

# Modélisation des sources de variabilité et d'erreur dans les modèles de réponse fonctionnelle. Exemple des perdrix

V. Bansaye, S. Billiard, M. Delattre, E. Vergu  
projet ANR ABIM

- 1 Modèle stochastique de réponse fonctionnelle
- 2 Objectifs
- 3 Formulation générique d'un modèle mixte
- 4 Analyse de sensibilité
- 5 Choix de modèle
- 6 Simulations

# Réponse fonctionnelle

**Réponse fonctionnelle** : nombre d'interactions entre individus de plusieurs espèces en fonction de la densité d'une espèce d'intérêt.

Exemples : nombre de proies consommées par un prédateur par unité de temps en fonction de la densité des proies.

Approche historique : phénoménologique, échelle macroscopique.

Billiard, Bansaye, Chazottes, *J Roy Soc Interface* 2018 : **approche mécaniste stochastique**; dérivation de la réponse fonctionnelle à l'échelle de la population par approximation des mécanismes décrits à l'échelle des individus.

# Modèle stochastique de réponse fonctionnelle : cadre générique

Sous l'hypothèse que le nombre d'interactions est grand et que la densité des pop.  $x, y$  ne change pas, la réponse fonctionnelle  $R(x, y)$  suit une loi normale (approximation dans le cadre de la théorie du renouvellement) :

$$R(x, y) \sim \mathcal{N}\left(\frac{1}{\mathbb{E}(T(x, y))}, \frac{\text{Var}(T(x, y))}{\Delta \mathbb{E}(T(x, y))^3}\right)$$

où  $\Delta$  est l'intervalle de temps et  $T(x, y)$  représente le temps (aléatoire) entre deux interactions.

# Modèle stochastique de réponse fonctionnelle : exemple

**Exemple** : recherche de nourriture dans un espace de dimension 2 avec manipulation de la proie

$$\mathbb{E}(R) = \frac{\sqrt{y} - 1}{C_{2E}\lambda + h(\sqrt{y} - 1)}$$

et

$$\text{Var}(R) = \frac{C_{2V} - C_{2E}^2}{\Delta} \frac{\lambda^2(\sqrt{y} - 1)}{(C_{2E}\lambda + h(\sqrt{y} - 1))^3}$$

avec  $y$  la densité de proies,  $C_{2E}$  et  $C_{2V}$  deux constantes,  $\lambda$  un paramètre d'échelle de l'espace ( $= L/2v$ , où  $L^2$  est l'aire considérée et  $v$  la vitesse de recherche),  $h$  le temps de manipulation.

# Objectifs

Prise en compte de différentes sources de variabilité et d'erreur dans la modélisation de la réponse fonctionnelle

- stochasticité intrinsèque
- variabilité interindividuelle
- erreur d'observation

Etude de robustesse de l'estimation (algorithme SAEM) des paramètres du modèle par analyse de sensibilité.

Sélection de modèle : quelle forme pour la variance de la réponse fonctionnelle ?

# Formulation générique d'un modèle mixte

## Notations:

- $Y_i$ : vecteur des  $n_i$  observations pour l'individu  $i$
- $c_i$ : vecteur de caractéristiques observées sur l'individu  $i$  (covariables)

## Formulation hiérarchique:

- 1 Observations  $\rightarrow$  même modèle paramétrique conditionnellement aux  $\psi_i$

$$Y_i | \psi_i \sim p(\cdot | \psi_i), \quad i = 1, \dots, N$$

- 2 Paramètres individuels

$$\psi_i \sim p(\cdot | \theta; c_i), \quad i = 1, \dots, N$$

$\theta$ : vecteur de paramètres appelés **paramètres de population**

Remarque: les  $\psi_i$  sont non observés

# Formulation générique d'un modèle mixte

Usuellement, les paramètres individuels sont tels que:

$$\psi_i = h^{-1}(\phi_i)$$

où

$$\phi_i = C_i\beta + D_i\eta_i$$

$$\eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$$

et  $C_i = C(c_i)$ ,  $D_i = D(c_i)$ .



# Exemples : modèles de réponse fonctionnelle (M1)

$Y_{ij}$  : réponse fonctionnelle de la  $i$ e perdrix pour une densité de graines  $d_{ij}$

## 1 Observations

$$Y_{ij} | \psi_i = (\lambda_i, h_i) \sim \mathcal{N}(m_{ij}, s_{ij}^2)$$

où

$$m_{ij} = \frac{\sqrt{d_{ij}} - 1}{C_{2E}\lambda_i + h_i(\sqrt{d_{ij}} - 1)}$$

et

$$s_{ij}^2 = \frac{C_{2V} - C_{2E}^2}{\Delta} \frac{\lambda_i^2 (\sqrt{d_{ij}} - 1)}{(C_{2E}\lambda_i + h_i(\sqrt{d_{ij}} - 1))^3}$$

## 2 Paramètres individuels

$$\log(\lambda_i) \sim \mathcal{N}(m_\lambda, s_\lambda^2), \log(h_i) \sim \mathcal{N}(m_h, s_h^2)$$

# Analyse de sensibilité

Analyse exploratoire pour appréhender si/comment les performances de l'estimation dépendent des valeurs numériques des paramètres du modèle, via différents régimes espérance/variance de la réponse fonctionnelle.

## Questions

- Quels sont les paramètres qui influencent le plus les sorties du modèle, i.e. la moyenne et la variance de la réponse fonctionnelle ?
- Quels sont les paramètres qui influencent le plus l'estimation, i.e. le biais et la variance des estimateurs des paramètres ?

# Analyse de sensibilité du modèle

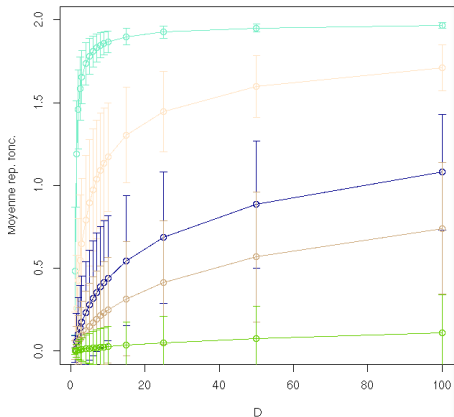
## Démarche

- On fait varier les paramètres du modèle, plan factoriel complet
  - ▶ séparément  $(m_\lambda, m_h)$ , modèle à effets fixes : **analyse univariée**
  - ▶ simultanément les 5 paramètres  $(m_\lambda, s_\lambda, m_h, s_h, \Delta)$ , modèle à effets mixtes : **analyse multivariée**
- On évalue la part dans la variabilité des sorties du modèle due à la variabilité des paramètres en entrée (calcul **d'indices de sensibilité**).

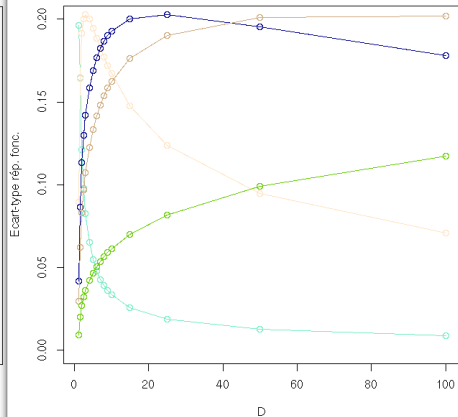
# Analyse de sensibilité univariée, modèle effets fixes, $\lambda$

$\lambda \in \{0.1, 1, 5, 10, 100\}$  (du bleu au vert)

Analyse sensibilité lambda

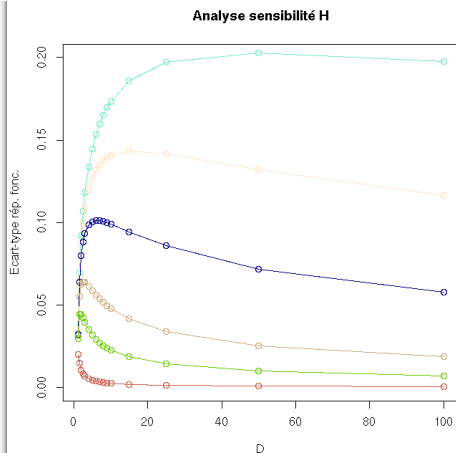
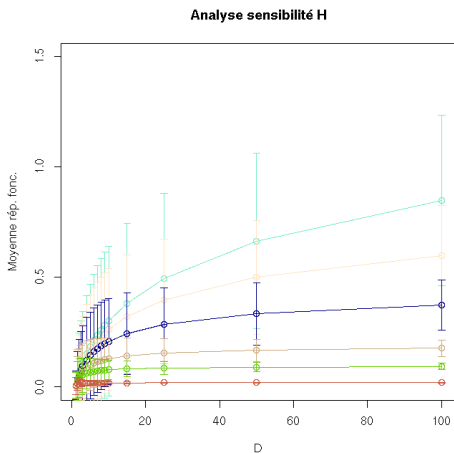


Analyse sensibilité lambda



# Analyse de sensibilité univariée, modèle effets fixes, $H$

$H \in \{0.5, 1, 2, 5, 10, 50\}$  (du bleu au rouge)

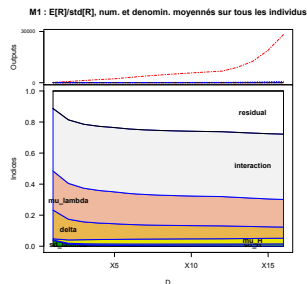
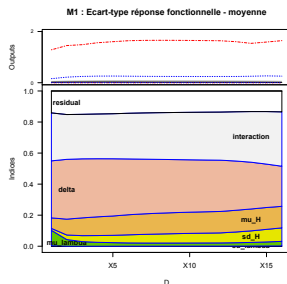
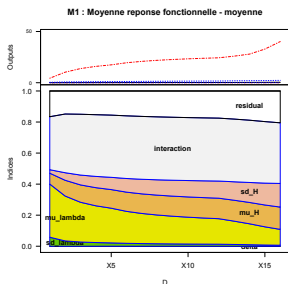


# Analyse de sensibilité multivariée, modèle à effets aléatoires

Les 5 paramètres ( $m_\lambda$ ,  $s_\lambda$ ,  $m_h$ ,  $s_h$ ,  $\Delta$ ) sont faits varier, selon un plan factoriel complet :

- $m_\lambda \in \{0.1, 1, 5, 10, 100\}$
- $s_\lambda \in \{0.1, 0.25, 0.5, 1, 2, 5\}$
- $m_H \in \{0.5, 1, 2, 5, 10, 50\}$
- $s_H \in \{0.1, 0.25, 0.5, 1, 2, 5\}$
- $\Delta \in \{0.1, 1, 10, 100\}$

# Analyse de sensibilité multivariée, modèle à effets aléatoires



- Paramètres les plus influents :  $m_\lambda$ ,  $\Delta$ ,  $m_h$
- Influence différente
  - ▶ sur la moyenne et l'écart-type de la réponse fonctionnelle
  - ▶ en fonction de la valeur de l'abscisse ( $D$ )

# Analyse de sensibilité sur l'estimation (par SAEM)

## Démarche

- On fait varier simultanément les paramètres  $(m_\lambda, s_\lambda, m_h, s_h)$  du modèle à effets mixtes et variance résiduelle : plan factoriel complet
  - ▶  $m_\lambda \in \{1, 5, 10, 20\}$
  - ▶  $cv_\lambda = s_\lambda / m_\lambda \in \{0.25, 0.5, 0.75, 1\}$
  - ▶  $m_H \in \{0.5, 2, 5, 10\}$
  - ▶  $cv_H = s_H / m_H \in \{0.25, 0.5, 0.75, 1\}$
- Pour chaque combinaison de valeurs de paramètres on simule des données et on estime les paramètres (estimateur ponctuel et écart-type - inverse de la matrice d'information de Fisher).
- On évalue l'impact de la variation des valeurs des paramètres sur le biais et la précision des estimateurs : analyse de variance.



# Analyse de sensibilité sur l'estimation : résultats

## Biais

	bias_m_lambda	bias_m_H	bias_s2_lambda	bias_s2_H
m_lambda	0.011	0.001	0.022	0.024
cv_lambda	0.015	0.005	0.008	0.008
m_H	0.003	0.028	0.261	0.048
cv_H	0.006	0.003	0.043	0.003
m_lambda:cv_lambda	0.029	0.030	0.017	0.022
m_lambda:m_H	0.025	0.033	0.029	0.104
m_lambda:cv_H	0.055	0.048	0.046	0.011
cv_lambda:m_H	0.018	0.030	0.052	0.042
cv_lambda:cv_H	0.028	0.015	0.015	0.045
m_H:cv_H	0.005	0.022	0.075	0.011

## Précision

	s_m_lambda	s_m_H	s_s2_lambda	s_s2_H
m_lambda	0.001	0.075	0.000	0.059
cv_lambda	0.768	0.000	0.742	0.000
m_H	0.112	0.054	0.099	0.028
cv_H	0.002	0.783	0.002	0.809
m_lambda:cv_lambda	0.002	0.000	0.000	0.000
m_lambda:m_H	0.008	0.032	0.002	0.027
m_lambda:cv_H	0.003	0.004	0.006	0.018
cv_lambda:m_H	0.063	0.002	0.081	0.001
cv_lambda:cv_H	0.001	0.001	0.001	0.002
m_H:cv_H	0.003	0.019	0.006	0.018

## RMSE = biais<sup>2</sup> + variance

	RMSE_m_lambda	RMSE_m_H	RMSE_s2_lambda	RMSE_s2_H
m_lambda	0.005	0.102	0.024	0.092
cv_lambda	0.511	0.001	0.292	0.002
m_H	0.191	0.088	0.229	0.074
cv_H	0.001	0.540	0.009	0.410
m_lambda:cv_lambda	0.009	0.007	0.029	0.005
m_lambda:m_H	0.010	0.079	0.047	0.121
m_lambda:cv_H	0.009	0.025	0.007	0.070
cv_lambda:m_H	0.106	0.007	0.169	0.006
cv_lambda:cv_H	0.005	0.001	0.007	0.003
m_H:cv_H	0.007	0.039	0.024	0.065

- **Biais pas impacté** : estimateurs asymptotiquement sans biais ( $N = 100$  ici)
- **Précision influencée par les CV** ( $s_\lambda/m_\lambda$  et  $s_H/m_H$ )
- **RMSE influencé par  $m_H$**  en plus des CV

# Sélection de modèles

## Questions

- Quelle forme pour la variance des réponses fonctionnelles?
- Effets fixes ou effets aléatoires?

## Exemples de modèles candidats sur (M1)

- Absence de variabilité inter-individuelle :

$$\lambda_i \equiv \lambda, \quad h_i \equiv h$$

- ▶  $\mathcal{M}_1^{(f)} : V(Y_{ij}) = \sigma^2$
- ▶  $\mathcal{M}_2^{(f)} : V(Y_{ij}) = s_j^2 + \sigma^2$
- ▶  $\mathcal{M}_3^{(f)} : V(Y_{ij}) = s_j^2$

- Variabilité inter-individuelle sur  $\lambda$  et  $h$  :

$$\log(\lambda_i) \sim \mathcal{N}(m_\lambda, s_\lambda^2), \quad \log(h_i) \sim \mathcal{N}(m_h, s_h^2)$$

- ▶  $\mathcal{M}_1^{(m)} : V(Y_{ij}|\psi_i) = \sigma^2$
- ▶  $\mathcal{M}_2^{(m)} : V(Y_{ij}|\psi_i) = s_{ij}^2 + \sigma^2$
- ▶  $\mathcal{M}_3^{(m)} : V(Y_{ij}|\psi_i) = s_{ij}^2$

# Sélection de modèles

Choix du modèle minimisant

$$BIC(\mathcal{M}) = -2 \ell(y; \theta_{\mathcal{M}}) + \dim(\theta_{\mathcal{M}}) \log(N)$$

où  $\ell(y; \theta_{\mathcal{M}})$  : log-vraisemblance (non explicite ici, calculée via des techniques d'importance sampling)

# Choix d'une forme pour la variance des réponses fonctionnelles

- $\lambda = 8$ ,  $h = 0.5$
- $d_{ij} \in \{1.1, 2, 3, 5, 10, 15, 25, 50, 100, 200, 400\}$
- $N = 100$
- Données simulées sous  $\mathcal{M}_1^{(f)}$ ,  $\mathcal{M}_2^{(f)}$  ou  $\mathcal{M}_3^{(f)}$  en faisant varier le niveau de variabilité résiduelle ( $\sigma$ )

# Choix d'une forme pour la variance des réponses fonctionnelles / modèles effets fixes

**Table:** Données simulées sous  $\mathcal{M}_1^{(f)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$
0.03	<b>89</b>	11	0
0.05	<b>97</b>	3	0
0.1	<b>94</b>	6	0
0.5	<b>47</b>	<b>53</b>	0
0.8	<b>49</b>	<b>51</b>	0
1	<b>47</b>	<b>53</b>	0

Remarque :  $s_j$  entre 0.03 et 0.2, moy 0.15

# Choix d'une forme pour la variance des réponses fonctionnelles / modèles effets fixes

**Table:** Données simulées sous  $\mathcal{M}_2^{(f)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$
0.03	0	<b>100</b>	0
0.05	0	<b>100</b>	0
0.1	12	<b>88</b>	0
0.5	44	<b>56</b>	0
0.8	50	<b>50</b>	0
1	50	<b>50</b>	0

Remarque :  $s_j$  entre 0.03 et 0.2, moy 0.15

# Choix d'une forme pour la variance des réponses fonctionnelles / modèles effets fixes

**Table:** Données simulées sous  $\mathcal{M}_3^{(f)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$
0	2	<b>98</b>

Remarque :  $s_j$  entre 0.03 et 0.2, moy 0.15

# Choix d'une forme pour la variance des réponses fonctionnelles / modèles mixtes

**Table:** Données simulées sous  $\mathcal{M}_1^{(m)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	<b>78</b>	0	22
0.05	<b>56</b>	23	21
0.1	<b>44</b>	<b>52</b>	4
0.5	<b>43</b>	<b>57</b>	0
0.8	<b>45</b>	<b>55</b>	0
1	<b>47</b>	<b>53</b>	0

Remarques :

- $s_j$  entre 0.03 et 0.2, moy 0.15
- Coefficients de variation pour les  $\lambda_i$  et  $h_i$  à 0.25  $\Rightarrow$  variance des effets aléatoires à 0.06



# Choix d'une forme pour la variance des réponses fonctionnelles / modèles mixtes

**Table:** Données simulées sous  $\mathcal{M}_2^{(m)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	31	<b>42</b>	27
0.05	43	<b>40</b>	17
0.1	41	<b>58</b>	1
0.5	43	<b>57</b>	0
0.8	45	<b>55</b>	0
1	47	<b>53</b>	0

Remarques :

- $s_j$  entre 0.03 et 0.2, moy 0.15
- Coefficients de variation pour les  $\lambda_i$  et  $h_i$  à 0.25  $\Rightarrow$  variance des effets aléatoires à 0.06

# Choix d'une forme pour la variance des réponses fonctionnelles / modèles mixtes

**Table:** Données simulées sous  $\mathcal{M}_3^{(m)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
14	19	67

Remarques :

- $s_j$  entre 0.03 et 0.2, moy 0.15
- Coefficients de variation pour les  $\lambda_i$  et  $h_i$  à 0.25  $\Rightarrow$  variance des effets aléatoires à 0.06

# Identification de toutes les sources de variabilité

**Table:** Données simulées sous  $\mathcal{M}_2^{(f)}$  : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	0	100	0	0	0	0
0.05	0	100	0	0	0	0
0.08	2	98	0	0	0	0
0.10	16	84	0	0	0	0

# Identification de toutes les sources de variabilité

**Table:** Données simulées sous  $\mathcal{M}_2^{(m)}$ , cv à 0.25 pour les ea : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	0	<b>89</b>	0	0	2	9
0.05	0	<b>94</b>	0	0	0	6
0.08	0	<b>98</b>	0	0	0	2
0.10	0	<b>98</b>	0	0	0	2

**Table:** Données simulées sous  $\mathcal{M}_2^{(m)}$ , cv à 0.50 pour les ea : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	0	<b>74</b>	0	2	5	10
0.05	0	11	0	36	<b>39</b>	14
0.08	0	<b>94</b>	0	0	4	2
0.10	0	<b>91</b>	0	0	5	4

# Identification de toutes les sources de variabilité

**Table:** Données simulées sous  $\mathcal{M}_1^{(m)}$ , cv à 0.25 pour les ea : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

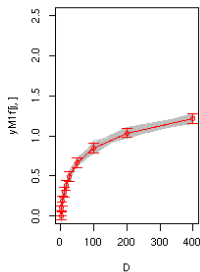
$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	0	7	0	73	1	19
0.05	0	89	0	0	2	9
0.08	0	99	0	0	0	1
0.10	0	99	0	0	0	1

**Table:** Données simulées sous  $\mathcal{M}_1^{(m)}$ , cv à 0.50 pour les ea : sur 100 répétitions, nombre de cas où chaque modèle est choisi par BIC.

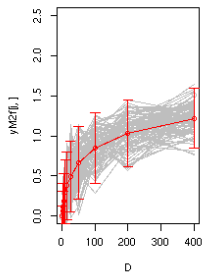
$\sigma$	$\mathcal{M}_1^{(f)}$	$\mathcal{M}_2^{(f)}$	$\mathcal{M}_3^{(f)}$	$\mathcal{M}_1^{(m)}$	$\mathcal{M}_2^{(m)}$	$\mathcal{M}_3^{(m)}$
0.03	0	0	0	80	4	16
0.05	0	0	0	52	27	21
0.08	0	48	0	8	34	10
0.10	0	80	0	0	16	4

# Trajectoires simulées, $\sigma = 0.03$ , $CV_{ea} = 0.25$

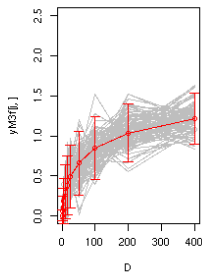
M1f - EF, var résid.



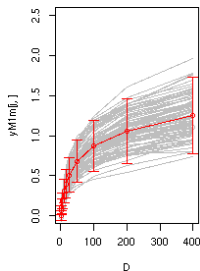
M2f - EF, var struct. & résid.



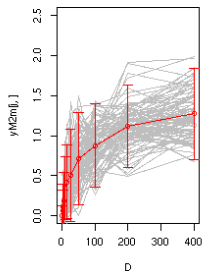
M3f - EF, var struct.



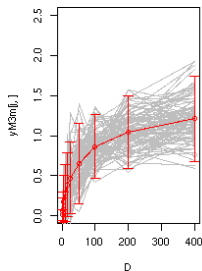
M1m - EM, var résid.



M2m - EM, var struct. & résid.

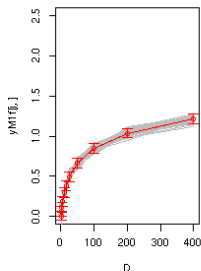


M3m - EM, var struct.

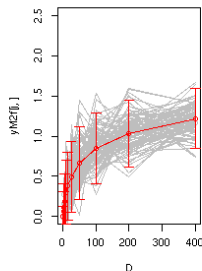


# Trajectoires simulées, $\sigma = 0.03$ , $CV_{ea} = 0.5$

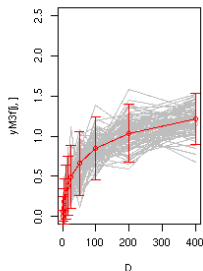
M1f - EF, var résid.



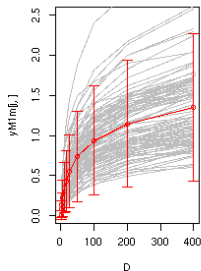
M2f - EF, var struct. & résid.



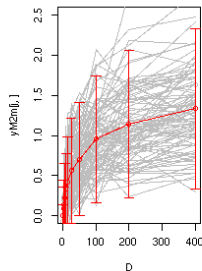
M3f - EF, var struct.



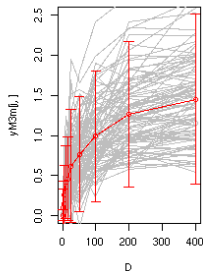
M1m - EM, var résid.



M2m - EM, var struct. & résid.

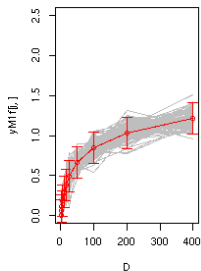


M3m - EM, var struct.

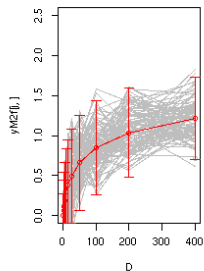


# Trajectoires simulées, $\sigma = 0.1$ , $CV_{ea} = 0.25$

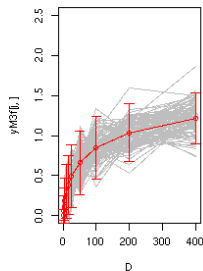
M1f - EF, var résid.



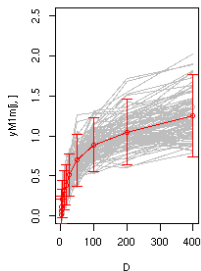
M2f - EF, var struct. & résid.



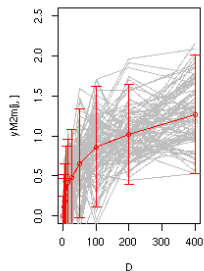
M3f - EF, var struct.



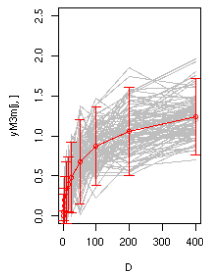
M1m - EM, var résid.



M2m - EM, var struct. & résid.



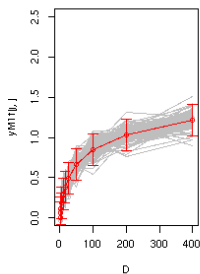
M3m - EM, var struct.



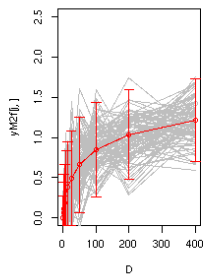


# Trajectoires simulées, $\sigma = 0.1$ , $CV_{ea} = 0.5$

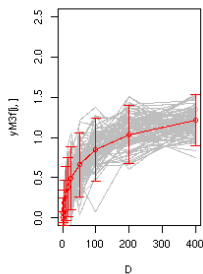
M1f - EF, var résid.



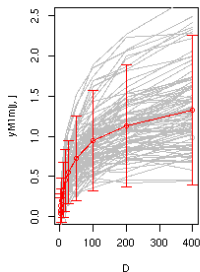
M2f - EF, var struct. & résid.



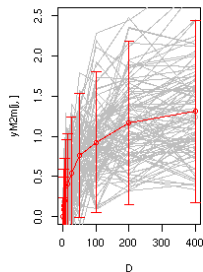
M3f - EF, var struct.



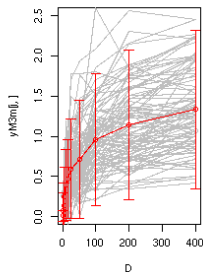
M1m - EM, var résid.



M2m - EM, var struct. & résid.



M3m - EM, var struct.



# Conclusion sélection de modèles

## Sensibilité au rapport signal/bruit

- on sait identifier la bonne forme de variance si la résiduelle (le bruit) n'est pas trop grande (la variabilité due au mécanisme de réponse fonctionnelle est "noyée" dans le bruit pour les plus grandes valeurs de  $\sigma$ )
- on sait en plus repérer la variabilité inter-individuelle si elle est suffisamment forte au regard de la résiduelle (s'il y a trop de bruit et si la variabilité inter-individuelle est trop faible, on ne choisit pas le bon modèle)