

Analyse de données écotoxicologiques par inférence bayésienne

E. Billoir¹, ML. Delignette-Muller², A. Péry³, S. Charles¹

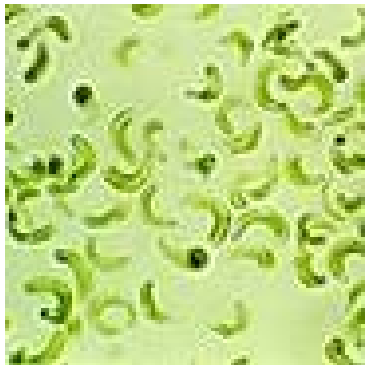
billoir@biomserv.univ-lyon1.fr

- ¹ Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Lyon 1, France
² Unité de Microbiologie Alimentaire et Prévisionnelle, Ecole Nationale Vétérinaire de Lyon, France
³ Laboratoire d'Ecotoxicologie, Cemagref, Lyon, France




Contexte

- des tests normalisés (ISO, OECD) effectués en routine dans les laboratoires
 - exposition constante à différentes concentrations
 - des tests aigus : LC50-48h, EC50-24h ...
 - des tests chroniques : NOEC, LOEC ...
 - croissance des algues (*Scenedesmus*, *Selenastrum* ou autre)
 - croissance des poissons (*Danio rerio*)
 - reproduction du zooplancton (*Daphnia magna*)



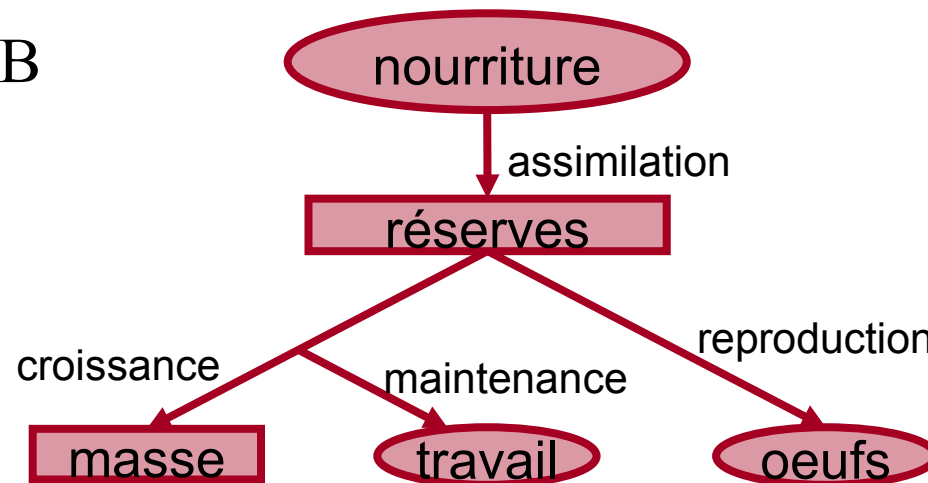
Données

- collectées au  Cemagref de Lyon
- daphnies < 24 h exposées durant 21 jours
- **Zn** (0, 0.074, 0.22, 0.66 mg/L) et **Cu** (0, 2.5, 5, 10 µg/L)
- mesure de la survie et de la reproduction quotidiennement
- mesure de la taille aux jours 7, 14, et 21



Analyse des données

- en général : LC, EC, NOEC, LOEC
- alternative : les modèles DEBtox (Dynamic Energy Budget in toxicology)
 - proposés par Kooijman et Bedaux (1996)
 - basés sur la théorie DEB



- gestion de l'énergie perturbée par un contaminant
- modèles mécanistes, paramètres avec signification biologique
- NEC

Analyse des données

Pour estimer les paramètres DEBtox

- logiciel DEBtox (Kooijman et Bedaux, 1996)



user-friendly, mais
requiert de fixer certains paramètres

- autres travaux

- T. Jager *et al.* (2004, 2006)
- E. Billoir *et al.* (2007)

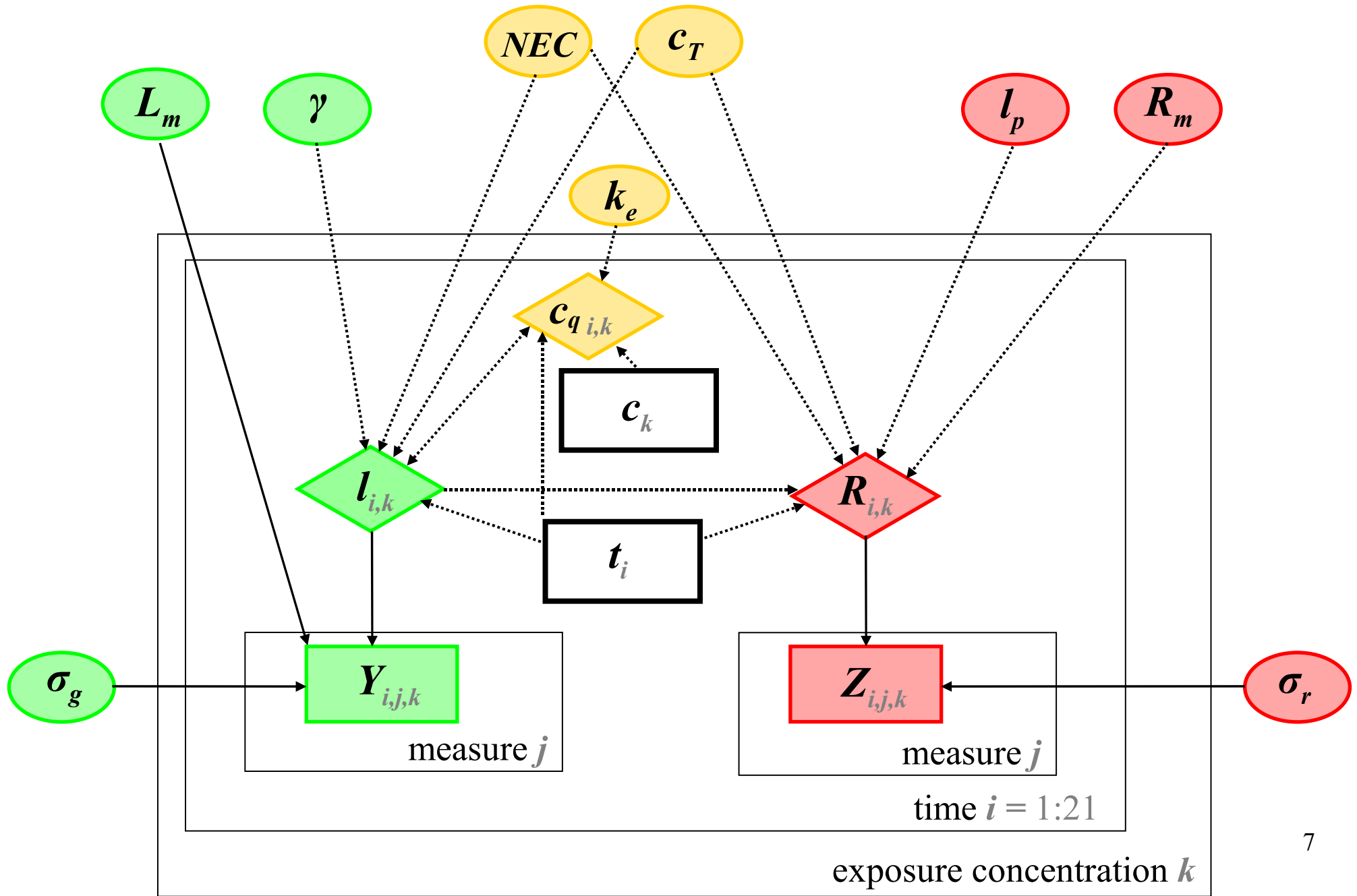
basés sur des approches moindres carrés,
font apparaître des problèmes statistiques:
- approximation linéaire erronée,
- erreurs standard très grandes,
- estimations aberrantes ...

Inférence bayésienne

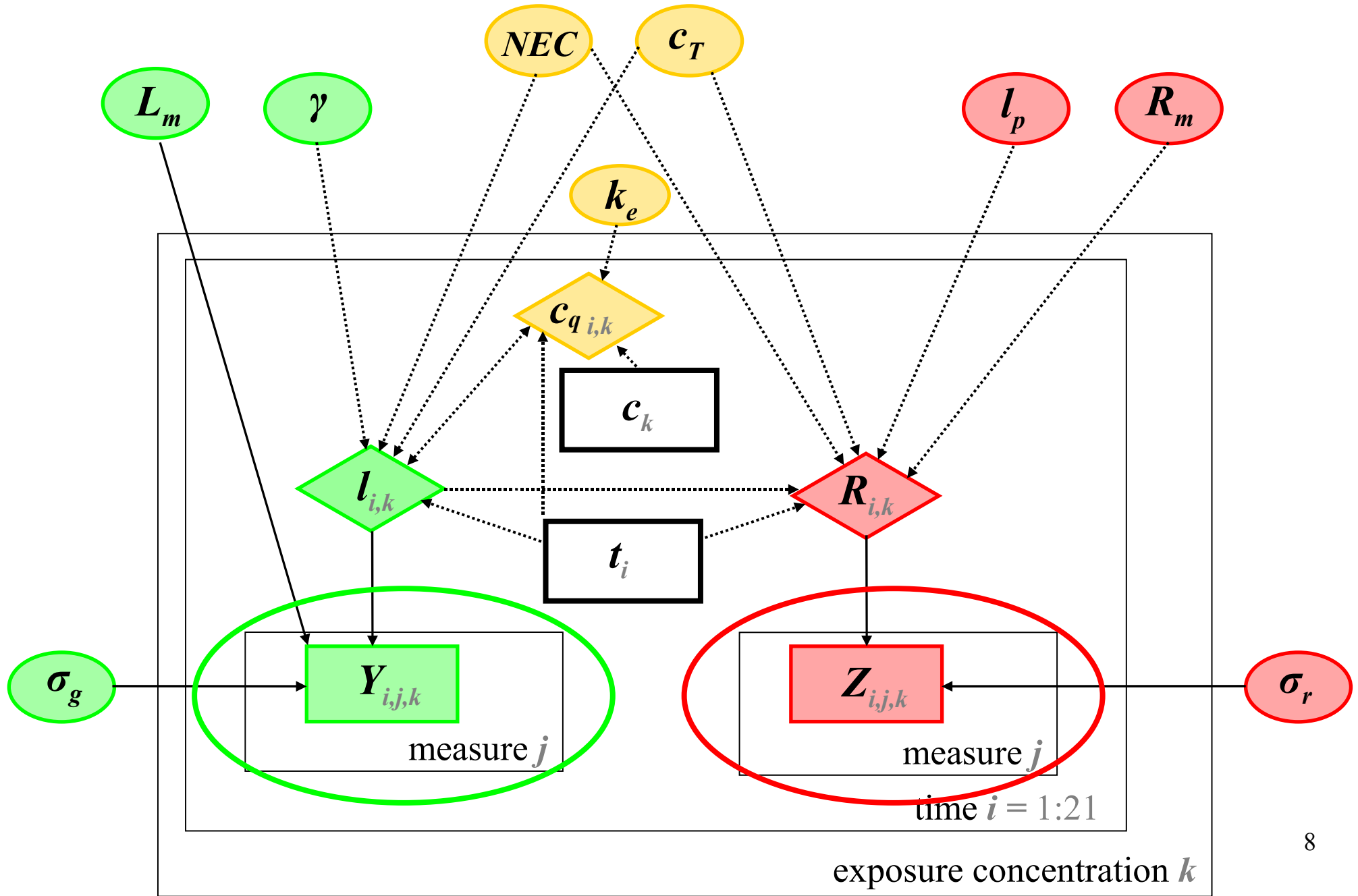
- permet de prendre en compte d'éventuelles connaissances *a priori* sur les valeurs des paramètres
dans notre cas, l'expertise sur les paramètres non-toxicologiques
- permet de faire de l'ajustement simultané sur différent types de données
dans notre cas, des données de croissance et de reproduction
- modèles complexes, donc Monte Carlo Markov Chain (MCMC)

Winbugs  + WBDiff

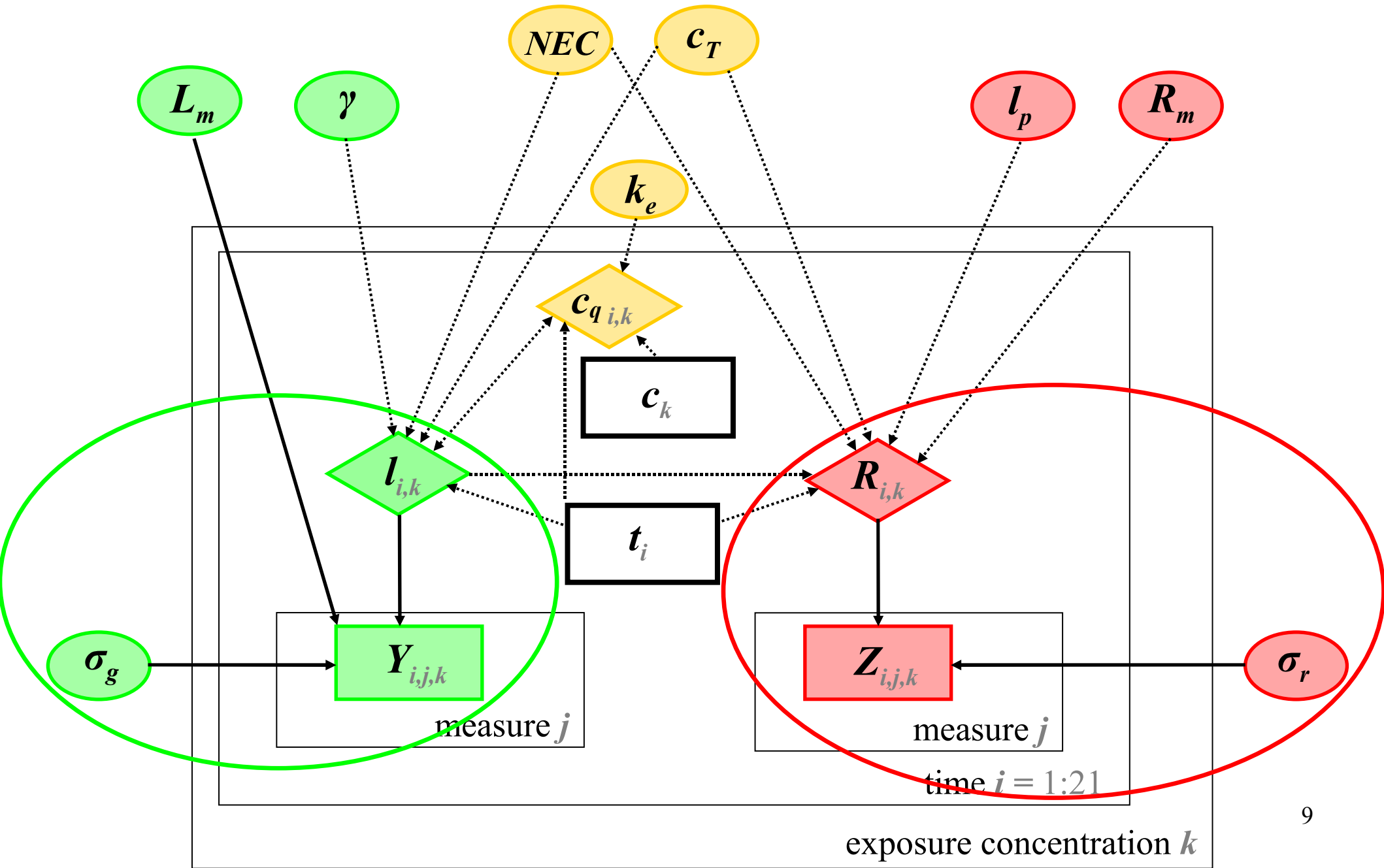
Modèle : graphe bayésien



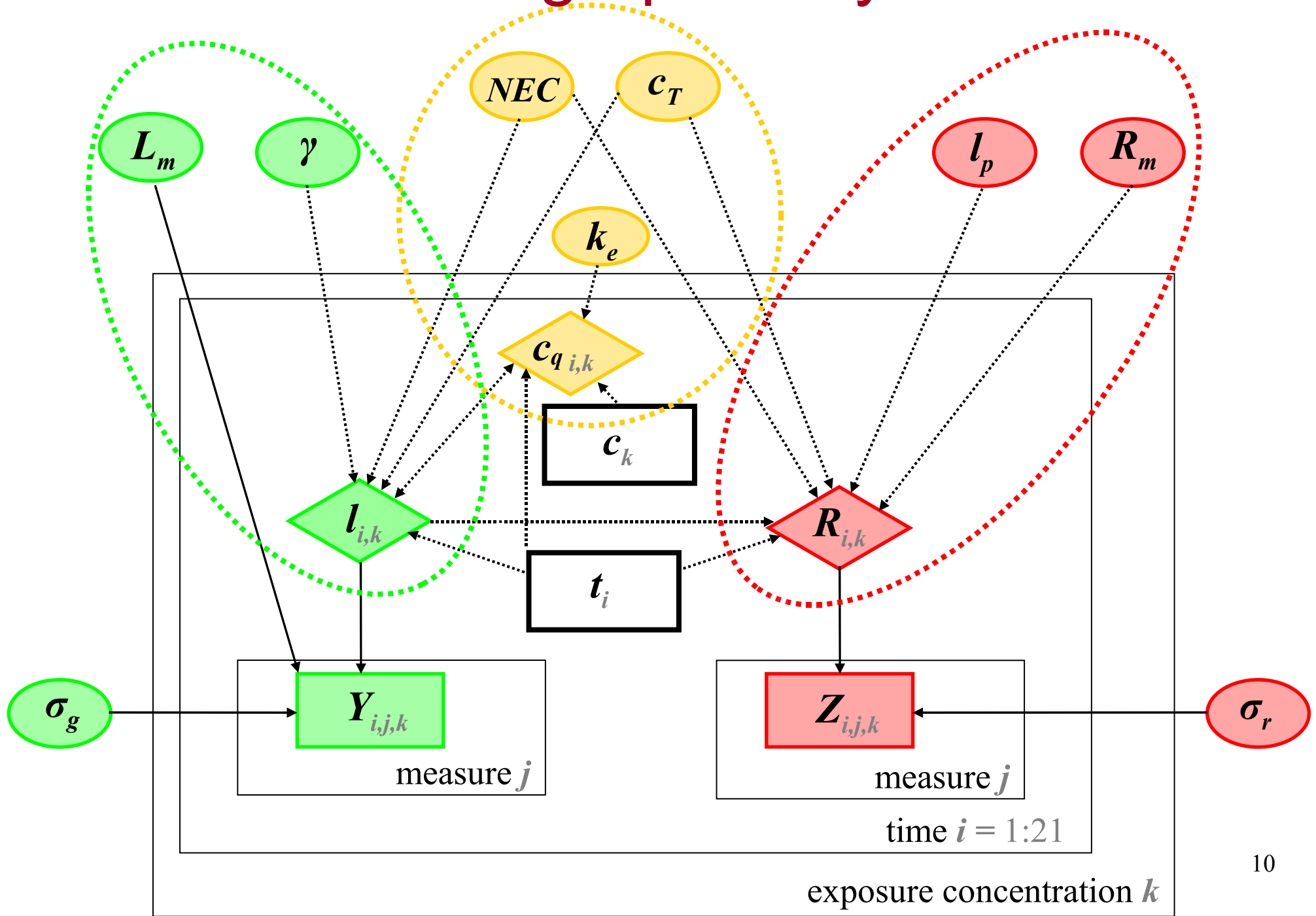
Modèle : graphe bayésien

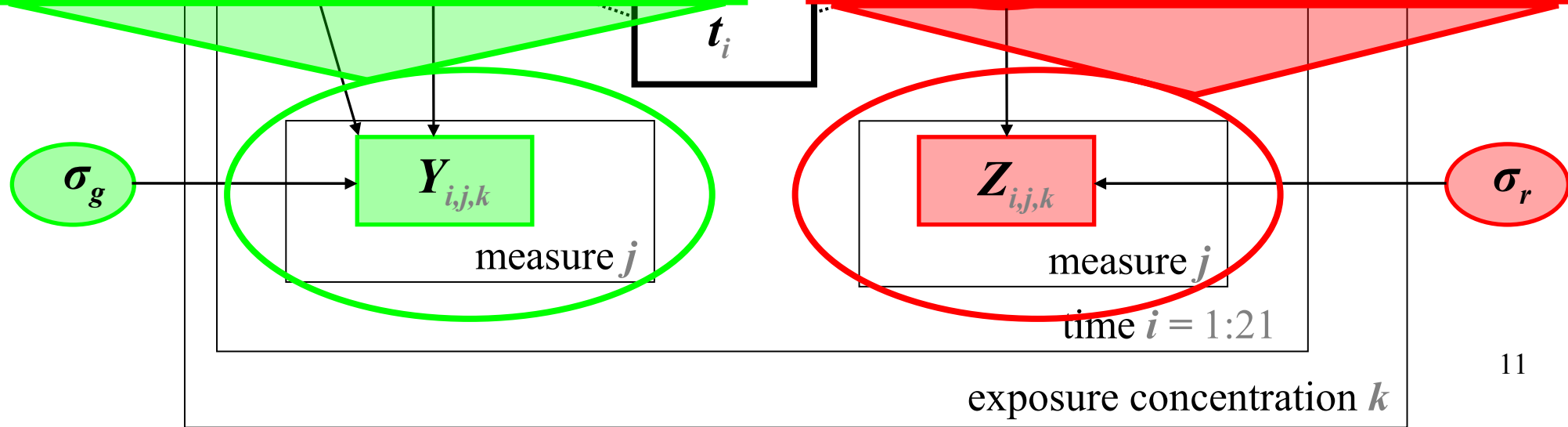
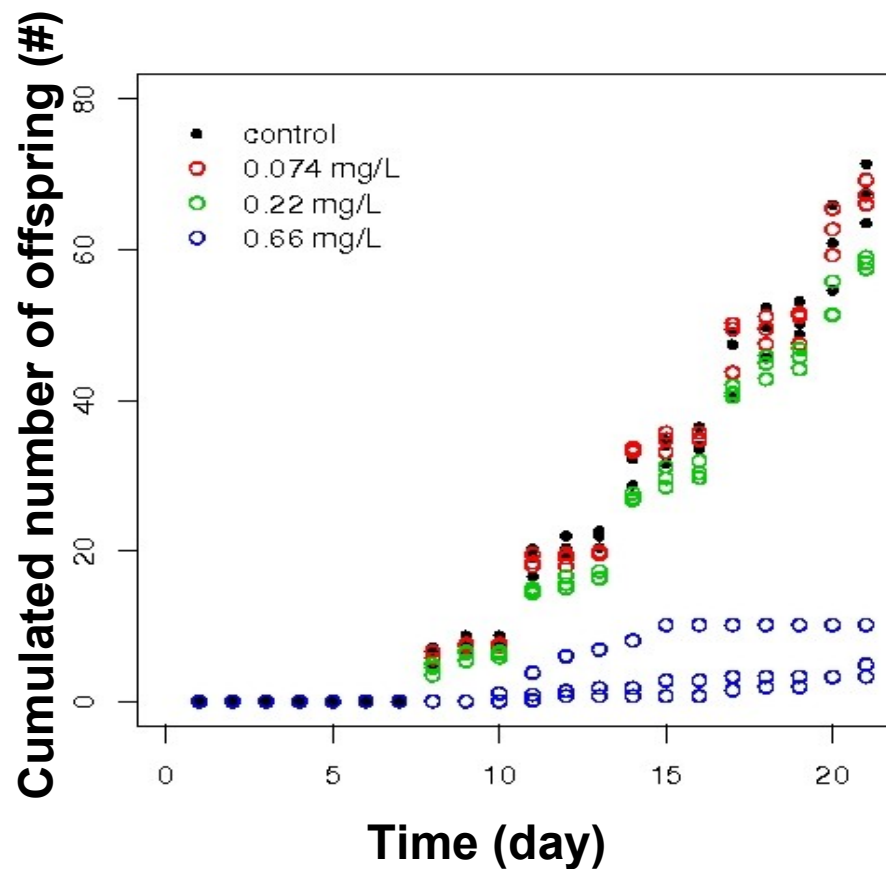
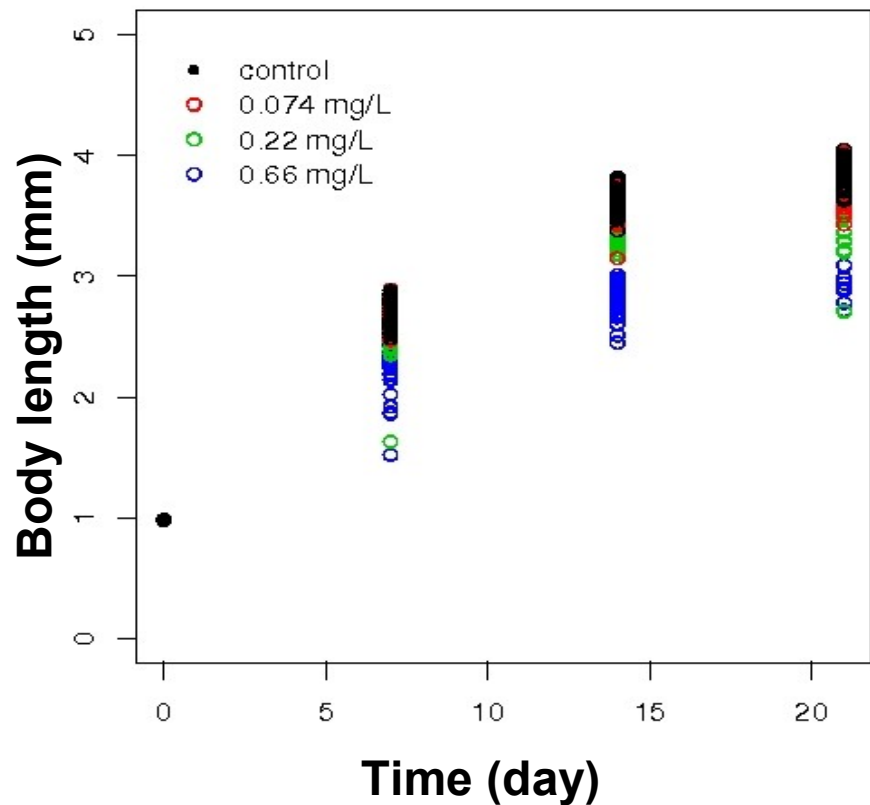


Modèle : graphe bayésien

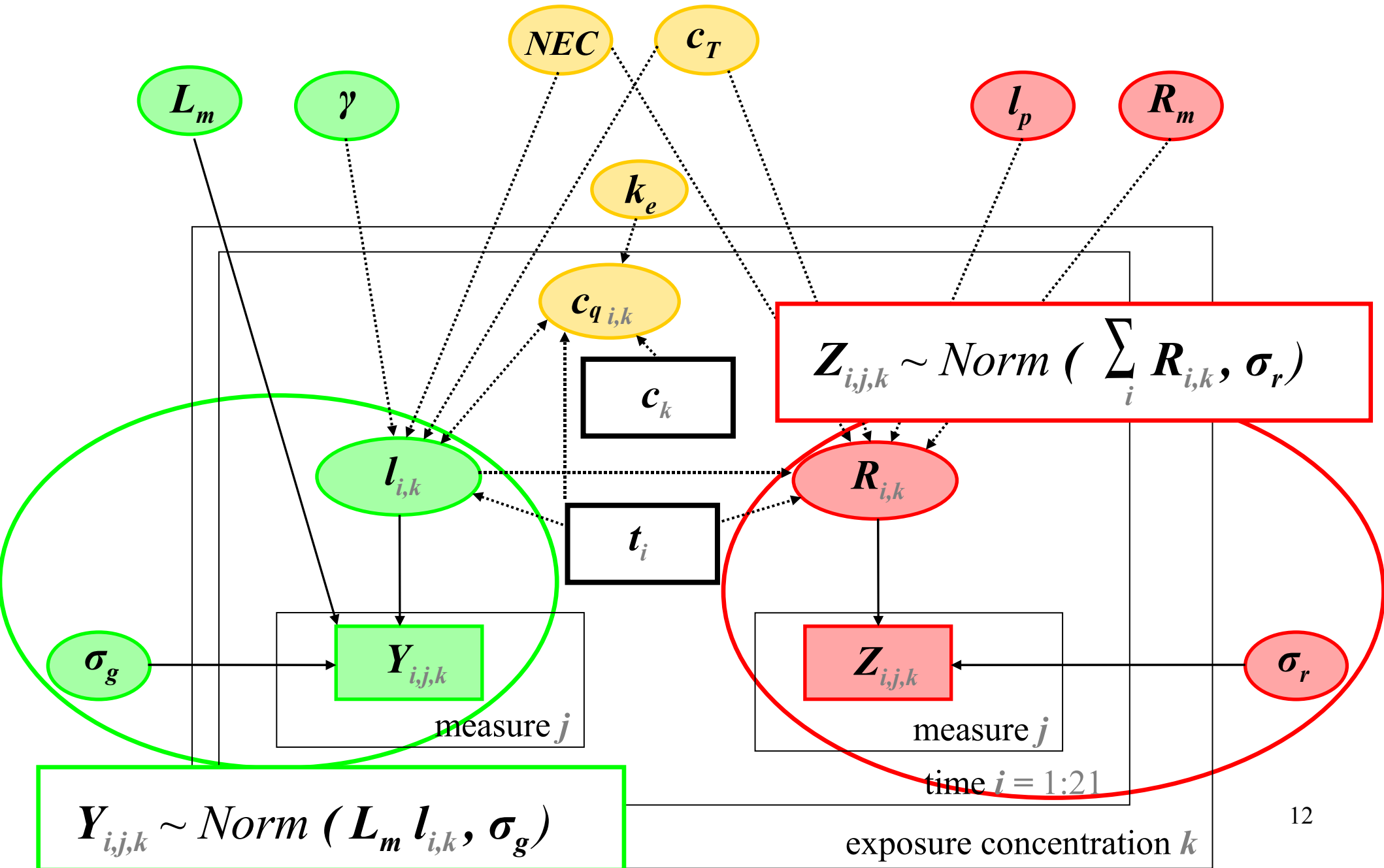


Modèle : graphe bayésien

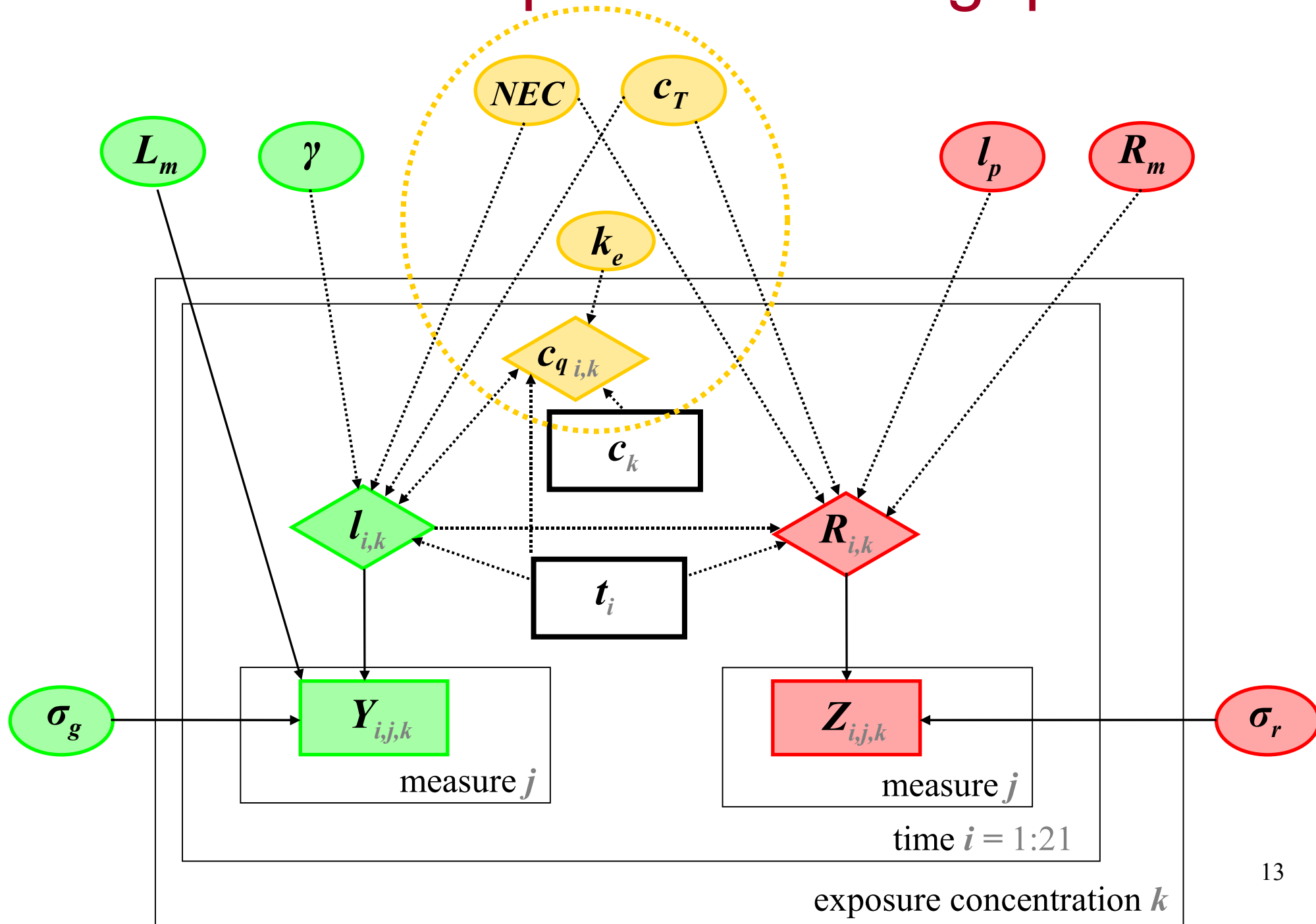




Modèle : liens stochastiques



Modèle : aspects toxicologiques



Modèle : aspects toxicologiques

- modèle de cinétique à un compartiment $\frac{dc_q}{dt} = \frac{k_e}{l} (c - c_q)$

Variables : t (temps), c (conc. d'exposition), c_q (conc. interne), l (taille normalisée)

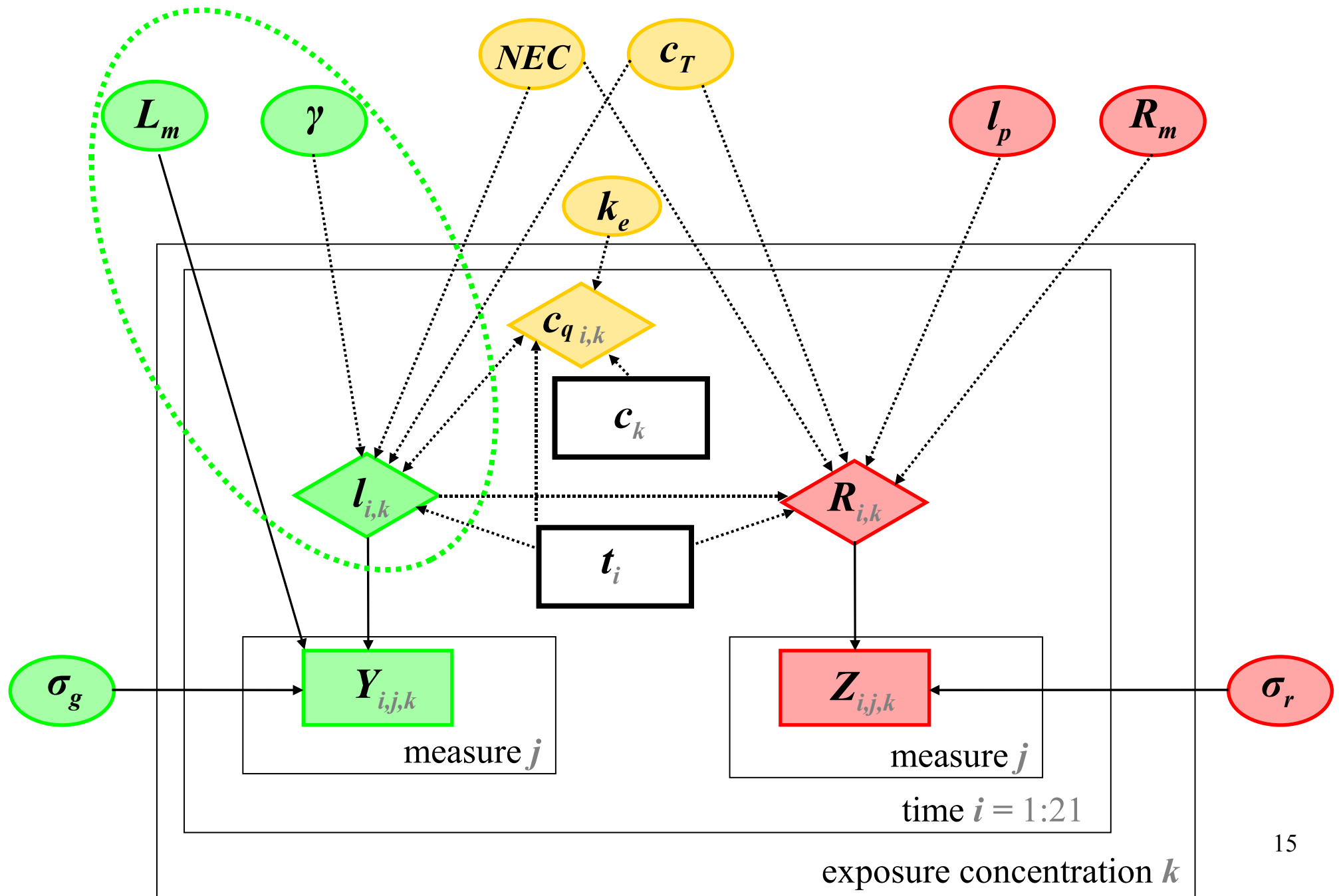
Paramètres : k_e (taux d'élimination)

- effets sur la reproduction et/ou sur la croissance considérés *via* une fonction de stress $s(c_q)$

$$s(c_q) = \begin{cases} c_T^{-1} (c_q - NEC) & \text{if } c_q > NEC \\ 0 & \text{if } c_q \leq NEC \end{cases}$$

Paramètres : c_T (concentration de tolérance), NEC (concentration sans effet)

Modèle : croissance



Modèle : croissance

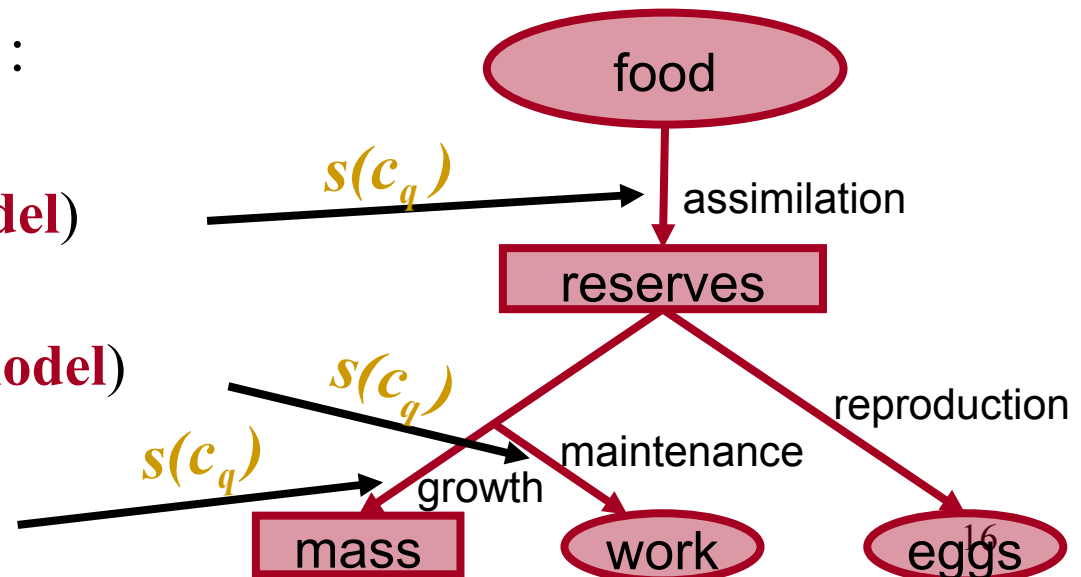
- témoins $\frac{dl}{dt} = \gamma (1 - l)$ $l = \frac{L}{L_m}$

Paramètres : γ (taux de croissance de von Bertalanffy), L_m (taille maximum)

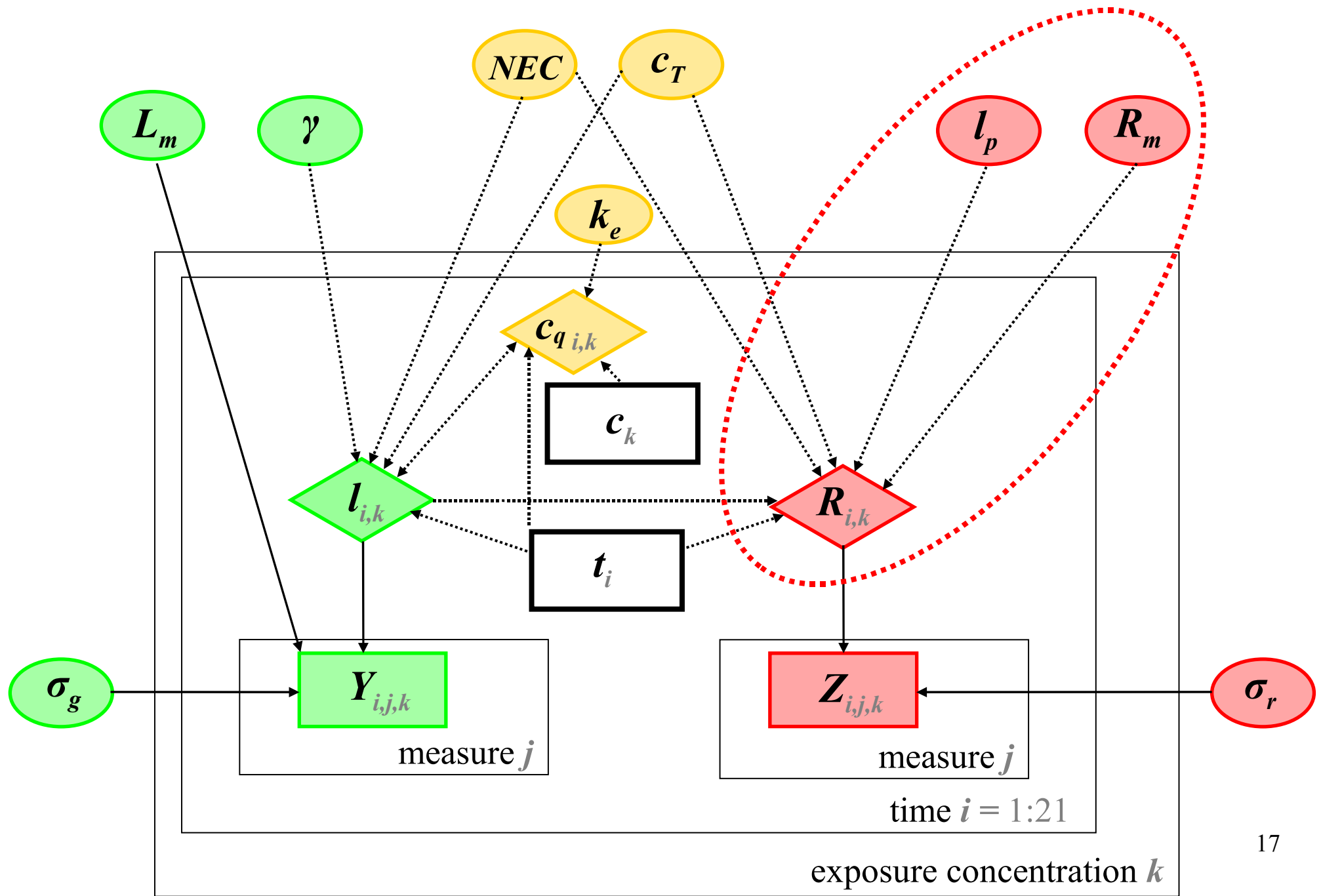
- 3 hypothèses sur le mode d'action du contaminant ...
... qui mènent à 3 équations différentes pour $\frac{dl(c)}{dt}$

fonction de stress $s(c_q)$ appliquée :

- à l'assimilation (**Assimilation model**)
- à la maintenance (**Maintenance model**)
- à la croissance (**Growth model**)



Modèle : reproduction



Modèle : reproduction

• témoins
$$R(t) = \frac{R_m}{1 - l_p^3} \left(\frac{1+l}{2} l^2 - l_p^3 \right) \text{ for } l > l_p$$

Paramètres : R_m (taux de reproduction maximum), l_p (taille normalisée à la puberté)

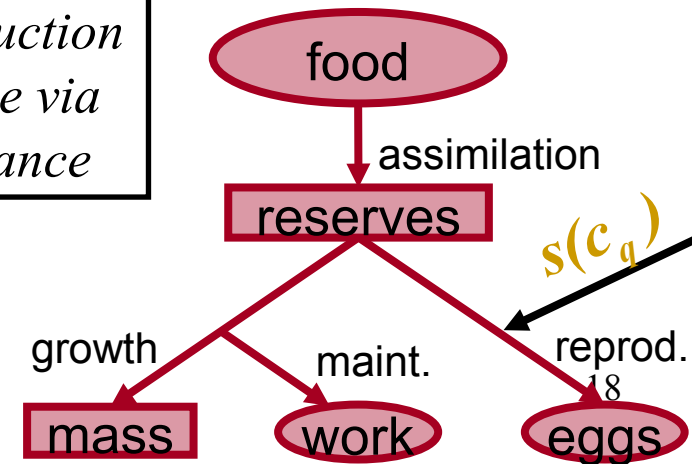
3 hypothèses sur le mode d'action du contaminant ...

... qui mènent à 3 équations différentes pour $R(t, c)$

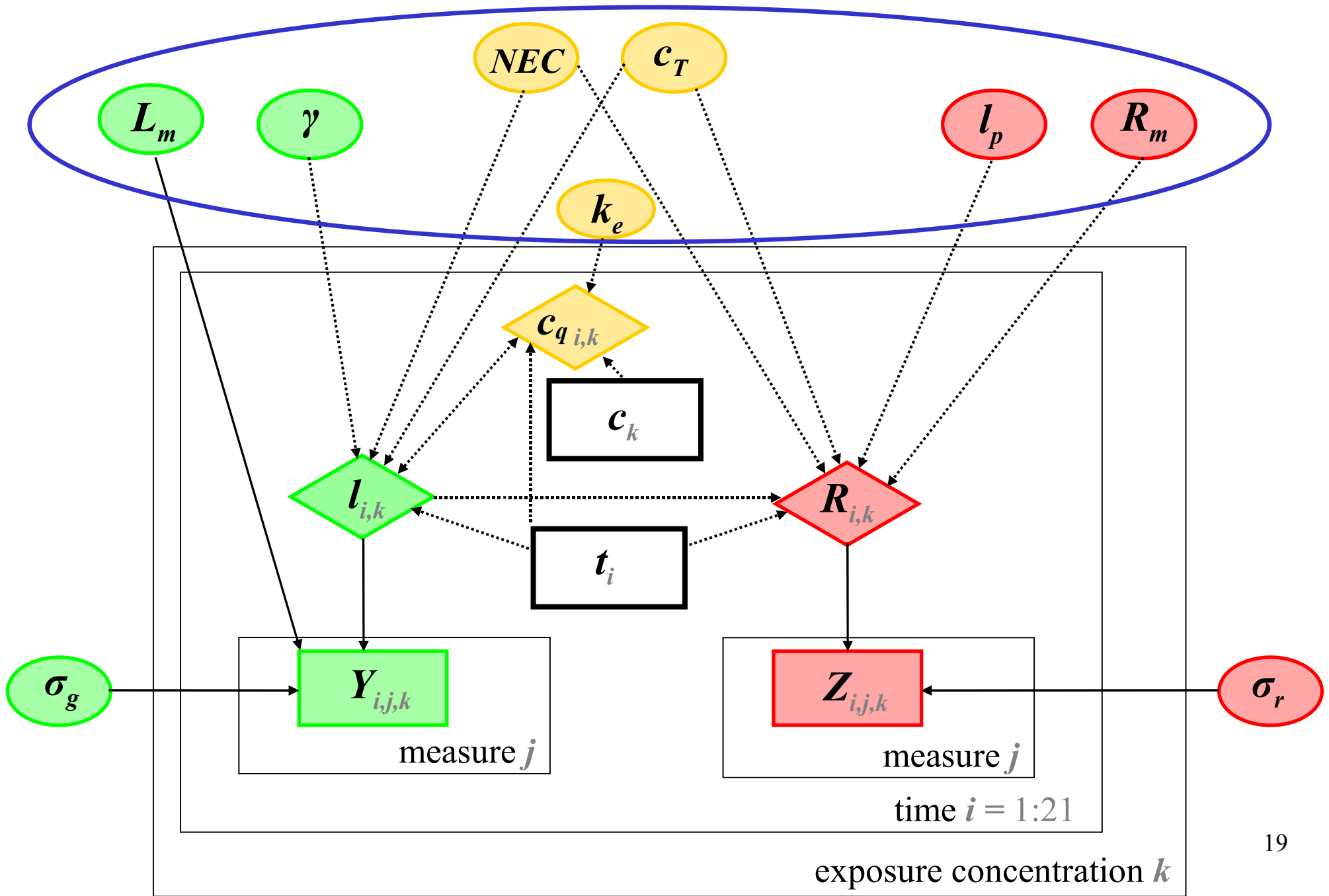
fonction de stress $s(c_q)$ appliquée :

- à l'assimilation (**Assimilation mod.**)
- à la maintenance (**Maintenance mod.**)
- à la croissance (**Growth mod.**)
- surmortalité durant l'ovogénèse (**Hazard mod.**)
- coût supplémentaire pour chaque oeuf (**Costs mod.**)

La reproduction est affectée via la croissance



Modèle : paramètres d'intérêt



Choix des distributions *a priori*

- informatives pour les paramètres non-toxicologiques

Expertise du  Cemagref, utilisation de toutes les données témoins disponibles

distributions normales pour L_m , γ , R_m , l_p

	μ	σ
L_m (mm)	4.77	0.710
γ (/d)	0.112	0.034
R_m (#/d)	10.7	3.6195
l_p (-)	0.487	0.0700

- informatif pour le paramètre de cinétique

expériences additionnelles Zn et Cu \implies cinétique rapide

$$k_e \sim Unif(1,5)$$

- non informatives pour les autres

contaminant-spécifiques

distributions uniformes très larges

Winbugs : les équations différentielles

```
model {
```

```
# system equations specified via BUGS language:
```

```
  solution[1:n.grid, 1:dim] <- ode(init[1:dim], grid[1:n.grid], D(C[1:dim], t), origin, tol)
```

```
  D(C[CQ0], t) <- (cext[1] - C[CQ0]) * ke * f / (C[LL0])
```

```
  D(C[LL0], t) <- gam * ( f+g ) / ( f*(1 - pow(ca,-1) * max(0,(C[CQ0] - nec) ) )+g) * ( f * (1 - pow(ca,-1) *  
max(0,(C[CQ0] - nec) ) ) - C[LL0])
```

```
  D(C[CQ1], t) <- (cext[2] - C[CQ1]) * ke * f / (C[LL1])
```

```
  D(C[LL1], t) <- gam * ( f+g ) / ( f*(1 - pow(ca,-1) * max(0,(C[CQ1] - nec) ) )+g) * ( f * (1 - pow(ca,-1) *  
max(0,(C[CQ1] - nec) ) ) - C[LL1])
```

```
  D(C[CQ2], t) <- (cext[3] - C[CQ2]) * ke * f / (C[LL2])
```

```
  D(C[LL2], t) <- gam * ( f+g ) / ( f*(1 - pow(ca,-1) * max(0,(C[CQ2] - nec) ) )+g) * ( f * (1 - pow(ca,-1) *  
max(0,(C[CQ2] - nec) ) ) - C[LL2])
```

```
  D(C[CQ3], t) <- (cext[4] - C[CQ3]) * ke * f / (C[LL3])
```

```
  D(C[LL3], t) <- gam * ( f+g ) / ( f*(1 - pow(ca,-1) * max(0,(C[CQ3] - nec) ) )+g) * ( f * (1 - pow(ca,-1) *  
max(0,(C[CQ3] - nec) ) ) - C[LL3])
```

```
# Initial conditions:
```

```
init[CQ0] <- 0; init[LL0] <- Lb;init[CQ1] <- 0; init[LL1] <- Lb;init[CQ2] <- 0; init[LL2] <- Lb;init[CQ3] <- 0; init[LL3]  
<- Lb;
```

Winbugs : les distributions *a priori*

```
ca ~ dunif(50,200)
ke ~ dunif(1,5)
nec ~ dunif(0,10)
Lm ~ dnorm(4.772,1.9848) I(1)
gam ~ dnorm(0.1116,865) I(0)
taug ~ dgamma(0.001,0.001) # non informatif
sigmag <- 1 / sqrt(taug)
Rm ~ dnorm(10.73927,0.07633) I(0)
Lp ~ dnorm(0.48695,204.64) I(Lb)
taur ~ dgamma(0.001,0.001) # non informatif
sigmar <- 1 / sqrt(taur)
```

Winbugs : les données

```
list(  
CQ0=1, LLD=2, CQ1=3, LL1=4, CQ2=5, LL2=6, CQ3=7, LL3=8,  
n.grid=22,  
dim=8,  
f = 1,  
g =1,  
origin = 0,  
tol = 1.0E-3,  
grid = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22),  
Lb = 0.2295,  
M=55,  
cext = c(0,2,5,5,10),
```

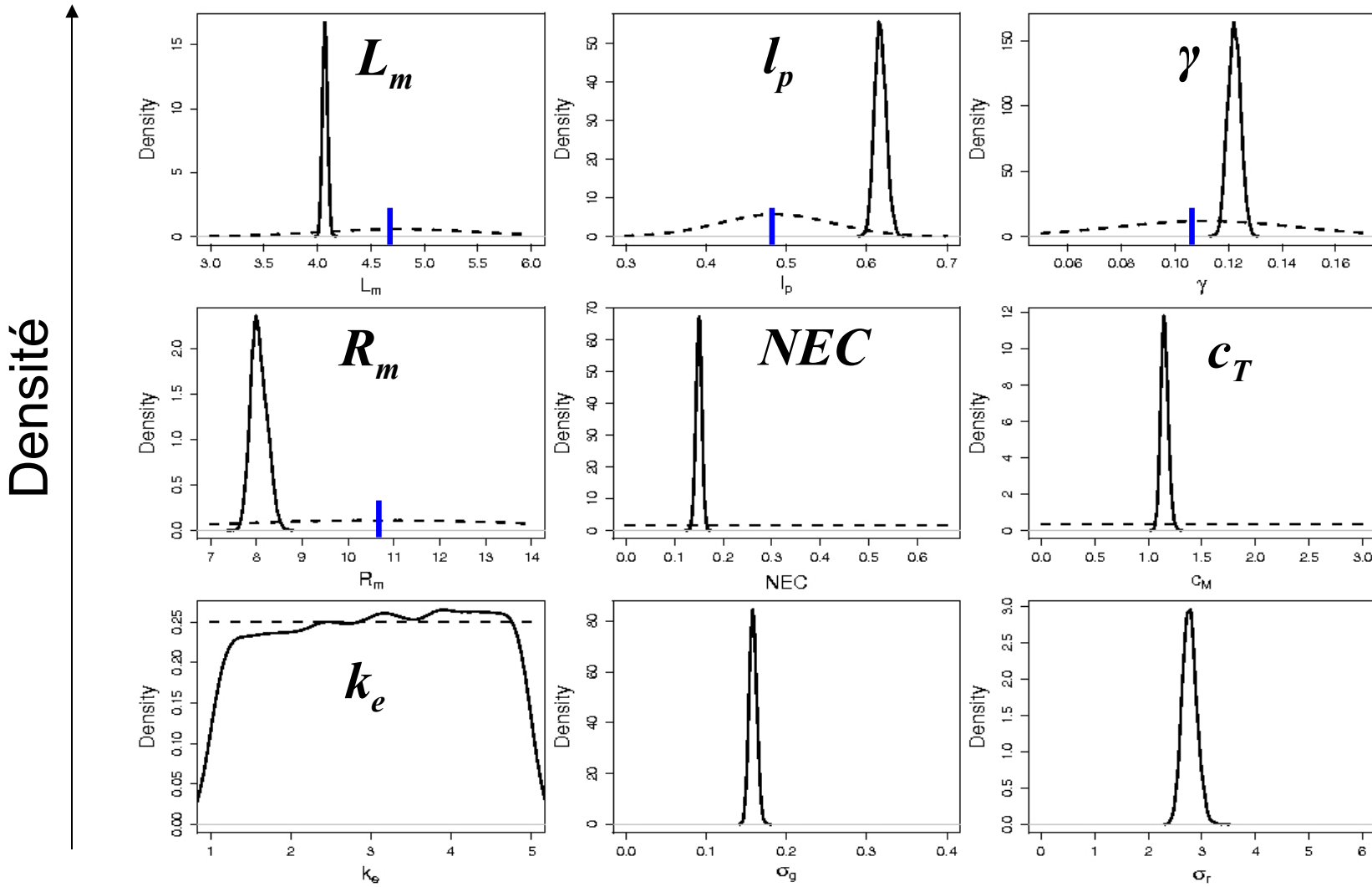
```
Y0=structure( Data =  
c(3.25,3.32,3.23,3.02,2.94,3.04,3.02,2.95,3.57,2.95,  
3.02,3.43,3.08,3.50,2.94,3.33,3.51,3.58,2.88,2.87,2.98,3.50,2.9  
8,3.01,3.51,3.09,3.40,3.49,2.93,3.47,2.96,3.23,3.02,3.52,2.92,2  
.95,2.76,3.00,3.02,2.85,3.06,3.53,3.00,3.41,3.34,3.06,2.90,3.52,2.96,2.95,2.92,2.99,2.92,3.58,NA,  
3.92,3.89,3.79,3.66,3.74,3.78,3.97,3.97,3.78,3.71,  
3.65,3.93,3.87,3.90,3.96,3.73,3.75,3.95,3.93,3.95,4.00,3.80,3.9  
3,3.80,4.07,3.92,3.92,3.91,3.89,3.95,3.87,4.00,3.86,3.92,3.93,3.71,  
3.82,4.01,3.92,3.80,3.96,3.86,3.81,3.99,3.89,3.86,3.81,3.94,3.72,3.82,3.84,3.79,3.81,4.04,3.72,  
4.05,4.07,3.91,4.13,4.09,4.19,4.12,4.19,4.11,4.03,4.08,4.26,4.1  
1,4.27,4.06,4.05,4.16,4.24,4.12,4.12,4.24,3.95,4.34,4.22,4.30,4  
.29,4.30,4.16,4.12,4.06,4.19,4.26,4.16,4.19,4.27,4.01,4.11,4.07  
,4.30,4.26,4.22,4.14,4.13,4.20,4.29,4.11,4.25,4.15,4.16,3.64,4.08,4.37,4.14,4.09,NA), Dim = c(3,55)),
```

```
Rep0=structure( Data =  
c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0.76,0.00,0.00,16.29,15.74,16.00,16.29,16.53,16.00,19.24,18.26,  
16.00,38.94,38.89,38.42,38.94,40.16,38.42,56.12,45.68,46.79,62.  
06,60.42,61.42,62.06,61.68,61.42,64.65,62.05,61.42,82.53,82.05,  
81.84,82.53,83.11,82.42,82.53,84.05,82.42,104.4,102.5,100.5,104.4,103.6,102.3), Dim = c(21,3)),
```

Winbugs : les problèmes rencontrés

- seul WBDiff permet de spécifier le modèle sous forme d'équations différentielles
- avec certains *a priori*, le logiciel tombe directement dans une trappe dès la première itération
- difficultés avec les boucles emboîtées : problème de définition multiple des nœuds obligeant à faire des définitions redondantes
- temps d'exécution très long (plusieurs heures)

Résultats (Zn) Modèle Maintenance



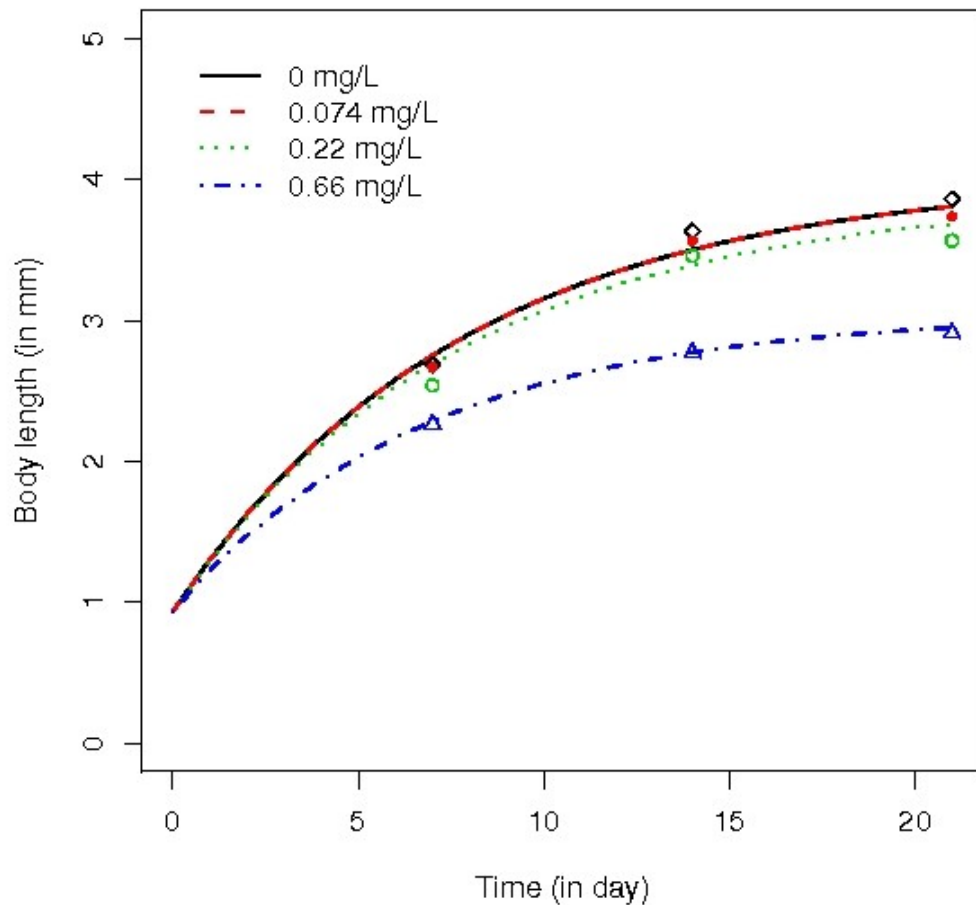
— distributions *a posteriori*

- - | - - distributions *a priori* (et leur moyenne quand informative)

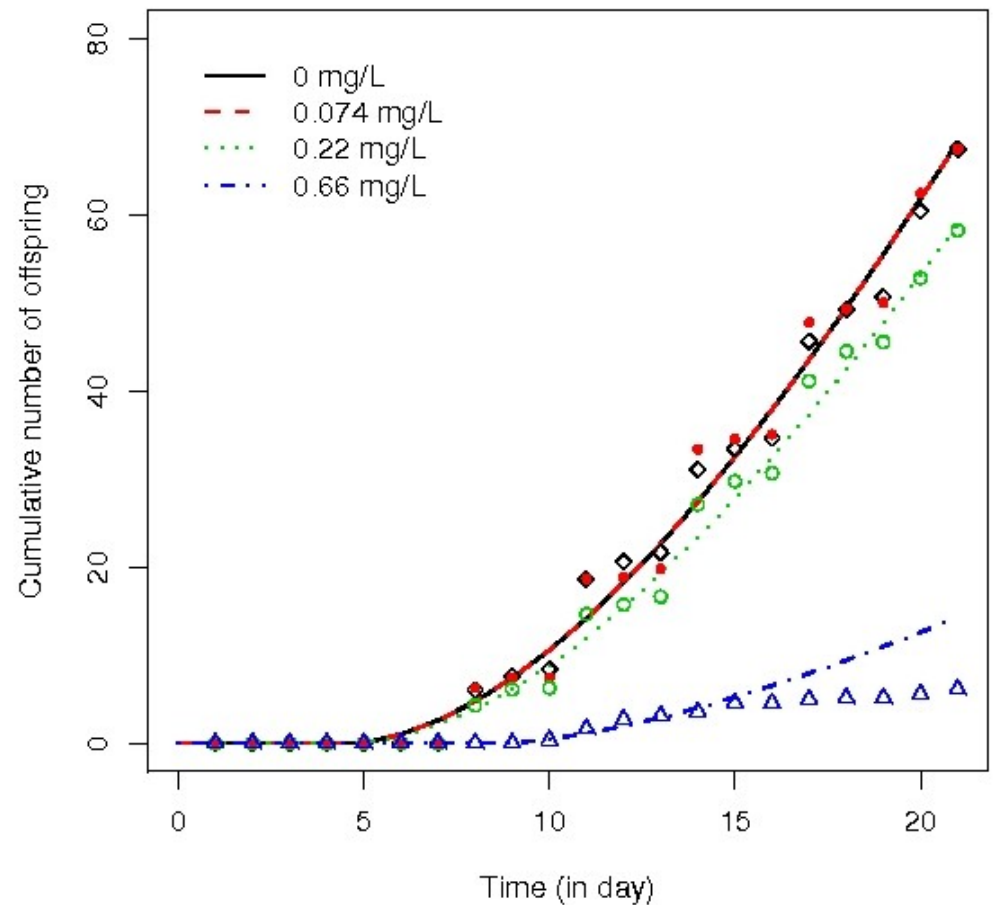
Résultats (Zn)

Meilleurs modèles = modèle Assimilation ou modèle Maintenance

Growth



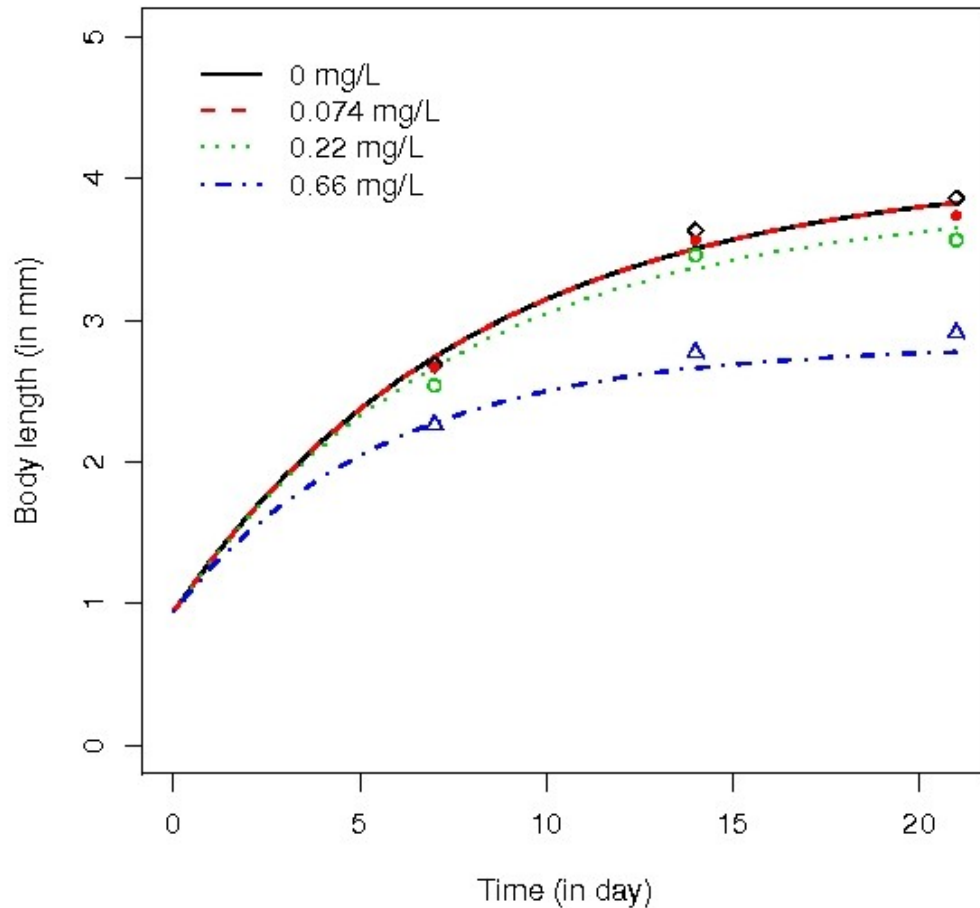
Reproduction



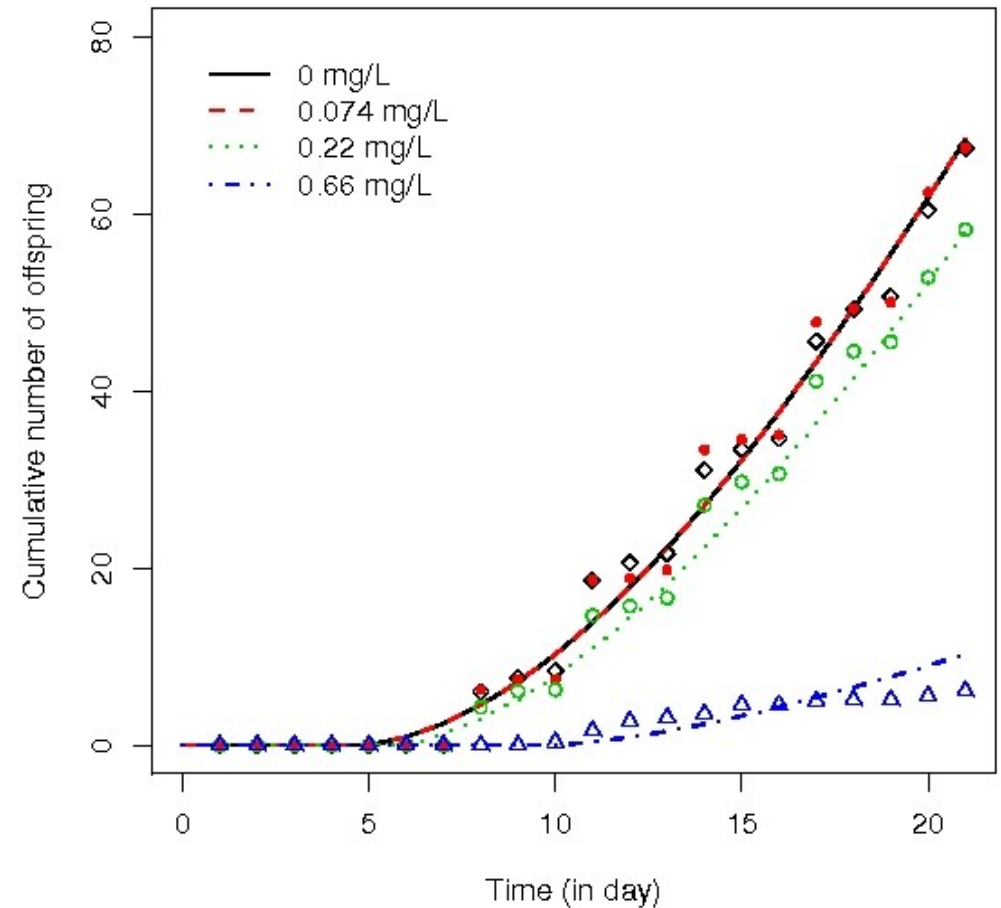
Résultats (Zn)

Meilleurs modèles = modèle Assimilation ou modèle Maintenance

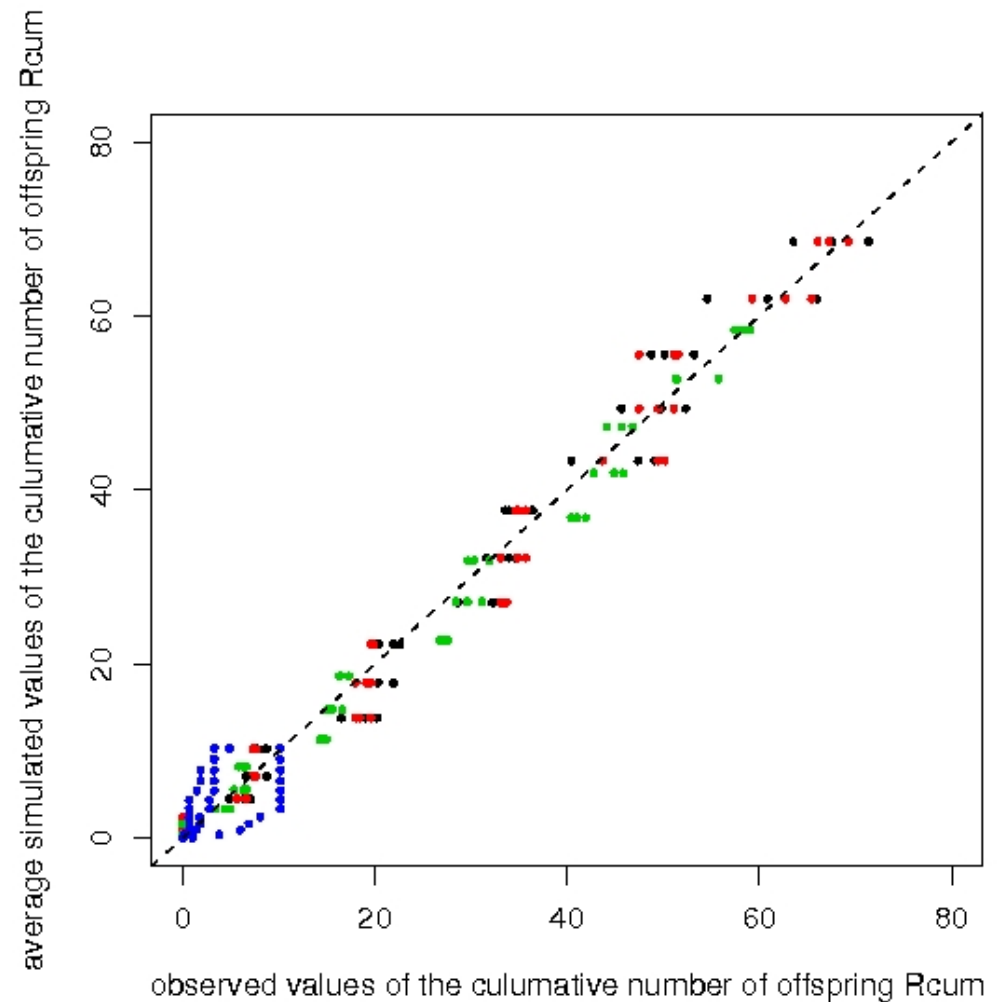
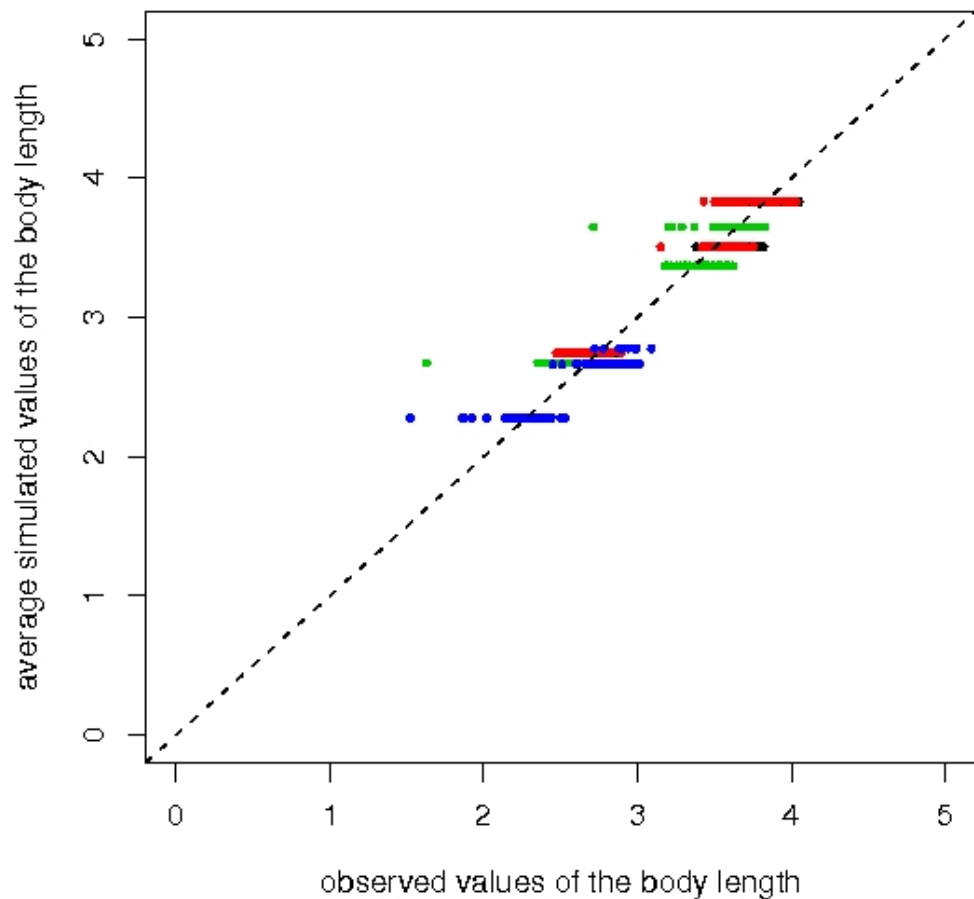
Growth



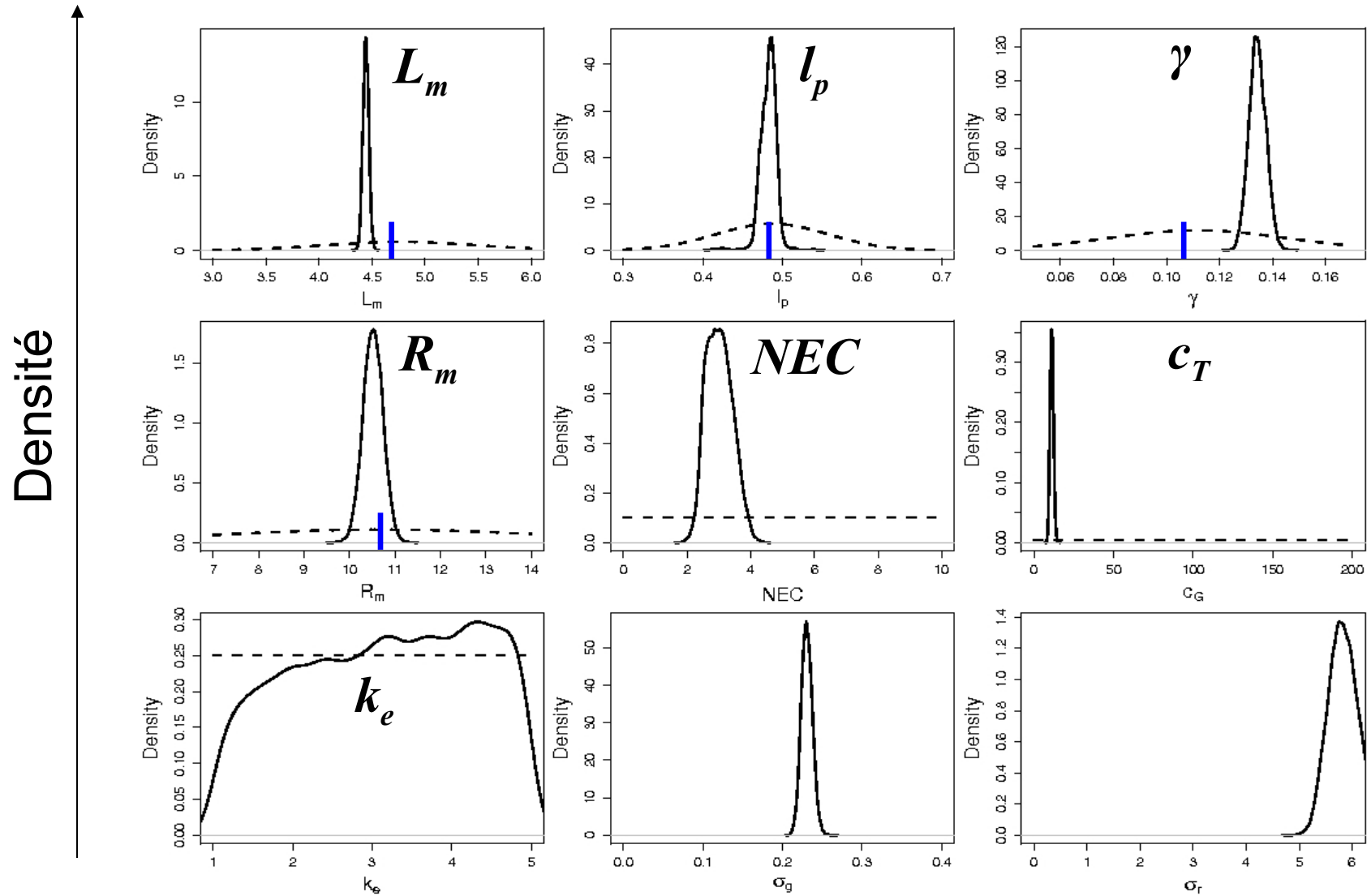
Reproduction



Résultats (Zn) Modèle Maintenance



Résultats (Cu) Modèle Growth



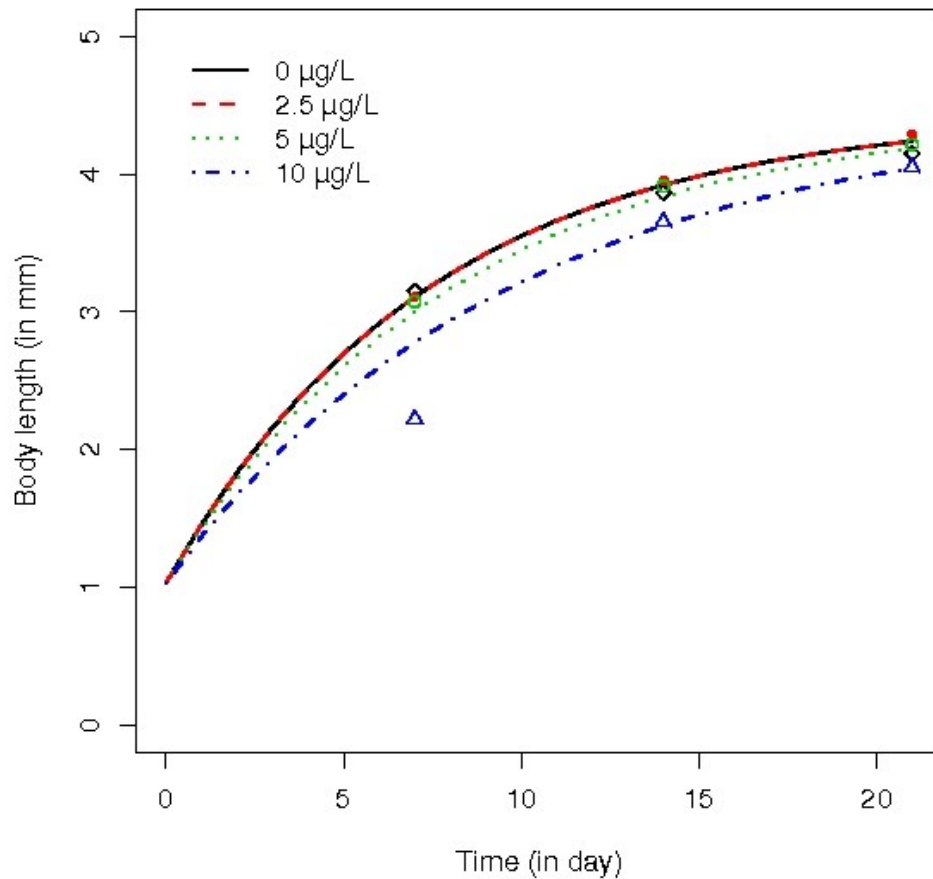
— distributions *a posteriori*

- - - | - - - distributions *a priori* (et leur moyenne quand informative)

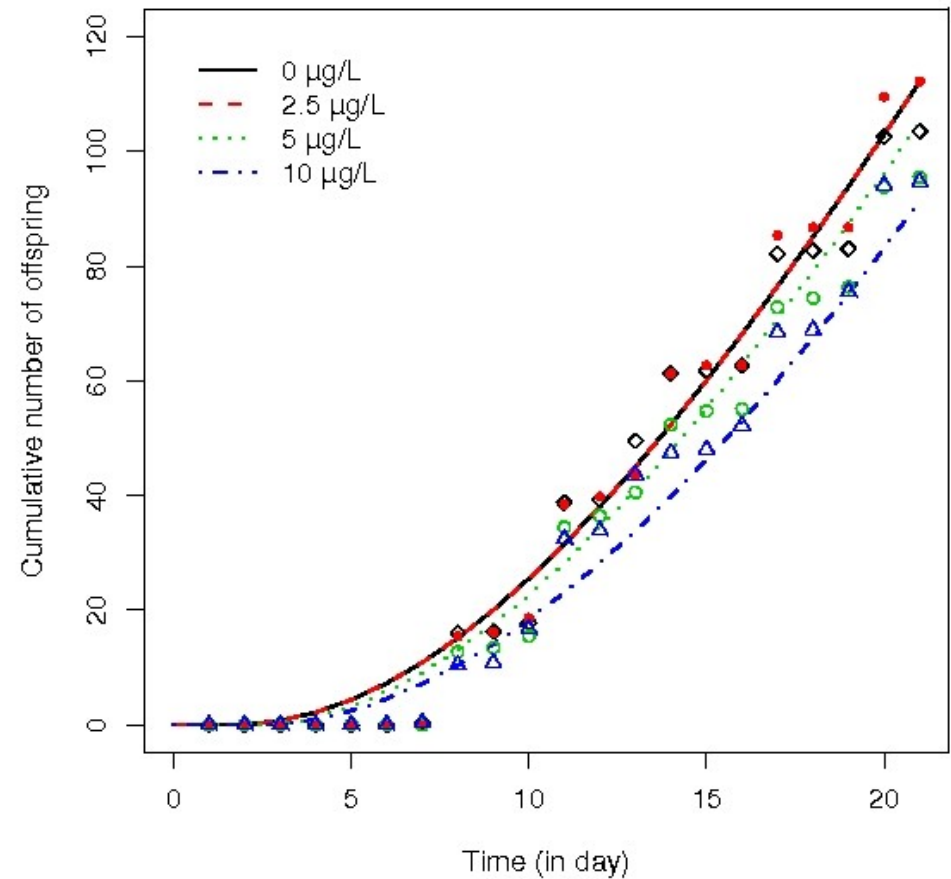
Résultats (Cu)

Meilleur modèle = modèle Growth

Growth



Reproduction



DIC

Modèle d'effet	Assimil.	Mainten.	Growth	Costs	Hazard
Zn	690.58	674.89	926.10	1324.41	1324.54
Cu	1558.61	1565.38	1476.69	1639.46	1641.13

Conclusions Questions Perspectives

- L'inférence bayésienne est bien adaptée dans le cas des modèles DEBtox
 - modèle mécanistes, connaissances *a priori* sur les paramètres
 - permet l'ajustement simultané à différents types de données
- L'inférence bayésienne donne de bons résultats dans le cas des modèles DEBtox
 - fournit des estimations cohérentes
 - fournit des distributions, pas seulement des estimations ponctuelles
- Savez-vous si quelqu'un a un projet de WBDiff dans BRugs ?
- Est-ce que quelqu'un a l'expérience du codage d'un algorithme bayésien pour un modèle contenant des équadifs ?