

# Modélisation hiérarchique de la distribution de sensibilité des espèces (SSD) : étude de cas sur des diatomées exposées à divers herbicides.

G. Kon Kam King, F. Larras, S. Charles,  
M.L. Delignette-Muller

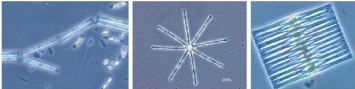


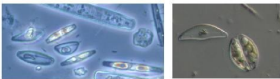
Applibugs, 19/12/2013



# Diatomées\*

\*Micro-algues unicellulaires utilisées comme bio-indicateurs pour diagnostiquer la qualité de l'eau.

Tableau I.6 : Synthèse des différentes formes de vie principales des diatomées.

FORMES DE VIE	GENRES
<p>Colonies</p> 	<p>Asterionella, Aulacoseira, Cymbella, Diatoma, Encyonema, Encyonopsis, Fragilaria, Melosira, Stauroneis</p>
<p>Tube muqueux</p> 	<p>Encyonema, Encyonopsis</p>
<p>Pédunculés</p> 	<p>Achnanthes, Achnantheskur, Diatoma, Fragilaria, Gomphonema, Fragilothidium</p>
<p>Solitaires</p> 	

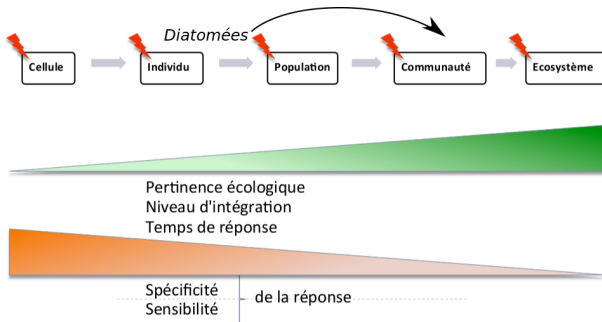
Crédits: F. Larras

Modèle développé sur des communautés de **diatomées benthiques** du lac Léman, exposées à divers herbicides.

6 polluants, 11 espèces de diatomées, mesure de la fluorescence (*endpoint*) en fonction de la concentration en herbicide.

# Objectif général et méthodologie

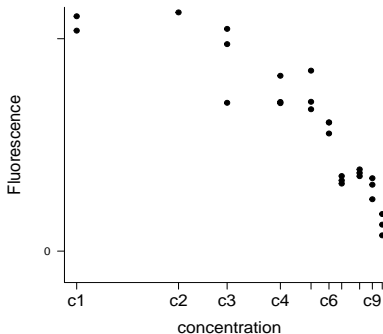
On veut prédire à partir de données de laboratoire le risque pour l'environnement lié à la présence d'un polluant.



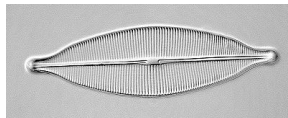
L'objectif général est de proposer des méthodes et modèles qui permettent de tirer le meilleur parti des données pour l'appréciation quantitative du risque.

# Bioessais, base de l'évaluation du risque.

On soumet un organisme à des concentrations croissantes de contaminant.

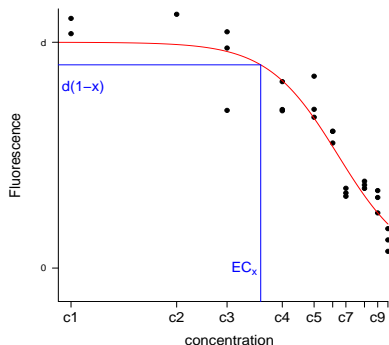


Effet sur la fluorescence de *Nitzschia Palea* d'un herbicide, le Diuron. La fluorescence est un indicateur de la biomasse.

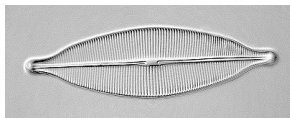


Courbe concentration-effet.

# Bioessais, base de l'évaluation du risque.



Une Critical Effect Concentration (CEC) est estimée à l'aide d'un modèle concentration-effet. La  $EC_x$  est un exemple de CEC.



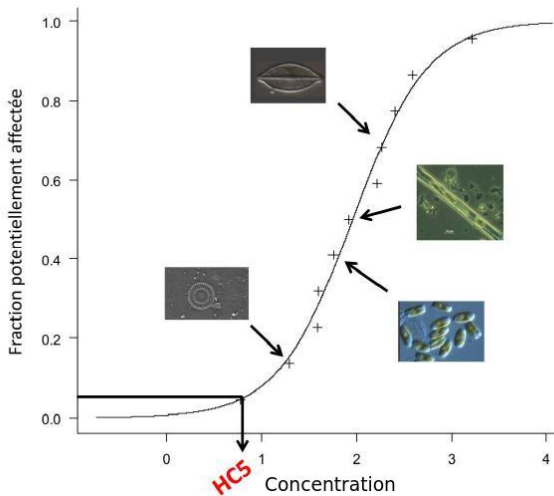
# Species Sensitivity Distribution (SSD)

La SSD est utilisée pour définir des concentrations en contaminants sans risque pour l'environnement.

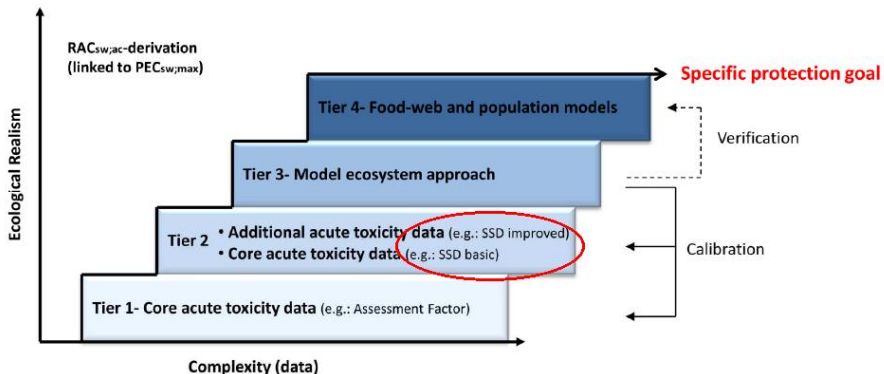
- La SSD se fonde sur des **seuils de toxicité** (CEC, Critical Effect Concentration) obtenus par des tests mono-spécifiques.
- La SSD consiste à **modéliser la distribution** des sensibilités de la communauté à partir des sensibilités d'un échantillon.
- On en extrait un **indicateur de risque** pour la communauté, comme la concentration affectant uniquement  $p\%$  de la communauté ( $HC_p$ )

L'utilisation de la SSD est recommandée par les instances gouvernementales de nombreux pays (US EPA, ANZECC, EFSA ... ) pour définir les niveaux de protection pour les milieux aquatiques.

# Une SSD

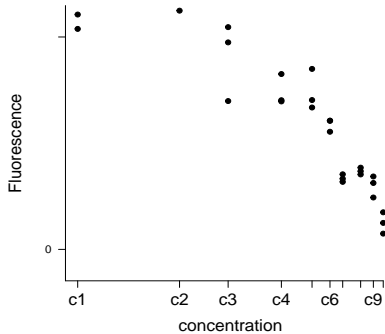


# Analyse du risque en écotoxicologie

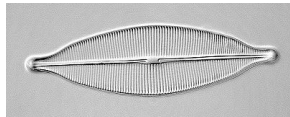




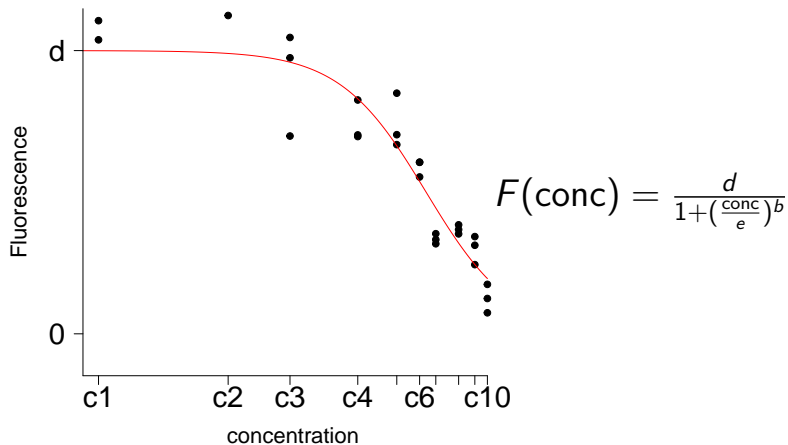
# Zoom sur la sensibilité de *Nitzschia Palea* au Diuron.



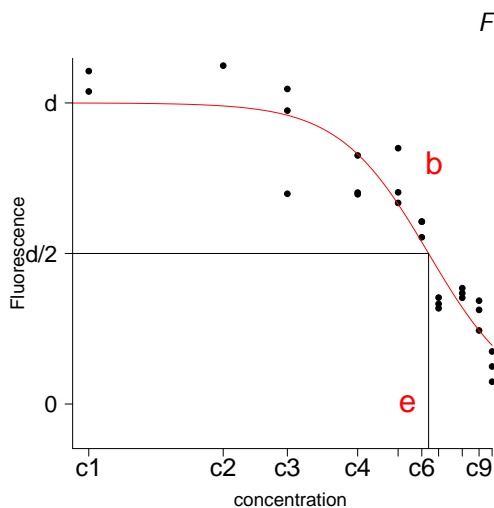
Effet sur la fluorescence de *Nitzschia Palea* d'un herbicide, le Diuron. La fluorescence est un indicateur de la biomasse.



## Ajustement d'un modèle log-logistique



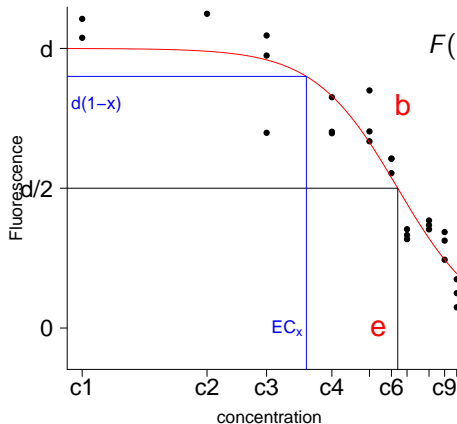
# Paramètres et concentration critique d'effet



$$F(\text{conc}) = \frac{d}{1 + \left(\frac{\text{conc}}{EC_{50}}\right)^b}$$

- $d$  est la fluorescence dans le contrôle.
- $e$  est la concentration où la fluorescence est réduite de moitié par rapport au témoin.
- $b$  est relié à la pente au niveau de  $e$ .
- $b$  et  $e$  sont les deux paramètres d'effet.

# Paramètres et concentration critique d'effet (CEC)

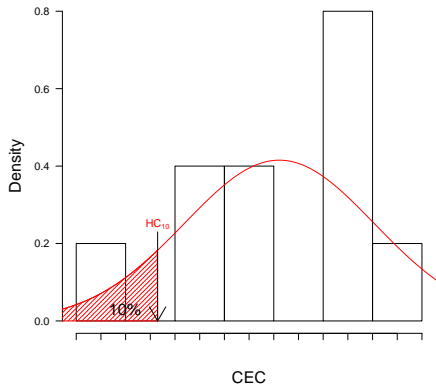


$$F(\text{conc}) = \frac{d}{1 + \left(\frac{\text{conc}}{EC_{50}}\right)^b}$$

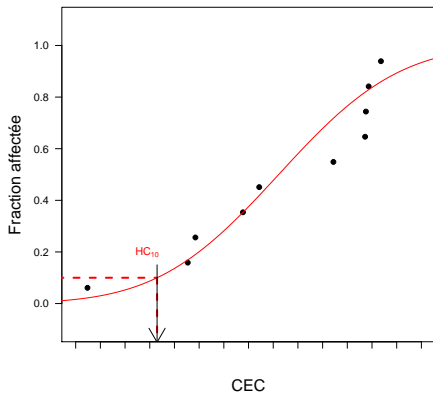
- Pour un seuil  $x$  choisi, on définit la  $EC_x$ .
- $e$  correspond à la  $EC_{50}$ .
- à une reparamétrisation près, la CEC peut-être considérée comme un des paramètres du modèle.

# Concentration sans risque, $HC_p$

Histogramme des données de sensibilité



Distribution cumulée (CDF)



Ajustement d'une distribution lognormale sur les estimations des CEC.

## Sources d'incertitude

Il y a deux sources d'incertitude dans l'approche SSD classique :

- Lors de la détermination des CEC, estimée à partir d'un modèle concentration-effet.
- Lors de l'ajustement de la distribution des CEC.

En règle générale, l'incertitude sur les CEC est tout bonnement **ignorée**.

## Quelques défauts de la SSD classique

- Dans l'approche SSD classique, l'utilisation de CEC aboutit à **l'abandon d'une grande quantité d'information** (autres paramètres d'effet).
- Le choix du seuil de toxicité ( $x$  de la  $EC_x$ ) est **arbitraire**.
- L'approche en deux temps revient à **négliger une partie de l'incertitude**.
- Enfin, la SSD ne fournit une  $HC_p$  basée sur une  $EC_x$ , c'est à dire une **proportion affectée** de la communauté. La réponse globale **relative au endpoint mesuré**, c'est à dire la réduction de la fluorescence globale, est au moins aussi importante, mais elle est perdue lors de la modélisation classique de la SSD.

# Objectifs

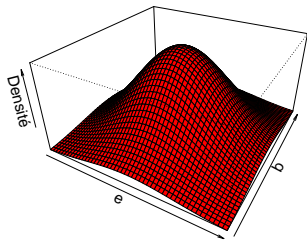
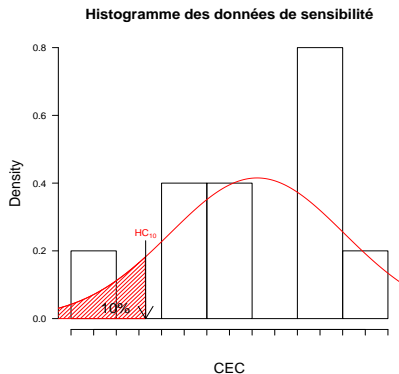
Notre approche est d'utiliser un modèle hiérarchique pour la SSD prenant en compte l'**ensemble des données** pour :

- **modéliser toute la variabilité biologique** de la communauté étudiée,
- **propager l'incertitude** sur les données originales jusqu'à l'indicateur de risque,
- afin de prédire la **réponse globale** d'une communauté à un polluant.



# SSD : distribution des autres paramètres.

La SSD classique est une modélisation de la distribution des CEC. On peut proposer une modélisation de la distribution jointe des deux paramètres  $b$  et  $e$  :

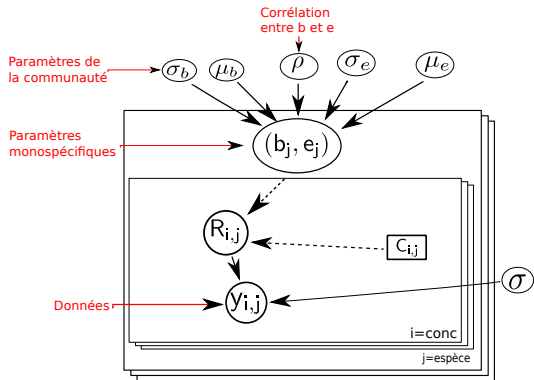


Distribution  $e$  :  $\mu_e, \sigma_e$

Distribution jointe  $b$  et  $e$ ,

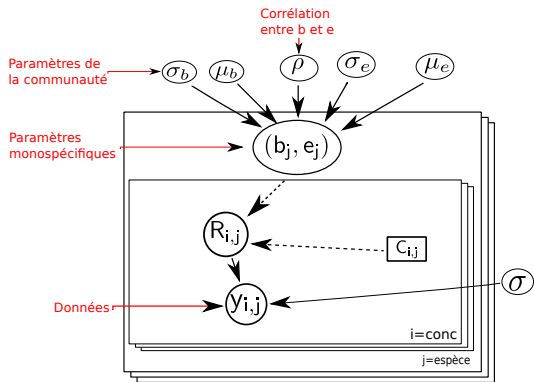
$\mu_e, \sigma_e, \mu_b, \sigma_b, \rho$

# Modélisation hiérarchique: Diagramme Acyclique Dirigé (DAG)



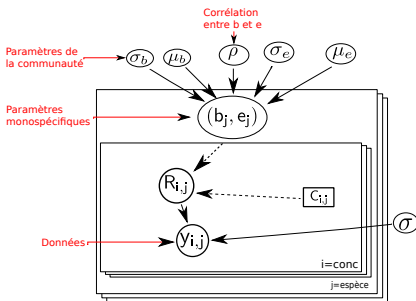
Pour un polluant, une espèce est représentée par une paire de paramètres, et la communauté est décrite par des hyper-paramètres (hors du cadre)

# Modélisation hiérarchique



- $d$  ne dépend que du contrôle, il est estimé séparément.
- $b$  et  $e$  caractérisent l'effet du polluant.

# Modélisation hiérarchique



$$(b_j, e_j) \sim \mathcal{N}_m(\mu, \Sigma)$$

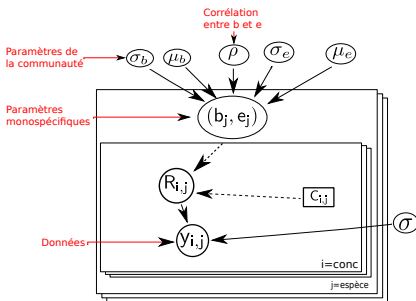
$$\mu = \begin{pmatrix} \mu_b \\ \mu_e \end{pmatrix}; \Sigma = \begin{pmatrix} \sigma_b^2 & \rho\sigma_e\sigma_b \\ \rho\sigma_e\sigma_b & \sigma_e^2 \end{pmatrix}$$


---


$$y_{i,j} \sim \mathcal{N}(R_{i,j}, \sigma)$$

Liens stochastiques du modèle.

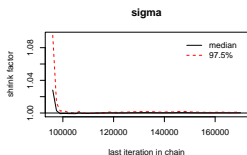
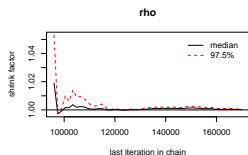
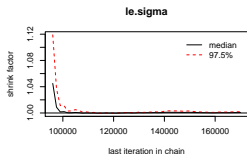
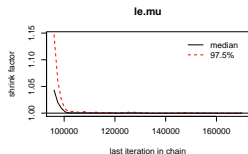
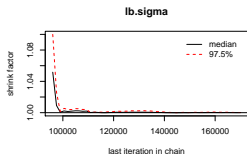
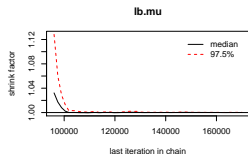
# Modélisation hiérarchique



$$\begin{aligned} \mu_{\ln b} &\sim \text{dnorm}(0., 1.) \\ \sigma_{\ln b} &\sim \text{dunif}(-1, 2) \\ \mu_{\ln e} &\sim \text{dnorm}(\overline{\ln(c_i)}, \max(\ln(c_i)) - \min(\ln(c_i))) \\ \sigma_{\ln e} &\sim \text{dunif}(0, 3) \\ \rho &\sim \text{dunif}(-1., 1.) \\ \sigma &\sim \text{dunif}(0., 2.) \end{aligned}$$

Priors sur les paramètres du modèle.

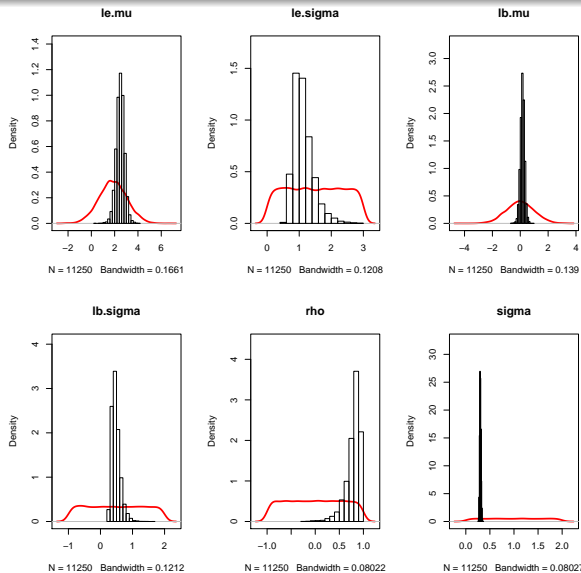
# Ajustement du modèle hiérarchique 1



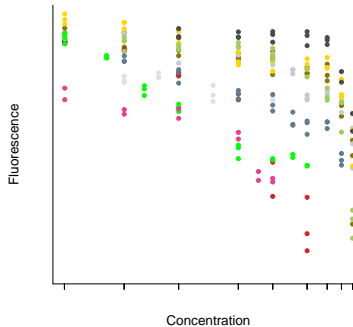
Modèle ajusté en utilisant JAGS

- burn in de 45000
- 75000 itérations (estimé avec Raftery-Lewis)
- thin de 20

# Ajustement du modèle hiérarchique 1



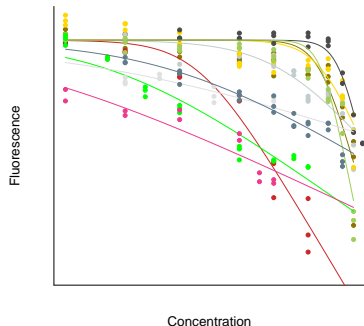
## Ajustement du modèle hiérarchique 3



Ajustement global pour un herbicide (Diuron)



# Ajustement du modèle hiérarchique 3

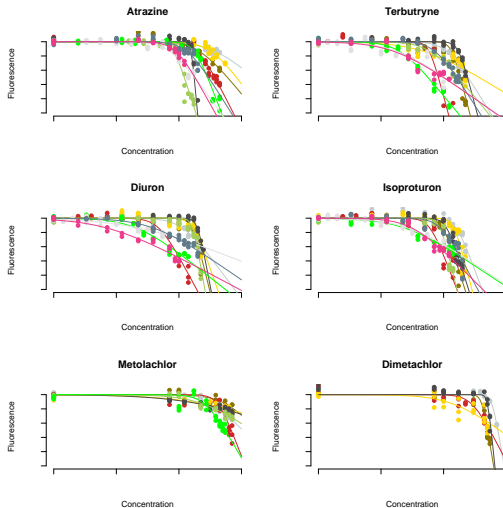


$\log_{10} \mu_b$	0.16[0.06, 0.45]
$\log_{10} \sigma_b$	0.46[0.39, 0.82]
$\log_{10} \mu_e$	2.49[2.27, 3.16]
$\log_{10} \sigma_e$	1.08[0.91, 1.9]
$\rho$	0.83[0.72, 0.95]

Paramètres du modèle et intervalles de crédibilité.

Visualisation de l'ajustement au niveau de chaque espèce (une couleur par espèce).

## Visualisation de l'ajustement du modèle sur chaque herbicide



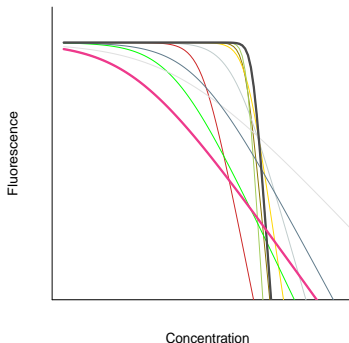
Chaque couleur correspond à une espèce différente.

# Evaluation de la corrélation

	$\rho$
Atrazine	-0.24[-0.47, 0.47]
Terburtyne	0.68[0.52, 0.91]
Diuron	0.83[0.72, 0.95]
Isoproturon	0.87[0.78, 0.97]
Metolachlor	0.41[0.14, 0.89]
Dimetachlor	0.85[0.64, 0.99]

Estimation des paramètres de corrélation avec leur intervalle de crédibilité à 95 %.

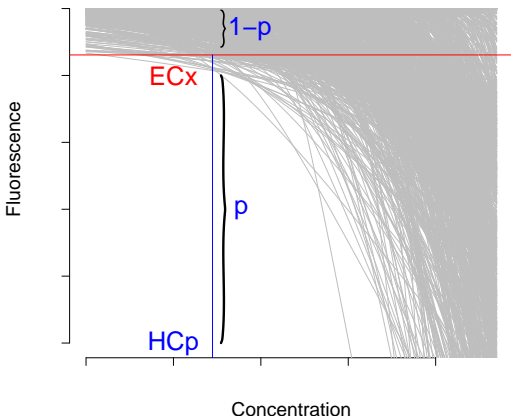
Dans le cas du Diuron, la corrélation se traduit par une pente douce commençant à faible conc., ou abrupte et commençant plus loin.



Courbes ajustées des 11 espèces soumises au Diuron

## Simulation d'une communauté complète et prédiction

On peut reconstruire la variété des réponses possibles dans la communauté complète en tirant au hasard des espèces fictives.



- Pour une concentration donnée, une espèce au dessus de la ligne rouge est affectée à moins de  $x\%$ , à plus de  $x\%$  au dessous.
- On peut mesurer la proportion d'espèces atteintes au delà du seuil de toxicité pour une concentration donnée.

## Modélisation de la réponse globale

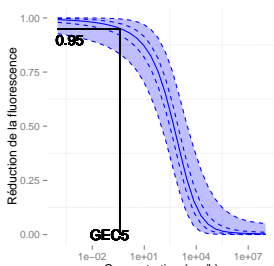
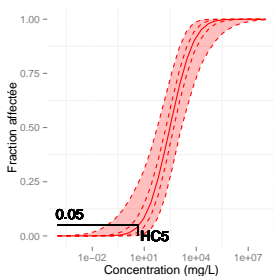
À partir de la distribution jointe, on peut simuler la réponse de communautés fictives. Une taille réaliste pour une communauté de diatomées est de 30 espèces.

En tirant des communautés de taille 30 dans la distribution postérieure du modèle hiérarchique, on obtient les réponses de 30 espèces.

On fait de plus l'hypothèse que chaque espèce de la communauté contribue de manière égale à la réponse globale. Pour calculer la réponse globale de la communauté, on somme les réductions individuelles de fluorescence.

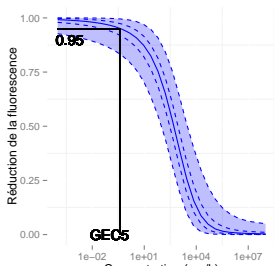
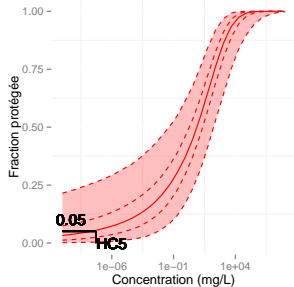
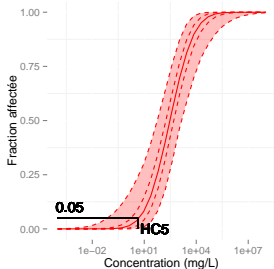
$$R_{\text{tot}} = \frac{1}{N_{\text{especes}}} \sum_i R_i \quad (1)$$

# Réponse en fraction affectée et réponse globale



- SSD construite sur des EC<sub>50</sub> (rouge) et HC<sub>5</sub>, et effet global du contaminant (bleu) avec la GEC<sub>5</sub>, la concentration préservant 95% de la fluorescence.
- À faible concentration, on peut protéger 95% des espèces sans parvenir à protéger 95% de la fluorescence.

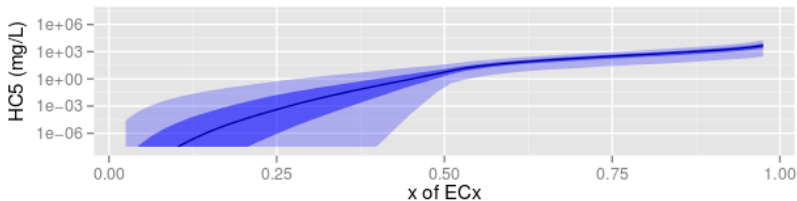
# Réponse en fraction affectée et réponse globale



On peut utiliser un autre seuil de toxicité, par exemple la  $EC_{10}$ .

# Évolution de la $HC_5$ en fonction du $x$ de la $EC_x$

On peut reconstruire une SSD classique pour n'importe quel  $x$  de la  $EC_x$ . Pour cela, on simule une communauté de grande taille ( $4 \cdot 10^6$ ) et on calcule la  $HC_5$  sur les  $EC_x$ .



$HC_5$  en fonction du  $x$  de la  $EC_x$  pour le Diuron. En trait plein la médiane, en sombre les quartiles centraux et en clair l'intervalle à 95%.



## Conclusion et perspectives

La nouvelle modélisation de la SSD :

- prend en compte toute l'information présente dans les données.
- fait apparaître des corrélations biologiques qui n'avaient pas été mises en évidence auparavant.

Elle permet de :

- prédire une réponse globale de la communauté en terme de biomasse.
- jeter un regard critique sur le traitement de l'incertitude dans l'approche SSD classique.

Pour aller plus loin :

- Discussion de l'hypothèse d'équi-contribution des espèces.
- Application à d'autres endpoints et choix d'un autre résumé.