



➤ La « sampled posterior p-value »: un autre outil Bayésien de diagnostic de modèles aux propriétés mathématiques intéressantes

Frédéric Gosselin¹, Thierno Diallo¹, Didier Chauveau²

¹UR EFNO, INRAE, Nogent-sur-Vernisson

²Institut Denis Poisson, Université d'Orléans

frederic.gosselin@inrae.fr

➤ La critique des modèles

- Reconnu par les statisticiens appliqués:

Box (1980) *JRSS A*: une des deux étapes principales

Cox (1997) *Int. Stat. Rev.*: un des dix outils principaux

McCullagh et Nelder (1989) *Chapman*: un des trois principes



➤ La critique des modèles

- Reconnu par les statisticiens appliqués:

Box (1980) *JRSS A*: une des deux étapes principales

Cox (1997) *Int. Stat. Rev.*: un des dix outils principaux

McCullagh et Nelder (1989) *Chapman*: un des trois principes

- Aussi défendu par des écologues:

Hilborn & Mangel (1997) *Princeton Univ. Press*: un des quatre outils principaux du détective écologique



➤ La critique des modèles

- Un de ces outils: les « (internal) goodness-of-fit (GOF) p-values »:

Ingrédients:

* un point de vue : une fonction de « discrédance » $T(x, \theta)$

dépendant de données \mathcal{X} et des paramètres du modèle statistique θ

* une loi $m(., \theta)$ pour générer des jeux de données « répliqués » X

Recette:

* générez des données « répliquées » X suivant la loi $m(., \theta)$

* mélangez tout cela intimement pour obtenir la p-value

$$P^{m(.)} [T(X, \theta) > T(x_{obs}, \theta)]$$



➤ La critique des modèles

- Un de ces outils: les « (internal) goodness-of-fit (GOF) p-values » :

Interprétation:

$$P^{m(\cdot)}[T(X, \theta) > T(x_{obs}, \theta)] = 0.4$$

Est-ce compatible avec le modèle ou pas?

Est-ce surprenant ou pas ?

➤ La critique des modèles

- Un de ces outils: les « (internal) goodness-of-fit p-values »:

Interprétation:

$$P^{m(\cdot)}[T(X, \theta) > T(x_{obs}, \theta)] = 0.4$$

Est-ce compatible avec le modèle ou pas?

Est-ce surprenant ou pas ?

↳ Besoin de méthodes pour lesquelles on connaît – à peu près – la distribution de proba de

$$P^{m(\cdot)}[T(X, \theta) > T(x_{obs}, \theta)]$$

quand on répète la méthode sur des jeux de données x_{obs} générés aléatoirement avec le bon modèle

➤ (Internal) goodness-of-fit p-values
Points abordés

1. Résultats mathématiques

2. Codage

3. Résultats de simulations

4. Mise en œuvre en écologie

5. Discussion



➤ Résultats mathématiques

* Parmi de nombreuses GOF p-values disponibles...

Table 1. Reference Densities for the Various Candidate p Values

Method	Reference density
Plug-in (p_{plug})	$m_{\text{plug}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = f(\mathbf{x}; \hat{\theta}_{\text{obs}})$
Prior predictive (p_{prior})	$m_{\text{prior}}(\mathbf{x}) = \int f(\mathbf{x}; \theta)\pi(\theta) d\theta$
Posterior predictive (p_{post})	$m_{\text{post}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = \int f(\mathbf{x}; \theta)\pi_{\text{post}}(\theta \mathbf{x}_{\text{obs}}) d\theta$
Partial posterior predictive (p_{ppost})	$m_{\text{ppost}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = \int f(\mathbf{x}; \theta)\pi_{\text{ppost}}(\theta \mathbf{x}_{\text{obs}}) d\theta$
Conditional predictive (p_{cpred})	$m_{\text{cpred}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = \int f(\mathbf{x} \hat{\theta}_{\text{cMLE[s]}}; \theta)\pi_{\text{cpred}}(\theta \mathbf{x}_{\text{obs}}) d\theta$
Conditional plug-in (p_{cplug})	$m_{\text{cplug}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = f(\mathbf{x}; \hat{\theta}_{\text{cMLE,obs}})$
Discrepancy (p_{dis})	$m_{\text{dis}}(\mathbf{x}, \theta \mathbf{x}_{\text{obs}}) = f(\mathbf{x}; \theta)\pi_{\text{post}}(\theta \mathbf{x}_{\text{obs}})$

Robins et al. (2000) JASA



➤ Résultats mathématiques

... les deux plus fréquemment utilisées

Table 1. Reference Densities for the Various Candidate p Values

Method	Reference density
Plug-in (p_{plug})	$m_{\text{plug}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = f(\mathbf{x}; \hat{\theta}_{\text{obs}})$
Prior predictive (p_{prior})	$m_{\text{prior}}(\mathbf{x}) = \int f(\mathbf{x}; \theta)\pi(\theta) d\theta$
Posterior predictive (p_{post})	$m_{\text{post}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = \int f(\mathbf{x}; \theta)\pi_{\text{post}}(\theta \mathbf{x}_{\text{obs}}) d\theta$
Partial posterior predictive (p_{ppost})	$m_{\text{ppost}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = \int f(\mathbf{x}; \theta)\pi_{\text{ppost}}(\theta \mathbf{x}_{\text{obs}}) d\theta$
Conditional predictive (p_{cpred})	$m_{\text{cpred}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = \int f(\mathbf{x} \hat{\theta}_{\text{cMLE[s]}}; \theta)\pi_{\text{cpred}}(\theta \mathbf{x}_{\text{obs}}) d\theta$
Conditional plug-in (p_{cplug})	$m_{\text{cplug}}(\mathbf{x} \mathbf{x}_{\text{obs}}) = f(\mathbf{x}; \hat{\theta}_{\text{cMLE,obs}})$
Discrepancy (p_{dis})	$m_{\text{dis}}(\mathbf{x}, \theta \mathbf{x}_{\text{obs}}) = f(\mathbf{x}; \theta)\pi_{\text{post}}(\theta \mathbf{x}_{\text{obs}})$

Robins et al. (2000) JASA



➤ Résultats mathématiques

Des GOF p-values au comportement connu **quand l'espérance de la fonction de discrédance $T(x)$ ne dépend asymptotiquement pas de θ**

- prior trop dépendante du prior
- posterior predictive, plug-in, partial predictive et conditional predictive *asymptotiquement fréquentistes* (i.e. ont asymptotiquement une distribution uniforme si le modèle est le bon)

Robins *et al.* (2000) *JASA*



➤ Résultats mathématiques

Des GOF p-values au comportement connu **quand la fonction de discrédance $T(x)$ ne dépend pas de θ mais son espérance dépend asymptotiquement de θ** :

- prior trop dépendante du ... prior
- posterior predictive et plug-in conservatives
- partial predictive et conditional predictive asymptotiquement fréquentistes

Robins *et al.* (2000) *JASA*



➤ Résultats mathématiques

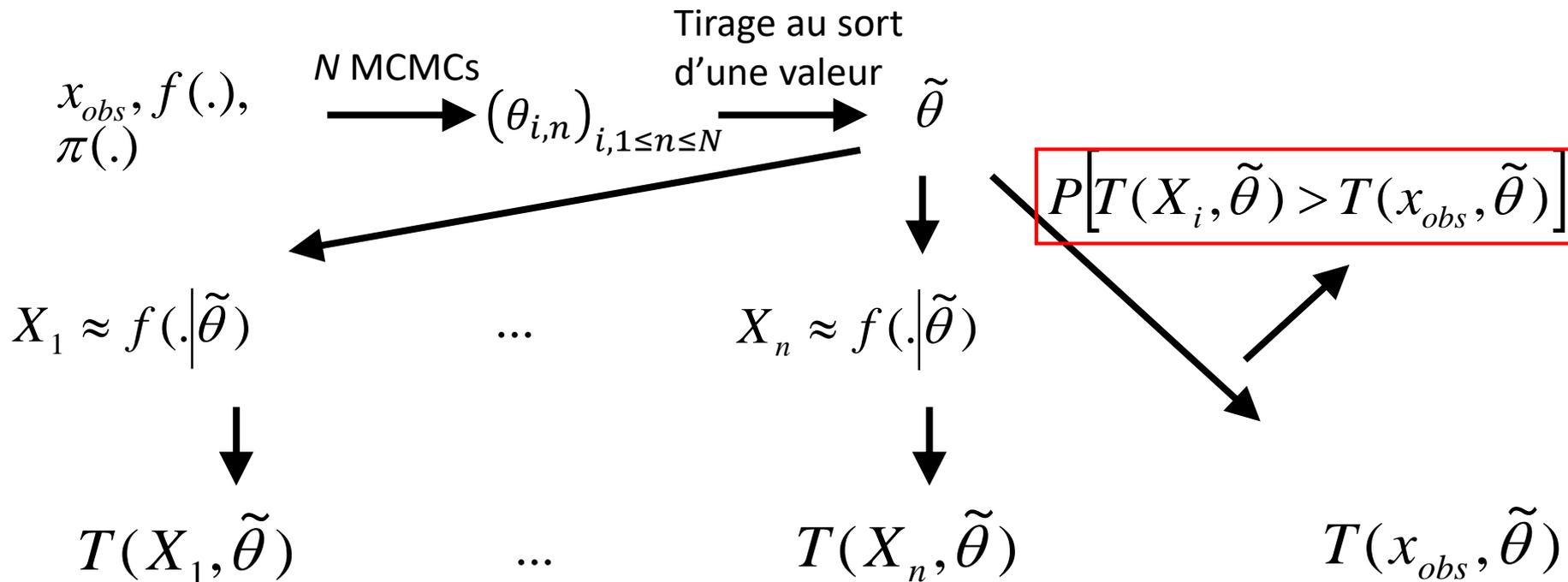
Des GOF p-values au comportement peu connu **quand la fonction de discrédance $T(x, \theta)$ dépend de θ** :

- contorsions nécessaires, qui ne seront pas toujours souhaitables (Bayarri & Castellanos 2007 *Stat. Science*)
- on passe outre (Sinharay & Stern 2003 *JSPI*)



➤ Résultats mathématiques

* Inspiré par Johnson (2004 *Ann. Stat.* et 2007 *Bayes. Anal.*), proposition d'utiliser la « sampled posterior p-value » (Gosselin 2011 *Plos One*)



➤ Résultats mathématiques

* La « sampled posterior p-value » a la propriété fréquentiste pour toute fonction de discrédance $T(x, \theta)$ quand le prior du modèle statistique correspond au prior utilisé pour générer les données (Gosselin 2011 *Plos One*, Diallo et al. *In Prep.*)

NB: quand le prior du modèle statistique ne correspond pas au prior utilisé pour générer les données, la démonstration de son caractère asymptotiquement fréquentiste (Gosselin 2011 *Plos One*) est à étayer



➤ Résultats mathématiques

* Extension de la sampled posterior p-value ("extended sampled posterior p-value") aux modèles hiérarchiques : tirage au sort non seulement des données observées X_i mais aussi d'effets aléatoires ou variables latentes ξ_i



➤ Résultats mathématiques

* La « sampled posterior p-value » **étendue** a la propriété d'être fréquentiste pour toute fonction de discrédance $T(x, \theta)$ quand le prior du modèle statistique correspond au prior utilisé pour générer les données (Diallo et al. *In Prep.*)

> Codage

* Le **coût numérique** des `sampled posterior` p-values **est plus faible** que leurs équivalents `posterior predictive` (Zhang 2014 *Comput. Stat. Data Anal.*, Diallo et al. *In Prep*)

* Par contre, intégration aux codes MCMC moins naturelle avec `sampled` qu'avec `posterior`. **Proposition de code** pour faire ces calculs à la fin des MCMC *nimble* lancés avec le package `runMCMCbtadjust` ([code DIC & sampled posterior](#)) (à adapter à chaque cas de figure – notamment en terme de fonction de discrédance)

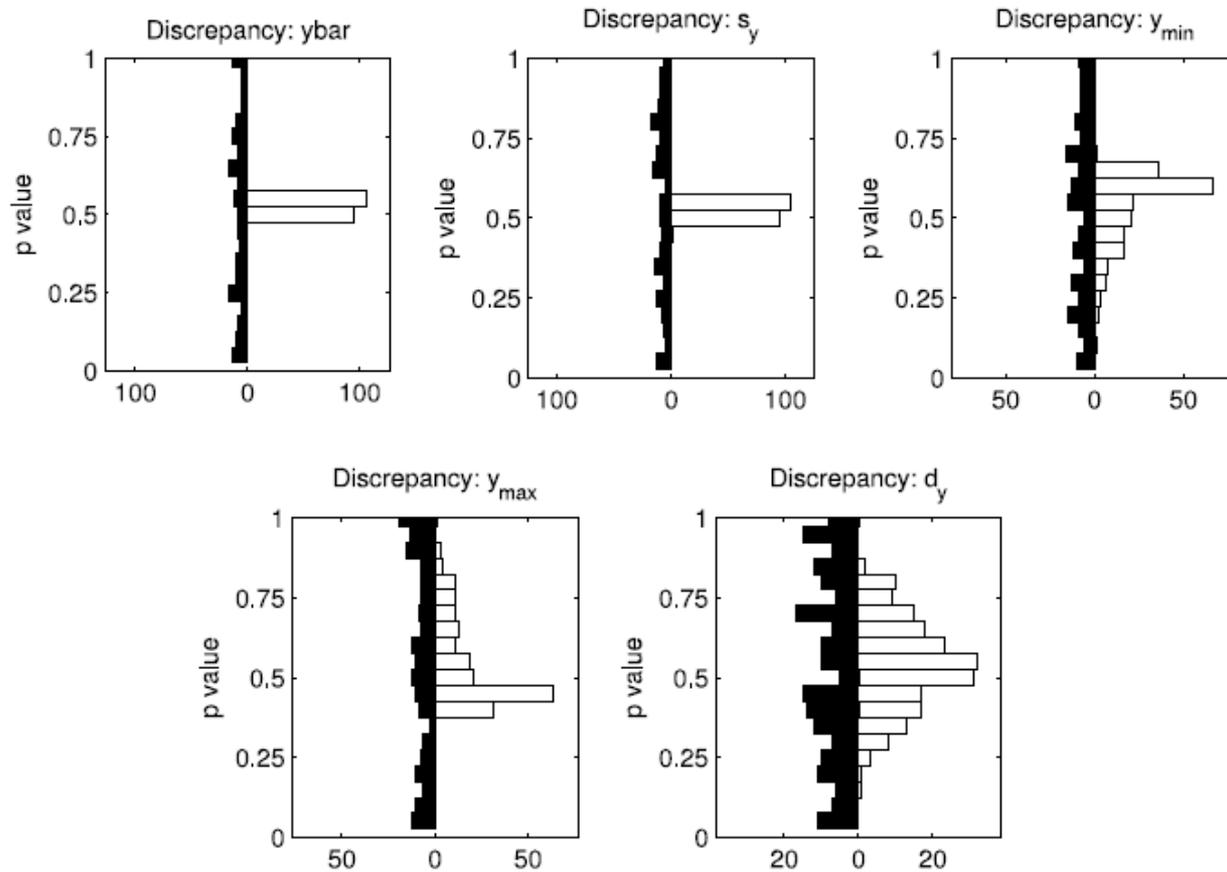


➤ Résultats de simulations

- * Premier exemple issu de Zhang (2014) *Comput. Stat. Data Anal.*
- * Mêmes modèles statistiques hiérarchiques **double-gaussiens** que Sinharay & Stern (2003)
- * Le **modèle faux** génère les effets aléatoires avec une **distribution exponentielle** (alors qu'une distribution gaussienne est utilisée dans le modèle statistique)
- * Diagnostic des données observées (avec différentes fonctions de discrédance)
 - * Autres exemples dans Gosselin (2011)



➤ Bonne vraisemblance



) CSDA

Sampled posterior → spp ppp ← Posterior predictive

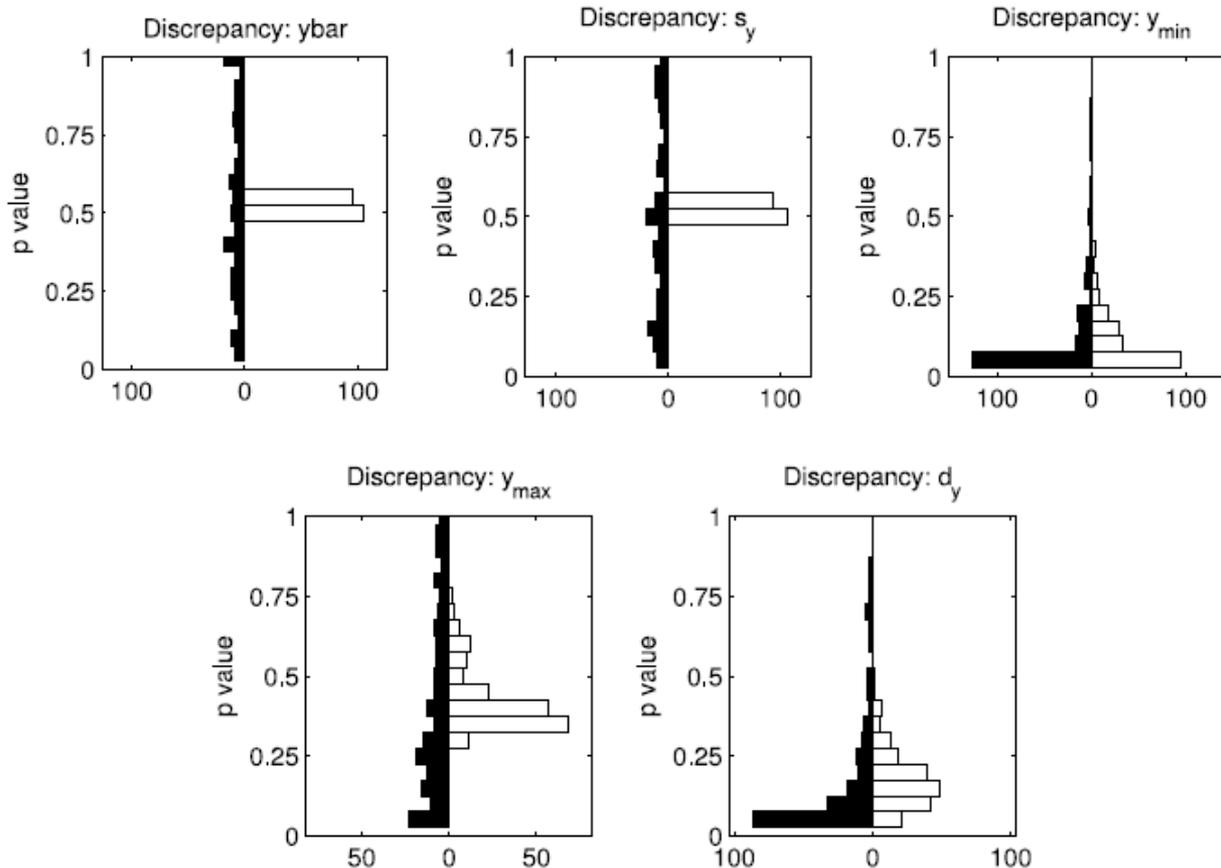


INRAE

Sampled posterior p-values

12 décembre 2024 / Applibugs / Université Paris Cité / F. Gosselin, Th. Diallo, D. Chauveau

➤ Mauvaise vraisemblance



Zhang (2014) CSDA

Sampled posterior \longrightarrow spp ppp \longleftarrow Posterior predictive

➤ Résultats de simulation

* Codage des (**classical**) posterior (resp. sampled) predictive p-value (**c**ppp, resp. **c**spp) – comme dans Zhang (2014) – et des **extended** posterior (resp. sampled) predictive p-value (**e**ppp, resp. **e**spp) sur les mêmes modèles (Diallo et al., *In Prep.*)



➤ Résultats de simulation

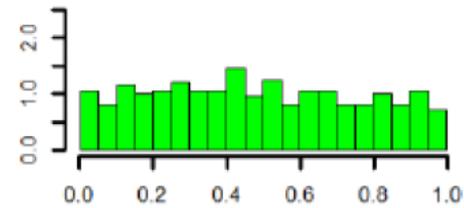
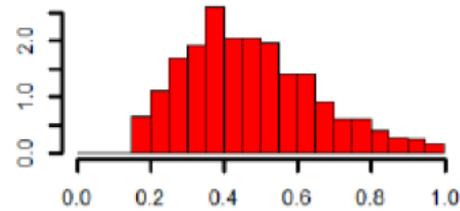
$$T(\xi) = \min(\xi)$$

■ *eppp*-values

■ *espp*-values

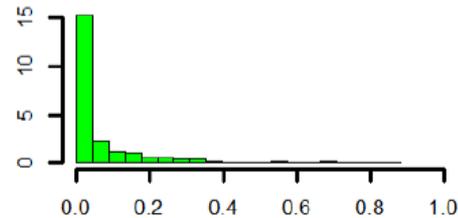
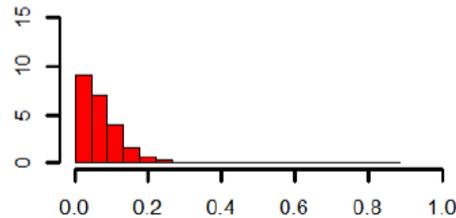
- (P1)

True Model



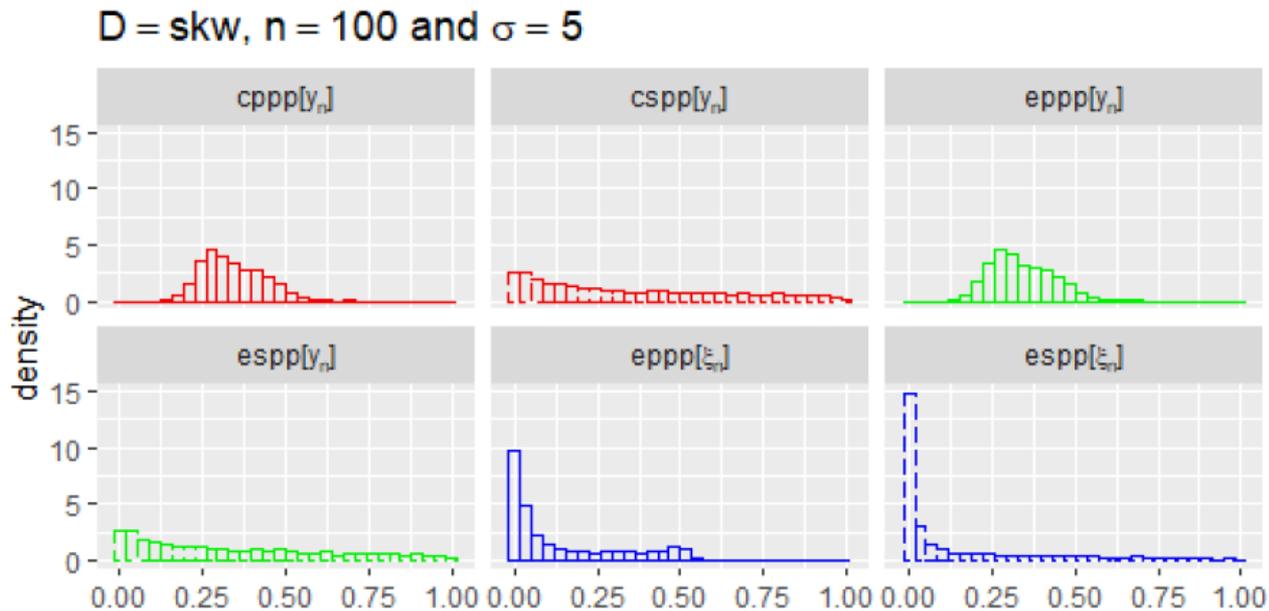
- (P2)

False Model



➤ Résultats de simulation

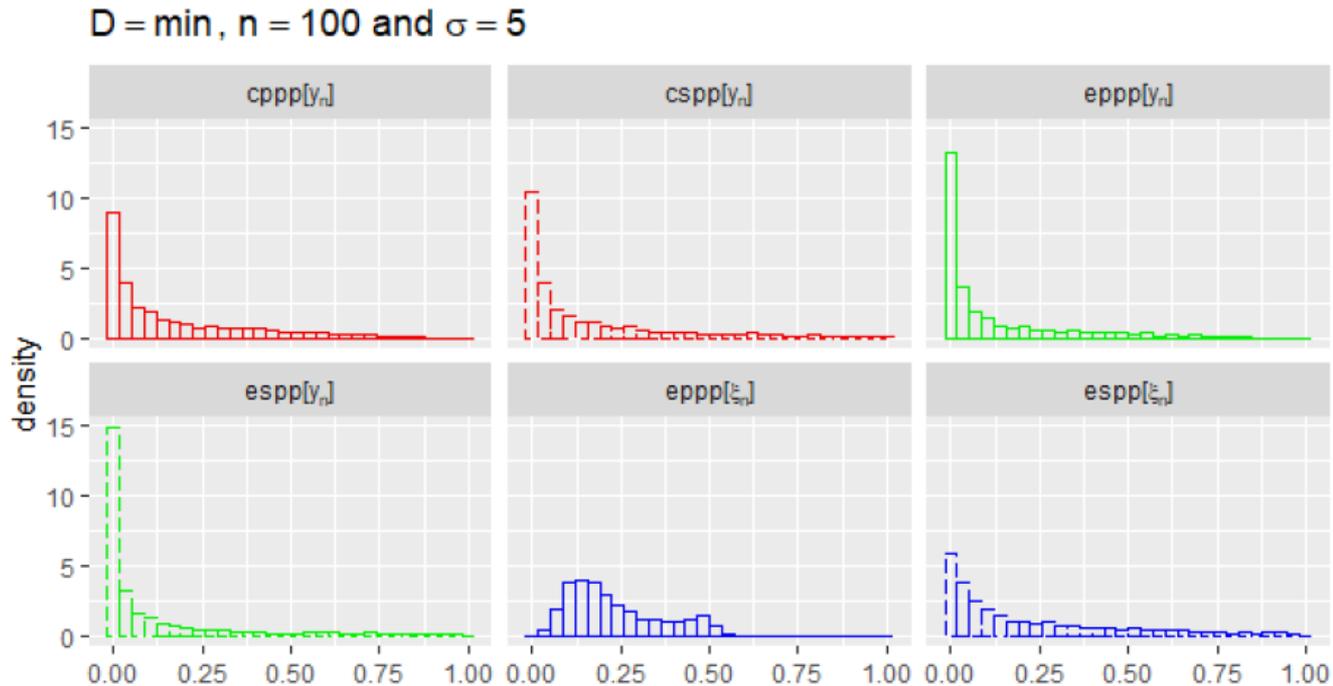
False model



ppp = posterior predictive p-value; spp = sampled posterior p-value
c = classical (ξ n'est pas répliqué); e = extended (ξ est répliqué)

➤ Résultats de simulation

False model



ppp = posterior predictive p-value; *spp* = sampled posterior p-value
c = classical (ξ n'est pas répliqué); *e* = extended (ξ est répliqué)

INRAE

Sampled posterior p-values

12 décembre 2024 / Applibugs / Université Paris Cité / F. Gosselin, Th. Diallo, D. Chauveau

➤ Résultats de simulation

↳ Résultats souvent mauvais pour $p_{\text{plug-in}}$ et $p_{\text{posterior}}$

↳ Comportement presque fréquentiste pour p_{sp}

↳ p_{sp} plus puissante que $p_{\text{plug-in}}$ et $p_{\text{posterior}}$

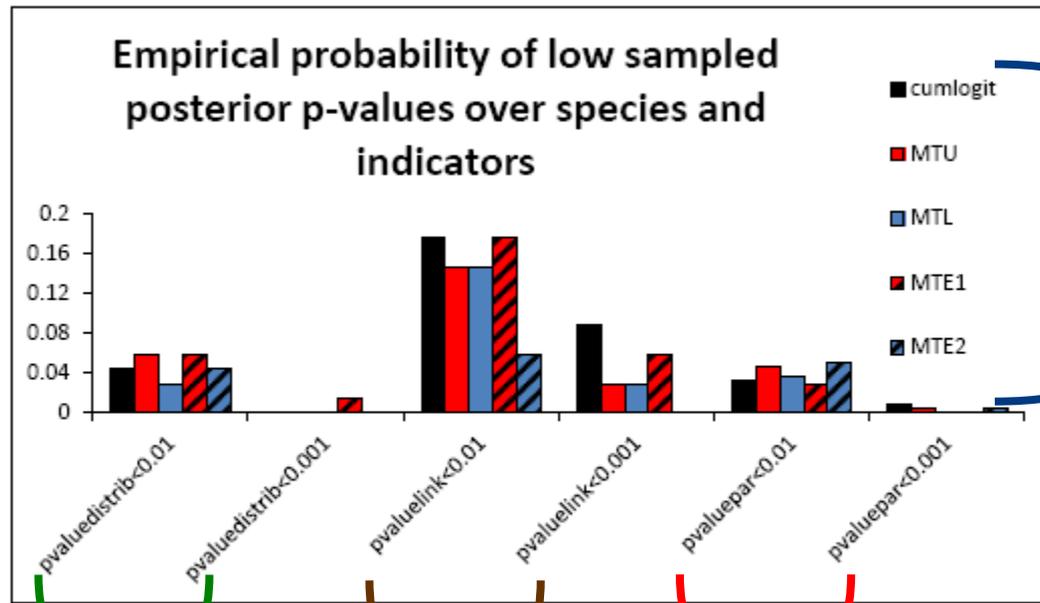
↳ Comportement de p_{sp} dépendant des priors et des fonctions de disrépance

➤ Mise en œuvre en écologie (I)

- * Exemple issu de [Herpigny et Gosselin \(2015\) *Ecol. Info.*](#).
- * Comparaison de différentes formes probabilistes pour modéliser des données catégorielles ordonnées (ici: classes Braun-Blanquet pour données floristiques)
- * Modèle écologique impliquant des variables abiotiques (climat, sol...) et biotiques (couvert arbre)
- * Modèle proche des GAMs, sauf pour la partie probabiliste.
- * Estimation par un MCMC adaptatif de type parallel tempering.



➤ Mise en œuvre en écologie (I)



Différentes formes probabilistes des modèles

Herpigny & Gosselin (2015) *Ecol. Info.*

Distribution résidus normalisés

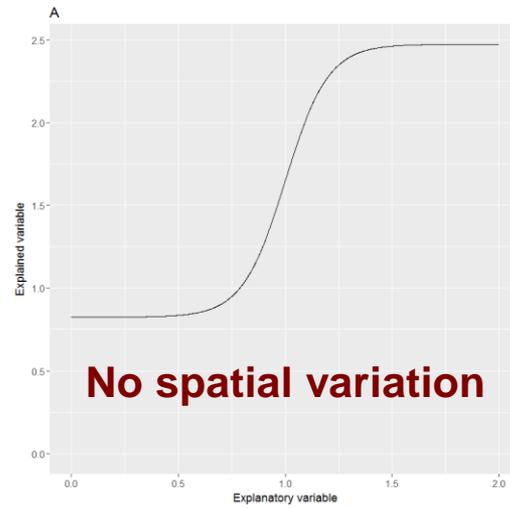
Diagnostic fonction lien (« linéarité »)

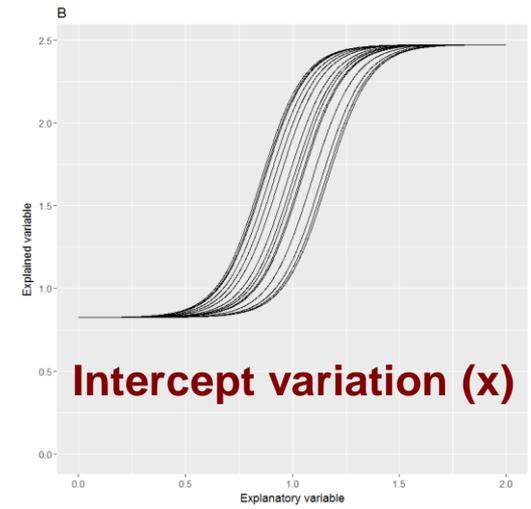
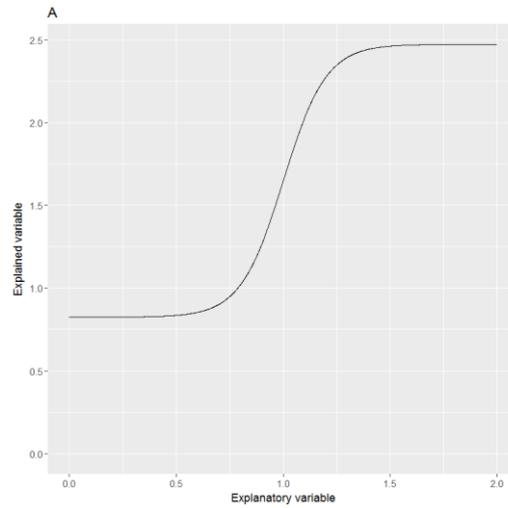
Diagnostic lien aux variables écologiques

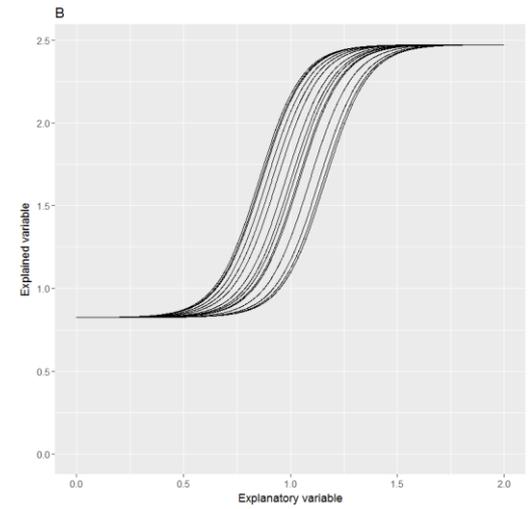
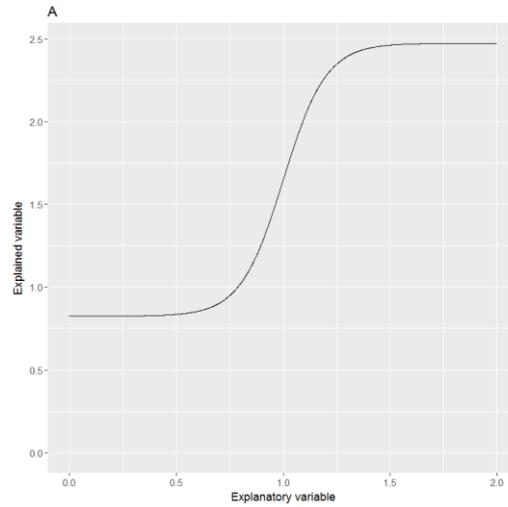
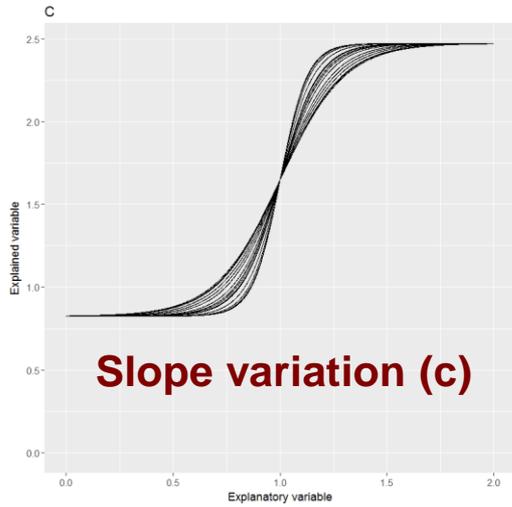
➤ Mise en œuvre en écologie (II)

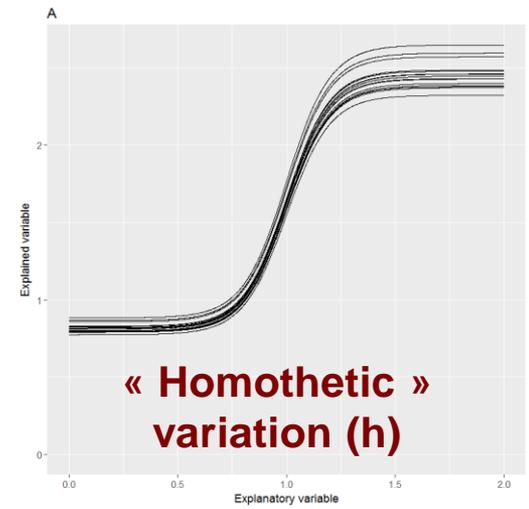
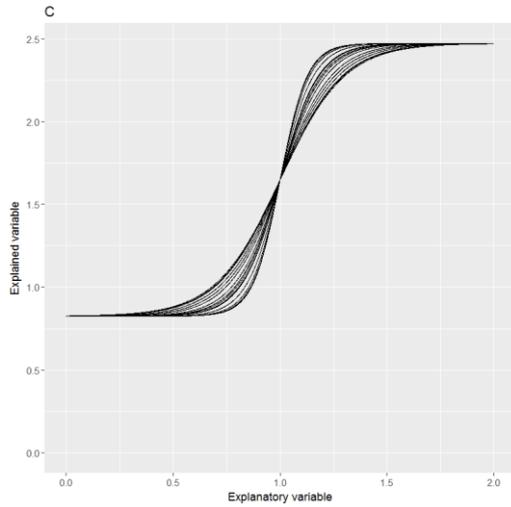
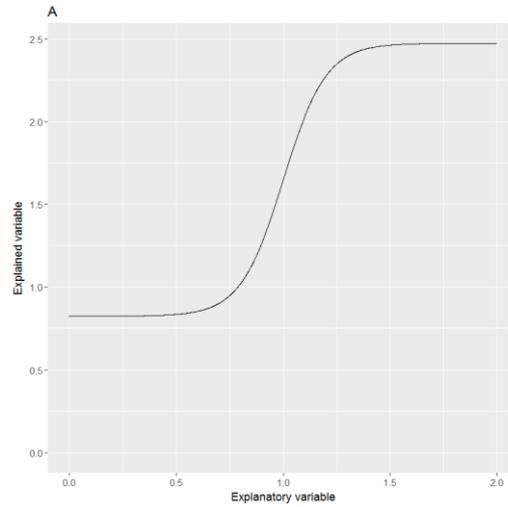
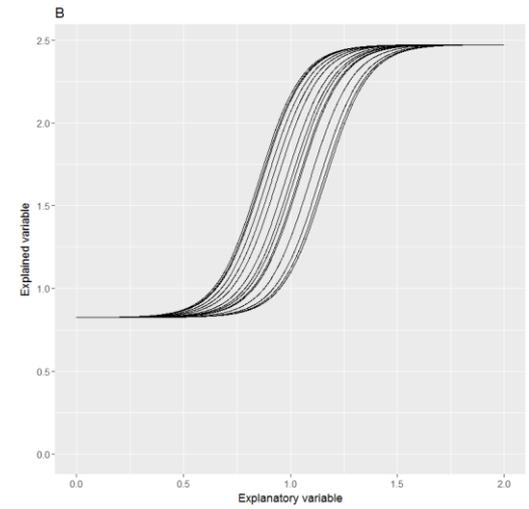
- * Exemple issu de [Godeau et al. \(2020\) *Forest Ecology & Management*](#).
- * Comparaison de différentes fonctions de lien (dont sigmoïdes) pour modéliser la relation entre volume de bois mort et richesse spécifique de Coléoptères saproxyliques
- * 589 placettes dans 33 forêts en France
- * Incorporation ou non de l'effet aléatoire forêt sur les différents paramètres des fonctions de lien











INRAE

Sampled posterior p-values

12 décembre 2024 / Applibugs / Université Paris Cité / F. Gosselin, Th. Diallo, D. Chauveau

➤ Mise en œuvre en écologie (II)

- * Exemple issu de [Godeau et al. \(2020\) *Forest Ecology & Management*](#).
- * Comparaison de différentes fonctions de lien (dont sigmoïdes) pour modéliser la relation entre volume de bois mort et richesse spécifique de Coléoptères saproxyliques
- * 589 placettes dans 33 forêts en France
- * Incorporation ou non de l'effet aléatoire forêt sur les différents paramètres des fonctions de lien
- * Estimation par MCMC avec le package *nimble*
- * Utilisation de sampled posterior p-values avec différentes fonctions de discrédance pour diagnostiquer les modèles (en parallèle d'outils de comparaisons de modèles)



Axes: proportions de fois – pour différents tirages au sort des paramètres – où les spp sont inférieures à 0.005

logis4p 2hr (best)
logis5p 2hr (best)
logis4p 3hra (full)
logis5p 4hrae (full)

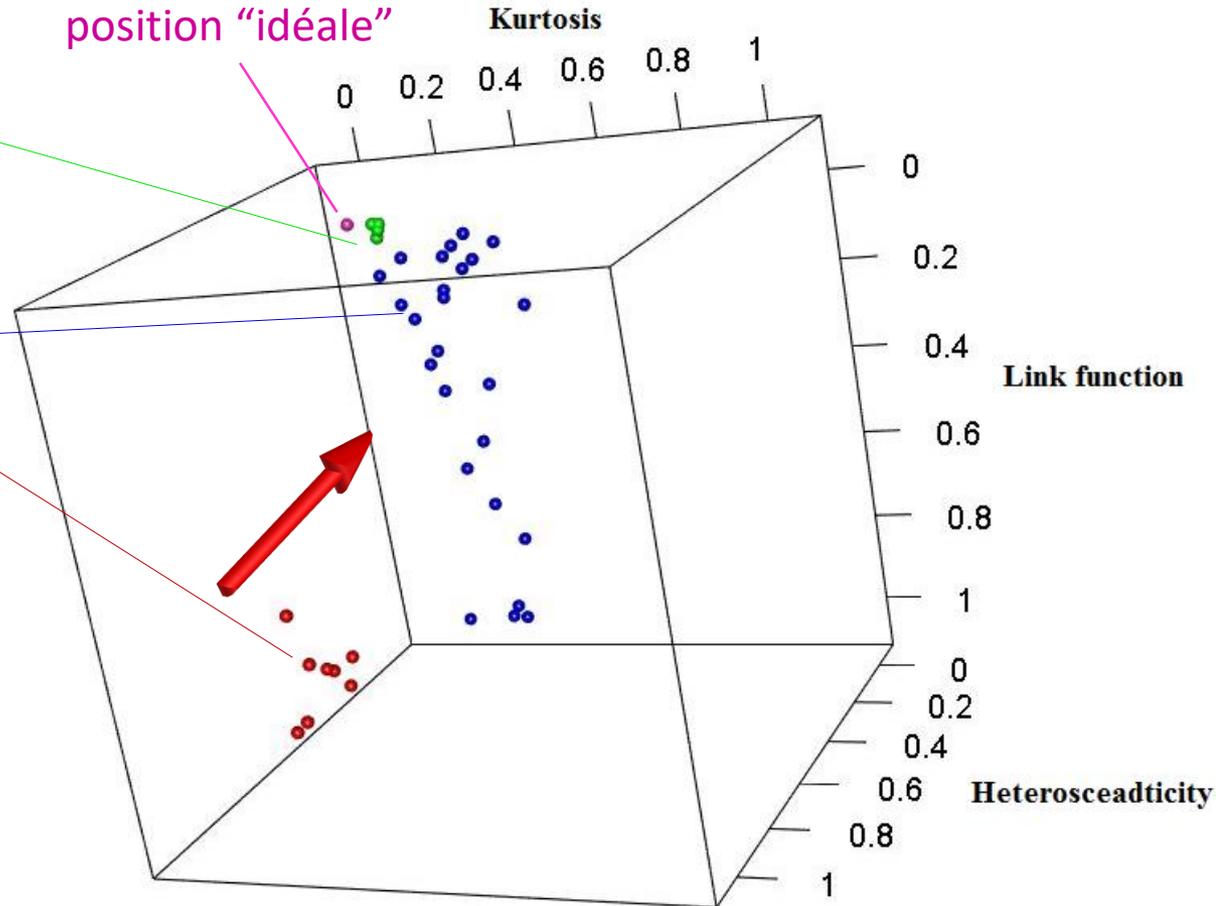
Autres modèles avec au moins un effet aléatoire

no.re

Ajout d'au moins 1 effet aléatoire → élimine les problèmes d'hétéroscédasticité

Choix d'effets aléatoires appropriés (best ou full) → élimine en plus les problèmes de kurtosis et de la fonction de lien

position "idéale"



➤ Discussion

* p_{sp} et le débat sur la double utilisation des données

↳ utilise deux fois les données et devrait donc succomber sous les mêmes critiques que p_{post}

↳ pourtant, ça marche : pourquoi?

“As we are working on sampled data to fit statistical models, we should also agree to work on sampled parameters to criticize the model. Indeed, this double sampling allowed us to make the roles of data and parameters symmetrical, enabling us to prove our mathematical results. ... The problem lies less in that a GOF p-value uses data twice, but more in how it uses the data twice ” (Gosselin 2011 Plos One)



> Discussion

* Sampled posterior p-value (p_{sp}) a priori la seule p-value numériquement "légère" fréquentiste pour toute fonction de discrédance $T(x, \theta)$ quand les deux priors correspondent

↳ capacité à s'adapter à chaque modèle et à chaque fonction de discrédance (p-value « omnibus »)

↳ capacité à aller jauger les différentes composantes (e.g. d'un modèle hiérarchique)



➤ Discussion

* Limite associée à la notion de significativité statistique:

*une p-value faible ne signifie pas qu'il y aura un impact fort pour
l'estimation / la décision*



➤ Discussion

* Limite associée à la notion de significativité statistique:

une p-value faible ne signifie pas qu'il y aura un impact fort pour l'estimation / la décision

* p_{sp} a un coût numérique plus faible que p_{post}

* p_{sp} est un peu moins naturelle à intégrer à un MCMC que p_{post} , mais code disponible pour la calculer à la fin d'un MCMC estimé avec *nimble* sous *runMCMCbtadjust*.





Remerciements: F. Mortier (ANR Gambas)

Contact: frederic.gosselin@inrae.fr