettant de respecter les critères des stratégies à

Laboratoire	EC ₅₀		EC ₁₀	
	"Palier"	"Minimum"	"Palier"	"Minimum"
1	31	35	42	38
2	-	53	-	53
3	31	45	56	49
4	35	45	-	56
5	38	45	52	56

Tableau 3.11 – Nouvelles durées d'exposition (en jour) obtenues après application des stratégies du "palier" et du "minimum" pour la EC₅₀ et la EC₁₀ de chaque laboratoire.

la fois sur la EC₁₀ et sur la EC₅₀. Nous pouvons alors voir l'impact de ce choix de durée sur les autres concentrations critiques d'effet des autres laboratoires. Le laboratoire 1 permet de réduire le plus la durée de l'expérience pour l'estimation de la EC₁₀ et de la EC₅₀ pour les deux stratégies. Comme nous pouvons le voir dans le Tableau 3.11, la stratégie du "palier" permet de réduire l'exposition à une durée de 42 jours, tandis que la stratégie du "minimum" permet de réduire la durée à 38 jours. La première remarque qui peut être faite, est que les deux stratégies donnent des durées d'expérience assez proches (4 jours de différence) et permettent de réduire l'expérience de 14 jours à 18 jours, et donc de réduire d'au moins un quart la durée totale de l'expérience.

Comparaison des estimations des différentes durées d'expérience

Afin d'évaluer l'impact de la durée d'exposition de l'expérience sur l'estimation des EC₅₀ et EC₁₀ pour chaque variable (nombre de pontes et d'oeufs par individu-jour), les intervalles de crédibilité à 95% autour des estimations ont été représentés sur la Figure 3.19 à 38 jours, 42 jours et 56 jours d'exposition. Globalement, les variations des estimations entre durées d'exposition sont moins importantes que celles observées entre les différents laboratoires. Concernant les résultats des estimations, les plus fortes différences entre les nouvelles durées d'exposition et la durée initiale de 56 jours sont observées pour le laboratoire 2. Toutefois, il est possible de constater qu'en plus du laboratoire 1, qui a permis de déterminer les nouvelles durées d'exposition, les intervalles de crédibilité des EC₅₀ obtenus pour les nouvelles durées d'exposition sont très proches pour les laboratoires 3 et 4. Pour les EC₁₀, en plus de ces laboratoires, les résultats sont également proches pour le laboratoire 5. Comme c'est pour le laboratoire 1 que la réduction de la durée d'expérience est la plus importante, ces résultats permettent de visualiser la perte d'information maximale qu'il peut y avoir. La durée de 38 jours peut raisonnablement être retenue comme nouvelle durée d'exposition compte tenu des résultats obtenus et de la différence inter-laboratoire observée.

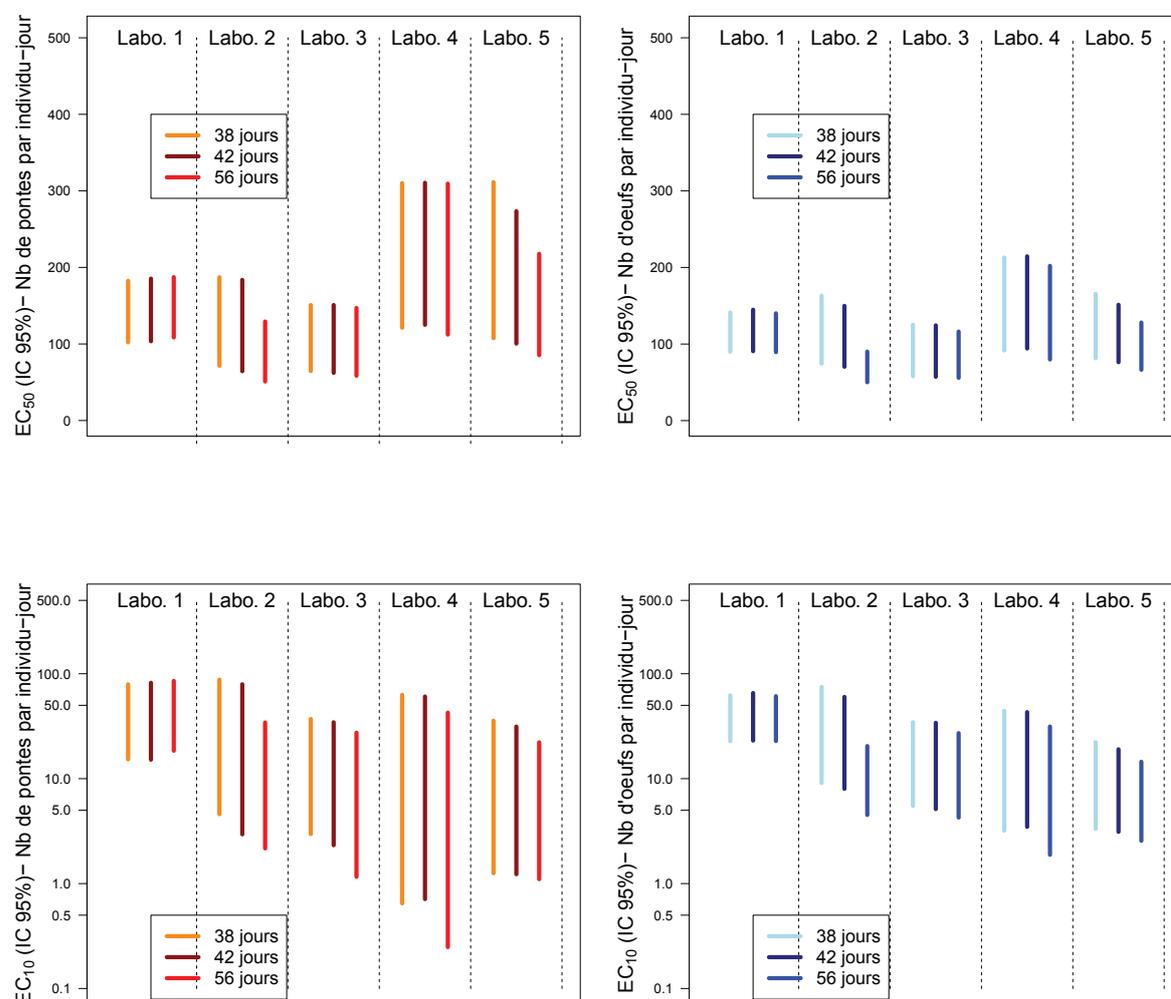


FIGURE 3.19 – Intervalles de crédibilités à 95% autour des EC₅₀ (graphes du haut) et autour des EC₁₀ (graphes du bas) pour le nombre de pontes par individu-jour (graphes de gauche) et pour le nombre d'oeufs par individu-jour (graphes de droite). Ces intervalles sont présentés pour chaque laboratoire, pour la durée d'exposition déterminée avec la stratégie du "minimum" (38 jours), pour celle déterminée avec la stratégie du "palier" (42 jours) et pour la durée totale de l'expérience (56 jours).

3.2.3 Réduction du nombre d'individus utilisés dans l'expérience

L'objectif est ici de réduire le nombre total d'individus utilisés pour l'expérience pour une durée de 56 jours puis pour la durée réduite retenue (38 jours). Nous avons choisi de réduire le nombre de réplicats par concentration (avec cinq individus par réplicat) tout en mettant le même nombre de réplicats par concentration pour l'ensemble des laboratoires. Pour cela, nous estimons les EC_{50} et les EC_{10} pour le nombre d'oeufs par individu-jour puis pour le nombre de pontes par individu-jour dans les cas où seules les données d'un, deux, trois, quatre ou cinq réplicat(s) par concentration (y compris pour le témoin) sont disponibles. Pour chacune des concentrations testées, nous avons fait 50 tirages avec remise de un, deux, trois, quatre ou cinq réplicats. Dans un premier temps, nous avons choisi de faire 100 tirages mais le temps de calcul nous a fait diviser ce nombre par deux. Les premiers résultats obtenus ne montraient pas de différence entre 100 et 50 tirages. Nous avons ainsi 50 jeux de données différents pour chaque nombre de réplicats testés. Pour chacun des jeux de données, nous obtenons des estimations des concentrations critiques d'effet et des intervalles de crédibilité à 95% autour de ces estimations. Nous avons retenu les valeurs moyennes des bornes des intervalles de crédibilité à 95% et des médianes obtenus pour en déduire l'impact que peut avoir le nombre de réplicats sur l'estimation des concentrations critiques d'effet. Afin d'estimer la précision des estimations, nous avons calculé le rapport des moyennes des 95^e percentiles sur les moyennes des 5^e percentiles des 50 intervalles de crédibilité appelé $R_{m95/m5}$. Ce rapport a été utilisé par [Rufi et Springer \(2011\)](#) pour apprécier la précision des estimations dans leurs travaux pour diminuer le nombre de poissons utilisés dans le test de toxicité aiguë de l'OCDE. Nous nous sommes basés uniquement sur l'étude de la précision des estimations obtenues pour la variable nombre d'oeufs par individu-jour, qui semble être la plus sensible au cadmium.

Les résultats à 56 jours sur les estimations et les bornes des intervalles de crédibilité des EC_{50} et des EC_{10} sont présentés sur la Figure 3.20 pour le nombre d'oeufs par individu-jour et sur la Figure 3.21 pour le nombre de pontes par individu-jour pour l'ensemble des laboratoires. Dans un premier temps, il est possible de constater que les variations observées en fonction de la réduction du nombre de réplicats sont globalement moins importantes que celles observées dans le cadre d'une réduction de la durée de l'expérience (ex : labo.2 Figure 3.18 et Figure 3.20). En fonction des laboratoires, l'influence du nombre de réplicats sur les valeurs estimées n'est pas la même. En effet, nous voyons que pour le laboratoire 1, l'impact de la réduction à 1 seul réplicat n'affecte que l'intervalle de crédibilité qui devient plus large, mais n'affecte pas la valeur moyenne estimée de la EC_{50} sur le nombre d'oeufs par individu-jour (Figure 3.20), ce qui n'est pas du tout le cas pour le laboratoire 3 où l'intervalle et la valeur moyenne sont affectés. Globalement, les résultats observés sont similaires à ceux obtenus dans la partie sur la survie : pour un nombre de concentrations identique, les estimations sont plus précises quand le nombre d'individus exposés augmente. Les valeurs des $R_{m95/m5}$, représentant la précision de l'estimation des EC_{50} pour la variable nombre d'oeufs par individu-jour, sont présentées dans le Tableau 3.12. Les $R_{m95/m5}$ obtenus pour l'expérience avec 6 réplicats par concentration (expérience initiale) se situent entre 1.56 et

2.45 suivant les laboratoires. Compte tenu de cette variabilité inter-laboratoire sur ce rapport $R_{m95/m5}$, nous avons choisi de retenir (de manière un peu arbitraire) comme valeur seuil $R_{m95/m5} = 3.00$. Ainsi quand la valeur de $R_{m95/m5}$ dépassera ce seuil, la précision perdue sera considérée comme trop importante. Ce seuil peut être modifié en fonction des exigences des écotoxicologues. Ainsi, pour que la perte de précision soit sous la valeur seuil pour tous les laboratoires, il faut retenir un nombre de quatre réplicats par concentration. Seul le laboratoire 3 nécessite une réduction à 4 réplicats par concentration. Le nombre de réplicats par concentration pour les autres laboratoires peut être réduit à trois, voire deux réplicats par concentration. Nous avons vérifié les valeurs des $R_{m95/m5}$ pour la variable nombre de pontes par individu-jour dans le cas d'une réduction du nombre de réplicats à quatre à 56 jours. Toutes les valeurs obtenues étaient inférieures au seuil que nous avons fixé (trois).

$R_{m95/m5}$ pour	Labo.1	Labo.2	Labo.3	Labo.4	Labo.5
6 replicats	1.56	1.81	2.03	2.45	1.93
5 replicats	1.61	1.86	2.14	2.49	2.07
4 replicats	1.70	1.94	2.57	2.60	2.23
3 replicats	1.98	2.17	3.21	3.00	2.57
2 replicats.	2.31	3.01	3.94	3.22	3.26
1 replicat	3.48	4.64	3.97	3.79	4.09

Tableau 3.12 – Valeurs des $R_{m95/m5}$ obtenues à 56 jours pour les estimations des EC_{50} en ce qui concerne le nombre d'oeufs par individu-jour pour les différents laboratoires et pour un nombre de réplicats par concentration réduit de six à un. La valeur de $R_{m95/m5}$ en bleu correspond au nombre de réplicats le plus petit pour lequel cette valeur est inférieure ou égale à 3.

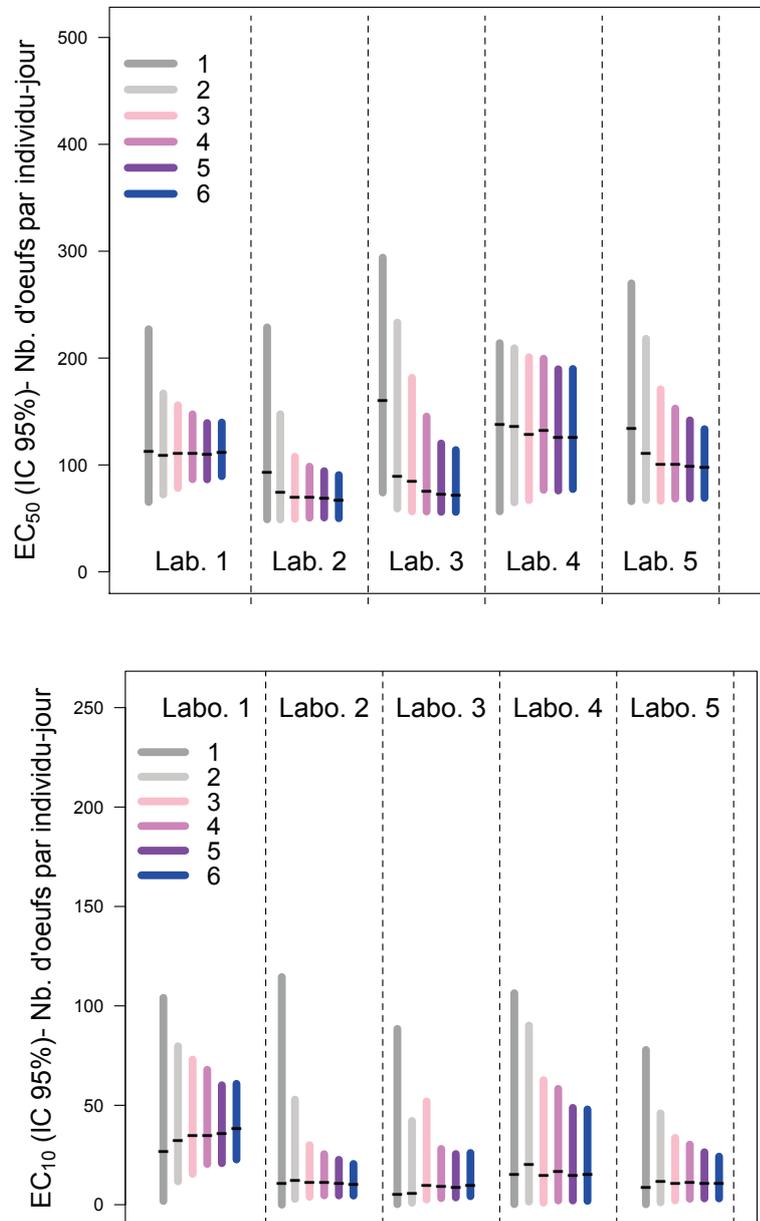


FIGURE 3.20 – Moyennes des estimations (tirets) et des bornes des intervalles de crédibilité à 95% (sur 50 jeux de données après tirage au sort pour chaque nombre de réplicat) des EC_{50} (graphes du haut) et des EC_{10} (graphes du bas) pour le nombre d'oeufs par individu-jour obtenus pour chaque laboratoire à 56 jours et pour un nombre de réplicats (de 1 à 6) codés par une couleur (expérience initiale en bleu).

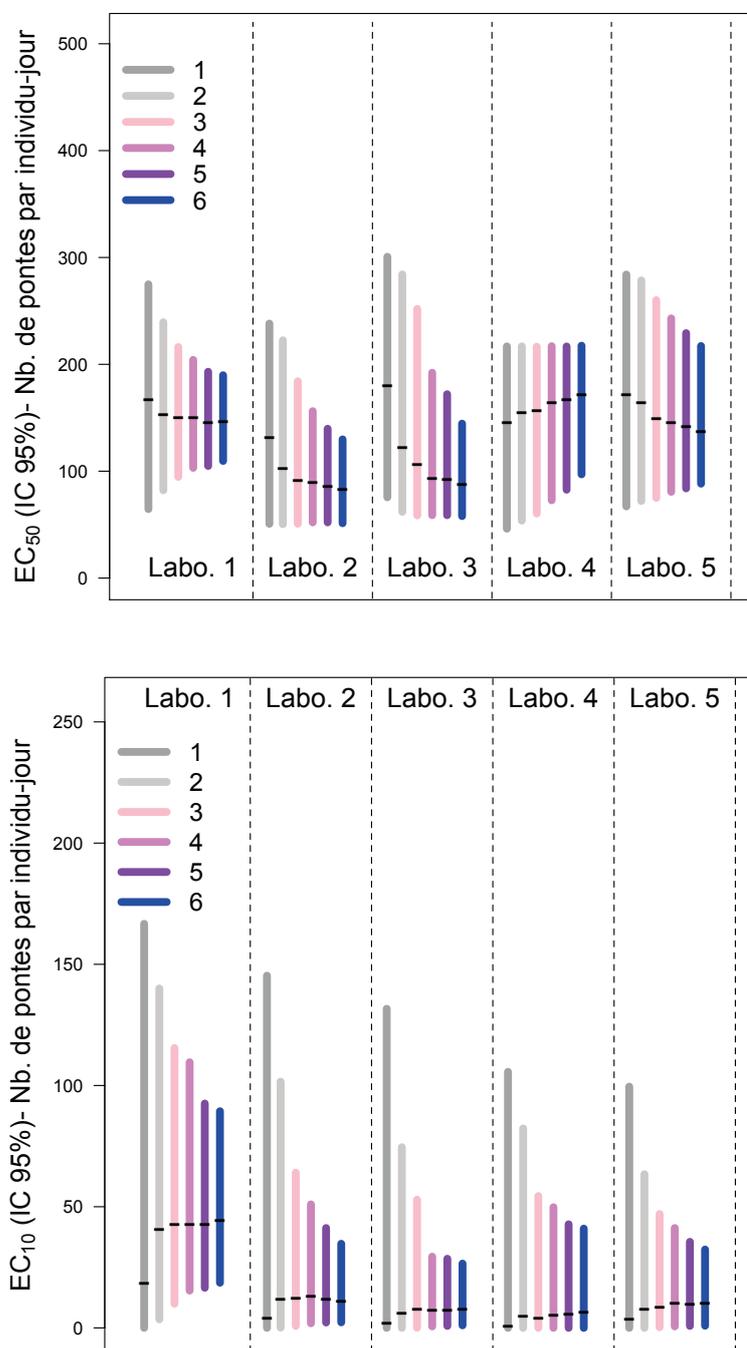


FIGURE 3.21 – Moyennes des estimations (tirets) et des bornes des intervalles de crédibilités à 95% (sur 50 jeux de données après tirage au sort pour chaque nombre de réplicat) des EC_{50} (graphes du haut) et des EC_{10} (graphes du bas) pour le nombre de pontes par individu-jour obtenus pour chaque laboratoire à 56 jours et pour un nombre de réplicats (de 1 à 6) codés par une couleur (expérience initiale en bleu).

Nous avons également réduit le nombre de réplicats pour la nouvelle durée d'expérience (38 jours), afin de voir les effets à la fois de la réduction du temps et de la réduction du nombre de réplicats sur les estimations et leurs intervalles de crédibilité à 95%. Les résultats sont présentés dans les Figures 3.22 pour le nombre d'oeufs par individu-jour et 3.23 pour le nombre de pontes par individu-jour. L'effet de la durée d'exposition sur les estimations est plus important que celui de la réduction du nombre de réplicats (hormis pour le laboratoire 3). De plus, les différences inter-laboratoires semblent moins marquées quand nous considérons les résultats obtenus avec quatre réplicats.

Concernant l'impact de la réduction du nombre de réplicats sur la précision des estimations pour le nombre d'oeufs par individu-jour, les résultats obtenus à 38 jours sont similaires à ceux obtenus à 56 jours (Tableau 3.13). Nous avons vérifié les valeurs des $R_{m95/m5}$ pour la variable nombre de pontes par individu-jour dans le cas d'une réduction du nombre de réplicats à quatre à 38 jours. Toutes les valeurs obtenues étaient inférieures au seuil que nous avons fixé (trois). Ainsi, un nombre de quatre réplicats par concentration peut être retenu pour l'expérience.

$R_{m95/m5}$ pour	Labo.1	Labo.2	Labo.3	Labo.4	Labo.5
6 réplicats	1.57	2.12	2.10	2.12	2.00
5 réplicats	1.60	2.31	2.46	2.13	2.18
4 réplicats	1.76	2.56	2.84	2.32	2.46
3 réplicats	1.91	2.87	3.46	2.58	2.72
2 réplicats.	2.11	3.46	4.19	3.33	3.45
1 rreplicat	3.55	4.19	3.97	3.99	4.13

Tableau 3.13 – Valeurs des $R_{m95/m5}$ obtenues à 38 jours pour les estimations des EC_{50} en ce qui concerne le nombre d'oeufs par individu-jour pour les différents laboratoires et pour un nombre de réplicats par concentration réduit de six à un. La valeur de $R_{m95/m5}$ en bleu correspond au nombre de réplicats le plus petit pour lequel cette valeur est inférieure ou égale à 3.

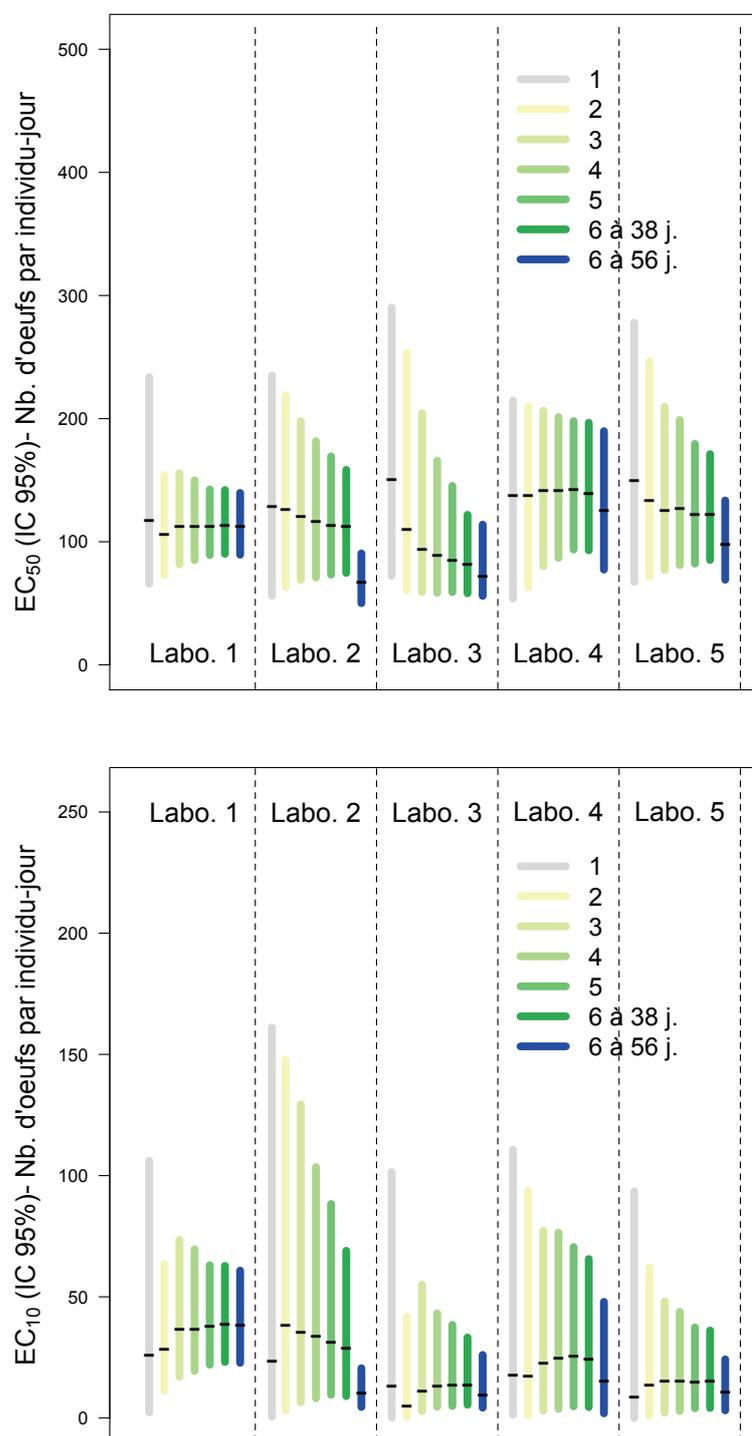


FIGURE 3.22 – Moyennes des estimations (tirets) et des bornes des intervalles de crédibilités à 95% (sur 50 jeux de données après tirage au sort pour chaque nombre de réplicat) des EC_{50} (graphes du haut) et des EC_{10} (graphes du bas) pour le nombre d'oeufs par individu-jour obtenus pour chaque laboratoire à 38 jours et pour un nombre de réplicats (de 1 à 6) codés par une couleur (expérience initiale en bleu). L'expérience avec 6 réplicats durant 56 jours est en bleu et en vert foncé pour la durée de 38 jours.

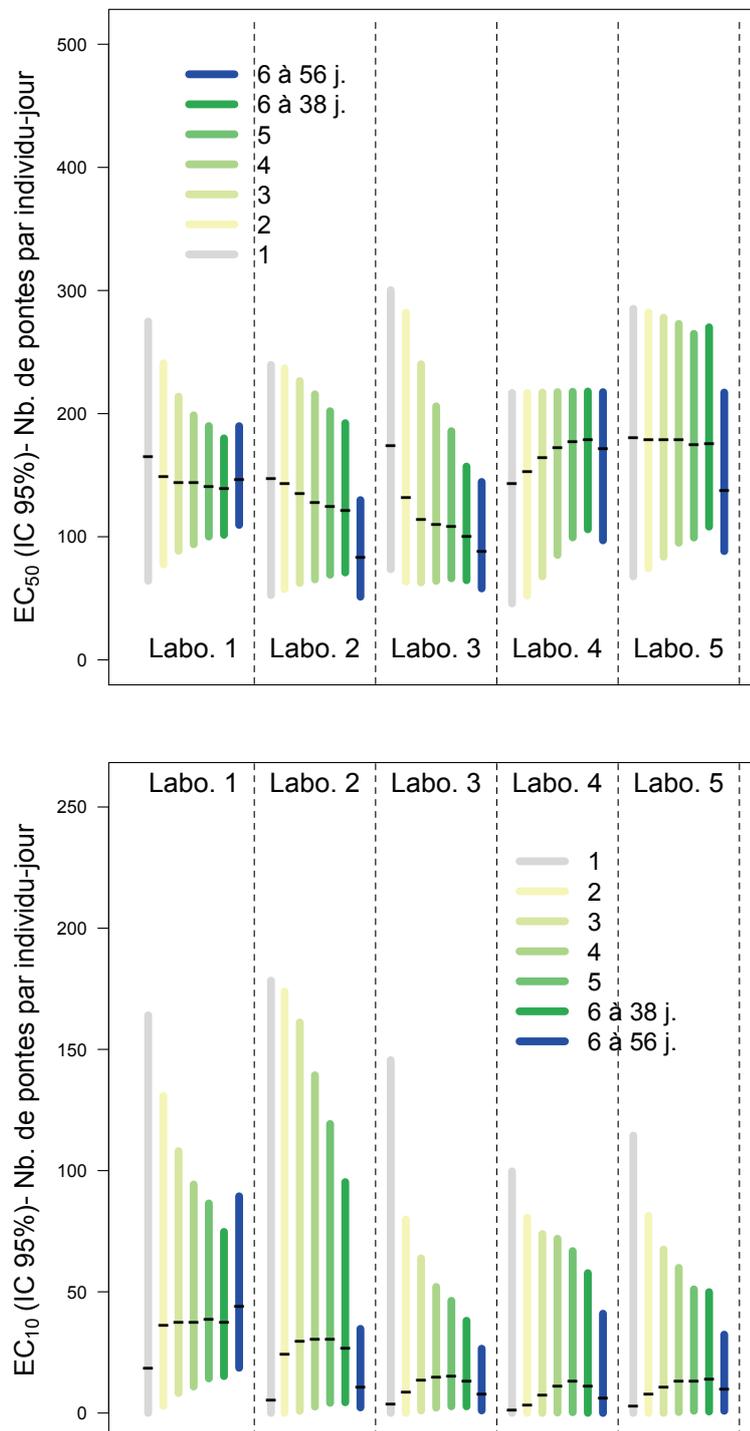


FIGURE 3.23 – Moyennes des estimations (tirets) et des bornes des intervalles de crédibilité à 95% (sur 50 jeux de données après tirage au sort pour chaque nombre de réplicat) des EC_{50} (graphes du haut) et des EC_{10} (graphes du bas) pour le nombre de pontes par individu-jour obtenus pour chaque laboratoire à 38 jours et pour un nombre de réplicats (de 1 à 6) codés par une couleur (expérience initiale en bleu). L'expérience avec 6 réplicats durant 56 jours est en bleu et en vert foncé pour la durée de 38 jours.

3.2.4 Conclusion sur l'optimisation d'un plan d'expérience en vue d'une standardisation chez la limnée.

Dans cette partie sur l'optimisation d'un plan d'expérience en reproduction, nous avons travaillé avec des données réelles puisque nous avons accès à cinq jeux de données au cours du temps pour une même expérience réalisée par cinq laboratoires différents. Ainsi, l'approche utilisée n'a pas été la même que celle pour les plans d'expérience dans le cadre des données de survie. Toutefois, nous aurions pu utiliser les gains de connaissance par exemple pour tester la réduction du nombre d'individus total en utilisant des informations *a priori*, issues des jeux de données que nous avons, pour créer des jeux de données simulées pour une durée d'exposition donnée.

En ce qui concerne la réduction du temps d'exposition, nous avons utilisé deux stratégies, celle "du pallier" et celle du "minimum", afin de déterminer un nouveau temps d'exposition pour lequel l'estimation des concentrations critiques d'effet ne serait pas très différente de celle à 56 jours. Les nouvelles durées d'exposition peuvent être différentes selon les laboratoires. Si nous choisissons la durée la plus courte, les nouvelles estimations obtenues sont tout à fait acceptables pour certains laboratoires (laboratoires 1, 3 et 4) et moins pour d'autres (laboratoires 2 et 5). Hormis pour le laboratoire 2, une réduction de la durée d'exposition à 38 jours peut être envisagée.

Concernant la réduction du nombre d'individus testés, nous avons choisi de réduire le nombre de réplicats par concentration. Pour que le rapport $R_{m95/m5}$, caractérisant la précision des estimations des EC_{50} , soit au maximum égal à 3.0 pour tous les laboratoires alors le nombre de réplicats par concentration peut être réduit à quatre au lieu de six que ce soit pour la durée de 56 jours que pour la durée réduite de 38 jours. Sachant que chaque réplicat est constitué de 5 individus, il est donc possible de réduire de 10 individus chaque concentration ainsi que le témoin, soit au total 60 individus pour l'expérience qui en comptait initialement 180.

Globalement, que ce soit pour la réduction de la durée de l'expérience ou pour la réduction du nombre d'individus testés, nous avons dû faire des choix concernant la perte d'information qui peut être acceptable ou non. Nous avons dû fixer des seuils de manière plus ou moins arbitraire en se basant sur les différences entre les laboratoires puisque nous sommes dans des conditions homogènes de test. Ces choix méritent d'être discutés et finalement, d'autres seuils peuvent être fixés de manière collégiale par le groupe de travail OCDE afin de pouvoir optimiser le plan d'expérience.

Conclusion :

Dans cette deuxième partie sur l'optimisation des plans d'expérience dans le cadre des données de reproduction, nous avons pu voir que :

- hormis pour un des laboratoires, une réduction du temps d'exposition de 56 jours à 38 jours est possible sans que les estimations ne soient trop affectées
- à 56 jours et à 38 jours, le nombre de réplicats par concentration peut être réduit de six à quatre.

Globalement, les résultats obtenus dépendent de la perte d'information considérée comme "acceptable", ils peuvent évoluer en fonction des décisions prises par les écotoxicologues.

Chapitre 4

Conclusion et Perspectives

Ce travail de thèse sur l'estimation des concentrations critiques d'effet en vue d'une évaluation quantitative du risque a été réalisé dans un contexte de débat animé sur les méthodes à utiliser. D'ailleurs, certains titres de publications montrent à quel point les avis sont tranchés sur l'utilisation de certaines concentrations critiques d'effet : c'est le cas des travaux de [Warne et van Dam \(2008\)](#) intitulés "NOEC and LOEC data should no longer be generated or used", de ceux de [Jager \(2011\)](#) intitulés "Some good reasons to ban EC_x and related concepts in ecotoxicology" et ceux de [Green et al \(2012\)](#) intitulés "The drive to ban NOEC/LOEC in favor of EC_x is misguided and misinformed". Nous avons décidé de comparer les méthodes et modèles proposés pour aboutir, nous l'espérons, à une meilleure analyse des données ou du moins pour faire ressortir les messages importants concernant l'estimation des concentrations critiques d'effet de manière plus claire que ce qui est présent dans le guide [OECD \(2006\)](#).

Pour cela, nous avons tout d'abord dressé une liste non exhaustive des principaux reproches faits aux différentes concentrations critiques d'effet. Puis, nous avons décidé de nous intéresser à deux traits d'histoire de vie pour lesquels la nature des données récoltées est particulière : les données de survie de type binaire (dichotomique) et les données de reproduction de type comptage. En effet, la nature des données nécessite de les traiter de manière appropriée ([Bailer and Oris, 1997](#)). Cependant, alors que les données binaires sont nettement distinguées dans le guide OCDE, celles de comptage ne le sont pas puisqu'il est recommandé de les traiter comme des données continues ([OECD, 2006](#)). Quand les conditions de normalité ne sont pas vérifiées, une transformation des données en logarithme est réalisée, c'est ce qui est fait par exemple dans [De Schamphelaere et Janssen \(2004\)](#).

L'amélioration de l'estimation des concentrations critiques d'effet peut être obtenue en améliorant les méthodes d'analyse des données. Cette amélioration peut aussi être obtenue en générant des données plus adaptées aux concentrations critiques d'effet à estimer et ceci en optimisant les plans d'expérience. En effet, le choix des concentrations, de leur nombre, de leurs valeurs affecte la précision des estimations ([Chapman et al, 1996](#)). Ainsi, pour les deux traits d'histoire de vie (survie et reproduction), nous nous sommes intéressés à ces deux aspects sur lesquels il est possible agir pour améliorer l'estimation des concentrations critiques d'effet : l'amélioration de l'analyse des données et l'optimisation des plans d'expérience. Ainsi en analysant des données de type survie,

nous avons pu comparer les différentes méthodes et modèles permettant de déterminer les concentrations critiques d'effet (NOEC/LOEC et LC_x). Nous avons pu conclure que le choix de la méthode est important pour la détermination des NOEC, ce qui avait été montré par [Isnard et al \(2001\)](#) mais pour des données de croissance et de reproduction. La méthode qui nous semblait la plus protectrice pour l'environnement était celle de Cochran-Armitage associée à une procédure "step-down". Pour les LC_x , c'est le modèle de Weibull que nous conseillons, même si le modèle log-logistique est également très performant. Ces résultats rejoignent également ceux obtenus par [Isnard et al \(2001\)](#) mais toujours pour des données de croissance et de reproduction. Concernant la NEC, nous avons choisi de prendre un modèle simple à seuil, qui puisse être flexible et prendre en compte la mortalité naturelle ou un éventuel effet maximal en utilisant toutes les données temporelles disponibles. Notre stratégie consiste à utiliser un modèle, le plus simple possible, et de ne le complexifier que si les données le nécessitent. Finalement, notre modèle se rapproche du modèle simplifié parmi les modèles possibles offerts par le modèle GUTS (General Unified threshold model of Survival) ([Jager et al, 2011](#)), qui est un modèle de toxicocinétique et toxicodynamique. Au niveau du compartiment toxicocinétique, notre modèle équivaut à GUTS mais avec une cinétique très rapide qui fait que la concentration du contaminant à l'intérieur de l'organisme est égale à la concentration externe. Au niveau toxicodynamique, notre modèle équivaut à GUTS mais en ne prenant pas en compte les dommages liés au contaminant et en considérant que l'effet est uniquement la mort de l'individu. Le but étant de pouvoir analyser des jeux de données très simples qui ne permettent pas forcément d'estimer des paramètres de toxicocinétique et de toxicodynamique. La comparaison des différentes concentrations critiques d'effet nous a permis de voir l'intérêt de l'utilisation du modèle NEC que nous proposons, notamment en ce qui concerne la prise en compte de la mortalité naturelle comme n'étant pas imputable à la substance étudiée. Sur cette analyse des données de survie, que ce soit pour estimer des LC_x ou des NEC, nous n'avons pas pris en compte les informations apportées par les réplicats, comme c'est généralement le cas, ce qui est critiquable ([Green et al, 2012](#)). En effet, ces informations apportées par les réplicats ne sont pas toujours disponibles et quand on a des jeux de données complets dans la littérature (et non pas uniquement des moyennes), les données issues des réplicats sont regroupées et pas distinguées en fonction du réplicat. L'analyse de ces données avec réplicats pourraient être une des perspectives intéressantes concernant l'amélioration de l'analyse des données avec le souci toujours en filigrane de ne pas gaspiller de l'information recueillie pendant l'expérimentation.

Dans un deuxième temps, nous avons choisi de travailler sur les plans d'expérience en vue d'estimer des LC_x dans le cadre de la survie. Comme [Bell et al \(2005\)](#), nous avons utilisé des jeux de données simulés pour cette étude tout en choisissant de tester au final des plans d'expérience réalisables et proches de ce qui est préconisé actuellement. Pour comparer les plans d'expérience entre eux, nous avons repris une approche de comparaison des plans d'expérience par le calcul des gains de connaissance. Cette approche utilisée par [Albert et al \(2012\)](#) présente à la fois l'avantage d'être calculable quelle que soit la LC_x ou EC_x que l'on souhaite estimer, mais elle renseigne aussi directement sur la plus-value que l'on peut attendre par rapport à l'information initiale que l'on avait sur la concentration critique d'effet d'intérêt. Afin de valider les résultats

obtenus par une méthode plus classique, nous avons en parallèle utilisé une démarche plus traditionnelle en optimisation des plans d'expérience qui est l'utilisation de la D-optimalité (Chernoff, 1953; Chaloner et al, 1989; King et Wong, 2000). Ainsi, les principaux résultats sont :

- qu'il vaut mieux centrer la gamme de concentrations sur la valeur de la LC_{50} attendue, même si on veut estimer une LC_5 . Et pour cela, la récolte d'information *a priori* sur la LC_{50} est importante.
- que sur une même étendue de gamme de concentrations pour un même nombre d'individus exposés, il est préférable d'augmenter le nombre de concentrations, c'est-à-dire de diminuer l'écart entre les concentrations (diminuer la raison de la progression géométrique). Etant donné que nous sommes à nombre d'individus fixés cela revient à diminuer le nombre de réplicat par concentration. Les travaux de Muller et Schmitt (1990) avaient également conclu qu'une estimation plus précise des LC_{50} pouvait être obtenue en utilisant plus de concentrations et moins de réplicats. Cependant, ces travaux montraient qu'un nombre de concentrations supérieur à 12 n'apportait pas un gain de précision plus important. Ce résultat rejoint celui que nous avons obtenu avec notre étude théorique utilisant la D-optimalité localement. Nous avons pu voir qu'au delà de 20 concentrations, toutes choses étant égales par ailleurs, les plans d'expérience donnent des valeurs de déterminants quasiment identiques et qu'entre 10 et 20 concentrations les valeurs sont encore très proches. Les recommandations du guide OCDE sur la reproduction de la daphnie (test chronique) (OCDE, 2012) ont évolué dans le sens des résultats que nous avons obtenus : pour l'estimation des EC_x , il est recommandé de tester plus de concentrations (10) avec moins de réplicats (5).
- qu'il est préférable d'avoir une étendue de gamme de concentrations la plus restreinte possible. Toutefois, si cela n'est pas possible le fait de mettre moins de poids sur les concentrations extrêmes est favorable au gain de connaissance sur les LC_x .

Notre deuxième partie portait sur les données de reproduction issues d'essais sur la limnée en vue d'une standardisation du test de reproduction chez les mollusques. Wang et Smith (2000) avaient souligné dans leurs travaux le problème que pouvait présenter la mortalité pour interpréter les résultats de données de reproduction. Ils ont proposé de différencier le "zéro de reproduction" dû à un effet réel sur la reproduction, du "zéro de reproduction" dû à la mort de la mère au cours du test. Une façon d'éviter le problème de la mortalité pendant les tests est de ne pas prendre en compte les résultats obtenus pour les mères qui ne survivent pas jusqu'à la fin du test, comme cela est recommandé par exemple dans le guide sur la reproduction de la daphnie récemment modifié (OCDE, 2012). Dans le travail que nous avons présenté, nous avons pu voir que la mortalité survenant au cours de l'expérience pouvait être traitée autrement qu'en enlevant les réplicats avec mortalité, mais en calculant un nombre d'individu-jours pour l'expérience. Le nombre de pontes ou d'oeufs rapporté au nombre d'individu-jours de l'expérience permet d'éviter le gaspillage de données. Ce qui permet même d'ajuster une courbe concentration-réponse. En effet, pour les fortes concentrations, il n'est pas rare d'observer de la mortalité dans les réplicats et jusqu'alors, cela pouvait engendrer une perte totale d'information pour ces concentrations et donc réduire le nombre

de concentrations de l'expérience et rendre inutilisable l'ensemble des données. Nous avons pu également prendre en compte le fait que les données de reproduction sont des données de comptage et non pas des données continues en utilisant des parties stochastiques suivant une loi de Poisson, de gamma-Poisson ou une loi binomiale négative au lieu d'utiliser une loi normale avec ou sans transformation des données, ceci va dans le sens de ce que préconise O'Hara and Kotze (2010). La comparaison des différents modèles nous a permis de déterminer que le meilleur modèle semblait être un modèle de Weibull avec une partie stochastique de type gamma-Poisson. Le choix de la partie déterministe du modèle s'est porté sur le modèle de Weibull qui était le plus protecteur pour l'environnement en ce qui concerne l'estimation des EC_{10} . Ce résultat va dans le sens de ce qu'avait conclu Scholze et al (2001) sur la comparaison des modèles logit, probit et Weibull. En effet, pour les faibles x , les EC_x sont surestimées par les modèles logit et probit alors que le modèle de Weibull aurait tendance à les sous-estimer et donc à être plus protecteur pour l'environnement. Nous n'avons pas déterminé de NOEC ni de LOEC pour ces données, puisque dans notre première partie sur les données de survie, nous avons vu qu'elles étaient les concentrations critiques d'effet les moins pertinentes. Nous n'avons pas non plus abordé le calcul de la NEC (commune à tous les temps d'exposition) pour les données de reproduction, car c'est un problème complexe si on veut prendre en compte la reproduction comme ce qu'elle est réellement, c'est-à-dire un processus non continu au cours du temps. Ce processus peut être affecté à différents niveaux : l'âge de la première ponte, la durée séparant chaque production de nouveaux individus, le nombre d'individus produits à chaque ponte... Et même le modèle complet DEBtox prend en compte la reproduction comme un processus continu dans le temps (Billoir et al, 2008b). Il serait donc tout à fait intéressant d'envisager de proposer un nouveau modèle permettant de décrire la reproduction au cours du temps de façon mécaniste et d'essayer d'estimer une NEC.

Concernant les plans d'expérience, dans le cadre d'un travail réalisé pour proposer un test standard de reproduction sur les mollusques, nous avons pu travailler sur la réduction de la durée d'un test de toxicité ainsi que sur la réduction du nombre de réplicats par concentration. En ce qui concerne la réduction de la durée d'exposition, nous avons pu constater que pour certains jeux de données, quand l'évolution des EC_x est tracée en fonction du temps, il n'est pas rare de voir apparaître un "palier" au bout d'une certaine durée. C'est à partir de cette constatation que nous avons développé deux stratégies permettant de réduire la durée d'expérience, l'une basée sur l'apparition d'un palier et l'autre sur un minimum atteint au cours de l'expérience dans le cas où il n'y aurait pas de palier. Nous avons pu voir qu'une réduction à 38 jours pouvait être envisagée sans que cela n'affecte de manière trop importante les estimations des EC_x d'intérêt. D'ailleurs, il est difficile de déterminer ce qu'on considère comme étant un niveau acceptable de différence sur les estimations. Nous avons décidé de considérer que moins de 10% de différence pouvait être acceptable, mais nous devons à l'avenir nous en remettre aux écotoxicologues pour un choix moins arbitraire. De même, nous avons pu présenter des résultats sur les estimations des paramètres d'intérêt dans le cas où on réduit le nombre de réplicats par concentration à la fois à 56 jours et à 38 jours. Nous avons pu constater que l'impact de la durée d'exposition semblait plus important que la réduction du nombre de réplicats. Nous avons calculé le ratio $R_{95/5}$,

comme l'ont fait Ruffi et Springer (2011) et nous avons fixé comme critère une valeur maximale égale à trois pour ce ratio. Ainsi, il est possible de passer de six à quatre réplicats par concentration à 56 jours comme à 38 jours, ce qui réduit de 180 à 120 le nombre d'individus à utiliser pour réaliser ce test de reproduction.

Pour terminer sur des perspectives plus générales en écotoxicologie, nous pouvons dire que cette science est relativement récente et qu'elle est née et a évolué à des fins d'évaluation du risque. Ainsi, l'évaluation du risque et l'écotoxicologie sont très intimement liées et c'est ce qui pose problème. De nouvelles méthodes, de nouveaux modèles pourraient faire évoluer rapidement l'écotoxicologie, cependant elle est retenue par la nécessité de rentrer dans le "moule" de l'évaluation du risque. En effet, les recommandations liées aux tests standards évoluent lentement et les écotoxicologues, faisant de l'évaluation du risque, doivent donc toujours suivre ces recommandations même si d'autres outils sont à leur disposition. Ainsi, malgré les inconvénients qu'elles représentent, les NOEC et des LOEC sont toujours présentes dans la nouvelle recommandation sur le test de reproduction chez la daphnie d'octobre 2012 (OCDE, 2012). Il faut donc contribuer à mettre à disposition des écotoxicologues des outils simples et intuitifs pour qu'ils puissent estimer d'autres concentrations critiques d'effet. L'une des perspectives de ce travail de thèse, est de proposer, via une interface web, un outil permettant d'utiliser les méthodes proposées dans cette thèse. Concernant les données de survie, notre modèle NEC permettant de prendre en compte la mortalité naturelle et les données au cours du temps pourrait être mis à disposition via cette interface. Pour l'estimation des LC_x , un package *drc* (Ritz et Streibig, 2005) permet d'analyser de manière correcte les données au temps final. Nous pouvons aussi envisager de proposer un outil d'aide à la mise en place de plans d'expérience basé sur le calcul des gains de connaissance. Cet outil permettrait à l'utilisateur de rentrer des paramètres comme le nombre maximal d'individus qu'il peut tester, la concentration critique d'effet qu'il veut estimer, les informations *a priori* qu'il a sur cette concentration etc...pour obtenir le plan d'expérience optimal par rapport à ses contraintes. Concernant la reproduction, il pourrait être intéressant de proposer un outil permettant le calcul du nombre d'individu-jours ainsi que l'utilisation d'un modèle concentration-réponse avec une partie stochastique de type gamma-Poisson basée sur l'inférence bayésienne (package *rjags* Plummer (2010)). C'est dans ce but qu'au sein de l'équipe MEPS (Modélisation et Ecotoxicologie Prédictives) du Laboratoire de Biométrie et Biologie Evolutive, il a été choisi de développer un groupe d'outils appelé MOSAIC (MOdeling and StAtistical tools for ecotoxICology), dont le premier module SSD a été mis en ligne récemment. Il est prévu à l'horizon 2014 un module "repro" pour l'analyse des données de reproduction et un module "plan" pour l'aide à la planification expérimentale <http://pbil.univ-lyon1.fr/software/mosaic/>.

Bibliography

- Adema DMM (1978) *Daphnia magna* as a test animal in acute and chronic toxicity tests. *Hydrobiologia* 59 :125–134
- Albert C, Ashauer R, Künsch HR, Reichert P (2012) Bayesian experimental design for a toxicokinetic-toxicodynamic model. *Journal of statistical planning and inference* 142 :263–275
- Altman D and Bland J (1995) Absence of evidence is not evidence of absence. *British Medical Journal* 311 :485–485
- Andersen JS, Bedaux JJM, Kooijman SALM and Holst H (2000) The influence of design characteristics on statistical inference in non-linear estimation; a simulation study based on survival data and hazard modeling. *Journal of Agricultural, Biological and Environmental Statistics* 5 :323–343
- Armitage P and Colton T (2005) Trend test for counts and proportions in *Encyclopedia of Biostatistics* second edition, vol 8 (5516-5522). Wiley, Chichester, England
- Baas J, Jager T and Kooijman SALM (2009) Estimation of no effect concentrations from exposure experiments when values scatter among individuals. *Ecological modelling* 220 :411–418
- Bailer AJ and Oris JT (1997) Estimating inhibition concentrations for different response scales using generalized linear models. *Environmental Toxicology and Chemistry* 16 (7) :1554–1559
- Bedaux J and Kooijman SALM (1994) Statistical analysis of bioassays, based on hazard modeling. *Environmental and Ecological statistics* 1 :303–314
- Bell BR, Bailer AJ and Wright SE (2005) A simulation study comparing different experimental designs for estimating uptake and elimination rates. *Environmental Toxicology and Chemistry* 25 :248–252
- Billoir E, Delignette-Muller ML, Péry A, Charles S (2008a) A Bayesian approach to analyzing ecotoxicological data. *Environmental Science and Technology* 42 :8978–84
- Billoir E, Delignette-Muller ML, Péry A, Geffard O and Charles S (2008b) Statistical cautions when estimating DEBtox parameters. *Journal of Theoretical Biology* 254 :55–64

- Billoir E, Delhaye H, Forfait C, Clément B, Triffault-Bouchet G, Charles S and Delignette-Muller ML (2012) Comparison of bioassays with different exposure time patterns : The added value of dynamic modelling in predictive ecotoxicology. *Ecotoxicology and Environmental Safety* 75 :80–86
- Biron PA, Massarin S, Alonzo F, Garcia-Sanchez L, Charles S and Billoir E (2012) Population-Level Modeling to Account for Multigenerational Effects of Uranium in *Daphnia magna*. *Environmental Science and Technology* 46 :1136–43
- de Bruijn J and Hof M (1997) How to measure no effect. Part IV : How acceptable is the EC_x from an environmental policy point of view? *Environmetrics* 8 :263–267
- Cairns J (1992) The threshold problem in ecotoxicology. *Ecotoxicology* 1 :3–16
- Chapman PM, Caldwell R and Chapman PF (1996) A warning : NOECs are inappropriate for regulatory use. *Environmental Toxicology and Chemistry* 15 :77–79
- Chapman PF, Crane M, Wiles J, Noppert F and McIndoe E (1996) Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* 5 :169–186
- Chaloner K and Larntz K (1989) Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference* 21, 191–208
- Chernoff H (1953) Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics* 24, 586–602
- Chèvre N (2000) Etude et modélisation des effets écotoxiques d'un micropolluant organique sur *Daphnia magna* et *Pseudokirchneriella subcapitata*. Ph.D. Thesis 2117, Swiss Federal Institute of Technology, Lausanne
- Chèvre N, Brazzale AR, Becker-van Slooten K, Behra R, Tarradellas J and Guettinger H (2005) Modelling the concentration-response function of the herbicide dinoseb on *daphnia magna* (survival time, reproduction) and *Pseudokirchneriella subcapitata* (growth rate). *Ecotoxicology and Environmental Safety* 62, 17–810
- Chèvre N and Brazzale AR (2000) Cost-effective experimental design to support modeling of concentration-response functions. *Chemosphere* 72, 803–810
- Crane M and Newman M (2000) What level of effect is a no observed effect? *Environmental Toxicology and Chemistry* 19 :516–519
- van Dam R A, Harford A J and Warne M St J (2012) Time to get off the fence : the need for definitive international guidance on statistical analysis of ecotoxicity data. *Integrated Environmental Assessment and Management* 8(2) :242–245
- van Dam R A, Harford A J and Warne M St J (2012) Canada showing the lead, however, we still have a NOEC problem : response to Van der Vliet et al. (2012) *Integrated Environmental Assessment and Management* 8(3) :399–400

- Delignette-Muller ML, Pouillot R, Denis JB and Dutang C (2010) `fitdistrplus` : help to fit of a parametric distribution to non-censored or censored data. R package version 0.1-3
- Delignette-Muller ML, Forfait C, Billoir E and Charles S (2011) A new perspective on the Dunnett procedure : filling the gap between NOEC/LOEC and EC_x concepts. *Environmental Toxicology and Chemistry* 30 :2888–28911
- Dilba G, Bretz F and Guiard (2006) Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal of Statistical Planning and Inference* 136 :2640–2658
- Dilba Djira G, Hasler M, Gerhard D and Schaarschmidt F (2011) `mratios` : Inferences for ratios of coefficients in the general linear model. R package version 1.3.16. <http://CRAN.R-project.org/package=mratios>
- Ducrot V, Billoir E, Péry A, Garric J and Charles S (2010) From individual to population level effects of toxicants in the tubicifid *Branchiura sowerbyi* using threshold effect models in a Bayesian framework. *Environmental Science and Technology* 44 :3566–71
- Ducrot V, Teixeira-Alves M, Lopes C, Delignette-Muller M L, Charles S and Lagadic L (2010) Development of partial life-cycle experiments to assess the effects of endocrine disruptors on the freshwater gastropod *Lymnaea stagnalis* : a case-study with vinclozolin. *Ecotoxicology* 19 :13126–1321
- Dunnett CW (1964). New Tables for Multiple Comparisons with a Control. *Biometrics* 20(3) :484–491
- Ebert D (2005) Ecology, Epidemiology, and Evolution of Parasitism in *Daphnia* [Internet]. Bethesda (MD) : National Center for Biotechnology Information (US). Available from : <http://www.ncbi.nlm.nih.gov/books/NBK2036/>
- ECB (2003) Technical guidance document on risk assessment. European Chemicals Bureau.
- European parliament and the council of european union (2007) Regulation (EC) No 1907/2006 of the european parliament and of the council of 18 December 2006 concerning the registration, Evaluation, Authorization and restriction of Chemicals (REACH). *Official Journal of the European Union* 29.5.2007 L 136/3
- Flammarion P, Devaux A et Garric J (2000) Marqueurs biochimiques de pollution dans les écosystèmes aquatiques continentaux. Exemples d'utilisation et perspectives pour le gestionnaire. *Bulletin Français de la Pêche et de la Pisciculture*. 357/358 : 209-226
- Forfait-Dubuc C, Charles S, Billoir E and Delignette-Muller ML (2012) Survival data analyses in ecotoxicology : critical effect concentrations, methods and models. What should we use ? *Ecotoxicology* 21 :1072–1083

- Fox D (2008) NECS, NOECS and the EC_x . *Australasian Journal of Ecotoxicology* 14 :7–9
- Fox D (2010) A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicology and Environmental Safety* 73 :123–131
- Fox D R, Billoir E, Charles S, Delignette-Muller M L and Lopes C (2012) What to do with NOECS/NOELS- prohibition or innovation ? *Integrated Environmental Assessment and Management* 8 :764–766
- Gardner MJ and Altman DG. 1986. Confidence intervals rather than P values : Estimation rather than hypothesis testing. *British Medical Journal* 292 :746– 750.
- Gelman A, Rubin D (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7 :457–511
- Green J W, Springer T A and Staveley J P (2012) The drive to ban NOEC/LOEC in favor of EC_x is misguided and misinformed. *Integrated Environmental Assessment and Management*. DOI 10.1002/ieam. 1367
- Hirano M, Ishibashi H, Matsumura N, Nagao Y, Watanabe N, Watanabe A, Onikura N, Kishi K and Arizono K (2004) Acute toxicity responses of two crustaceans, *Americamysis bahia* and *Daphnia magna*, to endocrine disrupters. *Journal of Health Science* 50(1) :97–100
- van der Hoeven N (1997) How to measure no effect. Part III : Statistical aspects of NOEC, EC_x and NEC estimates. *Environmetrics* 8 :255–261
- van der Hoeven N (2004) Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheoretica* 52 :201–217
- van der Hoeven N, Noppert F and Leopold A (1997) How to measure no effect. Part I : Towards a new measure of chronic toxicity in ecotoxicology. Introduction and workshop results. *Environmetrics* 8 :241–248
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6 :65–70
- Isnard P, Flammarion P, Roman G, Babut M, Bastien P, Bintein S, Esserméant L, Féraud J, Gallotti-Schmitt S, Saouter E, Saroli M, Thiébaud H, Tomassone R and Vindimian E (2001) Statistical analysis of regulatory ecotoxicity tests. *Chemosphere* 45 :659–69
- Jager T, Crommentuijn T, Van Geste CAM and Kooijman SALM (2004). Simultaneous modeling of multiple endpoints in life-cycle toxicity tests. *Environmental Science and Technology* 38 :2894–2900.

- Jager T, Albert C, Preuss T G and Ashauer R (2011). General unified threshold model of survival – a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental Science and Technology* 45, 2529–2540.
- Jager T (2011) Some good reasons to ban EC_x and related concepts in ecotoxicology. *Environmental Science and Technology* 45 :8180–8181
- Jager T and Zimmer E I (2012) Simplified Dynamic Energy Budget model for analysing ecotoxicity data. *Ecological Modelling* 225 :74–81
- Jager T (2012) Bad habits die hard : the NOEC's persistence reflects poorly on ecotoxicology. *Environmental Toxicology and chemistry* 31 :228–229
- Jonckheere A (1954) A distribution-free k-sample test against ordered alternatives. *Biometrika* 41 :133–145
- Jumel A, Coutellec M A, Cravedi J P and Lagadic L (2002) Nonylphenol polyethoxylate adjuvant mitigates the reproductive toxicity of fomesafen to the freshwater snail *Lymnaea stagnalis* in outdoor experimental ponds. *Environmental Toxicology and Chemistry*. 21 :1876–1888.
- Kim HY, Lee MJ, Yu SH and Kim SD (2012) The individual and population effects of tetracycline on *Daphnia magna* in multigenerational exposure. *Ecotoxicology* 21 :993–1002
- Kamaya Y, Tokita N and Suzuki K (2005) Effects of dehydroabietic acid and abietic acid on survival, reproduction, and growth of the crustacean *Daphnia magna*. *Ecotoxicology and environmental safety* 61 :83–88
- King J and Wong WK (2000) Minimax D-optimal designs for the logistic model. *Biometrics* 56, 1263–1267
- Kooijman SALM and Bedaux J (1996) *The Analysis of Aquatic Toxicity Data*. VU University Press, Amsterdam, The Netherlands
- Lagadic L, Coutellec M-A and Caquet T (2007) Endocrine disruption in aquatic pulmonate molluscs : few evidences, many challenges. *Ecotoxicology* 16 :45–59
- Landis WG (2012) Comment on Fox et al. (2012) : What to do with NOECs/NOELs. *Integrated Environmental Assessment and Management* 8 :764–766
- Li MH (2010) Chronic effects of perfluorooctane sulfonate and ammonium perfluorooctanoate on biochemical parameters, survival and reproduction of *Daphnia magna*. *Journal of Health Science* 56(1), 104–111
- Lopes C, Charles S, Vollat B and Garric J (2009) Toxicity of ivermectin on cladocerans : comparison of toxic effects on *Daphnia* and *Ceriodaphnia* species. *Environmental Toxicology and Chemistry* 28 :2160–2166

- Manar R, Bessi H and Vasseur P (2009) Reproductive effects and bioaccumulation of chlordane in *Daphnia magna*. *Environmental Toxicology and Chemistry* 28 :2150–2159
- Matthiessen P (2008) An Assessment of Endocrine Disruption in Mollusks and the Potential for Developing Internationally Standardized Mollusk Life Cycle Test Guidelines. *Integrated Environmental Assessment and Management* 4(3) :274–284
- Molle L (1984) Eloge du Professeur René Truhaut. *Revue d'histoire de la pharmacie* 262 :340–348
- Moore DRJ and Caux P-Y (1997) Estimating low toxic effects. *Environmental Toxicology and Chemistry* 16 :794–801
- Muller HG and Schmitt T (1990) Choice of number of doses for maximum likelihood estimation of the EC50 for quantal dose-reponse data. *Biometrics* 46 :117–129.
- Nelder J (1999) From statistics to statistical science. *Journal of the Royal Statistical Society Series D (The Statistician)* 48 :257–267
- Newman MC(1995) *Quantitative methods in aquatic ecotoxicology*, Chelsea, MI : CRC/Lewis Press, Inc.
- Newman MC (2008) "What exactly are you inferring?" a closer look at hypothesis testing. *Environmental Toxicology and Chemistry* 27 :1013–1019
- Newman MC and Clements W H(2008) *Ecotoxicology : A comprehensive treatment*. Boca Raton (FL) :CRC Press. 852 p.
- Newman MC (2012) Reluctant comment on Fox et al. (2012) on being dragged into the NOEC squabble. *Integrated Environmental Assessment and Management* 8 :767
- Nisbet R M, Muller E B, Lika K and Kooijman S A L M (2000) From molecules to ecosystems through dynamic energy budget models. *Journal of Animal Ecology* 69 :913–926
- Ntzoufras I (2009) *Bayesian modeling using WinBugs*. Wiley Series in Computational Statistics, Hoboken, USA
- OECD (1984) Ligne directrice l'OCDE pour les essais de produits chimiques n° 206 : Oiseaux, essai de reproduction. Tech. rep., Organisation for Economic Cooperation and Development
- OECD (1998) Guideline for testing of chemicals n° 211 : *Daphnia magna* reproduction test. Tech. rep., Organisation for Economic Cooperation and Development
- OECD (2004) Guideline for testing of chemicals n° 202 : *Daphnia* sp. Acute Immobilisation Test. Tech. rep., Organisation for Economic Cooperation and Development

- OECD (2004) Lignes directrices l'OCDE pour les essais de produits chimiques n° 220 : Essai de reproduction chez l'enchytrée. Tech. rep., Organisation for Economic Cooperation and Development
- OECD (2006) Current approaches in the statistical analysis of ecotoxicity data : a guidance to application. Tech. rep., Organisation for Economic Cooperation and Development
- OECD (2008) Lignes directrices l'OCDE pour les essais de produits chimiques n° 226 : Essai de reproduction d'un acarien prédateur (*Hypoaspis (Geolaelaps) aculeifer*) dans le sol. Tech. rep., Organisation for Economic Cooperation and Development
- OECD (2009) Guideline for testing of chemicals n° 229 : Fish short term reproduction assay. Tech. rep., Organisation for Economic Cooperation and Development
- OECD (2010) Series on testing and Assessment n° 121 : Detailed review paper (DRP) on mollusks life-cycle toxicity testing. Organisation for Economic Cooperation and Development
- OECD (2012) Guideline for testing of chemicals n° 211 : *Daphnia magna* reproduction test. Tech. rep., Organisation for Economic Cooperation and Development
- O'Hara RB and Kotze DJ (2010) Do not log-transform count data. *Methods in Ecology and Evolution* 1 :118–122
- Parent E et Bernier J (2007) Le raisonnement bayésien. Modélisation et inférence. Springer-Verlag France, Paris
- Pires A, Branco J, Picado A, and Mendonca E (2002) Models for the estimation of a 'no effect concentration'. *EnvironMetrics* 13(1) :15–27
- Plummer M (2010) R package version 2.1.0-10
- R Development Core Team (2010) R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria,
- Ramade F (2010) Introduction à l'écotoxicologie. Fondements et applications. Lavoisier, Paris
- Ritz C and Streibig JC (2005) Bioassays analysis using R. *Journal of Statistical Software* 12 (5)
- Ritz C (2010) Toward a unified approach to dose-response modeling in ecotoxicology. *Environmental Toxicology and Chemistry* 29 :220–229
- Ruffi H and Springer TA (2011) Can we reduce the number of fish in the OECD acute toxicity test ?. *Environmental Toxicology and Chemistry* 30(4) :1006–1011
- Russo J and Lagadic L (2004) Effects of environmental concentrations of atrazine on hemocyte density and phagocytic activity in the pond snail *Lymnaea stagnalis* (Gastropoda, Pulmonata). *Environmental Pollution* 127 :303–311

- Russo J, Lefeuvre-Orfila L and Lagadic L (2007) Hemocyte-specific responses to the peroxidizing herbicide fomesafen in the pond snail *Lymnaea stagnalis* (Gastropoda, Pulmonata). *Environmental Pollution* 127 :303–311.
- Sanchez-Bayo Francisco (2012) Should we forget NOECs? *Integrated Environmental Assessment and Management* 8 :564–565
- Santojanni A, Gorbi G and Sartore F (1995) Prediction of mortality in chronic toxicity tests on *Daphnia magna*. *Water Research* 29 :1453–1459
- De Schamphelaere KAC and Janssen CR (2004) Effects of chronic dietary copper exposure on growth and reproduction of *Daphnia magna*. *Environmental Toxicology and Chemistry* 23(8) :2038–2047
- Scholze M, Boedeker W, Faust M, Backhaus T, Altenburger R and Grimme LH (2001) Effects of chronic dietary copper exposure on growth and reproduction of *Daphnia magna*. *Environmental Toxicology and Chemistry* 23(8) :2038–2047
- Spiegelhalter D, Best N, Carlin B and van der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society series B* 64 :583–639
- van Straalen N (1997) How to measure no effect. Part II : Threshold effects in ecotoxicology. *Environmetrics* 8 :249–253
- Van der Vliet L, Taylor, L N and Scroggins R (2012) NOEC : Notable oversight of enlightened Canadians : A response to van Dam et al. (2012) *Integrated Environmental Assessment and Management* 8(3) :397–398
- Wang SCD and Smith EP (2000) Adjusting for mortality effects in chronic toxicity testing : mixture model approach. *Environmental Toxicology and Chemistry* 19(1) :204–209
- Warne MSJ and van Dam R (2008) NOEC and LOEC data should no longer be generated or used. *Australasian Journal of Ecotoxicology* 14 :1–5
- Wennig R (2000) Threshold values in toxicology-useful or not? *Forensic Science International* 113 :323–330
- Wiench K, Wohlleben W, Hisgen V, Radke K, Salina E, Zok S and Landsiedel R (2009) Acute and chronic effects of nano- and non-nano-scale TiO₂ and ZnO particles on mobility and reproduction of the freshwater invertebrate *Daphnia magna*. *Chemosphere* 76 :1356–1365
- Wollenberger L, Halling-Sørensen B and Kusk K (2000) Acute and chronic toxicity of veterinary antibiotics to *Daphnia magna*. *Chemosphere* 40, 723–730
- Wright SE and Bailer AJ (2006) Optimal experimental design for a nonlinear response in environmental toxicology. *Biometrics* 62 :886–892

-
- Zeman FA, Gilbin R, Alonzo F, Lecomte-Pradines C, Garnier-Laplace J, Aliaume C (2008) Effects of waterborne uranium on survival, growth, reproduction and physiological processes of the freshwater cladoceran *Daphnia magna*. *Aquatic toxicology* 86, 370–378
- Zhang J, Bailer AJ and Oris JT (2012) Bayesian approach to potency estimation for aquatic toxicology experiments when a toxicant affects both fecundity and survival. *Environmental Toxicology and Chemistry* 31(8) :1920–1930

ANNEXE 1 : Jonckheere-Terpstra

Test de Jonckheere-Terpstra

Test non paramétrique, équivalent du test de Kruskal-Wallis mais en supposant que la réponse est monotone.

Principe (Jonckheere, 1954)

Le test non paramétrique de Jonckheere-Terpstra permet de tester l'ordonnement de plusieurs moyennes sur des échantillons indépendants alors que le test de Kruskal-Wallis permet de tester la différence entre plusieurs moyennes sans préjuger de leur ordre. Ce test se fait sur les rangs des observations.

Les hypothèses :

H_0 : Les réponses moyennes pour chaque dose sont équivalentes.

H_1 : Les réponses moyennes croissent (ou décroissent) de manière monotone en fonction des doses.

On ordonne tous les traitements suivant la dose croissante de 1 à K . Soient deux traitements i et $j = i + 1$ consécutifs.

Soit n_i (resp. n_j) le nombre d'observations du traitement i (resp. j).

On observe $X_{i,\gamma}$, avec γ prenant les valeurs de 1 à n_i et $X_{j,\beta}$, avec β prenant les valeurs de 1 à n_j .

On calcule J en comptant le nombre $U_{i,j}$ de paires (γ, β) pour lesquelles on a $X_{i\gamma} < X_{j\beta}$:

$$J = \sum_{i=1}^{K-1} \sum_{j=i+1}^K U_{i,j}$$

A partir de cette valeur J , on détermine S de la manière suivante :

$$S = 2 \times J - \sum_{i=1}^{K-1} \sum_{j=i+1}^K n_i n_j$$

On peut calculer un coefficient de corrélation de rang :

$$\tau = \frac{S}{\sum_{i=1}^{K-1} \sum_{j=i+1}^K n_i n_j}$$

Plus τ est proche de 1, plus les données sont en conformité avec l'hypothèse alternative d'ordonnement des effets. On rejette l'hypothèse nulle si $P[S \geq S_0]$, que l'on peut lire dans une table (Jonckheere, 1954), est inférieur à un risque α généralement fixé à 5%, mais cette table est très rarement présente dans les ouvrages. On peut utiliser, dans le cas de grands échantillons, la variable Z qui suit asymptotiquement une loi normale de moyenne $E[J]$ et de variance $Var[J]$:

$$Z = \frac{J - E[J]}{\sqrt{Var[J]}}$$

avec

$$E[J] = \frac{n^2 - \sum_{j=1}^K n_j^2}{4}$$

et

$$Var[J] = \frac{n^2(2n+3) - \sum_{j=1}^K n_j^2(2n_j+3)}{72}$$

Si $Z > u_{0,95} = 1,64$ alors on rejette H_0 .

Nous faisons un test unilatéral car la relation entre la réponse et la dose est supposée monotone (croissante ou décroissante).

Exemple

Prenons un exemple avec différents traitements pour lesquels on a un ordonnancement présumé des effets.

Traitement					
A	133	139	149	160	184
B	111	125	143	148	157
C	99	114	116	127	146

Calculons le nombre $U_{B,A}$ de paires pour lesquelles on a $X_{B\gamma} < X_{A\beta}$:

$$U_{B,A} = 2 + 2 + 4 + 5 + 5 = 18$$

En effet, prenons la mesure 1 du traitement A (133) et comparons la avec le traitement B, seules deux mesures (mesure 1 et mesure 2) sont inférieures à la mesure 1 du traitement A d'où le premier terme de l'addition permettant de calculer $U_{B,A}$. Et il en va de même pour les autres termes : la mesure 3 du traitement A (149) a quatre valeurs de mesures qui lui sont inférieures dans le traitement B d'où le 4 en troisième position dans l'addition. La même opération est également effectuée pour $U_{C,B}$ et $U_{C,A}$, on obtient :

$$U_{C,B} = 1 + 3 + 4 + 5 + 5 = 18$$

$$U_{C,A} = 4 + 4 + 5 + 5 + 5 = 23$$

Ainsi, la statistique J de Jonckheere peut être calculée :

$$J = 18 + 18 + 23 = 59$$

et

$$S = 2 \times 59 - (3 \times 5 \times 5) = 43$$

Chaque échantillon du traitement A est comparé à la fois avec les 5 échantillons du traitement B ($5 \times 5 = 25$) et avec les 5 échantillons du traitement C ($5 \times 5 = 25$). Il reste à comparer les échantillons du traitement B à ceux 5 échantillons du traitement C ($5 \times 5 = 25$). Nous obtenons ainsi au total un nombre de comparaisons de $3 \times 5 \times 5$. On peut lire dans la table de Jonckheere que pour $k = 3$ échantillons de taille $m = 5$ que $P[S \geq 43] = 0.012$ ainsi

$$P[S \geq S_0] = P[S \geq 43] = 0.012 < 5\%$$

On peut ainsi conclure au rejet de H_0 , il y a une différence significative entre les traitements avec une relation monotone entre les traitements et la réponse.

ANNEXE 2 : Liste des membres du groupe de travail sur le développement et la validation d'un test standard de reproduction chez les mollusques

1. AstraZeneca : Laboratoire environnemental de Brixham (Royaume Uni)
2. BASF – Ecotoxicologie et protection des cultures (Allemagne)
3. CEFAS - Centre pour l'environnement, la pêche et l'aquaculture (Royaume Uni)
 - Laboratoire de Weymouth
 - Laboratoire de Lowestoft
4. INRA – Institut National de la Recherche Agronomique (France)
 - Equipe Ecotoxicologie et qualité des milieux aquatiques
 - Unité expérimentale d'écologie et d'écotoxicologie aquatique
5. Université Goethe de Franckort-sur-le-Main – Laboratoire d'écotoxicologie aquatique (Allemagne)
6. Université du Danemark du Sud – Institut de Biologie (Danemark)
7. Laboratoires Galab GmbH (Allemagne)
8. Université Lyon 1 – Equipe Modélisation et Ecotoxicologie Prédictives (France)

ANNEXE 3 : Correction de Bonferroni-Holm

Corrections à apporter en cas de comparaisons multiples. Correction de Bonferroni-Holm

Principe

Lorsque plusieurs tests statistiques sont réalisés simultanément pour chercher à répondre à la question de l'efficacité d'un traitement par exemple, le risque global d'erreur de première espèce (α) s'accroît.

Le risque global de conclure à tort à un effet du traitement est ainsi augmenté, ce risque est aussi appelé le "Family Wise Error Rate" (FWE). Soit k le nombre de tests effectués et le risque d'erreur initial $\alpha = 5\%$ on a :

$$FWE = 1 - (1 - \alpha)^k$$

Pour éviter d'avoir une inflation du risque global d'erreur, il existe plusieurs méthodes :

Correction de Bonferroni

La correction de Bonferroni consiste à prendre le risque d'erreur au niveau de chacune des k comparaisons individuelles (appelé "Comparison Error rate", CER) avec :

$$CER = \frac{\alpha}{k}$$

Par exemple pour cinq comparaisons, il faut tester chacune d'elle au risque individuel (CER) de 1%

Cette méthode n'est pas optimale au sens où elle protège "trop" en assurant un risque global inférieur. En effet, dans cet exemple le $FWE = 1 - (1 - 0.01)^5 = 0.049$ et plus le nombre de comparaisons est important plus le risque global est inférieur à 5%.

Correction de Bonferroni-Holm (Holm, 1979)

Prenons la plus petite des p-value calculées sans correction et comparons la au seuil $\frac{\alpha}{k}$ si cette p-value est $< \frac{\alpha}{k}$ alors on rejette l'hypothèse H_0 testée. Ensuite nous prenons la plus petite p-value parmi celles restantes et on la compare de la même manière au seuil $\frac{\alpha}{k-1}$ et ainsi de suite jusqu'à ce qu'on accepte H_0 . Cette correction est la plus souvent utilisée, elle est moins conservatrice que la correction de Bonferroni et offre un bon contrôle du FWE elle est également préconisée par l'OCDE.

Exemple

On a 4 valeurs de p issues de 4 tests de comparaison :

$$p_1 = 0.01 ; p_2 = 0.03 ; p_3 = 0.04 ; p_4 = 0.005$$

Prenons la plus petite valeur qui est p_4 , comme elle est inférieure à $\frac{0.05}{4} = 0.0125$ l'hypothèse H_{0_4} est rejetée et nous prenons la plus petite valeur suivante qui est p_1 . Comme p_1 est inférieure à $\frac{0.05}{3} = 0.016$ l'hypothèse H_{0_1} est rejetée. Ensuite, nous trouvons que p_3 est supérieur à $\frac{0.05}{2} = 0.025$ l'hypothèse H_{0_3} n'est pas rejetée. Nous en déduisons alors que l'hypothèse avec une p-value plus élevée ne sera pas rejetée non plus.

ANNEXE 4 : Matrice d'information de Fisher

Démonstration de l'écriture de la matrice d'information de Fisher telle qu'utilisée par King et Wong (2000), pour le modèle logistique dans le cas de données de survie.

On écrit ici le modèle logistique comme suit :

$$p(x, a, b) = \frac{1}{1 + \exp(-b(x - a))}$$

avec x la dose ou la concentration (éventuellement en log), a est la dose/concentration pour laquelle la probabilité de la réponse est égale à 0.5 (i.e., $a = ED_{50}$) et b est la pente en échelle logit.

La figure 1 montre une courbe représentative du modèle logistique pour des valeurs arbitraires des paramètres a et b .

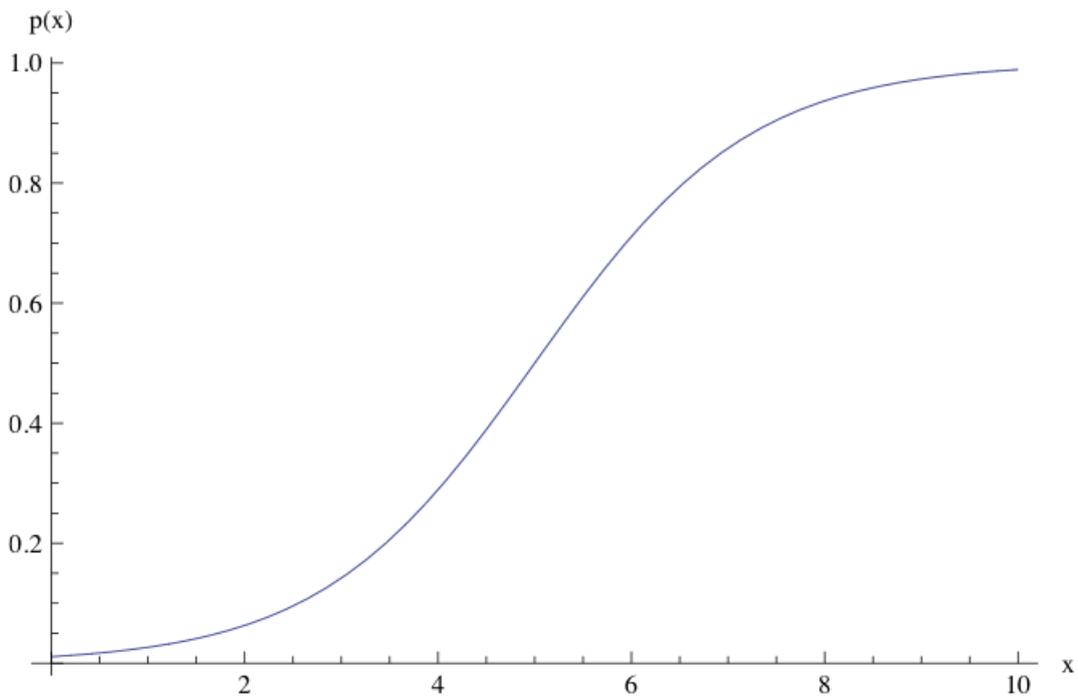


FIGURE 4.1 – Courbe représentative du modèle logistique pour $a = 5$ et $b = 0.9$

On cherche ici à expliciter la matrice d'information de Fisher définie par :

$$\mathbf{F} = - \begin{bmatrix} E \left[\frac{\partial^2 \ell}{\partial a^2} \right] & E \left[\frac{\partial^2 \ell}{\partial a \partial b} \right] \\ E \left[\frac{\partial^2 \ell}{\partial b \partial a} \right] & E \left[\frac{\partial^2 \ell}{\partial b^2} \right] \end{bmatrix}$$

avec ℓ la log-vraisemblance.

Sous l'hypothèse d'un modèle d'erreur binomial, la vraisemblance s'écrit :

$$L = \prod_{i=1}^K C_{n_i}^{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$$

avec K le nombre de concentrations, $p_i = p(x_i, a, b)$ la probabilité de réponse à la $i^{\text{ème}}$ concentration, y_i le nombre d'individus qui répondent à la concentration i et n_i le nombre total d'individus à la concentration i .

Ainsi :

$$\ell = \ln(L) = cte + \sum_{i=1}^K (y_i \ln p_i + (n_i - y_i) \ln(1 - p_i))$$

Pour simplifier un peu le calcul des dérivées secondes, on introduit :

$$p(x, a, b) = \frac{1}{u(x, a, b)}$$

avec $u(x, a, b) = 1 + \exp(-b(x - a))$.

Il vient :

$$\exp(-b(x - a)) = u(x, a, b) - 1 = \frac{1 - p(x, a, b)}{p(x, a, b)}$$

Commençons par les dérivées premières.

D'après l'expression de ℓ , il vient :

$$\frac{\partial \ell}{\partial a} = \sum_{i=1}^K \left(y_i \frac{\partial \ln p_i}{\partial a} + (n_i - y_i) \frac{\partial \ln(1 - p_i)}{\partial a} \right)$$

On obtient aisément :

$$\frac{\partial \ln p_i}{\partial a} = \frac{1}{p_i} \frac{\partial p_i}{\partial a} = -\frac{1}{u_i} \frac{\partial u_i}{\partial a}$$

et

$$\frac{\partial \ln(1 - p_i)}{\partial a} = -\frac{1}{1 - p_i} \frac{\partial p_i}{\partial a} = -\frac{p_i}{1 - p_i} \frac{\partial \ln p_i}{\partial a}$$

On a immédiatement :

$$\frac{\partial u_i}{\partial a} = b(u_i - 1) = b \frac{1 - p_i}{p_i}$$

d'où :

$$\frac{\partial \ln p_i}{\partial a} = -b(1 - p_i)$$

et

$$\frac{\partial \ln(1 - p_i)}{\partial a} = bp_i$$

Et finalement :

$$\frac{\partial \ell}{\partial a} = b \sum_{i=1}^K (n_i p_i - y_i)$$

De la même façon, on obtient :

$$\frac{\partial \ell}{\partial b} = \sum_{i=1}^K \left(y_i \frac{\partial \ln p_i}{\partial b} + (n_i - y_i) \frac{\partial \ln(1 - p_i)}{\partial b} \right)$$

avec

$$\frac{\partial \ln p_i}{\partial b} = \frac{1}{p_i} \frac{\partial p_i}{\partial b} = -\frac{1}{u_i} \frac{\partial u_i}{\partial b}$$

et

$$\frac{\partial \ln(1 - p_i)}{\partial b} = -\frac{1}{1 - p_i} \frac{\partial p_i}{\partial b} = -\frac{p_i}{1 - p_i} \frac{\partial \ln p_i}{\partial b}$$

On a immédiatement :

$$\frac{\partial u_i}{\partial b} = -(x_i - a) \frac{1 - p_i}{p_i}$$

d'où

$$\frac{\partial \ln p_i}{\partial b} = (x_i - a) (1 - p_i)$$

et

$$\frac{\partial \ln(1 - p_i)}{\partial b} = -(x_i - a) p_i$$

Et finalement

$$\frac{\partial \ell}{\partial b} = \sum_{i=1}^K (x_i - a) (y_i - n_i p_i)$$

Passons aux dérivées secondes.

$$\frac{\partial^2 \ell}{\partial a^2} = b \sum_{i=1}^K n_i \frac{\partial p_i}{\partial a}$$

Sachant que

$$\frac{\partial p_i}{\partial a} = -b p_i (1 - p_i)$$

il vient

$$\frac{\partial^2 \ell}{\partial a^2} = -b^2 \sum_{i=1}^K n_i p_i (1 - p_i)$$

Depuis l'expression

$$\frac{\partial \ell}{\partial a}$$

on obtient :

$$\frac{\partial^2 \ell}{\partial a \partial b} = \sum_{i=1}^K \left(n_i p_i + n_i b \frac{\partial p_i}{\partial b} - y_i \right)$$

Sachant que

$$\frac{\partial p_i}{\partial b} = (x_i - a) p_i (1 - p_i)$$

il vient

$$\frac{\partial^2 \ell}{\partial a \partial b} = \sum_{i=1}^K (n_i p_i - y_i + n_i b (x_i - a) p_i (1 - p_i))$$

Enfin

$$\frac{\partial^2 \ell}{\partial b^2} = \sum_{i=1}^K -n_i (x_i - a) \frac{\partial p_i}{\partial b}$$

Sachant que

$$\frac{\partial p_i}{\partial b} = (x_i - a) p_i (1 - p_i)$$

il vient

$$\frac{\partial^2 \ell}{\partial b^2} = - \sum_{i=1}^K n_i (x_i - a)^2 p_i (1 - p_i)$$

Pour finir, il faut calculer moins l'espérance des différentes dérivées secondes :

$$\begin{aligned} -E \left(\frac{\partial^2 \ell}{\partial a^2} \right) &= b^2 \sum_{i=1}^K n_i p_i (1 - p_i) \\ -E \left(\frac{\partial^2 \ell}{\partial b^2} \right) &= \sum_{i=1}^K n_i (x_i - a)^2 p_i (1 - p_i) \\ -E \left(\frac{\partial^2 \ell}{\partial a \partial b} \right) &= -E \left(\sum_{i=1}^K n_i p_i - y_i + n_i b (x_i - a) p_i (1 - p_i) \right) \\ &= \sum_{i=1}^K E(y_i) - n_i p_i - n_i b (x_i - a) p_i (1 - p_i) \end{aligned}$$

Or

$$E(y_i) = n_i p_i$$

Donc

$$-E \left(\frac{\partial^2 \ell}{\partial a \partial b} \right) = - \sum_{i=1}^K n_i b (x_i - a) p_i (1 - p_i)$$

Finalement, il vient :

$$\mathbf{F} = \begin{pmatrix} b^2 \sum_{i=1}^K n_i p_i (1 - p_i) & - \sum_{i=1}^K n_i b (x_i - a) p_i (1 - p_i) \\ - \sum_{i=1}^K n_i b (x_i - a) p_i (1 - p_i) & \sum_{i=1}^K n_i (x_i - a)^2 p_i (1 - p_i) \end{pmatrix}$$

Pour faire apparaître les poids, et donc retrouver l'expression de King et Wong (2000), il suffit d'écrire $n_i = \omega_i N$ et de remplacer dans l'expression de \mathbf{F} . On trouve alors effectivement que \mathbf{F} est proportionnelle à la matrice \mathbf{M} suivante :

$$\mathbf{M} = \begin{pmatrix} A & -B \\ -B & C \end{pmatrix}$$

avec

$$A = b^2 \sum_{i=1}^K \omega_i p_i (1 - p_i)$$

$$B = \sum_{i=1}^K b \omega_i (x_i - a) p_i (1 - p_i)$$

$$C = \sum_{i=1}^K \omega_i (x_i - a)^2 p_i (1 - p_i)$$

ANNEXE 5
Articles publiés

Survival data analyses in ecotoxicology: critical effect concentrations, methods and models. What should we use?

Carole Forfait-Dubuc · Sandrine Charles ·
Elise Billoir · Marie Laure Delignette-Muller

Accepted: 12 January 2012 / Published online: 3 February 2012
© Springer Science+Business Media, LLC 2012

Abstract In ecotoxicology, critical effect concentrations are the most common indicators to quantitatively assess risks for species exposed to contaminants. Three types of critical effect concentrations are classically used: lowest/no observed effect concentration (LOEC/NOEC), LC_x ($x\%$ lethal concentration) and NEC (no effect concentration). In this article, for each of these three types of critical effect concentration, we compared methods or models used for their estimation and proposed one as the most appropriate. We then compared these critical effect concentrations to each other. For that, we used nine survival data sets corresponding to *D. magna* exposition to nine different contaminants, for which the time-course of the response was monitored. Our results showed that: (i) LOEC/NOEC values at day 21 were method-dependent, and that the Cochran–Armitage test with a step-down procedure appeared to be the most protective for the environment; (ii) all tested concentration–response models we compared gave close values of LC_{50} at day 21, nevertheless the Weibull model

had the lowest global mean deviance; (iii) a simple threshold NEC-model both concentration and time dependent more completely described whole data (i.e. all time-points) and enabled a precise estimation of the NEC. We then compared the three critical effect concentrations and argued that the use of the NEC might be a good option for environmental risk assessment.

Keywords NOEC/ LC_x /NEC · Hypothesis testing · Concentration–response curves · Bayesian inference · Risk assessment

Introduction

Ecotoxicological bioassays are widely performed to determine critical effect concentrations also named “summary statistics” (van der Hoeven 1997; Crane and Newman 2000; Baas et al. 2009) or “thresholds” (Cairns 1992), which are the most common indicators to quantitatively assess risks for species exposed to contaminants. Indeed, the main goal often comes down to the estimation of the exposure level (e.g., concentration) above which adverse effects can occur on organisms, and below which adverse effects are unlikely, i.e., which cannot be distinguished from background noise (OECD 2006). For measuring such a “no effect” level or “acceptable effect” level (van der Hoeven 2004), at least three types of critical effect concentrations can be distinguished (van der Hoeven 1997): the no observed effect concentration (NOEC), the $x\%$ effective/lethal concentration (EC_x/LC_x) and the no effect concentration (NEC).

The NOEC is the highest concentration among the tested concentrations at which no statistically significant difference from the control can be highlighted. The LOEC is the

C. Forfait-Dubuc · S. Charles · E. Billoir ·
M. L. Delignette-Muller
Université de Lyon, 69000 Lyon, France

C. Forfait-Dubuc (✉) · S. Charles · E. Billoir ·
M. L. Delignette-Muller
CNRS, UMR5558, Laboratoire de Biométrie et Biologie
Evolutive, Université Lyon1, 69622 Villeurbanne, France
e-mail: carole.forfait@univ-lyon1.fr

E. Billoir
Pôle de Recherche ROVALTAIN en Toxicologie
Environnementale et Ecotoxicologie, Bâtiment Rhovalparc,
1 Avenue de la gare, BP 15173, 26958 Valence, France

M. L. Delignette-Muller
VetAgro Sup Campus Vétérinaire de Lyon, 1 Avenue Bourgelat,
69280 Marcy l'Etoile, France

tested concentration just above the NOEC (OECD 2006). Both LOEC and NOEC are obtained after hypothesis testing. Even if LOEC/NOEC values have proved a tremendous success with risk managers, they have been strongly criticized (van der Hoeven 1997; Nelder 1999; Isnard et al. 2001; Fox 2008; Warne and van Dam 2008) mainly for the following reasons: (i) LOEC/NOEC values depend on the experimental design (sample size, exposure duration and concentrations); and (ii) from a statistical point of view, the strongest argument against their interpretation comes from the improper use of p -values to establish biological thresholds (Newman 2008; Delignette-Muller et al. 2011).

The EC_x/LC_x , with x typically varying between 5 and 50, is the concentration that causes $x\%$ change in response (e.g., mortality, reproduction) during a specified time interval (OECD 2006). It is estimated by fitting a concentration (or dose)–response curve to experimental data. EC_x/LC_x values were first proposed as a parametric alternative to LOEC/NOEC, with a recommended x value equal to 5 or 10 (van der Hoeven et al. 1997; Warne and van Dam 2008). Nevertheless, EC_x/LC_x values themselves drawbacks: (i) as with LOEC/NOEC values, they depend on the exposure duration; (ii) any x value other than zero may imply that a certain percentage of effect is acceptable, so that the choice of x appears critical; and (iii) the extreme EC_x/LC_x values (e.g., x less than 5%) may depend on the concentration–response model (van der Hoeven 1997; Jager 2011).

The concept of NEC was proposed as another parametric alternative to LOEC/NOEC (Kooijman and Bedaux 1996). The principle is to fit models in which one parameter (NEC) is defined as a threshold for the effect occurrence. Such a NEC appears as an ideal parameter to estimate (van der Hoeven 1997; van der Hoeven 2004), however, the choice of an appropriate model to estimate NEC values impedes its use.

The OECD guideline, published in 2006, puts up as the main objective to provide practical guidance on how to statistically analyze experimental data from ecotoxicity tests (OECD 2006). In particular, this guideline recommends the use of concentration–response models to describe effects as functions of concentration only (response = $f(c)$) and the use of NEC-models to describe effects as functions of both concentration and exposure time (response = $f(c, t)$) when several exposure timepoints are available. Nevertheless, for each critical effect concentration and type of data, ecotoxicologists have to choose among several potential methods. By way of example, the following two sentences were extracted from the OECD guideline: “[concerning the described statistical methods] ... the choice is left to the reader” (OECD 2006, p. 34);

“[The concentration–response model] can be any function that is potentially suitable for describing a particular dataset” (OECD 2006, p. 62). It thus appears essential to clarify potential impacts of the choice of one particular method/model on the critical effect concentration which is intended to be estimated.

For this purpose, we focused on quantal data, in particular survival data, and the three critical effect concentrations mentioned above. For the determination of LOEC/NOEC and LC_x , methods/models as recommended by OECD were compared. For NEC estimation, we proposed a simple threshold model describing the response as a function of both concentration and time. The different methods/models were compared on the same battery of nine experimental data sets, where the freshwater invertebrate *Daphnia magna* was exposed to nine different contaminants. Based on previous works (Billoir et al. 2008a; Ducrot et al. 2010; Fox 2010), we chose Bayesian inference to estimate model parameters, in particular LC_{50} and NEC values and their uncertainty. One of the main advantages of Bayesian inference is the use of prior information (coming from expert knowledge or literature) expressed as probability distributions on parameters, which are combined with data in order to provide posterior distributions. Moreover, assessment of uncertainty in parameter estimates is obtained directly from posterior distributions. At the end, our aim was two-fold: (i) for each type of critical effect concentration, a comparison of the different methods/models in order to propose one as the most appropriate; (ii) a comparison between the critical effect concentrations as obtained with the most appropriate method/model (step (i)).

Materials and methods

Data sets

We used nine survival data sets all concerning the freshwater invertebrate *Daphnia magna* under exposure to nine different contaminants: three metals [copper and zinc (Billoir et al. 2008b), cadmium (Santojanni et al. 1995)], an oxidizing agent [potassium dichromate (Bedaux and Kooijman 1994)], an organochlorine insecticide [chlordan (Manar et al. 2009)], an anthelmintic used in veterinary medicine [ivermectin (Lopes et al. 2009)], and three veterinary antibiotics [oxolinic acid, streptomycin and sulfadiazine (Wollenberger et al. 2000)]. These data sets, showed in Fig. 1 (left graphs), came from chronic laboratory bioassays where survival was measured at several timepoints (depending on the contaminant) at least for 21 days. Table 1 gives the experimental design for each data set.

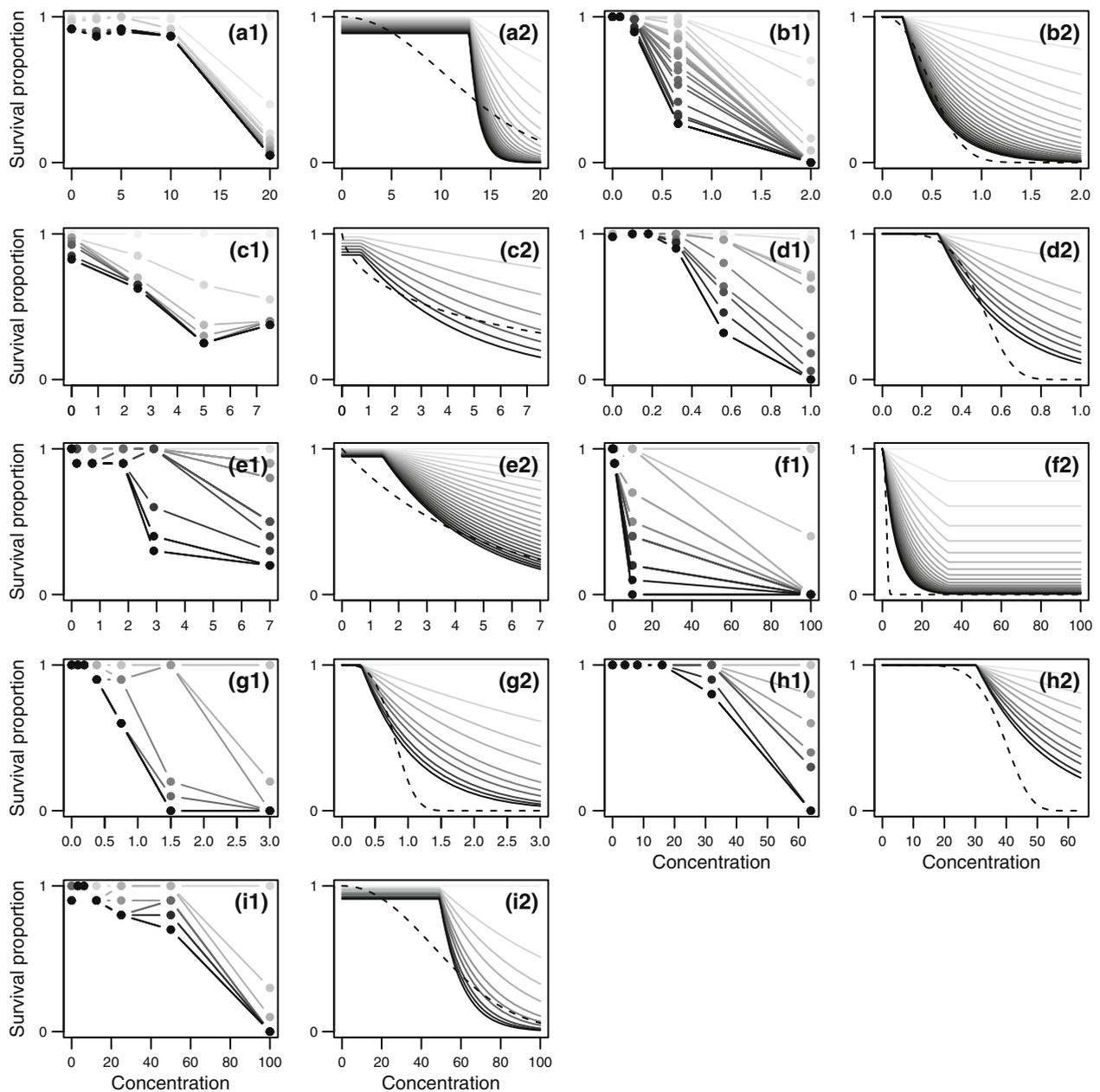


Fig. 1 Each couple of graphs with the *same letter* corresponds to one contaminant. Each graph depicts survival proportion versus contaminant concentration. *Points on left graphs* (noted 1) represent data with *darker and darker lines* corresponding to higher and higher exposure duration. Each *line* connects data of a same exposure duration. *Right graphs* (noted 2) depict fittings obtained by the most

adapted NEC-model for each exposure duration in *darker and darker plain lines* and by the Weibull model only at day 21 in *dotted line*. **a** Copper ($\mu\text{g/L}$), **b** zinc (mg/L), **c** cadmium (mg/L), **d** potassium dichromate (mg/L), **e** chlordan ($\mu\text{g/L}$), **f** ivermectin (pg/L), **g** oxolinic acid (mg/L), **h** streptomycin ($\mu\text{g/L}$), **i** sulfadiazine ($\mu\text{g/L}$)

Hypothesis-testing methods

We applied hypothesis-testing methods to determine LOEC/NOEC values. As stated in the introduction, LOEC/NOEC values depend on the exposure duration and these values are generally determined after a 21-day exposure for

D. magna. According to the OECD guideline (OECD 2006), the following methods are possible for survival data analyses.

First, if the response is expected to vary monotonically with the concentration, two tests can be used: the Cochran–Armitage test (Armitage and Colton 2005) or the Jonckheere–

Table 1 Experimental design for each data set

Contaminant	Number of tested concentrations (<i>k</i>)	Concentration range (<i>c_j</i>) [initial number of <i>D.magna</i> (<i>N_j</i>) with <i>j</i> = 1, <i>k</i>]	Number of measures = timepoints from 0 to 21 days
Copper (µg/L)	5	0, 2.5, 5, 10, 20 [60, 60, 60, 60, 60]	21
Zinc (mg/L)	5	0, 0.074, 0.22, 0.66, 2 [57, 56, 58, 60, 60]	21
Cadmium (µg/L)	4	0, 2.5, 5, 7.5 [40, 40, 40, 40]	8
Potassium dichromate (mg/L)	6	0, 0.1, 0.18, 0.32, 0.56, 1 [50, 50, 50, 50, 50, 50]	10
Chlordan (µg/L)	6	0, 0.18, 0.73, 1.82, 2.9, 7 [10, 10, 10, 10, 10, 10]	21
Ivermectin (pg/L)	5	0, 0.3, 1, 10, 100 [10, 10, 10, 10, 10]	21
Oxolinic acid (mg/L)	7	0, 0.094, 0.19, 0.38, 0.75, 1.5, 3 [10, 10, 10, 10, 10, 10, 10]	10
Streptomycin (mg/L)	6	0, 4, 8, 16, 32, 64 [10, 10, 10, 10, 10, 10]	10
Sulfadiazine (mg/L)	7	0, 3.1, 6.2, 12.5, 25, 50, 100 [10, 10, 10, 10, 10, 10, 10]	10

Terpstra test (Jonckheere 1954) both with a step-down procedure. The Cochran–Armitage test is a trend test for counts and proportions based on the chi-square statistics, while the Jonckheere–Terpstra test is based on the Wilcoxon rank-sum statistics. The step-down procedure works as follows: survival data at all tested concentrations are globally compared; if there is no significant difference between all the concentrations, the NOEC equals the highest tested concentration. Otherwise, the highest concentration and corresponding data are removed from the data set and a global test is run again on the remaining data. This operation is repeated until a NOEC value is determined.

Second, if there is no monotony assumption, comparisons are made by pairs (the control to each tested concentration) with the Fisher exact test using the Bonferroni–Holm correction (Holm 1979; OECD 2006). This correction is conventionally known as a step-down improvement of the classical Bonferroni adjustment method (Holm 1979).

All hypothesis tests were performed using the R software (R Development Core Team 2010). The Cochran–Armitage test and the Fisher exact test were performed respectively with the R-functions *prop.trend.test* and *fisher.test*, and we implemented our own functions (available upon request) for the Jonckheere–Terpstra test, the step-down procedure and the Bonferroni–Holm correction.

We estimated LOEC/NOEC values at day 21 with the three methods above for each data set. Following a failsafe

approach, we selected the most appropriate method as the one which gave the most often the smallest NOEC value, that is the most protective value for the environment.

Concentration–response models

Like LOEC/NOEC values, LC_x values are determined at a given time. In theory, numerous concentration–response models are possible but, practically, the most frequently used are the logit, the probit and the Weibull model (OECD 2006). Logit and probit models are classically written with the logarithm of the concentration, thus becoming the so-called log-logistic and log-probit models. Because of the small number of available data points at day 21, we used these models in their simplest form with only two parameters as described in the OECD guideline (OECD 2006). As also stated in the OECD guideline, control data were not taken into account, due to the fact that log-logistic and log-probit models were not defined in 0. In order to be homogenous and make the model comparison possible, we did the same for the Weibull model. Again, as recommended in the OECD guideline, all models were parameterized with the LC₅₀. Models are written as follows:

– Log-logistic:

$$S(c) = 1 - \frac{1}{1 + \exp\left(b \times \log_{10}\left(\frac{LC_{50}}{c}\right)\right)} \quad (1)$$

– Log-probit:

$$S(c) = 1 - F\left(b \times \log_{10}\left(\frac{c}{LC_{50}}\right)\right) \quad (2)$$

– Weibull:

$$S(c) = \exp\left(-\left(\frac{c}{LC_{50}}\right)^b \times \ln(2)\right) \quad (3)$$

with $S(c)$ the survival probability at concentration c , b a regression coefficient and F the normal cumulative distribution function. These models were fitted to data expressed as numbers of survivors at day 21 and concentration c . We assumed a binomial error model with parameters $S(c)$ for the success probability and the number of individuals initially introduced in the experiments (N_j , Table 1). For each data set, we compared the mean deviance of the three models, which denotes the quality of fit, and chose the most appropriate one as the one that provided the lowest mean deviance.

NEC-model

Like LC_x , several models including a threshold parameter were possible. The parameter of interest (the NEC) was introduced as a threshold concentration in a simple and general model. Note that the introduction of such a parameter NEC in the model does not presuppose the existence of a threshold, but just allows it to be estimated if data justify it (Fox 2010). A distinctive feature of this threshold model, called afterwards NEC-model, was its dependence on both concentration c and time t . Such a NEC-model can thus be fitted to larger data sets (including measures at all timepoints) compared to classical concentration–response models which only depend on concentration c . Consequently, more parameters may reasonably be estimated. The survival probability $S(c, t)$ depends on both exposure concentration c and time t . $S(c, t)$ is written itself as a function of the instantaneous mortality $h(c)$ which is assumed to be time-independent:

$$S(c, t) = e^{-h(c) \times t} \quad (4)$$

with

$$h(c) = m_0 + \sigma(c) \quad (5)$$

where m_0 (in/day) stands for the natural instantaneous mortality and $\sigma(c)$ is the so-called stress function, positive from the NEC (in concentration unit) and linearly increasing with a slope k_S (in 1/concentration unit/day) as c increases:

$$\sigma(c) = \min(k_S \times \max(0, c - NEC), m_{\max} - m_0) \quad (6)$$

Parameter m_{\max} (in /day) stands for a potential saturation of the instantaneous mortality, $h(c)$, at high concentrations.

The number of survivors at concentration c and time t was assumed to follow a conditional binomial distribution of parameters $1 - (S(c, t - 1) - S(c, t)) / S(c, t - 1)$ (i.e., 1 minus the probability of death between $t-1$ and t) and $N_{c,t-1}$ (i.e., the number of alive individuals for concentration c at time $t-1$). Our NEC-model thus involved four parameters: NEC, k_S , m_{\max} and m_0 . According to the data set, all parameters were not necessary to describe them. In some cases, the complete model with four parameters may be simplified to one of the nested models presented in Fig. 2: a two-parameter model (NEC, k_S) and two three-parameter models: (NEC, k_S and m_0) or (NEC, k_S and m_{\max}). Parameter m_0 was required only when mortality was observed in the control or when the response in function of the concentration was not monotonic, otherwise m_0 was fixed to zero. Parameter m_{\max} was required only when the number of survivors at the highest tested concentration was greater than or equal to the number of survivors at a lower concentration for at least one timepoint, otherwise m_{\max} was fixed to infinity and Eq. 6, becoming $\sigma(c) = k_S \times \max(0, c - NEC)$.

NEC values were estimated for each data set, by including all the timepoints until day 21, contrary to the other critical effect concentrations which were estimated only from data at day 21. To choose the most appropriate NEC-model for each data set, we first decided which parameters were necessary according to the above criteria; second, we checked if estimates of m_0 and m_{\max} parameters were precise enough (no more than three log10 between the lower and upper bounds of the 95% uncertainty interval). If not, we fitted a simpler model by removing the less precisely estimated parameter.

Bayesian inference

Concept

We used Bayesian inference to estimate all parameters of concentration–response models and NEC-models. Bayesian inference is a statistical method in which observations are used to update what is previously known about parameters. The Bayesian inference provides posterior distributions of parameters given (i) the data and (ii) prior information translated into prior distributions. Principles and details of Bayesian statistics can be found in Ntzoufras (2009). Some introductory ecotoxicological applications of Bayesian inference can be found in Billoir et al. (2008a), Ducrot et al. (2010) and Fox (2010).

The use of Bayesian inference requires the definition of prior distributions for model parameters. For some parameters, we elicited prior information from the range of tested concentrations, which were in fact themselves chosen by the experimenter from his or her prior knowledge.

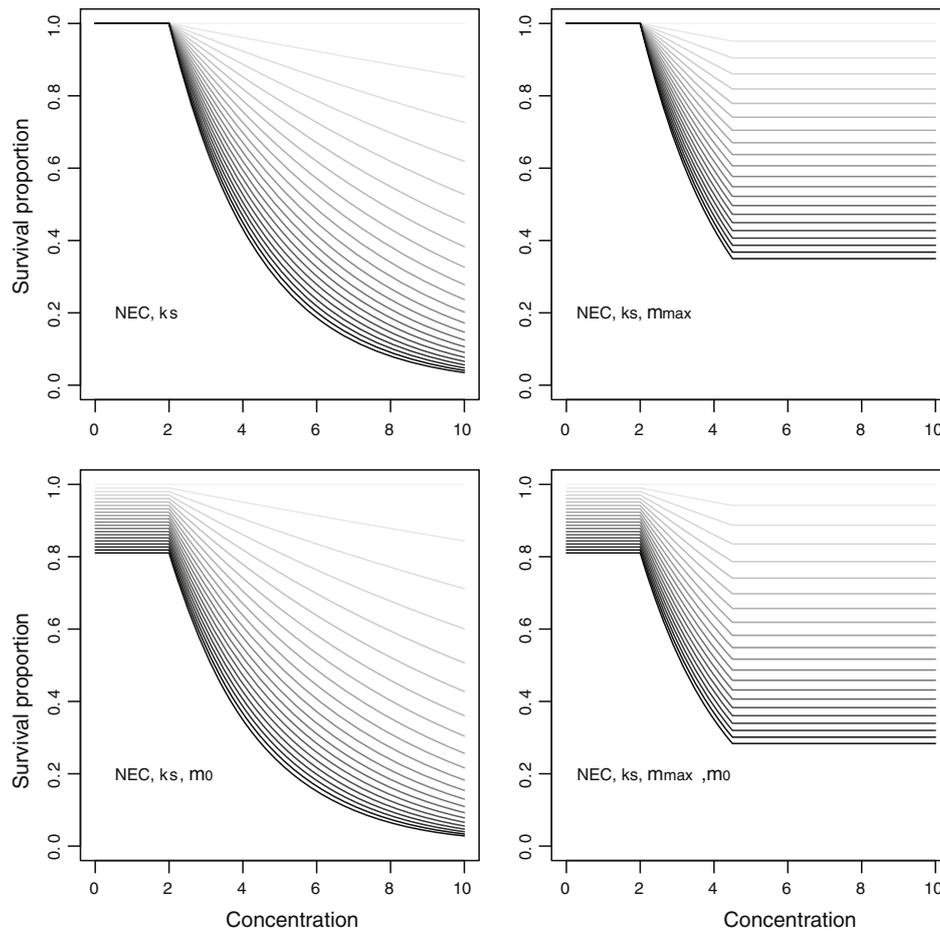


Fig. 2 NEC-models with two, three or four parameters are represented as the survival proportion versus an hypothetical contaminant concentration. The darker and darker lines depict 22 timepoints from

0 to 21 days. We chose arbitrary values of parameters: $NEC = 2 \text{ mg/L}$; $k_S = 0.02 \text{ L/mg/day}$; $m_{\max} = 0.05/\text{day}$ and $m_0 = 0.01/\text{day}$

Prior distributions for concentration–response models

For the concentration–response models, we defined an informative uniform prior distribution on parameter LC_{50} between 0 and the highest tested concentration; for parameter b in the log-logistic and the log-probit models we used a quasi non informative uniform distribution between 0 and 100; and for b in the Weibull model, we used a quasi non informative uniform distribution between 0 and 10.

Prior distributions for the NEC-models

For the NEC, we defined the same informative distribution as for LC_{50} : a uniform prior distribution between 0 and the highest tested concentration. According to the OECD guideline (OECD 1998), an experiment with *D. magna* must ensure a mortality below 20% in the control after a

21-day experiment. So, for parameter m_0 , the prior distribution of the $\log_{10}(m_0)$ was upper bounded by this accepted maximal mortality: if $S(21) = 1 - 0.2 = e^{-\max_{m_0} \times 21}$ then $\max_{m_0} = -\ln(0.8)/21 \simeq 0.01/\text{day}$. We finally used a uniform prior distribution for $\log_{10}(m_0)$ in $[-5; -2]$ for all data sets for which parameter m_0 was relevant. Concerning parameter k_S , we defined the upper bound of a uniform $\log_{10}(k_S)$ prior distribution by considering the lowest difference between two successive tested concentrations c_j and c_{j+1} . Hence, we assumed that k_S should be maximum in extreme or rare experimental situation in which concentration c_j would have no effect while concentration c_{j+1} would lead to a 99% effect after a 21-day exposure (99% was chosen instead of 100% to avoid any mathematical problem due to non defined functions). Such a case may be considered as extreme because the tested concentrations are usually chosen to induce intermediate responses strictly contained between 0 and 100% of effect. So, if

$S(c_{j+1} - c_j, 21) = 0.01 = e^{-\max_{k_S} \times (c_{j+1} - c_j) \times 21}$ then $\max_{k_S} = \frac{-\ln(0.01)}{(c_{j+1} - c_j) \times 21}$. For the lower bound of the uniform $\log_{10}(k_S)$ distribution, we arbitrarily chose: $\log_{10}(\min_{k_S}) = \log_{10}(\max_{k_S}) - 2$.

Computations

Bayesian inference was performed using JAGS and the *rjags* R-package (Plummer 2010). For each model, three independent MCMC chains were run in parallel. For each data set and each model, after an initial burn-in period of 20,000 iterations, the Bayesian algorithm was run between 30,000 and 60,000 iterations. This number of iterations was adapted to ensure convergence. Convergence was checked by both displaying MCMC chain traces and analyzing the Gelman and Rubin's statistics (Gelman and Rubin 1992).

Critical effect concentration comparisons and statement of uncertainty

After we selected a method/model for each type of critical effect concentration according to criteria previously described, we compared the NOEC, LC_x and NEC values and their uncertainty intervals for each data set.

The NOEC could be characterized by a point value (a tested concentration) and by an uncertainty interval bounded by 0 and the LOEC value. Such an interval is not a confidence interval accompanied with a level of confidence (Delignette-Muller et al. 2011). Nevertheless, as it is well-known that "absence of evidence is not evidence of absence" (Altman and Bland 1995), the only statistically well-founded conclusion one can draw from the results of hypothesis tests used to determine LOEC/NOEC values, is

that the LOEC value has an effect, and so that the true no effect concentration is below the LOEC value.

Bayesian inference provided the posterior distribution of model parameters among which LC_{50} or NEC. Credibility intervals (CI) can be easily extracted as an interval containing a chosen percentage of the posterior distribution. We used the 2.5 and 97.5 percentiles of the posterior distribution to define a 95% CI. The posterior distribution also provided credibility intervals of any combination of parameters, like for example LC_x with $x \neq 50\%$ as a combination of LC_{50} and b .

For the log-logistic model: $LC_x = LC_{50} / 10^{\frac{\ln(\frac{100-x}{x})}{b}}$;
 for the log-probit model: $LC_x = LC_{50} \times 10^{\left(\frac{F^{-1}\left(\frac{x}{100}\right)}{b}\right)}$;
 and for the Weibull model: $LC_x = LC_{50} \times \left(\frac{-\ln\left(1 - \frac{x}{100}\right)}{\ln(2)}\right)^{\frac{1}{b}}$.

We first calculated LC_5 values and their 95% credibility interval because the value $x = 5$ was chosen to approximate at best "no effect" concentrations. We also determined LC_{10} , LC_{20} and their 95% credibility intervals to highlight the impact of the chosen x and to sweep across the range of x values commonly used in chronic ecotoxicology in addition to LC_{50} (van der Hoeven et al. 1997).

Results

Hypothesis-testing methods

LOEC/NOEC values for the three tested methods are given in Table 2 for the nine data sets. We noticed that NOEC values were highly dependent on the method used. The Jonckheere–Terpstra test with the step-down procedure

Table 2 NOEC and LOEC values determined for each contaminant after a 21-day exposure by three hypothesis-testing methods: the Cochran–Armitage test (C–A) with a step-down procedure (SDP), the

Jonckheere–Terpstra test (J–T) with SDP and the Fisher exact test with the Bonferroni–Holm (BH) correction

Contaminant	Hypothesis-testing methods		
	C–A and SDP NOEC; LOEC	J–T and SDP NOEC; LOEC	Fisher with BH NOEC; LOEC
Copper (µg/L)	10; 20	20; UD	10; 20
Zinc (mg/L)	0.074; 0.22	0.66; 2	0.074; 0.22
Cadmium (µg/L)	0; 2.5	7.5; UD	0; 2.5
Potassium dichromate (mg/L)	0.18; 0.32	1; UD	0.32; 0.56
Chlordan (µg/L)	1.82; 2.9	7; UD	1.82; 2.9
Ivermectin (pg/L)	1; 10	10; 100	1; 10
Oxolinic acid (mg/L)	0.38; 0.75	3; UD	0.75; 1.5
Streptomycin (mg/L)	16; 32	64; UD	32; 64
Sulfadiazine (mg/L)	25; 50	100; UD	50; 100

UD Undefined value

was rarely able to detect a significant difference between the control and the tested concentrations thus providing NOEC values equal to the highest concentration, except in two cases (zinc and ivermectin). The Cochran–Armitage method was more efficient than the Fisher one to detect

significant effects because of its underlying hypothesis of a monotonic expected response. Hence, the Cochran–Armitage test with the step-down procedure was more protective for the environment with the smallest NOEC values; we decided to use results of the Cochran–Armitage

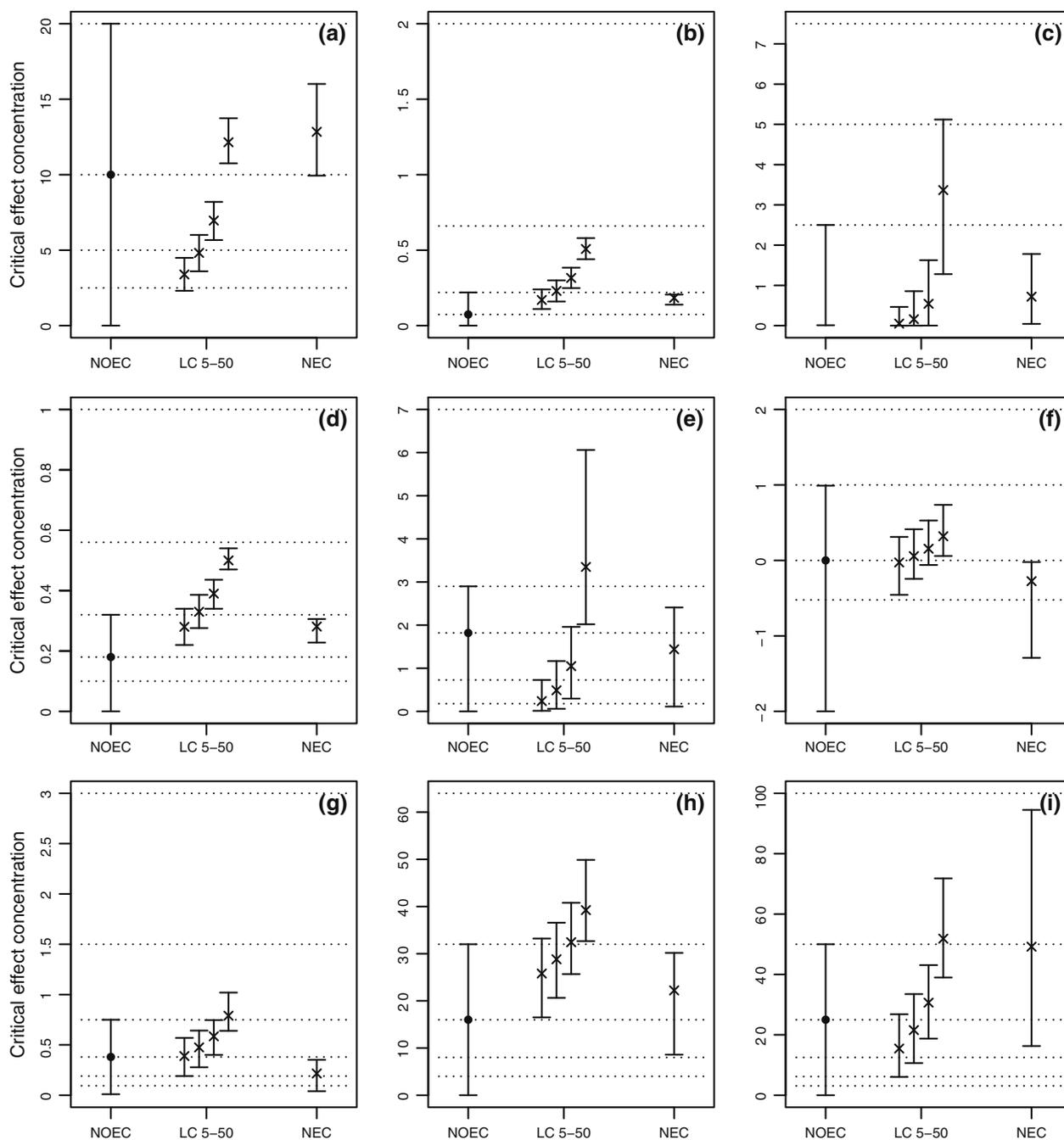


Fig. 3 Critical effect concentration values expressed as NOEC and [0, LOEC] intervals estimated from the Cochran–Armitage test with a step-down procedure; as LC₅, LC₁₀, LC₂₀, LC₅₀ and 95% CI estimated from the Weibull model; as NEC and 95% CI estimated from the most appropriate NEC-model. Horizontal dotted lines represent

the tested concentrations. For ivermectin, we chose to represent results in log₁₀ to gain visibility. **a** Copper (µg/L), **b** zinc (mg/L), **c** cadmium (mg/L), **d** potassium dichromate (mg/L), **e** chlordan (µg/L), **f** ivermectin in log₁₀ scale (log₁₀ pg/L), **g** oxolinic acid (mg/L), **h** streptomycin (µg/L), **i** sulfadiazine (µg/L)

method for further comparisons (Fig. 3). In Fig. 3, we noticed that the $[0, \text{LOEC}]$ interval was, as expected, sometimes very large and uninformative (copper (a)), or on the contrary very narrow (zinc (b)) depending on the experimental design. In the case of cadmium (c), the lowest tested concentration was significantly different from the control, so there was no point value of NOEC but only an uncertainty interval $[0; \text{LOEC}]$.

Concentration–response models

All parameter values and mean deviances of the three concentration–response models, estimated for each contaminant at day 21, are given in Table 3. Mean deviances were very close to each other, and we noticed a lack of convergence of MCMC chains with the log–probit model for chlordan and ivermectin. The three concentration–response models gave close LC_{50} values. Nevertheless, the Weibull model gave the best fit for 6 data sets out of 9, so it was preferred and used afterwards. The fits of the Weibull model at day 21 are shown in Fig. 1: the data sets are represented as black points on the left graphs with the darkest points corresponding to data at day 21 (a_1 – i_1) and the model fits as dashed lines on the right graphs (a_2 – i_2). For zinc, potassium dichromate, ivermectin, oxolinic acid and streptomycin, the Weibull model globally well described the data (Fig. 1b, d, f–h), whereas for copper, cadmium, chlordan and sulfadiazine, the Weibull model did not well capture the observed effects (Fig. 1a, c, e, i). This may be due to the natural mortality observed in the control but not described by the model. LC_x values ($x = 5, 10, 20, 50$) and their credibility intervals are represented in Fig. 3. As expected, we can see a strong effect of x on the LC_x point estimates. We might also have expected the LC_x uncertainty interval to be larger for the smaller x values, which is not the case. On the contrary, a larger uncertainty interval of LC_{50} values is observed for data sets where only a moderate mortality is reached at day 21 (cadmium (c) and chlordan (e)).

NEC-model

Posterior distribution median values and credibility intervals of the NEC-model parameters are summarized in Table 4. Fits of the most appropriate NEC-model for each data set are shown in Figure 1 (darker and darker plain lines, right graphs a_2 to i_2). In Figure 1, the parallel can be made between the trend observed in data (left graphs) and the trend described by the chosen NEC-model (right graphs). We noticed that the models well described the main features observed in the data sets. The natural mortality clearly observed for copper, cadmium, chlordan and sulfadiazine (Figure 1 (a1), (c1), (e1), (i1)) was well

described by the selected model (Figure 1 (a2), (c2), (e2), (i2)). The transition between range of concentrations without any effect and range of concentrations with increasing effects seems to be also well captured by the fitted models (see for example Figure 1 (d2), (h2)). Finally, the uncertainty intervals around the NEC values were relatively narrow, reflecting the high precision of these estimates in most of the cases (Table 4, Figure 3).

Critical effect concentration comparisons

All critical effect concentrations calculated with the most appropriate method or model are gathered together in Fig. 3. Except for cadmium, NOEC values, LC_5 and NEC median values were greater than or equal to the lowest tested concentration, highlighting the relevance of the experimental designs. The uncertainty was lower for LC_5 (6/9 data sets) than for NEC (3/9 data sets) or for NOEC (0/9 data sets). For the contaminant for which a model with two parameters was used, the NEC uncertainty was lower than the LC_5 uncertainty. Whereas for contaminant for which more than two parameters were used, the NEC uncertainty was higher than the LC_5 one. Moreover, NEC point estimates were always included in NOEC or LC_x intervals. In cases where the natural instantaneous mortality was negligible, NEC point estimates were close to or below LC_5 ones (Fig. 3 b, d, f–h) whereas in cases where the natural instantaneous mortality was not negligible, NEC point estimates were higher than LC_5 ones (Fig. 3a, c, e, i). For these latter cases, we previously observed that the Weibull model did not well capture the trend of observed data (Fig. 1a₂, c₂, e₂, i₂), so the corresponding low values of LC_5 may be underestimated due to not taking natural mortality into account.

Discussion

In this paper, we first compared the methods or models commonly used for calculation of NOEC, LC_x and NEC values determined after a 21-day exposure experiment. In contrast to LC_x and NEC, NOEC values are not model dependent, which could be considered as an advantage (van der Hoeven 1997). Nevertheless, in accordance with the literature (Chapman et al. 1996) we showed that NOEC values strongly depended on the hypothesis-testing method used. Indeed, the Cochran–Armitage test with a step-down procedure revealed itself as the method giving the most protective threshold for the environment. But one of the very often reported drawbacks of the NOEC is that it is inevitably one of the tested concentrations (de Bruijn and Hof 1997). At last, the uncertainty interval on the NOEC is also inevitably bounded by 0 and the

Table 3 Values of LC_{50} and b for the three concentration–response models were estimated for each tested contaminant after a 21-day exposure

Contaminant	Parameter	Log-logistic model	Log-probit model	Weibull model
Copper	LC_{50} ($\mu\text{g/L}$)	12.0 (10.3–14.2)	11.7 (9.94–14.4)	12.1 (10.7–13.7)
	b	5.18 (3.92–6.60)	2.70 (2.08–3.36)	2.04 (1.58–2.54)
	MD	65.8	71.6	52.5
Zinc	LC_{50} (mg/L)	0.465 (0.398–0.541)	0.458 (0.395–0.532)	0.508 (0.444–0.578)
	b	7.33 (5.63–9.48)	4.15 (3.24–5.25)	2.40 (1.76–3.26)
	MD	11.7	10.5	10.7
Cadmium	LC_{50} ($\mu\text{g/L}$)	3.32 (1.40–4.92)	3.32 (1.37–4.93)	3.36 (1.28–5.12)
	b	2.28 (0.561– 4.27)	1.39 (0.345–2.60)	0.627 (0.148–1.19)
	MD	19.2	19.3	19.7
Potassium dichromate	LC_{50} (mg/L)	0.483 (0.443–0.526)	0.479 (0.440–0.522)	0.504 (0.467–0.543)
	b	13.8 (10.4–18.4]	7.77 (5.97–10.0]	4.56 (3.27–6.40)
	MD	11.6	10.5	10.4
Chlordan	LC_{50} ($\mu\text{g/L}$)	3.13 (1.80–6.11)	No convergence	3.35 (2.02–6.06)
	b	2.88 (1.31–5.12)	No convergence	0.985 (0.459–1.64)
	MD	19.3	–	18.7
Ivermectin	LC_{50} (pg/L)	1.44 (1.04–4.28)	No convergence	2.09 (1.14–5.45)
	b	16.00 (4.27–89.72)	No convergence	3.36 (1.27–9.40)
	MD	3.41	–	3.38
Oxolinic acid	LC_{50} (mg/L)	0.761 (0.600–0.976)	0.745 (0.586–0.968)	0.794 (0.643–1.02)
	b	10.7 (5.85–20.6)	5.71 (3.29– 9.48)	3.59 (1.91–7.36)
	MD	8.53	8.41	7.18
Streptomycin	LC_{50} (mg/L)	35.0 (32.1–45.1)	34.2 (32.0–44.1)	39.2 (32.7–49.8)
	b	36.3 (11.0–94.9)	26.6 (6.62–93.9)	6.29 (3.00–9.75)
	MD	3.72	3.58	3.71
Sulfadiazine	LC_{50} (mg/L)	49.8 (35.7–75.9)	48.0 (34.2–76.5)	51.8 (39.0–71.8)
	b	5.66 (3.08–9.48)	3.08 (1.73–4.87)	2.14 (1.21–3.51)
	MD	15.2	15.1	13.0

The results correspond to the median values of the posterior distributions with their 95% credibility intervals. To evaluate the goodness of fit, the mean deviance (MD) of the three models for a given contaminant was determined. The model with the minimum mean deviance provided the best fitting (bold MD value)

LOEC value, what is not very informative for risk managers in terms of a concentration threshold where no hazard is observed.

The LC_{50} values obtained from the three concentration–response models were close to each other (Table 3), in accordance with van der Hoeven (1997) and Ritz (2010). It was also observed for LC_x values where $5 \leq x < 50$ (results not shown). Because of our findings, we selected the Weibull model as the most appropriate one, based on its small global mean deviance. Moreover, this model does not require the log conversion of concentrations and the control data do not have to be discarded (Ritz 2010). However, the log-logistic model would also be relevant, contrary to the log-probit model for which parameters can not always be estimated. These results were in agreement with the work of Isnard et al. (2001) who also highlighted a number of problems with parameter estimation using the log-probit model when comparing the same three models but using a

frequentist approach. Moreover, the strong impact of the x value on the LC_x is an inconvenience for risk managers who have to choose an x value to determine an acceptable effect level (van der Hoeven 1997, 2004). Contrary to what we might have thought, LC_5 estimates were globally as precise as LC_{50} and even more precise than LC_{50} estimates when there was not enough mortality in data sets at day 21. So, the lowest LC_x (i.e., LC_5) among those recommended to use (van der Hoeven et al. 1997; Warne and van Dam 2008) appeared to be relatively precisely estimated with all data sets we studied.

We estimated NEC values by using a threshold model where the instantaneous mortality is a function of both concentration and time, keeping in mind the sake of simplicity. The large number of available data from day 0 to day 21 allowed us to fit a model with at least two parameters and at most four parameters if necessary. Such a flexible model appeared to be relevant and led to a good

Table 4 Parameter median values and their 95% credibility intervals for the most appropriate NEC-model with two, three or four parameters depending on the contaminant

Contaminant	Parameters	Parameter values
Copper	NEC, k_S , m_0	NEC = 12.84 (9.94–16.01) $\mu\text{g/L}$ k_S = 0.0501 (0.0309–0.0842) L/ $\mu\text{g/day}$ m_0 = 0.00574 (0.00379–0.00824) /day
Zinc	NEC, k_S	NEC = 0.184 (0.139–0.206) mg/L k_S = 0.143 (0.112–0.180) L/mg/day
Cadmium	NEC, k_S , m_0	NEC = 0.720 (0.0427–1.78) $\mu\text{g/L}$ k_S = 0.0121 (0.00840–0.0176) L/ $\mu\text{g/day}$ m_0 = 0.00744 (0.0049–0.00985) /day
Potassium dichromate	NEC, k_S	NEC = 0.281 (0.228–0.306) mg/L k_S = 0.145 (0.112–0.184) L/mg/day
Chlordan	NEC, k_S , m_0	NEC = 1.44 (0.113–2.41) $\mu\text{g/L}$ k_S = 0.0145 (0.0123–0.0240) L/ $\mu\text{g/day}$ m_0 = 0.00261 (0.000144–0.00816) /day
Ivermectin	NEC, k_S , m_{max}	NEC = 0.531 (0.0513–0.952) pg/L k_S = 0.00990 (0.00744–0.00865) L/pg/day m_{max} = 0.185 (0.0833–0.342) /day
Oxolinic acid	NEC, k_S	NEC = 0.216 (0.0393–0.352) mg/L k_S = 0.0379 (0.0251–0.0587) L/mg/day
Streptomycin	NEC, k_S	NEC = 22.2 (8.6–30.2) mg/L k_S = 0.00126 (0.000617–0.00261) L/mg/day
Sulfadiazine	NEC, k_S , m_0	NEC = 49.2 (16.3–94.5) mg/L k_S = 0.00431 (0.00106–0.0370) L/mg/day m_0 = 0.00447 (0.00106–0.00900) /day

description of our nine data sets on the whole. By definition, NEC by definition, the NEC based model exhibited a greater flexibility than concentration–response models. Hence, when it was required, the natural instantaneous mortality was thus estimated easily and precisely. Even if it could be theoretically possible to add a natural mortality in the Weibull model as well, the number of data at day 21 solely was too small to properly estimate three parameters (Ritz 2010). In particular, in such cases with a non negligible natural mortality, LC_5 estimates (at day 21) were lower than NEC ones, because the concentration–response model mistook the background effect for the effect due to the contaminant. In these four cases, the credibility intervals around LC_5 were lower than the credibility intervals around NEC what is not surprising because of the greater complexity of NEC-model due to natural mortality.

The comparison between critical effect concentrations showed that: (i) none of the three types of critical effect concentrations systematically appeared as the most protective of the environment when only looking at point values for the nine contaminants we studied; (ii) credibility intervals around NEC values were in most cases relatively narrow; and (iii) NEC values belonged to NOEC or LC_x intervals. Therefore, it is reasonable to propose that the use of a NEC-model may be a better option in terms of precision and environment protectiveness, meeting Pires et al. (2002)

findings. The use of NEC, however, remains debatable because of biological reasons as summarized by van Straalen (1997) (e.g., some physiological reactions may rise to a graded, non-threshold response).

In support of the NEC-model, we would like to suggest to opponents that: (i) a threshold model (e.g., NEC-model) could simply be viewed as an approximation of a model without a threshold (e.g., the Weibull model) with the corresponding value used as a guidance rather as an absolute value (Wennig, 2000); (ii) the threshold method we used must be taken in context to the biological response under modeling consideration: survival probability is not necessarily as affected as are more sensitive physiological processes; (iii) the use of a “no effect” threshold in risk assessment is already a reality with NOEC and is a necessity for decision makers who still use NOEC despite its poor statistical meaning.

Conclusions

Our results showed that:

- (i) LOEC/NOEC values at day 21 were method-dependent, and that the Cochran–Armitage test with a step-down procedure appeared to be the most protective for the environment;

- (ii) all concentration–response models we compared gave close values of LC_{50} at day 21, nevertheless the Weibull model had the lowest global mean deviance;
- (iii) a simple threshold NEC-model both concentration and time dependent completely described whole data sets (i.e., all timepoints) and enabled a precise estimation of the NEC.

Moreover the use of LC_x are now clearly criticized (Jager 2011) and so we think that it would be useful for ecotoxicologists to develop convenient and user-friendly tools appropriate for testing concentration and time-based survival data using NEC and associated NEC-models. It would also be interesting to expand this approach when analyzing other types of contaminant response data associated with growth, reproduction, etc.

Acknowledgments We would like to thank Martyn Plummer for developing the useful JAGS and r-jags tools and David Fox for his precious comments. We also would like to thank the French Ministry of Higher Education and Research for providing financial support.

References

- Altman D, Bland J (1995) Absence of evidence is not evidence of absence. *Br Med J* 311:485–485
- Armitage P, Colton T (2005) Trend test for counts and proportions in encyclopedia of biostatistics, 2nd edn, vol 8 (5516–5522). Wiley, Chichester
- Baas J, Jager T, Kooijman SALM (2009) Estimation of no effect concentrations from exposure experiments when values scatter among individuals. *Ecol Model* 220:411–418
- Bedaux J, Kooijman SALM (1994) Statistical analysis of bioassays, based on hazard modeling. *Environ Ecol Stat* 1:303–314
- Billoir E, Delignette-Muller ML, Péry A, Charles S (2008) A Bayesian approach to analyzing ecotoxicological data. *Environ Sci Technol* 42:8978–84
- Billoir E, Delignette-Muller ML, Péry A, Geffard O, Charles S (2008) Statistical cautions when estimating DEBtox parameters. *J Theor Biol* 254:55–64
- de Bruijn J, Hof M (1997) How to measure no effect. Part IV: how acceptable is the EC_x from an environmental policy point of view?. *Environmetrics* 8:263–267
- Cairns J (1992) The threshold problem in ecotoxicology. *Ecotoxicology* 1:3–16
- Chapman P, Caldwell R, Chapman P (1996) A warning: NOECs are inappropriate for regulatory use. *Environ Toxicol Chem* 15:77–79
- Crane M, Newman M (2000) What level of effect is a no observed effect? *Environ Toxicol Chem* 19:516–519
- Delignette-Muller ML, Forfait C, Billoir E, Charles S (2011) A new perspective on the Dunnett procedure: filling the gap between-NOEC/LOEC and EC_x concepts. *Environ Toxicol Chem* 30: 2888–2891
- Ducrot V, Billoir E, Péry A, Garric J, Charles S (2010) From individual to population level effects of toxicants in the tubicifid *Branchiura sowerbyi* using threshold effect models in a Bayesian framework. *Environ Toxicol Chem* 44:3566–71
- Fox D (2008) NECS, NOECs and the EC_x . *Australas J Ecotoxicol* 14:7–9
- Fox D (2010) A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicol Environ Saf* 73:123–131
- Gelman A, Rubin D (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–511
- van der Hoeven N (1997) How to measure no effect. Part III : statistical aspects of NOEC, EC_x and NEC estimates. *Environmetrics* 8:255–261
- van der Hoeven N (2004) Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheor* 52:201–217
- van der Hoeven N, Noppert F, Leopold A (1997) How to measure no effect. Part I: towards a new measure of chronic toxicity in ecotoxicology. Introduction and workshop results. *Environmetrics* 8:241–248
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Isnard P, Flammarion P, Roman G, Babut M, Bastien P, Bintein S, Esserméant L, Féraud J, Gallotti-Schmitt S, Saouter E, Saroli M, Thiébaud H, Tomassone R, Vindimian E (2001) Statistical analysis of regulatory ecotoxicity tests. *Chemosphere* 45:659–69
- Jager T (2011) Some good reasons to ban EC_x and related concepts in ecotoxicology. *Environ Sci Technol* 45:8180–8181
- Jonckheere A (1954) A distribution-free k-sample test against ordered alternatives. *Biometrika* 41:133–145
- Kooijman SALM, Bedaux J (1996) The analysis of aquatic toxicity data. VU University Press, Amsterdam
- Lopes C, Charles S, Vولات B, Garric J (2009) Toxicity of ivermectin on cladocerans: comparison of toxic effects on *Daphnia* and *Ceriodaphnia* species. *Environ Toxicol Chem* 28:2160–2166
- Manar R, Bessi H, Vasseur P (2009) Reproductive effects and bioaccumulation of chlordane in *Daphnia magna*. *Environ Toxicol Chem* 28:2150–2159
- Nelder J (1999) From statistics to statistical science. *J R Stat Soc Ser D (The Statistician)* 48:257–267
- Newman M (2008) “What exactly are you inferring?” a closer look at hypothesis testing. *Environ Toxicol Chem* 27:1013–1019
- Ntzoufras I (2009) Bayesian modeling using WinBugs. Wiley Series in computational statistics. Hoboken, NJ
- OECD (1998) Guidelines for testing of chemicals n° 211: *Daphnia magna* reproduction test. Technical report, Organisation for Economic Cooperation and Development
- OECD (2006) Current approaches in the statistical analysis of ecotoxicity data: a guidance to application. Technical report, Organisation for Economic Cooperation and Development
- Pires A, Branco J, Picado A, Mendonca E (2002) Models for the estimation of a ‘no effect concentration’. *Environmetrics* 13(1): 15–27
- Plummer M (2010) R package version 2.1.0-10
- R Development Core Team (2010) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL <http://www.R-project.org/>. ISBN 3-900051-07-0
- Ritz C (2010) Toward a unified approach to dose-response modeling in ecotoxicology. *Environ Toxicol Chem* 29:220–229
- Santojanni A, Gorbi G, Sartore F (1995) Prediction of mortality in chronic toxicity tests on *Daphnia magna*. *Water Res* 29:1453–1459
- van Straalen N (1997) How to measure no effect. Part II: threshold effects in ecotoxicology. *Environmetrics* 8:249–253
- Warne MSJ, van Dam R (2008) NOEC and LOEC data should no longer be generated or used. *Australas J Ecotoxicol* 14:1–5
- Wennig R (2000) Threshold values in toxicology-useful or not? *Forensic Sci Int* 113:323–330
- Wollenberger L, Halling-Sørensen B, Kusk K (2000) Acute and chronic toxicity of veterinary antibiotics to *Daphnia magna*. *Chemosphere* 40:723–730

Short Communication

A NEW PERSPECTIVE ON THE DUNNETT PROCEDURE:
FILLING THE GAP BETWEEN NOEC/LOEC AND EC_x CONCEPTS

MARIE-LAURE DELIGNETTE-MULLER,*†§ CAROLE FORFAIT,† ELISE BILLOIR,†‡ and SANDRINE CHARLES†

†Université de Lyon, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France

‡Pôle de Recherche ROVALTAIN en Toxicologie Environnementale et Ecotoxicologie, Valence, France

§Université de Lyon, Lyon, and VetAgro Sup, Marcy l'Etoile, France

(Submitted 21 June 2011; Returned for Revision 17 August 2011; Accepted 8 September 2011)

Abstract—The no-observed-effect concentration (NOEC) is known to be based on a wrong usage of hypothesis tests, and the use of confidence intervals is preferred. The purpose of the present study is to provide an easy and proper way to interpret ecotoxicological tests based on simultaneous confidence intervals associated with the commonly used Dunnett procedure, and to show how these intervals may allow one to infer EC_x values (effective concentrations). *Environ. Toxicol. Chem.* 2011;30:2888–2891. © 2011 SETAC

Keywords—Ecotoxicological bioassays Simultaneous confidence intervals Toxicity endpoints Multiple comparisons

INTRODUCTION

In ecotoxicology, the effects of contaminants on living organisms are usually measured at the individual level, in the laboratory, and according to standards. This ensures the reproducibility of bioassays and the control of environmental factors, often confusing in an *in situ* approach. These bioassays, which test acute or chronic toxicity, generally apply to reproduction, growth, or survival of organisms. The most widely used analysis of standardized bioassays leads to the determination of the no-observed-effect concentration (NOEC) or the lowest-observed-effect concentration (LOEC) on the basis of null-hypothesis significance tests as recommended in the Organisation for Economic Co-operation and Development guideline [1]. However, the relevance of these toxicity endpoints has often been questioned [2–5]. From a statistical point of view, a strong argument against the NOEC is that it relies on an improper use of *p* values, which are not intended to establish biological thresholds. Indeed, the NOEC is defined as the higher experimentally tested concentration for which the difference to the control is not statistically significant, but it is widely used in risk assessment as a zero-effect concentration, which is an incorrect interpretation [2]. In fact, this misinterpretation of *p* values is very common and comes from the wrong assumption that absence of evidence is evidence of absence [6]. Many scientists reported this type of error in various research areas and advocated the use of confidence intervals instead of null-hypothesis significance tests [4,7,8]. Nevertheless, to our knowledge, no method based on confidence intervals has been proposed in ecotoxicology to replace the NOEC determination by using a Dunnett procedure, and a current belief seems to be that no confidence interval can be given on the NOEC.

To show that a tested concentration of a contaminant has no or at most a negligible effect on the response to an ecotoxicological bioassay means to show that the difference between this tested concentration and the control is null or negligible. This

task is closely linked to that of bioequivalence studies commonly performed in drug trials, for which a statistical procedure is now well established [9]. To ascertain bioequivalence between two drugs for a pharmacokinetic measure, the 90% confidence interval for the ratio of the two means should fall within a bioequivalence limit corresponding to an acceptable relative difference of 20%. In fact, one may not state equivalence from confidence intervals without defining a level of acceptable difference. In the same way, to show that a tested concentration of a contaminant has no or a negligible effect on the response of ecotoxicological bioassays, one needs to previously define what is called a negligible effect (e.g., 5%, 10%, 20%) as is done when classically estimating *x*% effective concentrations (EC_x) using concentration–response curves.

In a bioequivalence study, because only two treatments (two drugs) are compared, only one confidence interval should be determined, the one on the ratio of the two means obtained with the two treatments. In an ecotoxicological bioassay, several treatments (several concentrations of a contaminant) are compared with the control, using multiple comparisons, in the context of which determining confidence intervals is more difficult. However, simultaneous confidence intervals for relative effects between several treatments and a control were recently developed [10,11] and provided in packages associated with the *R* statistical environment [12]. The purpose of the present study is to illustrate how such packages may be used to analyze ecotoxicological bioassays and how simultaneous confidence intervals may be interpreted by analogy to what is performed in bioequivalence trials, thus naturally filling the gap between NOEC and EC_x values.

MATERIALS AND METHODS

In recent years, several methods have been developed to calculate simultaneous confidence intervals for ratios of means, especially ratio-to-control associated with the Dunnett procedure, the most common method used to assess the significance of differences between the control response and responses at all other concentrations [3]. We chose to calculate simultaneous 95% confidence intervals for ratios of each mean to the mean of the control using the plug-in method provided within the *R*

* To whom correspondence may be addressed
(ml.delignette@vetagro-sup.fr).

Published online 19 September 2011 in Wiley Online Library
(wileyonlinelibrary.com).

package *mratios* [10]. This method calculates approximate simultaneous confidence intervals from a multivariate *t* distribution, which correlation matrix is estimated from the maximum likelihood estimation of the ratios. The very simple *R* code required to calculate and vertically plot these confidence intervals is provided in the Appendix.

To help interpret the results, horizontal lines were added to each plot, corresponding to ratio-to-control values of 1 (higher line), 0.95, and 0.90, that is, to no effect, 5% effect, and 10% effect respectively. Considering an *x*% effect and its corresponding horizontal line, if the ratio-to-control confidence interval for one concentration is entirely below or above this line, we can say that the EC_{*x*} is, respectively, below or above this concentration. When the confidence interval overlaps the line, nothing can be concluded. Looking at the confidence intervals for all the tested concentrations, it is thus generally possible to bound the EC_{*x*} whatever *x*, including 0 (EC₀ corresponding to a zero-effect, often incorrectly replaced by the NOEC). Bounds obtained in this way do not strictly define confidence intervals, but they may be interpreted as bounds of uncertainty intervals: they represent what may be inferred on EC_{*x*} using confidence intervals on ratios to control.

Uncertainty intervals were thus determined for EC₀, EC₅, and EC₁₀. For the sake of comparison, EC₅ and EC₁₀ were also estimated from the fit of the three-parameter log-logistic model or the four-parameter Brain-Cousens model in case of observed hormesis [13], using the *R* package *drc* [14].

For our purpose, we analyzed four data sets collected during standard 21-d *Daphnia magna* bioassays. Daphnids were exposed either to Zn (0, 0.074, 0.22, 0.66 mg/L) [15], Cu

(0, 2.5, 5, 10 μg/L) [15], chlordane (insecticide) (0, 0.18, 0.73, 1.82, 2.9, 7 μg/L) [16], or Cd (0, 0.37, 0.86, 1.64, 4.36 μg/L) (H. Delhaye and B. Clément, Université de Lyon, personal communication). Observed data were body length measurements (in mm) at the end of the experiment (after 21 d) as shown in Figure 1.

RESULTS AND DISCUSSION

Simultaneous 95% confidence intervals (CI) for ratio of means to control mean were reported in Figure 2 for each studied contaminant and each tested concentration. Bounds for EC₀, EC₅, and EC₁₀ were then easily deduced from these plots. Let us take the example of the inference on EC₁₀ value for Zn: looking at the 95% CI for ratios of mean to control mean for each tested concentration, for 0.22 mg/L it is entirely above the horizontal line corresponding to a ratio of 0.90 (a 10% effect), whereas for 0.66 mg/L it is entirely below that line. It thus can be inferred that the EC₁₀ value lies between these two concentrations. For each dataset, Table 1 reports bounds for EC₀, EC₅, and EC₁₀ likewise deduced from 95% simultaneous intervals on ratio-to-control. Table 1 also shows 95% confidence intervals on EC₅ and EC₁₀ as estimated from the fitting of the 3-parameter log-logistic model or the Brain-Cousens model (for Cu).

For EC₀, all of the intervals are left bounded by 0 and right bounded by the LOEC. But with a lower bound at 0, such an interval cannot help regulators to define acceptable effect levels. To obtain a more informative interval from a Dunnett procedure, one must specify what is considered to be a negligible effect, and so infer on EC_{*x*} values, with a non-null value of *x*.

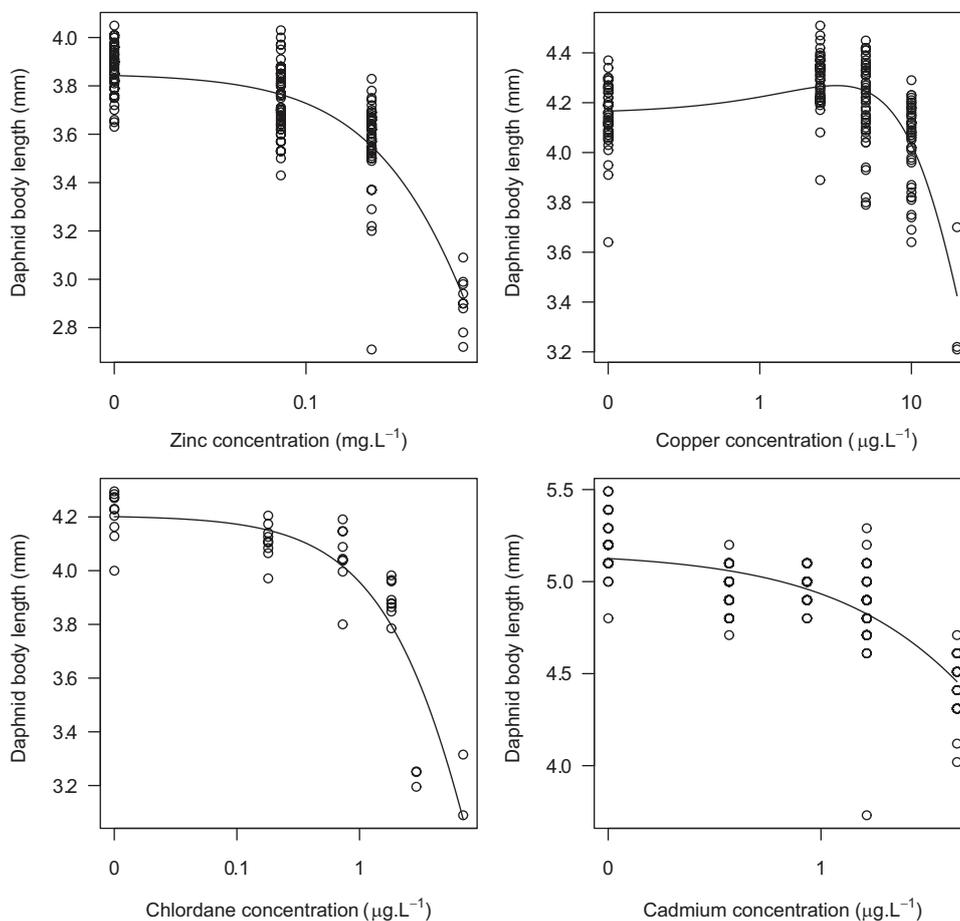


Fig. 1. Daphnid body length measurements as a function of concentration (log scale), with the log-logistic model or the Brain-Cousens model (for Cu) fitted to data.

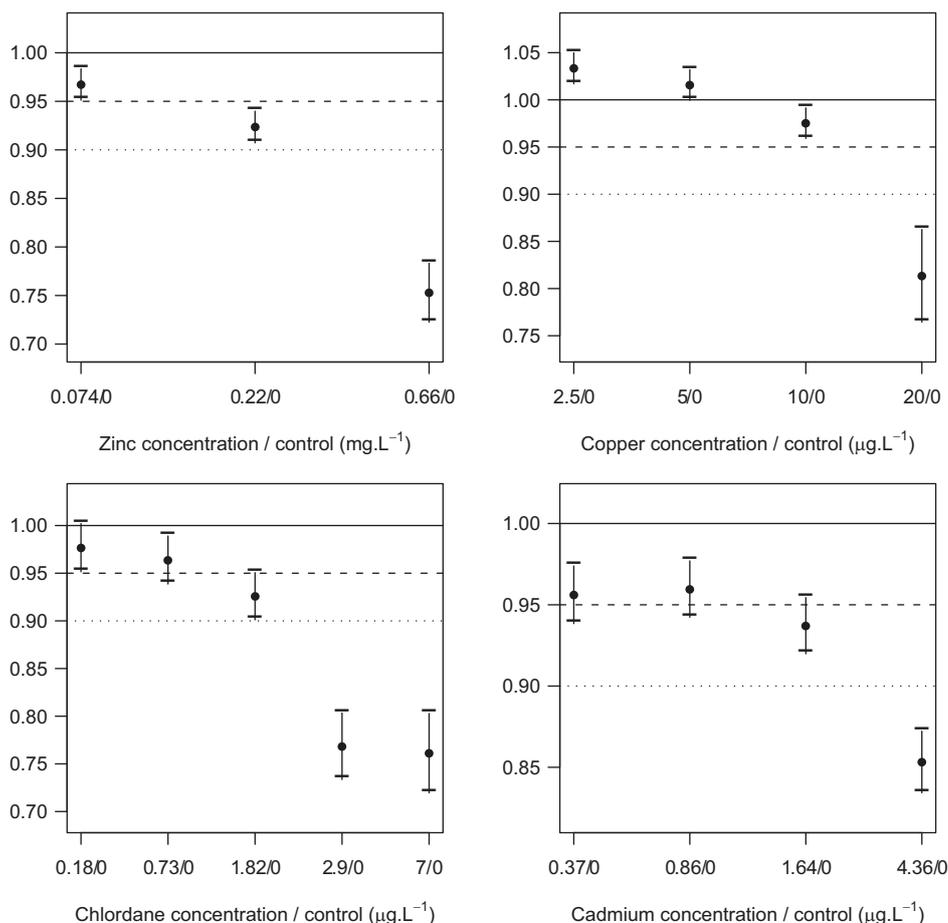


Fig. 2. Simultaneous 95% confidence intervals for ratios of body length mean m_k for each concentration c_k , to the mean m_0 of the control (concentration c_0 , generally 0), with lines indicating no effect (EC0 : ratio = 1), 5% effect (EC5 : ratio = 0.95), and 10% effect (EC10 : ratio = 0.90). On the x label the two concentrations c_k and c_0 are indicated as c_k/c_0 .

The EC5 and EC10 bounds given by fitting a dose–response model or defined from simultaneous confidence intervals are in agreement, always overlapping. That should be confirmed by comparing both methods on a greater number of data sets. In contrast to NOEC-based methods, the method using simultaneous confidence intervals can represent the uncertainty attributable to the number of observations or to the variability of observations. For example, with the Cd data set, the low number of replicates induces a great uncertainty on the EC5 (Figs. 1 and 2 and Table 1). Moreover, intervals proposed in the present study do not rely on any model assumption; this may be seen as an advantage over the classical ECx, sometimes criticized because of being model-dependent [2,17]. Nevertheless, that model-dependence may not be such a problem when toxicity endpoints are estimated by interpolation (without extrapolation)

[18] and, on the four tested data sets, the regression method gave smaller uncertainty intervals than those we proposed that are inevitably bounded by concentrations of the experimental design (Table 1). Thus, if the method proposed in the present study appears to be a good alternative to NOEC-based method for data designed to be analyzed with the Dunnett procedure, the fit of models may be a more efficient way to estimate ECx values, but it should be accompanied by an adaptation of experimental designs, including more closely spaced treatments and fewer replicates [18].

Following the same objective of preventing the wrong interpretation of the NOEC as a concentration that would cause no toxic effect, some authors also advised that the NOEC should be accompanied by the minimum significant difference at the NOEC [19]. Indeed, such a measure indicates how much

Table 1. Bounds for EC0, EC5, and EC10 values deduced from 95% simultaneous confidence intervals (CI) on ratio-to-control (Fig. 2) and 95% confidence intervals on EC5 and EC10 values estimated from the fitting of the log-logistic or the Brain-Cousens model (for copper) (Fig. 1)

	Zinc (mg/L)	Copper ($\mu\text{g/L}$)	Chlordane ($\mu\text{g/L}$)	Cadmium (mg/L)
Bounds for EC0 from 95% CI	0; 0.074	0; 10	0; 0.73	0; 0.37
Bounds for EC5 from 95% CI	0.074; 0.22	10; 20	0.18; 2.9	0; 4.36
95% CI on EC5 log-logistic model	0.11; 0.18	10.6; 12.4	0.41; 1.24	0.74; 1.61
Bounds for EC10 from 95% CI	0.22; 0.66	10; 20	1.82; 2.9	1.64; 4.36
95% CI on EC10 log-logistic model	0.24; 0.31	13.6; 16.1	1.28; 2.48	2.34; 3.40

EC = effective concentration.

difference in response should be observed minimally for a response to be found significantly different from that observed in the control. Unfortunately, this measure adds one more statistical concept, which may not be so simple to understand and to apply for regulators. Hence, the use of simultaneous confidence intervals proposed in the present study appears easier, because it is directly linked to well-known EC_x values.

CONCLUSIONS

The method we propose, based on the use of simultaneous confidence intervals on ratio-to-control, provides an easy and proper way to interpret the results of a Dunnett procedure and prevents the common misuse of *p* values associated with LOEC or NOEC calculations. It provides the result as an uncertainty interval for whatever EC_x values, even EC₀ (zero effect concentration). In fact EC₀ may only be upper-bounded by the LOEC value, which gives little useful information when the objective is to define an acceptable effect level [20]. Thus, to properly infer useful information from the commonly used Dunnett procedure, one must previously specify what is called a negligible effect, and to work on EC_x values (with a non-null value of *x*) as classically done with dose–response curves. Another interesting way to interpret such data, without defining an *x*% effect, is the estimation of the no effect concentration as a parameter of a model describing the occurrence of an effect above a threshold concentration [15,21,22].

Acknowledgement—We thank developers of the two *R* packages used in the present work, G. Dilba Djira and his collaborators for the *mratios* package and C. Ritz and J. Streibig for the *drc* package. We also thank R. Manard and his collaborators and B. Clément and H. Delhaye for kindly providing us with their data.

REFERENCES

1. Organisation for Economic Co-operation and Development. 2006. Series on Testing and Assessment Number 54. Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. OECD Environment Health and Safety Publications, Paris, France.
2. van der Hoeven N. 1997. How to measure no effect? Part III: Statistical aspects of NOEC, EC_x and NEC estimates. *Environmetrics* 8:255–261.
3. Fox DR. 2008. NECs, NOECs and the EC_x. *Aust J Ecotoxicol* 14:7–9.
4. Newman MC. 2008. “What exactly are you inferring?” A closer look at hypothesis testing. *Environ Toxicol Chem* 27:1013–1019.
5. Warne MSJ, van Dam R. 2008. NOEC and LOEC data should no longer be generated or used. *Aust J Ecotoxicol* 14:1–5.
6. Altman DG, Bland JM. 1995. Absence of evidence is not evidence of absence. *Br Med J* 311:495.
7. Armstrong JS. 2007. Significance tests harm progress in forecasting. *Int J Forecasting* 23:321–327.
8. Gardner MJ, Altman DG. 1986. Confidence intervals rather than P values: Estimation rather than hypothesis testing. *Br Med J* 292:746–750.
9. U.S. Food and Drug Administration. 2001. Guidance for industry: Statistical Approaches to Establishing Bioequivalence and hazardous concentration in ecotoxicology. Washington, DC.
10. Dilba G, Bretz F, Guiard V, Hothorn LA. 2004. Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods Inf Med* 43:465–469.
11. Hothorn T, Bretz F, Westfall P., 2008. Simultaneous inference in general parametric models. *Biometrical J* 50:346–363.
12. R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
13. Brain P, Cousens R. 1989. An equation to describe dose responses where there is stimulation of growth at low dose. *Weed Res* 29:93–96.
14. Ritz C, Streibig JC., 2005. Bioassay Analysis using R. *J Stat Softw* 12:1–22.
15. Billoir E, Delignette-Muller ML, Péry AR, Charles S. 2008. A Bayesian approach to analyzing ecotoxicological data. *Environ Sci Technol* 42:8978–8984.
16. Manar R, Bessy H, Vasseur P. 2009. Reproductive effects and bioaccumulation of chlordane in *Daphnia magna*. *Environ Toxicol Chem* 28:2150–2159.
17. Isnard P, Flammarion P, Roman G, Babut M, Bastien P, Bintein S, Esserméant L, Féraud JF, Gallotti-Schmitt S, Saouter E, Saroli M, Thiébaud H, Tomassone R, Vindimian E. 2001. Statistical analysis of regulatory ecotoxicity tests. *Chemosphere* 45:659–669.
18. Moore DRJ, Caux PY. 1997. Estimating low toxic effects. *Environ Toxicol Chem* 16:794–801.
19. van der Hoeven N. 2008. Calculation of the minimum significant difference at the NOEC using a non-parametric test. *Ecotoxicol Environ Saf* 70:61–66.
20. van der Hoeven N. 2004. Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheor* 52:201–217.
21. Kooijman S, Bedaux J., 1996. *The Analysis of Aquatic Toxicity Data*. VU University Press, Amsterdam, The Netherlands.
22. Fox DR. 2010. A Bayesian approach for determining the no effect concentration *Ecotoxicol Environ Saf* 73:123–131.

APPENDIX

Beginning users of *R* can find information in one of the multiple free manuals provided on the Comprehensive *R* Archive Network in the sections “Manuals” or “Contributed” (www.r-project.org).

First one must load the *mratios* package using the following command in *R*: `require(mratios)`

You may then use the function `sci.ratio` to calculate confidence intervals, by specifying in a formula (eg, `response ~ treatment`) the numerical response (`response`) and the grouping factor (`treatment`) coding for various concentrations and control, in a data frame the name of which must also be given as the second argument of the function.

$$s < -sci.ratio(\text{formula} = \text{response} \sim \text{treatment}, \\ \text{data} = \text{nameofthedataframe})$$

By default, two-sided 95% confidence intervals are calculated for the Dunnett contrast type, using the first level of the grouping variable as the control, but various other arguments of the function may be modified, such as the confidence level or the level of the grouping factor corresponding to the control. A complete list of `sci.ratio` arguments is provided in its help file accessible using the following *R* command:

```
?sci.ratio.
```

After being calculated, confidence intervals may be printed and plotted by simply applying the `print` and `plot` functions to the object returned by the `sci.ratio` function. In the following command using the `plot` function, the argument `CIvert` is fixed at `TRUE` so as to plot confidence intervals vertically; a vector of values defining the ratios of interest is affected to the argument `rho0`, so as to plot corresponding horizontal lines, as in Figure 2.

```
print(s)
```

```
plot(s, rho0 = c(1, 0.95, 0.90), CIvert = TRUE)
```



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

Comparison of bioassays with different exposure time patterns: The added value of dynamic modelling in predictive ecotoxicology

Elise Billoir^{a,b,*}, H el ene Delhaye^c, Carole Forfait^b, Bernard Cl ement^c, Ga elle Triffault-Bouchet^d,
Sandrine Charles^b, Marie Laure Delignette-Muller^{b,e}

^a P ole de Recherche ROVALTAIN en Toxicologie Environnementale et Ecotoxicologie, 1 avenue de la gare – BP 15173 – Alixan, F-26958, Valence Cedex 9, France

^b Universit  de Lyon, F-69000, Lyon; Universit  Lyon 1; CNRS, UMR5558, Laboratoire de Biom trie et Biologie Evolutive, F-69622, Villeurbanne, France

^c Universit  de Lyon, F-69000, Lyon; Universit  Lyon 1; ENTPE, F-69518, Vaulx-en-Velin; CNRS, UMR5023 Ecologie des Hydrosyst mes Naturels et Anthropis s, F-69622, Villeurbanne, France

^d Centre d'expertise en analyse environnementale du Qu bec, Minist re du d veloppement durable, de l'Environnement et des Parcs, 2700 rue Einstein, Bureau E-2-220, Qu bec (QC), G1P 3W8, Canada

^e Universit  de Lyon, F-69000, Lyon; VetAgro Sup Campus V t rinaire de Lyon, 1 avenue Bourgelat, F-69280 Marcy l'Etoile, France

ARTICLE INFO

Article history:

Received 12 May 2011

Received in revised form

1 August 2011

Accepted 6 August 2011

Available online 1 September 2011

Keywords:

Time-varying concentration

Cadmium

Daphnia magna

Bayesian analysis

No effect concentration

TKTD models (toxicokinetic–

toxicodynamic)

ABSTRACT

The purpose of this study was to compare *Daphnia magna* responses to cadmium between two toxicity experiments performed in static and flow-through conditions. As a consequence of how water was renewed, the two experiments were characterised by two different exposure time patterns for daphnids, time-varying and constant, respectively. Basing on survival, growth and reproduction, we addressed the questions of organism development and sensitivity to cadmium. Classical analysis methods are not designed to deal with the time dimension and therefore not suitable to compare effects of different exposure time patterns. We used instead a dynamic modelling framework taking all timepoints and the time course of exposure into account, making comparable the results obtained from our two experiments. This modelling framework enabled us to detect an improvement of organism development in flow-through conditions compared to static ones and infer similar sensitivity to cadmium for both exposure time patterns.

  2011 Elsevier Inc. All rights reserved.

1. Introduction

Experimental model ecosystems are useful tools for understanding ecological processes and evaluation of toxicant impacts on ecosystems (Chapman, 2002; Benton et al., 2007). These methods appear to be more relevant for ecotoxicological risk assessment than water only short-term single-species standard bioassays because they take into account ecological interactions between abiotic compartments, between species and between biotic and abiotic components (Taub, 1997; Cl ement and Cadier, 1998; Fleeger et al., 2003; Barry and Davies, 2004). Consequently, they allow to combine the study of exposure and toxicity. Outdoor mesocosms are adequate to mimic ecosystems but the experimental set-up, maintenance and interpretation of results are complex. Microcosms are laboratory small-scale systems, composed of different abiotic compartments (freshwater, sediment), with several species. Microcosms are an interesting compromise between simplified single-species tests and complex

outdoor large mesocosms for evaluating the chronic impacts of potentially contaminated matrices on lentic ecosystems, because they allow reproducibility and standardization (Cl ement and Cadier, 1998; Taub et al., 1986).

Most microcosm experiments are conducted in static conditions, mainly in order to simplify experimental procedures (Cl ement and Cadier, 1998; Taub et al., 1986; Triffault-Bouchet et al., 2005). Nevertheless, some researchers choose to renew water periodically (Roman et al., 2007), continuously (Brockway et al., 1979; Kottelat et al., 2010) or to introduce a recirculated flow (Kersting, 1997). Water renewal is expected to (i) remove ions and substances released from sediment and toxic minerals issued from organic matter decomposition such as NH₃ (Nebeker et al., 1986), (ii) introduce nutrients, major and oligo-elements and so stabilize their concentration and prevent depletion of chemical pool (Lauth et al., 1996; Kottelat et al., 2010) and (iii) stabilize colloidal concentrations (Kottelat et al., 2010) and algal density in the water column. Consequently we formulated the assumption that water renewal may improve the development of primary producers because of nutrient inputs, and of primary consumers due to primary producer abundance and a lower exposition to ammonia.

* Corresponding author at: Universit  Lyon 1, CNRS, UMR5558, Laboratoire de Biom trie et Biologie Evolutive, F-69622 Villeurbanne, France.
E-mail address: elise.billoir@univ-lyon1.fr (E. Billoir).

In a previous study (Billoir et al., 2011), we have proposed a modelling framework to understand *Daphnia magna* responses to cadmium (i.e. decrease in survival, growth and reproduction) observed in laboratory aquatic microcosms conducted in static conditions. Experiments in static conditions have resulted in time-decreasing cadmium concentration in water. Accordingly, models have been designed to account for time-varying conditions of exposure through a dynamic formalism. The models describe the time course of processes leading to toxic effects (exposure, bioaccumulation and its impacts), they are of the family of toxicokinetics/toxicodynamics models (Jager et al., 2011).

Since then, another similar experiment was performed this time in flow-through conditions, so that the cadmium concentration in water remained close to constant. Hence we aimed at comparing the new experiment to the previous one, given that their experimental design were similar except the procedure for water renewal. In the laboratory aquatic microcosms, daphnids interacted more or less with other organisms (algae, duckweeds, amphipods and chironomids). The modelling framework developed for the analysis of results did not explicitly deal with the biotic interactions that occurred in laboratory aquatic microcosms. Nevertheless, because toxicity experiments were actually conducted in microcosms, the analysis of experimental results took the biotic and abiotic interactions into account.

Based on *D. magna* responses, in the present paper we focused our purpose on the overall comparison between results obtained in flow-through conditions with those obtained in static conditions. In particular, we addressed the following questions: did the flow-through conditions improve the organism development (survival, growth and reproduction)? Did they influence the sensitivity of organisms to cadmium? Classical methods, that are the determination of No observed effect concentrations (NOECs) or lethal/effective concentrations (LC/ECs), are not designed to deal with exposure dynamics and therefore not suitable for the purpose of comparison of daphnid sensitivity to cadmium between the two experiments. In contrast, the dynamic modelling framework previously developed (Billoir et al., 2011) is adaptable to any exposure time pattern. It was applied for analysing separately the two experiments and enabled their comparison. The dynamic modelling approach was finally discussed with respect to its ability to provide answers to our biological questions.

2. Materials and methods

In this study, we will focus on *D. magna*, therefore we limit the description of the experimental monitoring to this species. Exposure is defined as the total cadmium concentration in water (including dissolved organic carbon concentration). We deal with two different exposure time patterns: a time-varying (TV) one and a constant (CST) one.

2.1. Experimental design

Data were collected at ENTPE, Vaulx-en-Velin, France. The TV experiment was run in June 2009 and the CST experiment in October 2009. Culture conditions of daphnids were strictly controlled during all the experimentations. Acute monospecific tests were also done with cadmium to ensure that the sensitivity of *D. magna* is stable over time. The two laboratory aquatic microcosm experiments aimed at testing chronic effects of cadmium on a small-scale model ecosystem composed of five freshwater pelagic and benthic species in freshwater/sediment systems. The main difference between the two experiments was water renewal and the resulting cadmium exposure time pattern. Cadmium was introduced through cadmium-spiked artificial freshwater. All cadmium was introduced initially in static microcosms (TV experiment) whereas it was introduced continuously in flow-through microcosms (CST experiment).

Tests were conducted in plastic beakers with 100 g of dry microcosm artificial sediment and 2 L of artificial freshwater. Sand (88 g), dried and sieved (1 mm) sediment from a pristine lake (Aiguebelette, Savoie, France) (10 g), α -cellulose

(1.5 g) and TetraMin® (0.15 g), as food for the benthic organisms, were mixed to prepare the artificial sediment. The artificial freshwater (pH=7.7, conductivity=290 $\mu\text{S cm}^{-1}$, hardness=60) was prepared with demineralized water and the macroelements solutions of OECD medium (OECD, 1993) (in modified concentrations) and oligo-elements and vitamins of Elendt M4 medium (Elendt and Bias, 1990). Beakers were filled seven days before the introduction of organisms (on day 7) with spiked (TV) or uncontaminated (CST) microcosm artificial freshwater. In static systems (TV), cadmium was introduced at nominal concentrations 0 (control treatment), 10, 20, 40 and 80 $\mu\text{g L}^{-1}$ (treatments T1–T4). In flow-through systems (CST), spiked water was introduced continuously starting 6 days before the introduction of organisms (on day 6) until the end of the experiment, with nominal concentrations cadmium concentrations of 0 (control treatment), 1.25, 2.5, 5 and 10 $\mu\text{g L}^{-1}$ (treatments T1–T4). A flow of 1 L per day was chosen so that abiotic conditions and algal concentration in the system were stabilized.

Four replicates (i.e. four microcosms) were set up for every treatment plus control. However, in the TV experiment, because of a mistake when making treatments, one microcosm of treatment T2 was not contaminated, and was therefore considered as a control. The microcosms were conditioned 7 days in the dark before introduction of tested species so as to reach a microbiological stabilization in the sediment while preventing development of micro-algae (Verrhiest et al., 2002). During this phase and the rest of the experiment, oxygen concentration was kept above 80 percent of saturation (TV systems were gently aerated). Experiments were conducted under constant conditions of temperature ($20 \pm 3^\circ\text{C}$) and illumination (16 hours per day). On day 0, 4×10^7 cells of *Pseudokirchneriella subcapitata*, 6 fronds of *Lemma minor*, 10 *D. magna* (24 h-old), 10 *Hyalella azteca* (7–14-days-old) and 10 *Chironomus riparius* (48 h-old) were introduced into each microcosm system. The 4×10^7 cells of *P. subcapitata* were added on day 2 in the flow-through experiment (CST). These quantities are based on the works of Clément and Cadier (1998) and Triffault-Bouchet et al. (2005), with slight modifications introduced by Hélène Delhaye in her Ph.D. research work on microcosm assay optimization. Organisms were cultured in the laboratory following internal protocols.

Systems were monitored during 21 days. Several abiotic and biotic factors were measured. For example, conductivity, pH and dissolved oxygen were measured twice a week and particle concentration in the water was monitored over the course of the experiment using a Coulter Counter Z1 (Beckman Coulter), giving an estimation of algal cell density. However, we focus the description of the experimental design on the material that is actually analysed in this study (exposure, survival, growth and reproduction of daphnids).

Water was sampled weekly in the TV experiment and twice a week in the CST experiment. Unfiltered water samples (10 mL) were taken in the middle of the water column of all microcosms using a 10 mL pipette, acidified (nitric acid), stored at 4°C for analysis of cadmium concentration by graphite furnace atomic absorption spectrometry (1500°C) (Hitachi, model Z-8 200, Tokyo, Japan). Daphnid survival, growth and reproduction were followed-up during the course of the experiments. We repeatedly observed how many first generation daphnids stayed alive on days 2, 3, 7, 10, 14, 17 and 20 in the TV experiment and on days 2, 4, 7, 8, 10, 11, 14, 17, 20 and 21 in the CST experiment. Body length of daphnids was measured using a binocular on days 7, 14 and 20 in the TV experiment and on days 7, 16 and 21 in the CST experiment. Neonates were counted and removed (when reproduction occurred) on days 2, 3, 7, 10, 14, 17 and 20 in the TV experiment and on days 2, 4, 6, 8, 10, 11, 14, 17, 20 and 21 in the CST experiment. Other organisms were not discarded before the end of the experiment, except chironomus adults after emergence. The main intervention of the operator on the systems were removal of neonates and water sampling, without any other major perturbation.

2.2. Dynamic modelling approach

2.2.1. Overview

We used the modelling framework proposed in Billoir et al. (2011) to analyse both TV and CST experiments. The TV case has already been analysed by Billoir et al. (2011). In the present study we were interested in the comparison of the TV and CST experiments that were analysed separately.

The modelling framework involved five connected modules: exposure, bioaccumulation, survival, growth and reproduction, and described quantities as functions of both exposure concentration and time. The exposure module was adapted according to the experimental conditions. Hence, the exposure models and their parameters were not the same in the TV and CST cases. In contrast, for the other modules the same models were used in both cases, taking exposure into account day after day and making possible the comparison of the TV and CST experiments.

In the microcosms, a prey–predator interaction took place between algae and daphnids, since no additional food for daphnids was introduced in the system. However, given the good development of algal population in both experiments, we assumed that daphnids were in good conditions regarding food level, as confirmed by results on control daphnids (high survival, normal adult size, high reproduction rate during the assays).

Exposure was defined as the total cadmium concentration in water, in both dissolved and suspended fractions (including algae). For daphnids, both waterborne and dietary exposure routes were therefore taken into account, as recommended in the literature (Barata et al., 2002; Goulet et al., 2007; Geffard et al., 2008).

2.2.2. Exposure

The TV and CST exposure time patterns are shown in Fig. 1. For the TV data set, following Billoir et al. (2011) we used an exponential decay model:

$$C_j(t) = C_j(-7) \exp(-b(t+7)) \tag{1}$$

where $C_j(t)$ is the exposure concentration ($\mu\text{g L}^{-1}$) at time t (day) for treatment j , and b the decrease rate (per day) of cadmium concentration in the microcosm water. In this model of TV exposure, according to the treatment, $C_j(-7)$ was fixed at 0, 10, 20, 40 or $80 \mu\text{g L}^{-1}$, that is the nominal concentration initially introduced (7 days before the introduction of organisms) for each treatment respectively, and b was fixed at 0.167 d^{-1} , as estimated in Billoir et al. (2011).

For the CST data set, we used a constant model:

$$C_j(t) = C_j \tag{2}$$

In this model of CST exposure, according to the treatment, C_j was fixed at 0, 0.37, 0.86, 1.64 or $4.36 \mu\text{g L}^{-1}$, that is the mean of measured concentrations in water over the exposure duration for each treatment respectively.

2.2.3. Impacts

Regarding bioaccumulation, survival, growth and reproduction, we used the same modelling framework to analyse the TV and CST data sets, the constant exposure being considered as a particular case of time-varying exposure. A full description of the model is given in Billoir et al. (2011). Here we just revisit main principles and assumptions.

The potentially active scaled internal cadmium concentration was considered to follow the external one through a one-compartment model with an elimination rate constant k_e (per day) (also *dominant rate constant*, Jager et al., 2011). Basing on a dynamic approach with a time step of one day, we then assumed that cadmium effects between times $t-1$ and t depended on the potentially active scaled internal

concentration at $t-1$. More precisely, effects were assumed to occur when the potentially active scaled internal concentration exceeded a threshold concentration called NEC_* (no effect concentration, $\mu\text{g L}^{-1}$), and to be proportional (coefficient k_* , $\mu\text{g}^{-1} \text{ L}$) to the excess above the NEC_* . Two so-called stress functions (Kooijman and Bedaux, 1996), $\sigma_* = \max(0, k_*(C_{int} - NEC_*))$, where C_{int} is the scaled internal concentration, were distinguished: one accounting for lethal effects—in this case $* = S$ for survival, and one accounting for sublethal effects—in this case $* = GR$ for growth and reproduction. Those stress functions were integrated into the equations describing the time course of survival, growth and reproduction of control organisms. As the lethal stress function σ_S (resp. the sublethal stress function σ_{GR}) increases, the survival rate (resp. growth and reproduction rate) decreases. All parameters involved in the modelling framework are summarized in Table 1. Thereafter, we name *non-toxicological* the parameters of control equations: blank mortality rate m (per day), body growth rate γ (per day), maximum body length L_m (mm), scaled length at puberty l_p (dimensionless) and maximum reproduction rate R_m (number of offspring per daphnid per day) and *toxicological* the parameters related to cadmium toxicity: k_e , NEC_S , k_S , NEC_{GR} and k_{GR} (interpretation and dimensions above).

2.2.4. Parameter estimation

The TV and CST experiments were analysed independently. For each data set (TV and CST), we used simultaneously the three types of data (survival, growth and reproduction), including all timepoints from day 0 to day 20 or 21 to estimate all parameters (Table 1). Data were related to model outputs through error models adapted to each endpoint: conditional binomial for survival data (number of survivors at a timepoint conditional on the number at the previous timepoint), normal for growth data (body lengths) and negative binomial for reproduction data (cumulated number of offspring per mother).

Parameters were estimated by the mean of Bayesian inference. Based on the Bayes theorem, the principle of this inference technique is to update chosen prior probability distributions using data to obtain posterior probability distributions. We refer interested readers for instance to Ntzoufras (2009) or Fox (2010). Except in very simple cases, carrying out Bayesian inference requires the use of Markov Chains Monte Carlo (MCMC) algorithms like the Gibbs sampling and other Metropolis–Hastings algorithms. In contrast to frequentist methods (e.g.

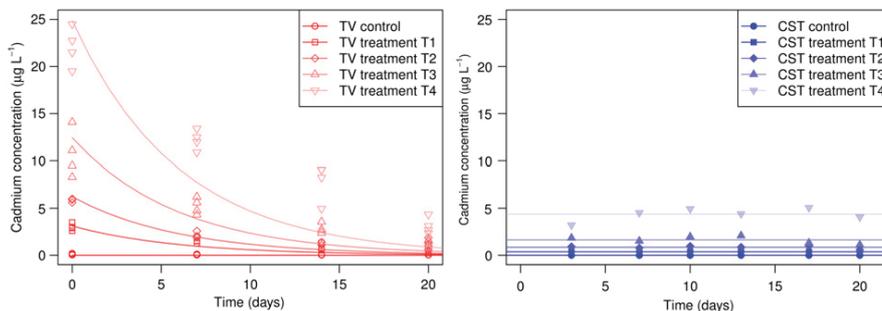


Fig. 1. Exposure time patterns in the TV (left) and CST (right) experiments. Symbols correspond to data while solid lines correspond to fitted models.

Table 1

Parameters involved in the dynamic modelling framework. The horizontal double line delineates the non-toxicological and toxicological parameters. Empirical posterior distribution statistics are given as median [2.5th–97.5th] percentiles. (–) means dimensionless and # means number of offspring per daphnid. $Norm(me, sd^2)$ denotes a normal distribution with a mean me , and a variance sd^2 , $T(a, b)$ denotes interval truncation between bounds a and b (see Plummer, 2010 for details). $\log Unif(inf, sup)$ denotes a log-uniform distribution, meaning that the natural logarithm of the random variable is uniformly distributed between inf and sup .

Symbol	Dimension	Interpretation	Prior distribution	Posterior distribution—TV	Posterior distribution—CST case
m	d^{-1}	Blank mortality rate	$\log Unif(-10, -4.5)$	0.0033 [0.0015–0.0062]	0.00064 [0.00011–0.0020]
L_m	mm	Maximum body length	$Norm(4.77, 0.59^2) T(1,)$	5.1 [4.8–5.5]	5.5 [5.1–5.9]
γ	d^{-1}	von Bertalanffy growth rate	$Norm(0.11, 0.03^2) T(0,)$	0.11 [0.097–0.14]	0.11 [0.086–0.14]
l_p	(–)	Scaled length at puberty	$Norm(0.49, 0.07^2) T(0.25,)$	0.54 [0.46–0.61]	0.48 [0.43–0.57]
R_m	$\#\text{d}^{-1}$	Maximum reproduction rate	$Norm(10.74, 3.62^2) T(0,)$	15 [12–19]	19 [16–23]
k_e	d^{-1}	Elimination rate	$\log Unif(-5, 1)$	0.90 [0.79–1.06]	0.81 [0.20–1.8]
NEC_S	$\mu\text{g L}^{-1}$	No effect concentration for survival	$\log Unif(-3, 3)$	1.8 [1.2–2.3]	2.2 [0.26–4.0]
k_S	$\mu\text{g}^{-1} \text{ L}$	Survival effect intensity coefficient	$\log Unif(-5, 5)$	13 [7.0–30]	19 [3.5–110]
NEC_{GR}	$\mu\text{g L}^{-1}$	No effect concentration for growth and reproduction	$\log Unif(-3, 3)$	0.15 [0.052–1.1]	0.14 [0.052–1.1]
k_{GR}	$\mu\text{g}^{-1} \text{ L}$	Growth and reproduction effect intensity coefficient	$\log Unif(-5, 5)$	0.032 [0.019–0.046]	0.038 [0.029–0.059]

maximum-likelihood algorithms), Bayesian inference algorithms, if they converge, provide samples from (i.e. approximate) the parameter posterior distributions. MCMC algorithms are implemented in softwares such as BUGS (Spiegelhalter et al., 1995) or JAGS (Plummer, 2010). We performed Bayesian analyses using the *rjags* R package (Plummer, 2011).

The TV and CST data sets were analysed starting from the same prior distributions (Table 1), corresponding to those used in Billoir et al. (2011). For each analysis, three independent MCMC chains were run in parallel. After an initial burn-in period of 5000 iterations, the Bayesian algorithm (that is the building of MCMC chains) was run 10,000 iterations in order to reach the convergence of MCMC chains and the corresponding sample of parameter posterior distributions was recorded. The convergence was checked using the Gelman and Rubin convergence diagnostics (Brooks and Gelman, 1997). In the Bayesian framework, all quantities except data (parameters, latent variables, model outputs) are random variables. Hence it is possible to express any of those quantities (not only parameters) as probability distributions. To do that, the wanted quantity is sampled throughout MCMC chains.

To evaluate the goodness-of-fit of the modelling framework, replicated data were also simulated and recorded for posterior predictive checking (Gelman, 2003). Posterior predictive checking is a graphical comparison of observed and replicated data, taking uncertainty associated with parameter estimates and error model into account. Indeed, replicated data are simulated (1) throughout MCMC chains, along which the parameters vary according to their uncertainty, and (2) with the same error model as the one assumed for observed data. Results were graphically represented using the R software (R Development Core Team, 2010).

3. Results and discussion

Bayesian analyses succeeded in simultaneously estimating all parameters both from TV and CST data sets. The posterior predictive checking (Fig. 2) showed a good accordance between observed and replicated data for the three biological endpoints considered (survival, growth and reproduction) and for both experiments, denoting a good quality of fit. For each parameter, the posterior distributions inferred from the TV and CST data sets are superimposed in Fig. 3 (corresponding empirical statistics are

given in Table 1). Their comparison involves two points: the point estimate and the uncertainty associated with the estimate. Both can be assessed from the parameter posterior distribution: the position of the distribution mode (the peak) on the *x*-axis gives an idea of the point estimate while the distribution width gives information about uncertainty.

For non-toxicological parameters, point estimates obtained from the TV and CST data sets differed (Fig. 3), while uncertainties associated with estimates were similar. We obtained a higher blank mortality rate (*m*), a lower maximum body length (*L_m*) and a lower maximum reproduction rate (*R_m*) in the TV case than in the CST one. These three results meant a better survival, growth and reproduction of daphnids in the CST experiment than in the TV one, independent of cadmium exposure. In accordance with our expectations, this may be attributable to the improvement of bioassay conditions due to water renewal in the CST experiment. Hence, to the first biological question addressed in this study, we would respond that indeed the flow-through conditions improved the organism development in terms of survival, growth and reproduction. In the CST case, jointly with a high maximum body length (*L_m*), we obtained a low scaled length at puberty (*l_p*). This directly stemmed from the model parameterization and the resulting correlations between parameters: by definition, the scaled length at puberty is the length at puberty divided by the maximum length, so the higher *L_m*, the lower *l_p*. Last, we obtained a similar growth rate (*γ*) in both cases.

The informative prior distributions of non-toxicological parameters were built by eliciting realistic values in *D. magna* (Billoir et al., 2008b). Estimates based on the posterior distributions remained within those likely values (Fig. 3). We noticed that posterior distributions of the maximum body length (*L_m*) and the maximum reproduction rate (*R_m*) were shifted to higher values compared to prior distributions. This may indicate a better

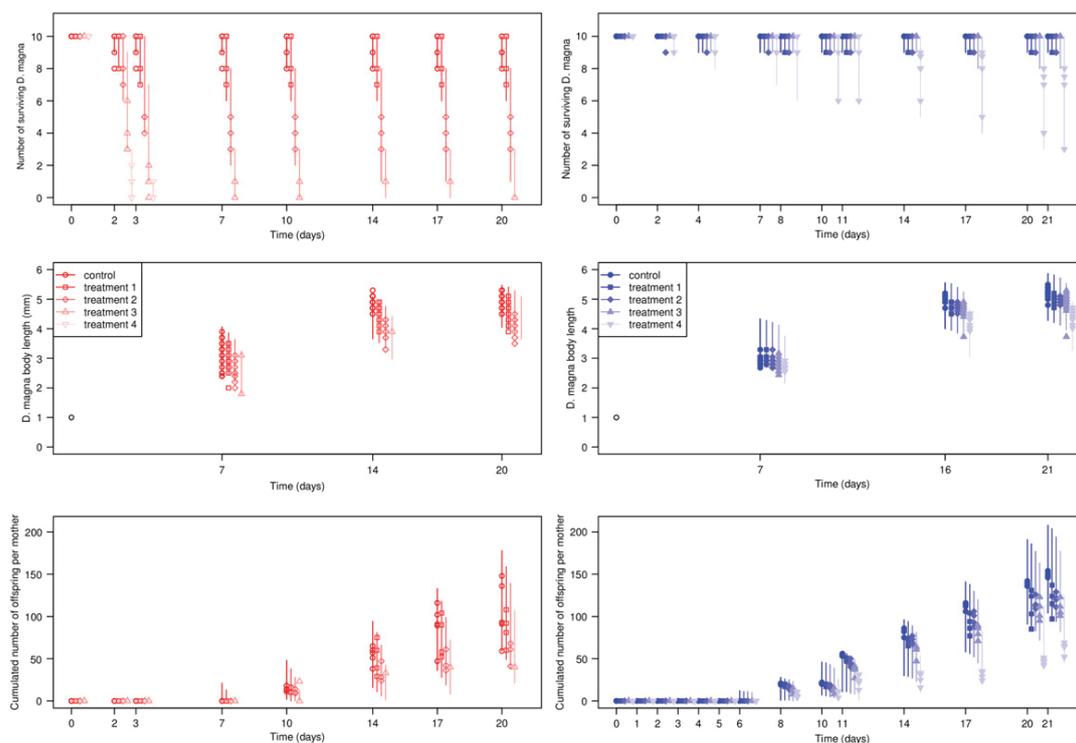


Fig. 2. Posterior predictive checking of survival (top), growth (middle) and reproduction (bottom) in *D. magna* in the TV (left, red) and CST (right, blue) experiments. Observed data (symbols) and corresponding 95 percent extreme results (calculated as [2.5th–97.5th] percentiles) of replicated data posterior distributions (segments) are superimposed for a visual assessment of the goodness of fit. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

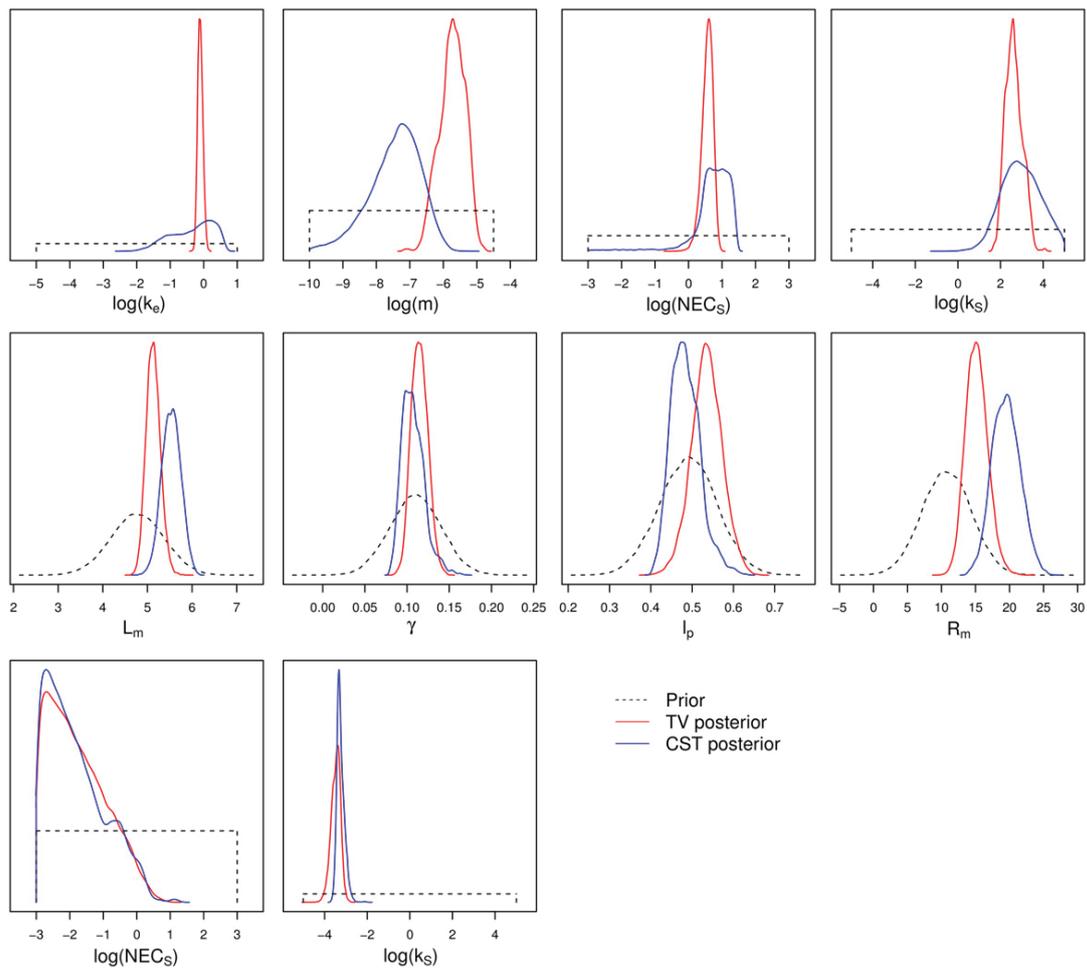


Fig. 3. For each parameter involved in the dynamic modelling framework, prior distributions (dashed lines), posterior distributions obtained from the TV experiment (red lines), and posterior distributions obtained from the CST experiment (blue lines). When the prior distribution applied to the natural log of parameters (for k_e , m , NEC_S , k_S , NEC_{GR} and k_{GR}), posterior distribution of the natural log of parameters was plotted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

development (in terms of growth and reproduction) of daphnids in our experiments than in those on which prior distributions were based (Billoir et al., 2008b), and was in accordance with our assumption of good conditions regarding food level.

In contrast to non-toxicological parameters, for some toxicological parameters, point estimates obtained from the TV and CST data sets differed less than the uncertainties associated with the estimates (Fig. 3). The elimination rate (k_e) was much more precisely estimated in the TV experiment than in the CST one, as well as lethal stress function parameters, that are the no effect concentration for survival (NEC_S) and the lethal effect intensity coefficient (k_S). Nevertheless, point estimates were in accordance. From both experiments, very similar posterior distributions were inferred for sublethal stress function parameters, that are the no effect concentration for growth and reproduction (NEC_{GR}) and the sublethal effect intensity coefficient (k_{GR}). In our modelling approach, so-called stress functions accounted for the impacts of cadmium in comparison to control conditions (toxic effects). Hence, results indicated that water renewal did not seem to influence the sensitivity of daphnids to cadmium.

As another way to address the second biological question of our study, that is what can be learned from the TV and CST exposure tests regarding daphnid sensitivity to cadmium, we compared stress functions inferred from the two experiments.

Fig. 4 shows the lethal (left) and sublethal stress functions (right) obtained in the TV and CST cases. Regarding lethal stress functions (Fig. 4, left), the TV and CST exposure tests led to close median (solid lines). The width of the 95 percent extreme results ([2.5th–97.5th] percentiles, dashed lines) for the CST experiment stemmed from the imprecision of NEC_S and k_S estimates in this case (Fig. 3). Regarding sublethal stress functions (Fig. 4, right), the TV and CST experiments led to very close results, in line with the similarity of NEC_{GR} and k_{GR} posterior distributions.

Globally, the uncertainty associated with the estimates was less in the TV case than in the CST one (Fig. 3). This result might be viewed as counterintuitive, since the CST case is simpler than the TV one where the exposure dynamics might be expected to be a confusing factor for data analysis. Although no general conclusion could be drawn from our particular study, our results show that TV exposure design sometimes turns out to be more informative than CST conditions to capture organism responses (more precise estimates of model parameters). This can be explained by the experimental design which highlighted limited effects on daphnid survival in the CST case (Fig. 2, top right). Besides, in the TV experiment, effects of cadmium on survival were actually especially reflective of the exposure dynamics, that enabled a precise estimation of the elimination rate (k_e) (Fig. 3) and subsequently improved the estimation of the other

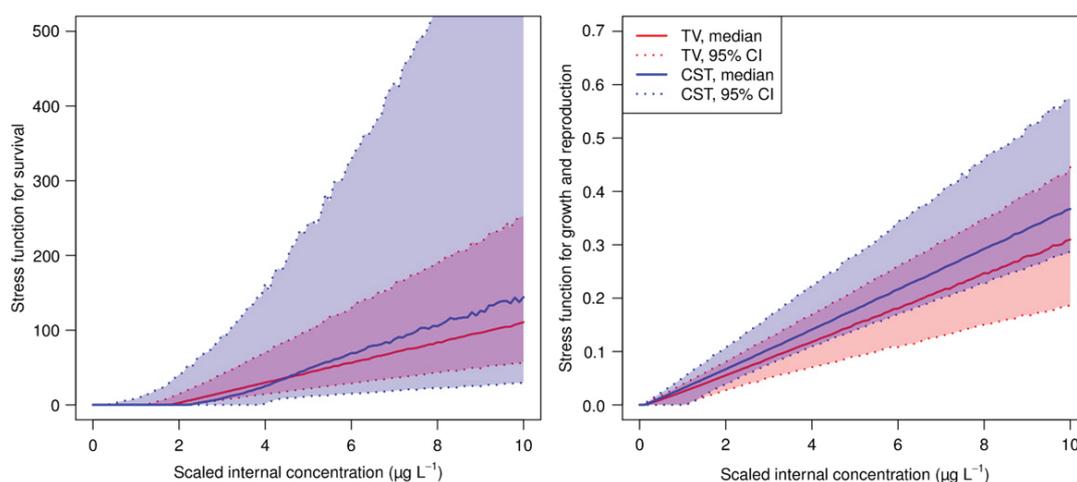


Fig. 4. Stress functions for survival (left) and for growth and reproduction (right) estimated from the TV experiment (red lines), and from the CST experiment (blue lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

toxicological parameters (Billoir et al., 2011). In contrast to the TV case, we obtained a very uncertain estimate of the elimination rate (k_e) in the CST case (Fig. 3), meeting previous studies (Jager et al., 2004; Billoir et al., 2008a) that highlighted the difficulty to estimate the elimination rate of the bioaccumulation kinetic model when no dedicated data were available.

The median estimates we obtained for k_e (0.90 d^{-1} in the TV case, 0.81 d^{-1} in the CST case, Table 1) are higher than the one estimated by Heugens et al. (2003) with the same bioaccumulation model (0.016 h^{-1} reported in their manuscript, corresponding to 0.38 d^{-1}) for cadmium in *D. magna*. The early equilibrium between scaled internal concentration and external concentration that we predicted (high k_e) is possibly due to the dietary exposure which did not take place in the study by Heugens et al. (2003). Other studies (Poldoski, 1979; Penttinen et al., 1995) have reported equilibrium between cadmium external and internal concentration in *D. magna* after about two days of exposure, which is consistent with our results. Besides, the NEC values we estimated (Table 1) are consistent with the literature (Biesinger and Christensen, 1972; van Leeuwen et al., 1985; Elnabarawy et al., 1986; Knowles and McKee, 1987; Borgmann et al., 1989; Kühn et al., 1989) where cadmium chronic effects have been observed with toxicity estimates in the range $[0-7.78 \mu\text{g L}^{-1}]$.

Due to different water renewal procedure, TV and CST experiments had subsequent differences in terms of biotic and abiotic factors. This means that many confounding factors might explain differences in responses that are bioaccumulation kinetics and toxic effects on survival, growth and reproduction. Nevertheless, bioaccumulation kinetics and toxic effects (stress functions) independently inferred from the TV and CST cases were in accordance, although control parameters differed. This result therefore reveals the predictive potential of our modelling approach to describe cadmium impacts, whatever the exposure time pattern (at least between the two exposure patterns tested in this work). Given control conditions, our models might be able to predict the impacts of any scenario of cadmium exposure in *D. magna*. This predictive potential remains to be tested experimentally. For this purpose, flow-through is a useful tool to simulate various exposure patterns (Kottelat et al., 2010).

In the present study, we focused on *D. magna*, mainly because it was the species for which we had most data. The way it was analysed we are aware of the fact that the system was more equal to a *D. magna* reproduction test than to a microcosm. Nevertheless, the dynamic modelling approach is extendable to the

whole microcosm and taking into account of ecological interactions, provided data sets under consideration are informative enough to feed such models.

4. Conclusions

The results obtained in this study confirm the potential of modelling to make sense of ecotoxicological test results. Dynamic modelling makes it possible to deal with various exposure time patterns. Hence it enables the analysis of responses observed in different exposure conditions and the comparison of results obtained from our two experiments. The answers to our biological questions might be roughly summarized as follows: indeed flow-through conditions improved *D. magna* development, and no, it did not make the organisms more or less sensitive to cadmium. Finally, the similarity of cadmium stress functions inferred from time-varying and constant exposure conditions underlines the potential of our approach in predictive ecotoxicology.

Acknowledgments

This work was financially supported by the French National Center for Scientific Research (CNRS) and the French Ministry of Ecology, Sustainable Development, Transport and Housing (MEDDTL). The authors thank Martyn Plummer for the development of JAGS.

References

- Barata, C., Markich, S., Baird, D., Soares, A., 2002. The relative importance of water and food as cadmium sources to *Daphnia magna* Straus. *Aquat. Toxicol.* 61, 143–154.
- Barry, M., Davies, W., 2004. Effects of invertebrate predators and a pesticide on temporary pond microcosms used for aquatic toxicity testing. *Environ. Pollut.* 131, 25–34.
- Benton, T., Solan, M., Travis, J., Sait, S., 2007. Microcosm experiments can inform global ecological problems. *Trends Ecol. Evol.* 22, 516–521.
- Biesinger, K., Christensen, G., 1972. Effects of various metals on survival, growth, reproduction, and metabolism of *Daphnia magna*. *J. Fish. Res. Board Can.* 29, 1691–1700.
- Billoir, E., Delignette-Muller, M., Péry, A., Geffard, O., Charles, S., 2008a. Statistical cautions when estimating DEBtox parameters. *J. Theor. Biol.* 254, 55–64.
- Billoir, E., Delignette-Muller, M., Péry, A., Charles, S., 2008b. A Bayesian approach to analyzing ecotoxicological data. *Environ. Sci. Technol.* 42, 8978–8984.

- Billoir, E., Delhaye, H., Clément, B., Delignette-Muller, M., Charles, S., 2011. Bayesian modelling of daphnid responses to time-varying cadmium exposure in laboratory aquatic microcosms. *Ecotoxicol. Environ. Saf.* 74, 693–702.
- Borgmann, U., Ralph, K., Norwood, W., 1989. Toxicity test procedures for *Hyallella azteca*, and chronic toxicity of cadmium and pentachlorophenol to *H. azteca*, *Gammarus fasciatus*, and *Daphnia magna*. *Arch. Environ. Contam. Toxicol.* 18, 756–764.
- Brockway, D., Hill, J., Maudsley, J., Lassiter, R., 1979. Development, replicability and modelling of naturally derived microcosms. *Int. J. Environ. Stud.* 13, 149–158.
- Brooks, S., Gelman, A., 1997. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7, 434–455.
- Chapman, P., 2002. Integrating toxicology and ecology: putting the “eco” into ecotoxicology. *Mar. Pollut. Bull.* 44, 7–15.
- Clément, B., Cadier, C., 1998. Development of a new laboratory freshwater/sediment microcosm test. *Ecotoxicology* 7, 279–290.
- Elenndt, B.P., Bias, W.R., 1990. Trace nutrient deficiency in *Daphnia magna* cultured in standard medium for toxicity testing. Effects of the optimization of culture conditions on life history parameters of *Daphnia magna*. *Water Res.* 24, 1157–1167.
- Elnabarawy, M., Welter, A., Robideau, R., 1986. Relative sensitivity of three *Daphnia* species to selected organic and inorganic chemicals. *Environ. Toxicol. Chem.* 5, 393–398.
- Fleeger, J., Carman, K., Nisbet, R., 2003. Indirect effects of contaminants in aquatic ecosystems. *Sci. Total Environ.* 317, 207–233.
- Fox, D., 2010. A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicol. Environ. Saf.* 73, 123–131.
- Geffard, O., Geffard, A., Chaumot, A., Volla, B., Alvarez, C., Tusseau-Vuillemin, M.-H., Garric, J., 2008. Effects of chronic dietary and waterborne cadmium exposures on the contamination level and reproduction of *Daphnia magna*. *Environ. Toxicol. Chem.* 27, 1128–1134.
- Gelman, A., 2003. A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *Int. Stat. Rev.* 71, 369–382.
- Goulet, R., Krack, S., Doyle, P., Hare, L., Vigneault, B., McGeer, J., 2007. Dynamic multipathway model of Cd bioaccumulation in *Daphnia magna* using waterborne and dietborne exposures. *Aquat. Toxicol.* 81, 117–125.
- Heugens, E., Jager, T., Creyghton, R., Kraak, M., Hendriks, A., Straalen, N.V., Admiraal, W., 2003. Temperature-dependent effects of cadmium on *Daphnia magna*: accumulation versus sensitivity. *Environ. Sci. Technol.* 37, 2145–2151.
- Jager, T., Crommentuijn, T., van Gestel, C., Kooijman, S., 2004. Simultaneous modeling of multiple end points in life-cycle toxicity tests. *Environ. Sci. Technol.* 38, 2894–2900.
- Jager, T., Albert, C., Preuss, T.G., Ashauer, R., 2011. General unified threshold model for survival—a toxicokinetic–toxicodynamic framework for ecotoxicology. *Environ. Sci. Technol.* 45, 2529–2540.
- Kersting, K., 1997. On the way to eternity; the success of an aquatic laboratory microecosystem. *Aquat. Ecol.* 31, 29–35.
- Knowles, C., McKee, M., 1987. Protein and nucleic acid content in *Daphnia magna* during chronic exposure to cadmium. *Ecotoxicol. Environ. Saf.* 13, 290–300.
- Kooijman, S., Bedaux, J., 1996. *The Analysis of Aquatic Toxicity Data*. VU University Press, Amsterdam.
- Kottelat, R., Vignati, D., Garcia-Bravo, A., Dominik, J., Ferrari, B., 2010. Characterization of trace element geochemistry in continuous flow-through microcosms: a preliminary step to environmentally meaningful ecotoxicological experiments. *Chemosphere* 81, 1407–1415.
- Kühn, R., Pattard, M., Pernak, K., Winter, 1989. Results of the harmful effects of water pollutants to *Daphnia magna* in the 21 day reproduction test. *Water Res.* 23, 501–510.
- Lauth, J.R., Dyer, S.D., Belanger, S.E., Cherry, D.S., 1996. A novel flow-through method for toxicity assessments using *Ceriodaphnia dubia*. *Environ. Toxicol. Water Qual.* 11, 335–343.
- Nebeker, A., Onjukka, S., Cairns, M., Krawczyk, D., 1986. Survival of *Daphnia magna* and *Hyallella azteca* in Cd-spiked water and sediment. *Environ. Toxicol. Chem.* 5, 933–938.
- Ntzoufras, I., 2009. *Bayesian Modeling Using WinBugs*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- OECD, 1993. OECD Guideline for Testing of Chemicals Number 201. Algal Growth Inhibition Test, Draft Proposal for a Guidance Document, Freshwater Lentic Field Tests. OECD Environment Health and Safety Publications.
- Penttinen, S., Kukkonen, J., Oikari, A., 1995. The kinetics of cadmium in *Daphnia magna* as affected by humic substances and water hardness. *Ecotoxicol. Environ. Saf.* 30, 72–76.
- Plummer, M., 2010. JAGS Version 2.2.0 User Manual.
- Plummer, M., 2011. rjags: Bayesian Graphical Models Using MCMC. R Package Version 2.2.0-3. <<http://CRAN.R-project.org/package=rjags>>.
- Poldoski, J., 1979. Cadmium bioaccumulation assays. Their relationship to various ionic equilibria in Lake Superior water. *Environ. Sci. Technol.* 13, 701–706.
- R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org>>.
- Roman, Y., Schampelaere, K.D., Nguyen, L., Janssen, C., 2007. Chronic toxicity of copper to five benthic invertebrates in laboratory-formulated sediment: sensitivity comparison and preliminary risk assessment. *Sci. Total Environ.* 387, 128–140.
- Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1995. *Bugs: Bayesian Inference Using Gibbs Sampling*, Version 0.50. MRC Biostatistics Unit, Cambridge.
- Taub, F., 1997. Unique information contributed by multispecies systems: examples from the standardized aquatic microcosm. *Ecol. Appl.* 7, 1103–1110.
- Taub, F., Kindig, A., Conquest, L., 1986. Preliminary results of interlaboratory testing of a standardized aquatic microcosm. In: *Community Toxicity Testing, ASTM STP 920*. American Society for Testing and Materials, pp. 93–120.
- Triffault-Bouchet, G., Clément, B., Blake, G., 2005. Assessment of contaminated sediments with an indoor freshwater/sediment microcosm assay. *Environ. Toxicol. Chem.* 24, 2243–2253.
- van Leeuwen, C., Luttmmer, W., Griffioen, P., 1985. The use of cohorts and populations in chronic toxicity studies with *Daphnia magna*: a cadmium example. *Ecotoxicol. Environ. Saf.* 9, 26–39.
- Verrhiest, G.J., Cortes, S., Clément, B., Montuelle, B., 2002. Chemical and bacterial changes during laboratory conditioning of formulated and natural sediments. *Chemosphere* 46, 961–974.

Résumé

En écotoxicologie, les effets des substances toxiques sur les organismes vivants sont classiquement mesurés au niveau individuel, en laboratoire et selon des normes, ce qui assure la reproductibilité des bioessais et le contrôle des facteurs environnementaux. Ces tests standardisés, en toxicité aiguë ou chronique, portent généralement sur la survie, la reproduction et la croissance d'organismes modèles de laboratoire ; leur analyse statistique conduit à l'estimation de concentrations critiques d'effet. Ce sont ces concentrations qui sont utilisées en analyse quantitative du risque en écotoxicologie. Cependant, pour l'estimation d'un même type de concentration critique d'effet, différentes méthodes/modèles peuvent être utilisés qui sont plus ou moins adaptés en fonction du type de jeux de données. Le premier objectif de ce travail de thèse est donc de sélectionner les méthodes/modèles les plus adaptés afin d'améliorer l'analyse des données issus des tests de toxicité et donc l'estimation des concentrations critiques d'effet. Habituellement, les jeux de données sont construits à partir de tests standards en fonction de l'organisme étudié : la durée du test est généralement fixée et des recommandations sont faites, par exemple sur le nombre minimal d'organismes à exposer à un nombre minimal de concentrations. Il est donc légitime de penser que ces recommandations ne sont pas forcément les plus adaptées pour toutes les concentrations critiques d'effet et les méthodes/modèles les plus adaptés. C'est pourquoi, le deuxième objectif de cette thèse est d'optimiser les plans d'expérience afin d'aller soit vers une amélioration des estimations des concentrations critiques d'effet à coût constant soit pour un même niveau de qualité des estimations, d'éviter le gaspillage en temps et en organismes.

Mots clés : Concentration critique d'effet ; Modélisation ; Optimisation des plans d'expérience

Abstract

In ecotoxicology, the effects of toxic compounds on living organisms are usually measured at the individual level, in the laboratory and according to standards. This ensures the reproducibility of bioassays and the control of environmental factors. Bioassays, in acute or chronic toxicity, generally apply to survival, reproduction and growth of organisms. The statistical analysis of standardized bioassays classically leads to the estimation of critical effect concentrations used in risk assessment. Nevertheless, several methods/models are used to determine a critical effect concentration. These methods/models are more and less adapted to the data type. The first aim of this work is to select the most adapted methods/models to improve data analysis and so the critical effect concentration estimation. Usually, data sets are built from standard bioassays and so follow recommendations about exposure duration, number and range of tested concentrations and number of individuals per concentration. We can think that these recommendations are not the most adapted for each critical effect concentration and each method/model. That's why, the second aim of this work is to optimize the experimental design in order to improve the critical effect concentration estimations for a fixed cost or at least to reduce the waste of time and organisms.

Keywords : Critical effect concentration ; Modeling ; Experimental design optimization