A Multinomial–Dirichlet Mixture Approach to Rethinking Conditional Dependence Modelling in Latent Class Models

Thibaut Lurier; Pierre Latouche, Pierre Druilhet, Myriam Garrido

UMR EpiA, VetAgro Sup; LMBP, CNRS, UCA

Evaluation de méthodes diagnostiques

- Considérons une maladie M et un test diagnostique T
- On cherche à estimer la Sensibilité (Se) et la spécificité (Sp) du test :

$$Se = P\left(T^{+} \middle| M^{+}\right) \qquad Sp = P\left(T^{-} \middle| M^{-}\right)$$

- Par rapport à un Gold Standard de Se et Sp = 100%

$$\hat{S}e = \frac{n_{T^{+}}}{n_{M^{+}}} \qquad \hat{S}p = \frac{n_{T^{-}}}{n_{M^{-}}}$$

Gold Standard?

- Généralement :
  - le Gold Standard **n'existe pas!**
  - Il est trop cher ou trop invasif
- Autres possibilités :
  - Utiliser un **test de référence de Se et Sp connues**

  $\Rightarrow$ Assez rare en médecine vétérinaire
  - Utiliser un **Gold standard imparfait** en tant que Gold standard

  $\Rightarrow$ Estimations biaisées
  - Utiliser des **patients de statuts connus** et souvent caricaturaux

  $\Rightarrow$ Estimations biaisées (surestimation des Se et Sp)

Modèle à classe latente *(Hui and Walter, 1980)*

- Estimation des caractéristiques de **plusieurs tests binaires** :
  - Considérons 2 tests appliqués à une population dont la prévalence de la maladie est P

	Test 2 positif	Test 2 négatif
Test 1 positif	$p_{11} = Se_1 \times Se_2 \times P + (1 - Sp_1) \times (1 - Sp_2) \times (1 - P)$	$p_{10} = Se_1 \times (1 - Se_2) \times P + (1 - Sp_1) \times Sp_2 \times (1 - P)$
Test 1 négatif	$p_{01} = (1 - Se_1) \times Se_2 \times P + Sp_1 \times (1 - Sp_2) \times (1 - P)$	$p_{01} = (1 - Se_1) \times (1 - Se_2) \times P + Sp_1 \times Sp_2 \times (1 - P)$

  - 3 degrés de liberté pour 5 paramètres à estimer ($P\ Se_1\ Sp_1\ Se_2\ Sp_2$)
  - En revanche, si on se place dans deux populations de prévalences différentes

$\Rightarrow$ **6 DL pour 6 paramètres** à estimer ($P_1\ P_2\ Se_1\ Sp_1\ Se_2\ Sp_2$)

  - Variantes possibles avec 3 tests et une population (8DL pour 7 paramètres)…

Modèle général

- Si on considère 2 tests appliqués à S populations de tailles respectives Ns et de prévalences Ps alors :

$$
\begin{aligned}
&n_S \sim multinomiale\left(N_S, p_{S11}, p_{S10}, p_{S01}, p_{S00}\right)\\
&p_{S11} = \pi_S \times Se_1 \times Se_2 + \left(1-\pi_S\right)\times\left(1-Sp_1\right)\times\left(1-Sp_2\right)\\
&p_{S10} = \pi_S \times Se_1 \times \left(1-Se_2\right) + \left(1-\pi_S\right)\times\left(1-Sp_1\right)\times Sp_2\\
&p_{S01} = \pi_S \times \left(1-Se_1\right)\times Se_2 + \left(1-\pi_S\right)\times Sp_1 \times\left(1-Sp_2\right)\\
&p_{S00} = \pi_S \times \left(1-Se_1\right)\times\left(1-Se_2\right) + \left(1-\pi_S\right)\times Sp_1 \times Sp_2
\end{aligned}
$$

- Inférence possible par maximum de vraisemblance ou par inférence bayésienne mais deux solutions sans priors informatifs
- Convergence du modèle vers un statut latent ⇔ consensus entre les deux tests

Hypothèses du modèle

- **Se et Sp constantes** dans chacune des sous populations
  - Les Se et Sp identiques dans les populations à prévalences différentes?
- **Indépendance conditionnelle** au statut vis-à-vis de la maladie

$$P\left(T_2^+ \middle| T_1^+, M^+\right) = P\left(T_2^+ \middle| M^+\right)$$

  - On considère que l'hypothèse est vérifiée quand les tests reposent sur des principes biologiques différents (PCR qualitative et bactériologie)
  - Possibilité de prendre en compte la dépendance conditionnelle mais alors avec plus de paramètres… *(Hadgu and Qu, 1998; Dendukuri and Joseph, 2001)*
- Le **statut latent est biologiquement similaire** pour les deux tests
  - Immunité pour des sérologies// présence du pathogène pour bactério et PCR
  - Possibilité de modéliser plus de 2 statuts latents *(Dendukuri et al. 2009)*

Données d’application

- Données issues du projet KitEval4500
  - Sous-échantillon d’une enquête épidémiologique *(Gache et al. 2017)*

  1500x3 sérums bovins, ovins et caprins issus d’environ 150x3 élevages dans 10 départements français
- Sérums analysés par le LNR fièvre Q avec les 3 tests sérologiques commercialisés en France
- Performances diagnostiques?
  - Pas estimées pour tous les tests dans toutes les espèces

  ⇒Globalement, en fonction des espèces et des études :
  - Se variant entre 70 et 100%
  - Sp variant entre 90 et 100%
- Des **biais méthodologiques** potentiellement importants
- Aucun d’entre eux ne peut être considéré comme **un Gold standard** à Se et Sp = 100%

Dépendance conditionnelle au statut vis-à-vis de la maladie

- Est-ce que les tests ont tendance à **se tromper en même temps** chez les vrais séropositifs et séronégatifs ?
- Chez les vrais « séropositifs » : **oui ?** (probablement)
  - Si le taux d'anticorps est bas, les DO seront probablement basses pour les trois tests
  - Si souches « particulières » : anticorps spécifiques détectés par certains tests (ex : deux réactifs donnant des résultats concordants et l'autre non)
- Chez les vrais « séronégatifs » : **non ?** (moins évident)
  - Seulement si réaction croisée commune entre les trois tests avec un autre agent infectieux
  - Sinon plutôt erreurs indépendantes d'un test à l'autre

Prise en compte de la dépendance conditionnelle

- Ajout de **termes correcteurs** sur les probabilités

	Réactif 2 positif	Réactif 2 négatif
Réactif 1 positif	$p_{11} = \left(Se_1 \times Se_2 \boxed{+\gamma_{Se}}\right) \times P + \left((1-Sp_1) \times (1-Sp_2) \boxed{+\gamma_{Sp}}\right) \times (1-P)$	$p_{10} = \left(Se_1 \times (1-Se_2) \boxed{-\gamma_{Se}}\right) \times P + \left((1-Sp_1) \times Sp_2 \boxed{-\gamma_{Sp}}\right) \times (1-P)$
Réactif 1 négatif	$p_{01} = \left((1-Se_1) \times Se_2 \boxed{-\gamma_{Se}}\right) \times P + \left(Sp_1 \times (1-Sp_2) \boxed{-\gamma_{Sp}}\right) \times (1-P)$	$p_{00} = \left((1-Se_1) \times (1-Se_2) \boxed{+\gamma_{Se}}\right) \times P + \left(Sp_1 \times Sp_2 \boxed{+\gamma_{Sp}}\right) \times (1-P)$

- **Plus de paramètres** donc le modèle est moins facilement identifiable
- Potentiellement non identifiable en fonction du niveau de dépendance et du statut latent mesuré

Dépendance conditionnelle et statut latent

- Face à deux tests très concordants réalisés sur des individus de statuts inconnu
  - Les deux tests identifient très bien les individus vrai positifs et vrai négatifs → Se et Sp élevées, dépendance conditionnelle faible
  - Les deux tests se trompent en même temps → Se et Sp faibles, dépendance conditionnelle élevée

⇒« Statut latent » ajusté = consensus entre les deux tests donc plutôt Se et Sp élevées

Dépendance conditionnelle et statut latent

- Face à deux tests très concordants réalisés sur des individus de statuts inconnu
  - Les deux tests identifient très bien les individus vrai positifs et vrai négatifs → Se et Sp élevées, dépendance conditionnelle faible
  - Les deux tests se trompent en même temps → Se et Sp faibles, dépendance conditionnelle élevée

  ⇒« Statut latent » ajusté = consensus entre les deux tests donc plutôt Se et Sp élevées
- Et si on ajoute un troisième test indépendant?
  - Identification de la dépendance conditionnelle entre les deux premiers tests?

  ⇒Changement de statut latent vers le test 3 et Se et Sp des tests 1 et 2 plus faible

  ⇒Pas de changement du statut latent et potentiels biais des Se et Sp du test 3

Contraintes sur les paramètres de dépendance conditionnelle

	Réactif 2 positif	Réactif 2 négatif
Réactif 1 positif	$p_{11} = \left(Se_1 \times Se_2 \boxed{+\gamma_{Se}}\right) \times P + \left(\left(1-Sp_1\right)\times\left(1-Sp_2\right) \boxed{+\gamma_{Sp}}\right) \times \left(1-P\right)$	$p_{10} = \left(Se_1 \times \left(1-Se_2\right) \boxed{-\gamma_{Se}}\right) \times P + \left(\left(1-Sp_1\right)\times Sp_2 \boxed{-\gamma_{Sp}}\right) \times \left(1-P\right)$
Réactif 1 négatif	$p_{01} = \left(\left(1-Se_1\right)\times Se_2 \boxed{-\gamma_{Se}}\right) \times P + \left(Sp_1 \times \left(1-Sp_2\right) \boxed{-\gamma_{Sp}}\right) \times \left(1-P\right)$	$p_{00} = \left(\left(1-Se_1\right)\times\left(1-Se_2\right) \boxed{+\gamma_{Se}}\right) \times P + \left(Sp_1 \times Sp_2 \boxed{+\gamma_{Sp}}\right) \times \left(1-P\right)$

$$p_{11} \in [0;1]$$

$$\left(Se_1 \times Se_2 + \gamma_{Se}\right) \in [0;1]$$

$$\left(\left(1-Se_1\right)\times\left(1-Se_2\right)+\gamma_{Se}\right) \in [0;1]$$

...

$$\gamma_{Se} \in \left[-\mathrm{Se}_1 \times \mathrm{Se}_2 + \max\left(0, \mathrm{Se}_1 + \mathrm{Se}_2 - 1\right); \min\left(\mathrm{Se}_1, \mathrm{Se}_2\right) - \mathrm{Se}_1 \times \mathrm{Se}_2\right]$$

$$\gamma_{Sp} \in \left[-\mathrm{Sp}_1 \times \mathrm{Sp}_2 + \max\left(0, \mathrm{Sp}_1 + \mathrm{Sp}_2 - 1\right), \min\left(\mathrm{Sp}_1, \mathrm{Sp}_2\right) - \mathrm{Sp}_1 \times \mathrm{Sp}_2)\right]$$

Pour 3 tests (Wang et al. 2017)!

Catégorie	Probabilité
$T_1^- T_2^- T_3^-$	$P \times \left(\left(1-Se_1\right)\times\left(1-Se_2\right)\times\left(1-Se_3\right)+\gamma_{Se000}\right)+$ $\left(1-P\right)\times\left(Sp_1\times Sp_2\times Sp_3+\gamma_{Sp000}\right)$
$T_1^- T_2^+ T_3^-$	$P \times \left(\left(1-Se_1\right)\times Se_2\times\left(1-Se_3\right)+\gamma_{Se010}\right)+$ $\left(1-P\right)\times\left(Sp_1\times\left(1-Sp_2\right)\times Sp_3+\gamma_{Sp010}\right)$
$T_1^- T_2^- T_3^+$	$P \times \left(\left(1-Se_1\right)\times\left(1-Se_2\right)\times Se_3+\gamma_{Se001}\right)+$ $\left(1-P\right)\times\left(Sp_1\times Sp_2\times\left(1-Sp_3\right)+\gamma_{Sp001}\right)$
$T_1^- T_2^+ T_3^+$	$P \times \left(\left(1-Se_1\right)\times Se_2\times Se_3+\gamma_{Se011}\right)+$ $\left(1-P\right)\times\left(Sp_1\times\left(1-Sp_2\right)\times\left(1-Sp_3\right)+\gamma_{Sp011}\right)$
$T_1^+ T_2^- T_3^-$	$P \times \left(Se_1\times\left(1-Se_2\right)\times\left(1-Se_3\right)+\gamma_{Se100}\right)+$ $\left(1-P\right)\times\left(\left(1-Sp_1\right)\times Sp_2\times Sp_3+\gamma_{Sp100}\right)$
$T_1^+ T_2^+ T_3^-$	$P \times \left(Se_1\times Se_2\times\left(1-Se_3\right)+\gamma_{Se110}\right)+$ $\left(1-P\right)\times\left(\left(1-Sp_1\right)\times\left(1-Sp_2\right)\times Sp_3+\gamma_{Sp110}\right)$
$T_1^+ T_2^- T_3^+$	$P \times \left(Se_1\times\left(1-Se_2\right)\times Se_3+\gamma_{Se101}\right)+$ $\left(1-P\right)\times\left(\left(1-Sp_1\right)\times Sp_2\times\left(1-Sp_3\right)+\gamma_{Sp101}\right)$
$T_1^+ T_2^+ T_3^+$	$P \times \left(Se_1\times Se_2\times Se_3+\gamma_{Se111}\right)+$ $\left(1-P\right)\times\left(\left(1-Sp_1\right)\times\left(1-Sp_2\right)\times\left(1-Sp_3\right)+\gamma_{Sp111}\right)$

- 8 catégories de résultats croisées des trois tests

$\Rightarrow$7 DL

- Minimum 7 paramétres (P, Se, Sp)
- Dépendance conditionnelle
  - En théorie 8 + 8 mais :

$$\gamma_{Se100} = \gamma_{Se011} + \gamma_{Se111} - \gamma_{Se000} \qquad \gamma_{Sp100} = \gamma_{Sp011} + \gamma_{Sp111} - \gamma_{Sp000}$$

$$\gamma_{Se101} = -\left(\gamma_{Se001} + \gamma_{Se011} + \gamma_{Se111}\right) \qquad \gamma_{Sp101} = -\left(\gamma_{Sp001} + \gamma_{Sp011} + \gamma_{Sp111}\right)$$

$$\gamma_{Se110} = \gamma_{Se000} + \gamma_{Se001} - \gamma_{Se111} \qquad \gamma_{Sp110} = \gamma_{Sp000} + \gamma_{Sp001} - \gamma_{Sp111}$$

$$\gamma_{Se010} = -\left(\gamma_{Se000} + \gamma_{Se001} + \gamma_{Se011}\right) \qquad \gamma_{Sp010} = -\left(\gamma_{Sp000} + \gamma_{Sp001} + \gamma_{Sp011}\right)$$

  - Donc 8 paramètres supplémentaire

Back to data : 3 tests (Wang et al. 2017)!

$$-Se_1 \times Se_2 \times Se_3 < \gamma_{Se111} < \min\left(Se_1, \min\left(Se_2, Se_3\right)\right) - Se_1 \times Se_2 \times Se_3$$

$$-\left(1-Se_1\right) \times Se_2 \times Se_3 < \gamma_{Se011} < \min\left(\left(1-Se_1\right), \min\left(Se_2, Se_3\right)\right) - \left(1-Se_1\right) \times Se_2 \times Se_3$$

$$-\left(1-Se_1\right) \times \left(1-Se_2\right) \times Se_3 < \gamma_{Se001} < \min\left(\left(1-Se_1\right), \min\left(\left(1-Se_2\right), Se_3\right)\right) - \left(1-Se_1\right) \times \left(1-Se_2\right) \times Se_3$$

$$-\left(1-Se_1\right) \times \left(1-Se_2\right) \times \left(1-Se_3\right) < \gamma_{Se000} < \min\left(\left(1-Se_1\right), \min\left(\left(1-Se_2\right), \left(1-Se_3\right)\right)\right) - \left(1-Se_1\right) \times \left(1-Se_2\right) \times \left(1-Se_3\right)$$

$$-Se_1 \times \left(1-Se_2\right) \times \left(1-Se_3\right) < \gamma_{Se100} < \min\left(Se_1, \min\left(\left(1-Se_2\right), \left(1-Se_3\right)\right)\right) - Se_1 \times \left(1-Se_2\right) \times \left(1-Se_3\right)$$

$$-Se_1 \times \left(1-Se_2\right) \times Se_3 < \gamma_{Se101} < \min\left(Se_1, \min\left(\left(1-Se_2\right), Se_3\right)\right) - Se_1 \times \left(1-Se_2\right) \times Se_3$$

$$-Se_1 \times Se_2 \times \left(1-Se_3\right) < \gamma_{Se110} < \min\left(Se_1, \min\left(Se_2, \left(1-Se_3\right)\right)\right) - Se_1 \times Se_2 \times \left(1-Se_3\right)$$

$$-\left(1-Se_1\right) \times Se_2 \times \left(1-Se_3\right) < \gamma_{Se010} < \min\left(\left(1-Se_1\right), \min\left(Se_2, \left(1-Se_3\right)\right)\right) - \left(1-Se_1\right) \times Se_2 \times \left(1-Se_3\right)$$

Back to data : Kiteval4500

- 1500x3 sérums bovins, ovins et caprins issus d'environ 150x3 élevages dans 10 départements français **= 10 population?**

Nombre de positifs/négatifs aux seuils constructeurs dans chaque département

dpt	$T_1^- T_2^- T_3^-$	$T_1^- T_2^- T_3^+$	$T_1^- T_2^+ T_3^-$	$T_1^- T_2^+ T_3^+$	$T_1^+ T_2^- T_3^-$	$T_1^+ T_2^- T_3^+$	$T_1^+ T_2^+ T_3^-$	$T_1^+ T_2^+ T_3^+$	Total
A	91	27	0	10	2	0	0	35	165
B	94	28	5	11	0	0	2	22	162
C	145	3	0	0	1	0	0	0	149
D	124	2	8	4	0	0	0	7	145
E	142	10	0	0	0	0	0	3	155
F	154	1	1	0	1	0	0	0	157
G	161	0	0	0	0	0	0	0	161
H	91	25	3	4	3	2	0	18	146
I	124	15	9	3	2	1	0	2	156
J	24	2	1	0	0	0	0	9	36

3 départements avec très peu de séropositifs

$\Rightarrow$Les trois tests semblent spécifiques

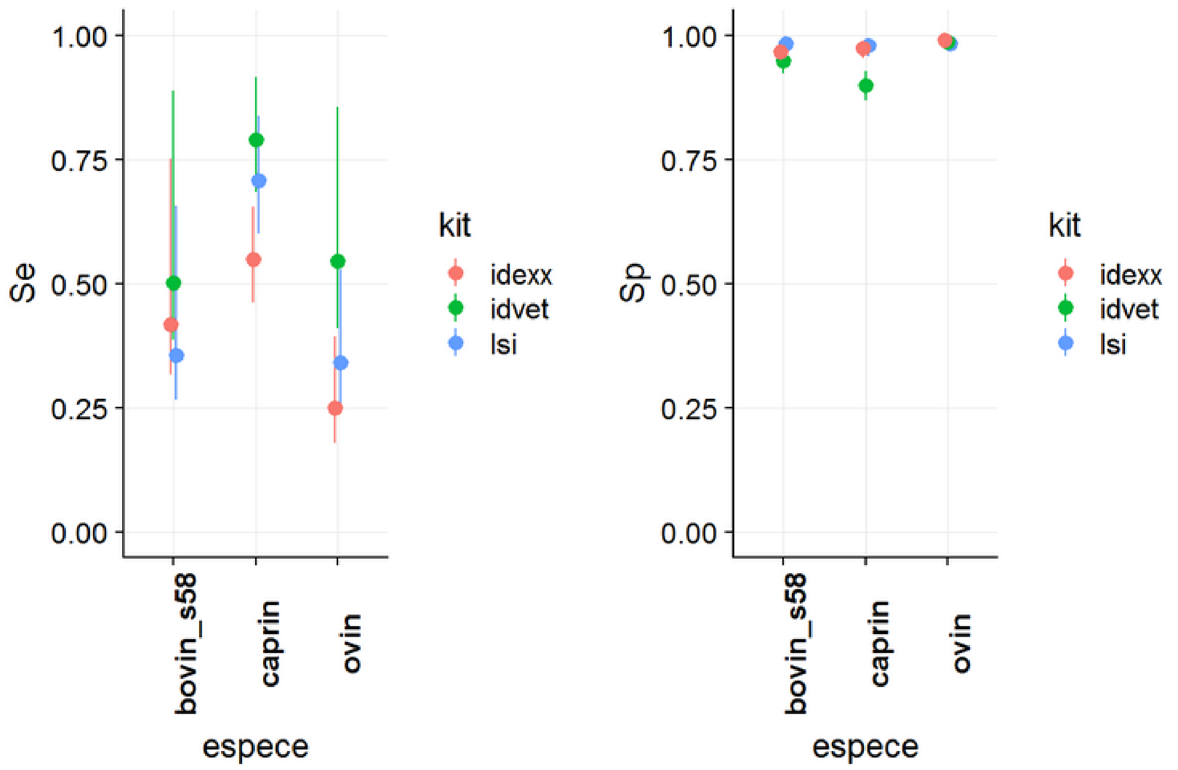
La plupart des positifs sont positifs avec le test 3 ou avec les trois tests

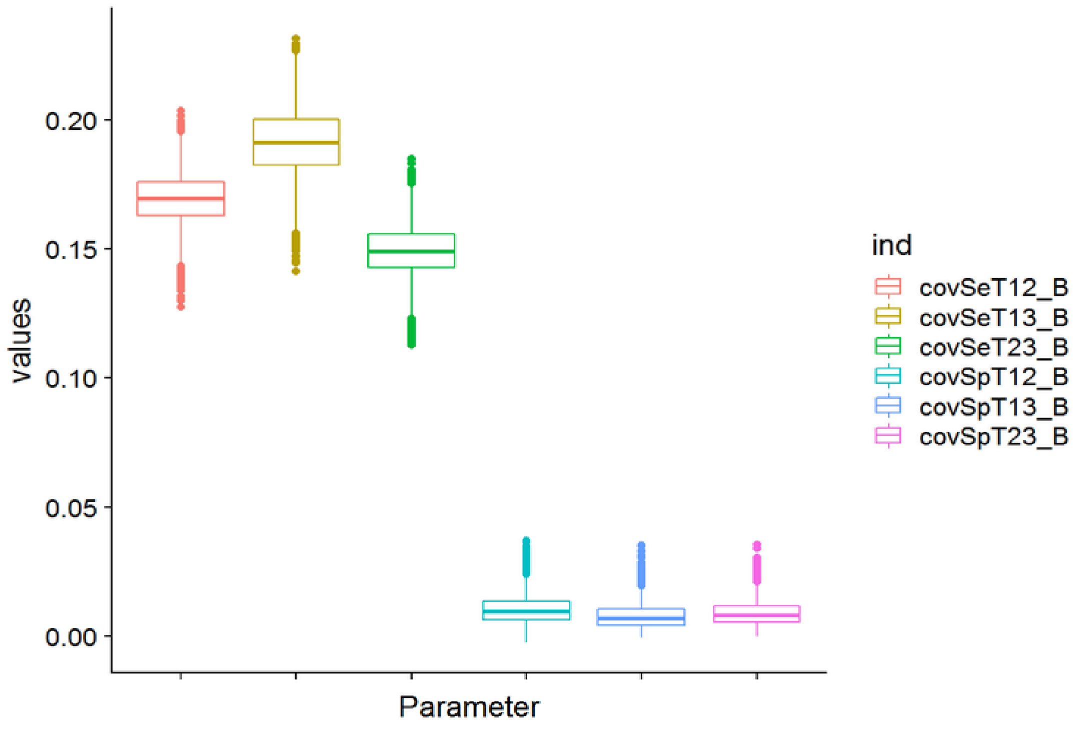
$\Rightarrow$Le test 3 semble plus sensible que les deux autres

Première inférence sous JAGS

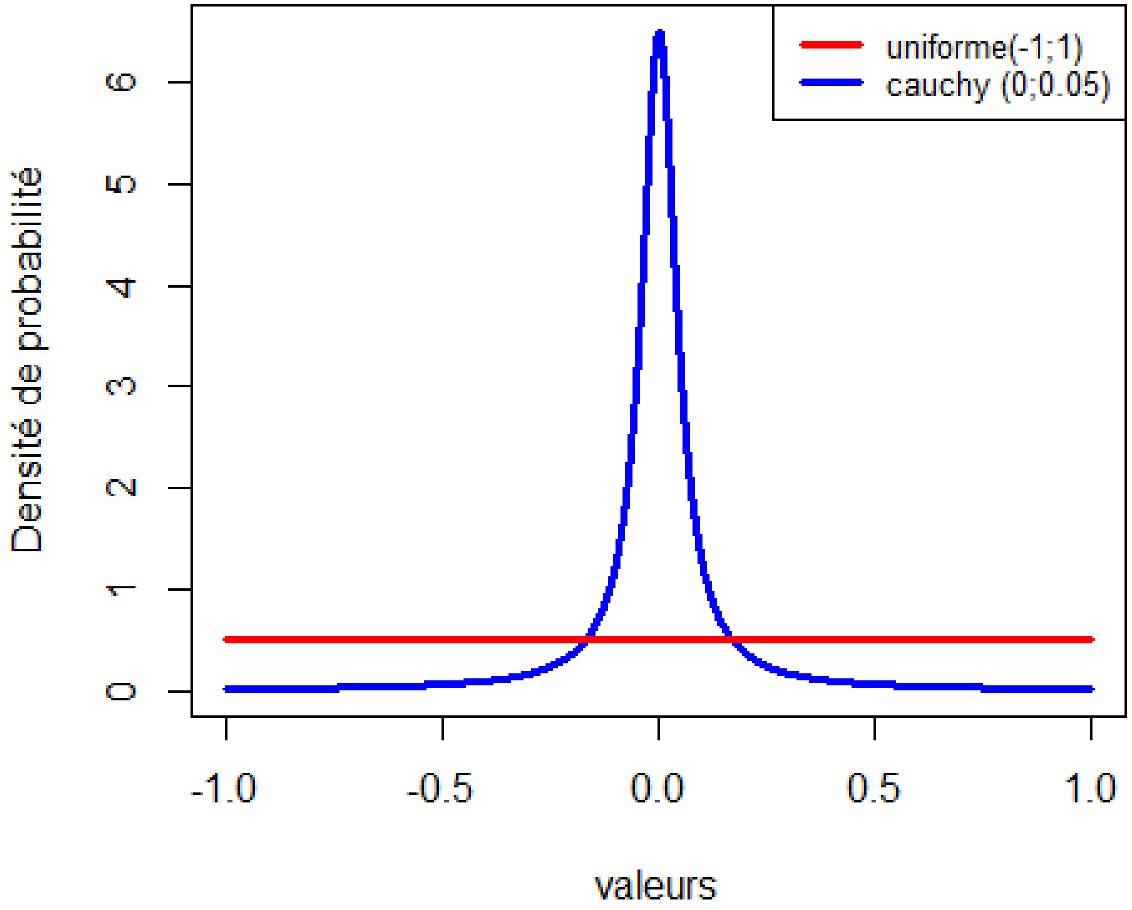
- Priors
  - Loi beta (0.5,0.5) sur les Se, Sp et P
  - Loi uniforme [-0.5;0.5] pour les dépendances conditionnelles

Se et Sp estimées toutes espèces



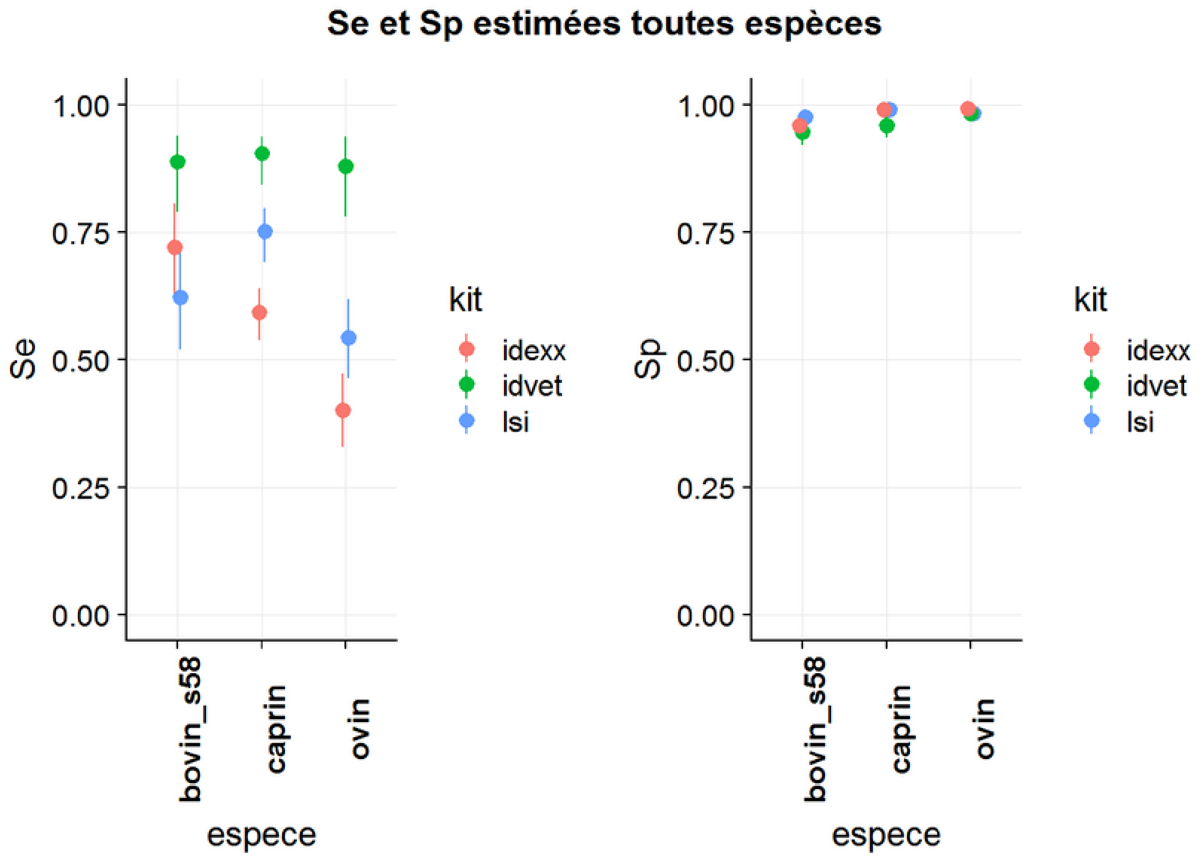


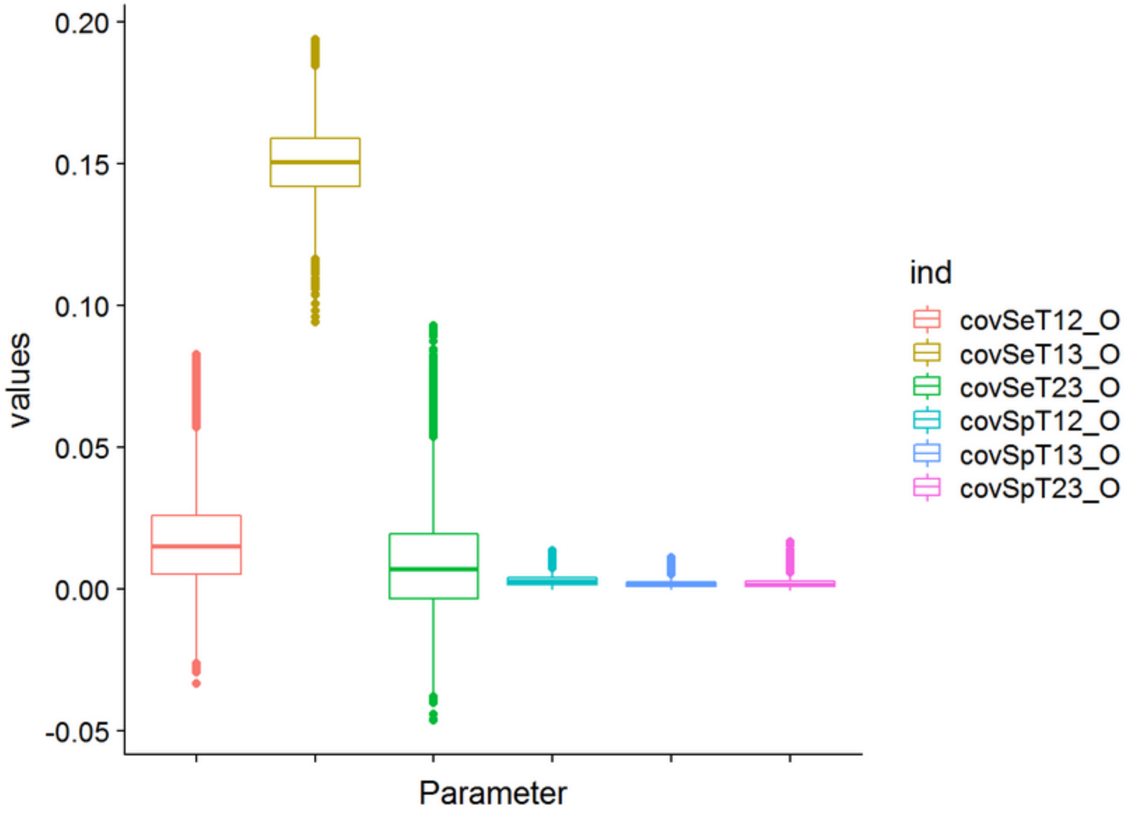
Priors cauchy vs loi uniforme???



Deuxième inférence sous JAGS

- Priors
  - Loi cauchy (0, 0.039) pour les dépendances conditionnelles





Back to data

- 1500x3 sérums bovins, ovins et caprins issus d'environ 150x3 élevages = **150 population?** dans 10 départements français
- 150 prévalences par modèle à estimer?
  - Modélisation de distribution des prévalences dans chaque élevages (WHP)
  - Avec possibilité que certains élevages soit indemne

$\Rightarrow$Modélisation de la proportion d'élevage non indemne dans chaque élevage (BHP)

Modèle à classes latentes

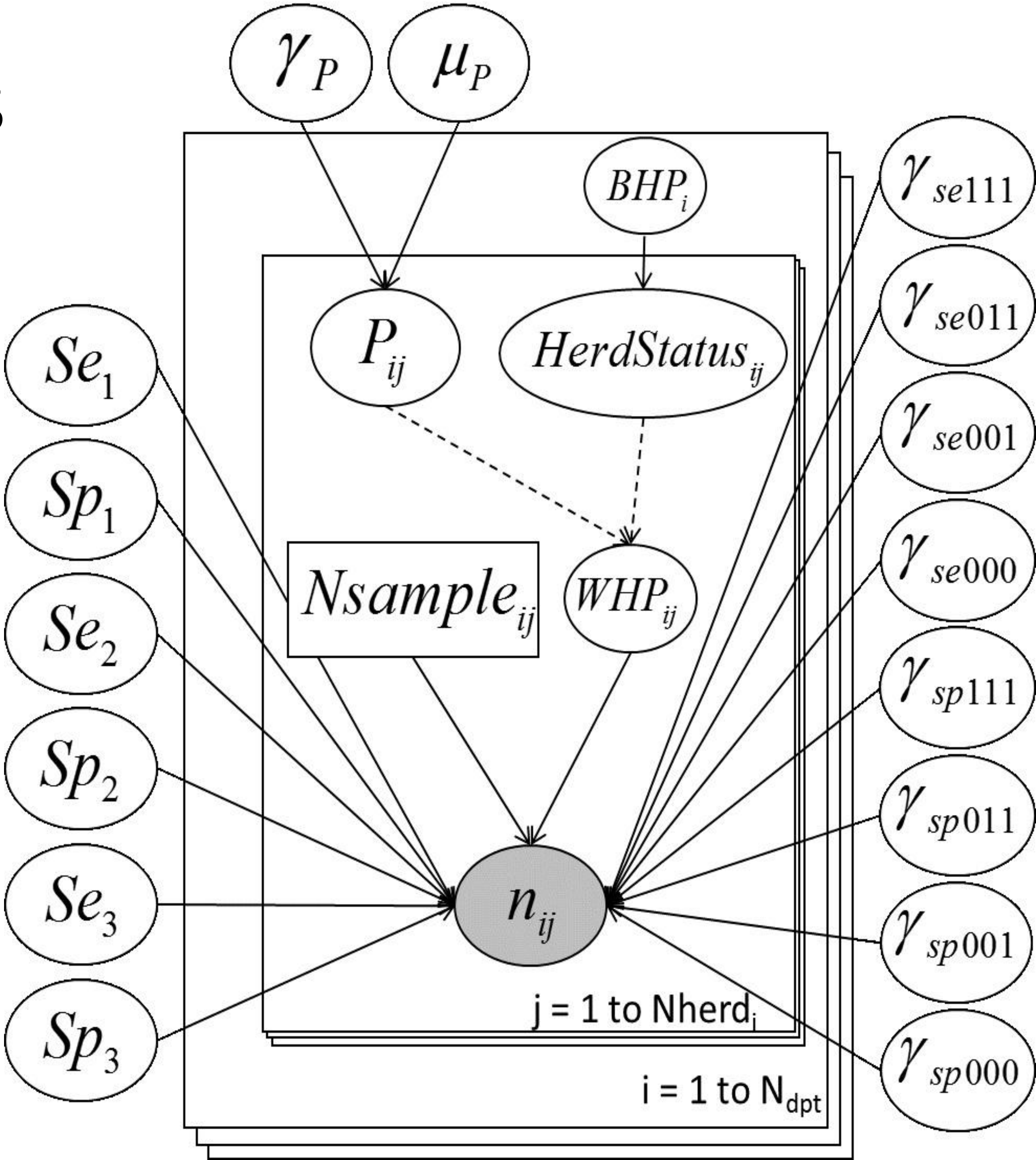
- Une population par élevage
- Modélisation de la prévalence inter-élevage par département
  - Avec possibilité que certains élevages soient indemnes
- Modélisation de la distribution des prévalences intra-élevages dans les élevages séropositifs

Modèle beta-binomiale hiérarchique sur les prévalence

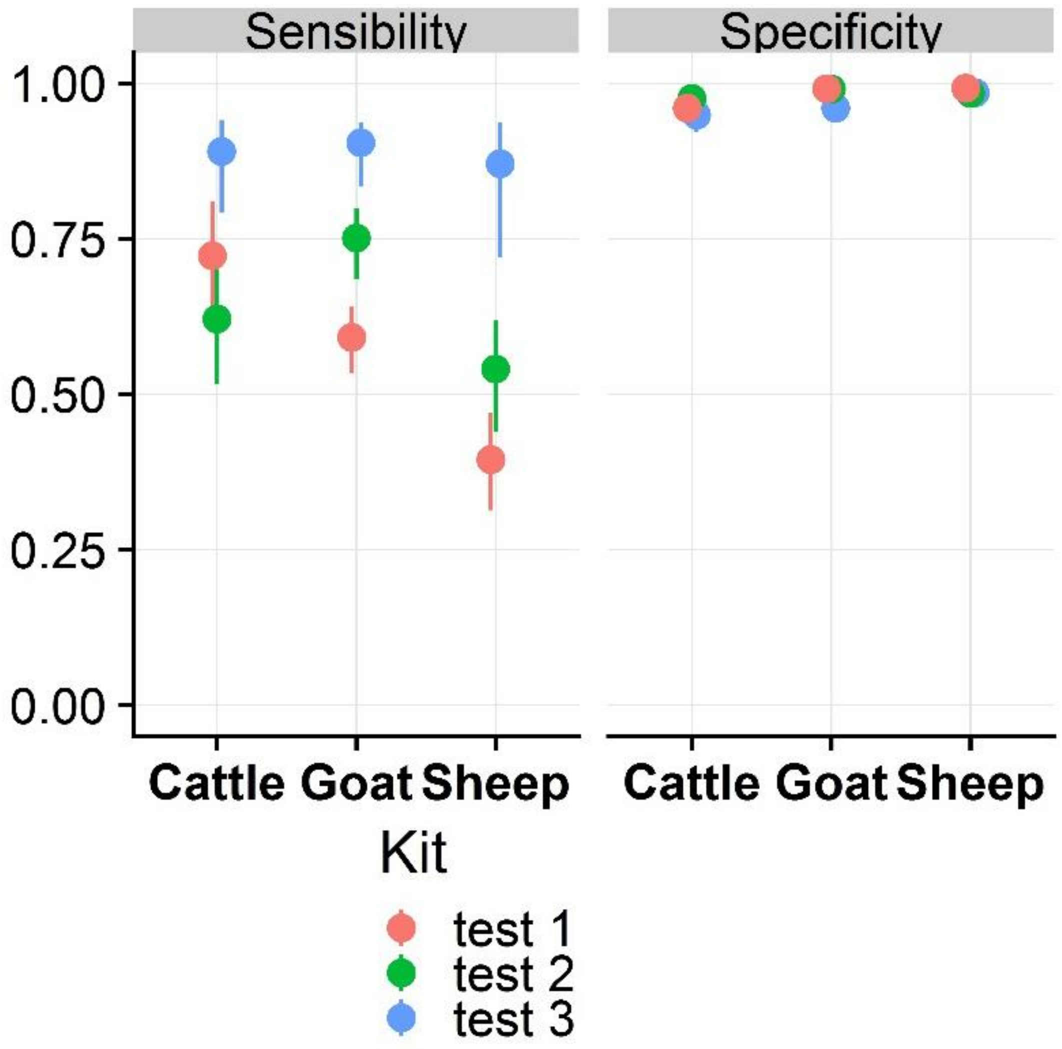
$$HerdStatus_{ij} \sim Bern(BHP_i)$$

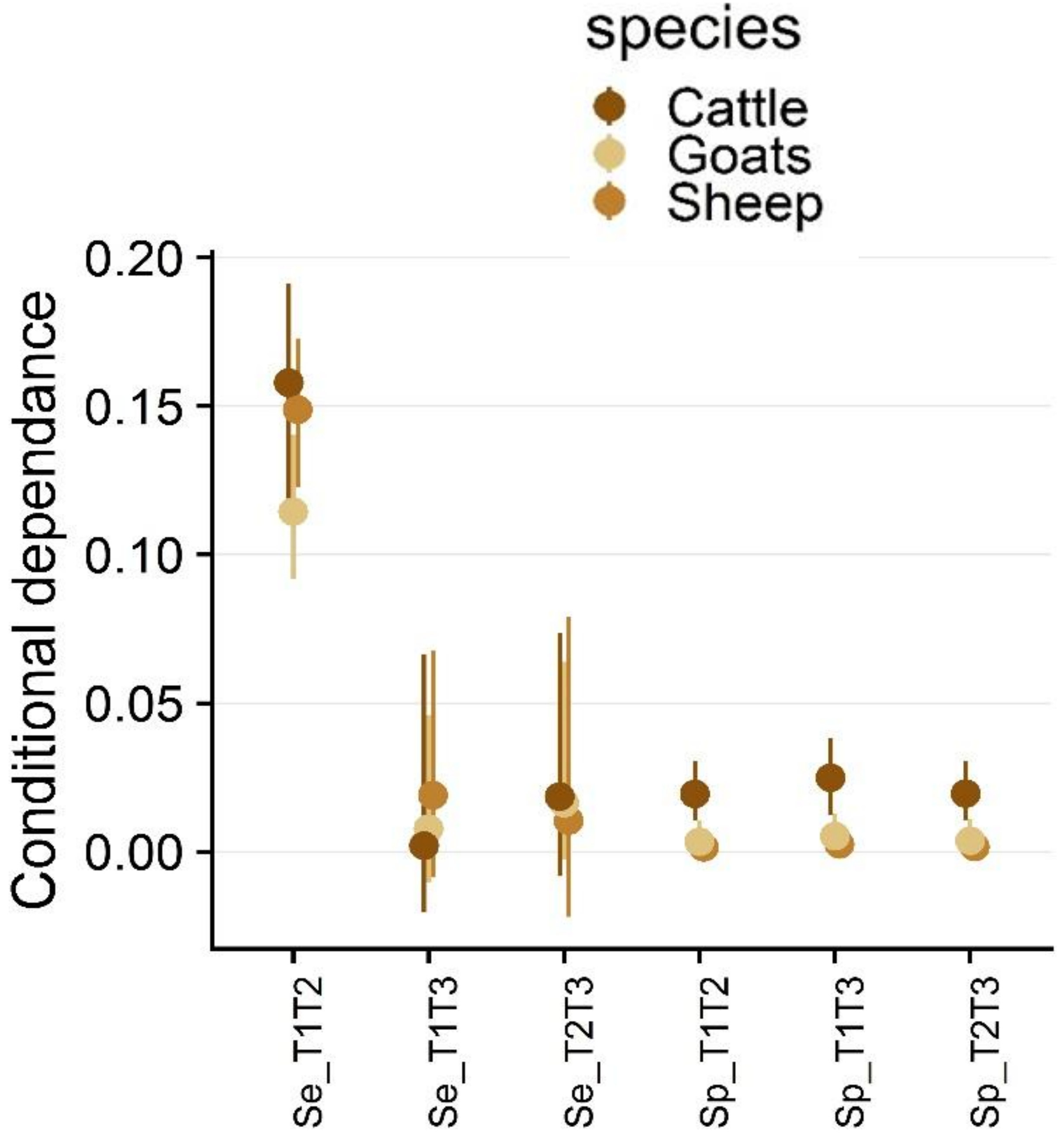
$$P_{ij} \sim beta\left(\mu_P \times \frac{(1-\gamma_P)}{\gamma_P}, (1-\mu_P) \times \frac{(1-\gamma_P)}{\gamma_P}\right)$$

$$WHP_{ij} = HerdStatus_{ij} \times P_{ij}$$



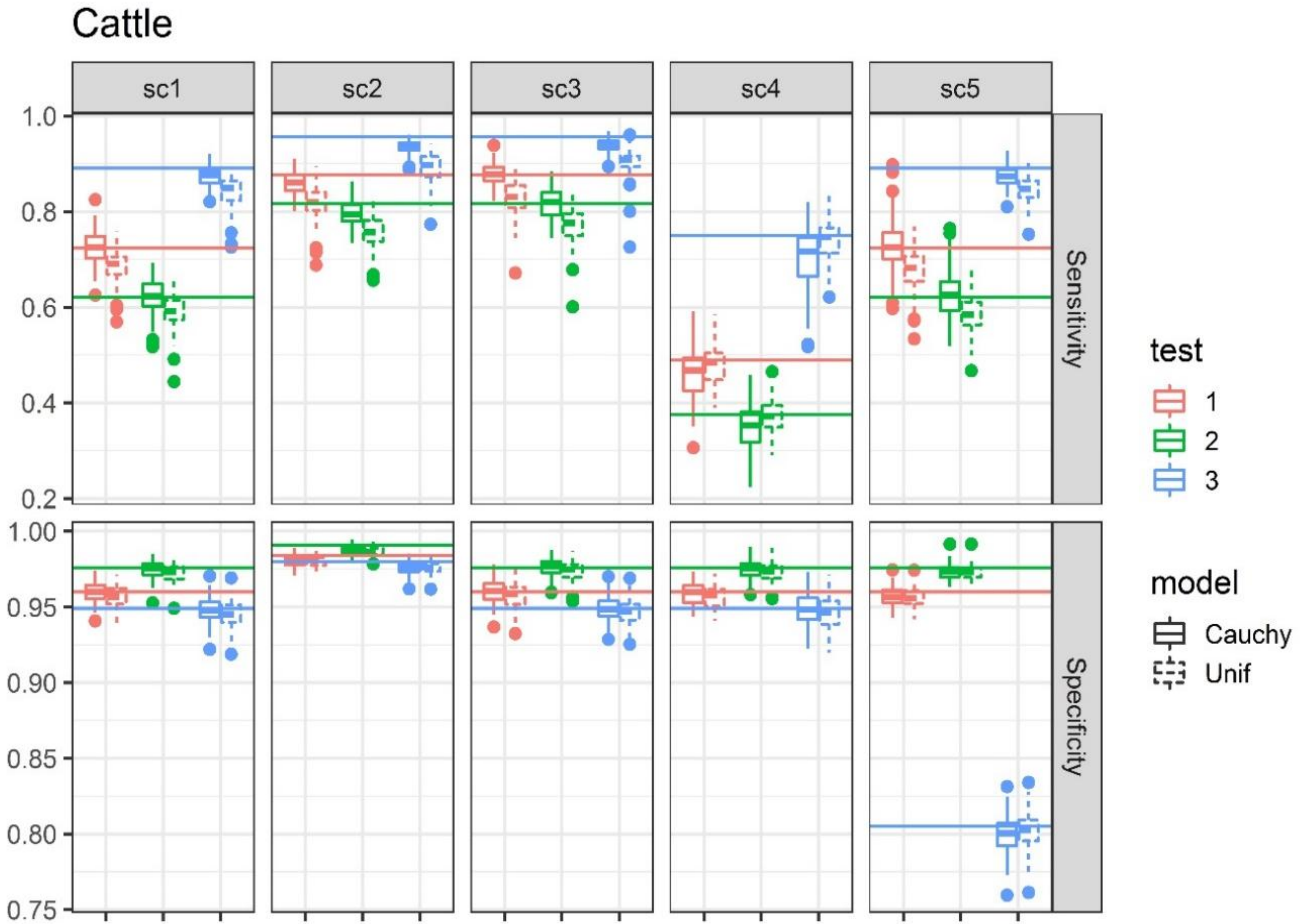
Résultats


Sensibility
Specificity
1.00
0.75
0.50
0.25
0.00
Cattle Goat Sheep
Cattle Goat Sheep
Kit
test 1
test 2
test 3


species
Cattle
Goats
Sheep
Conditional dependance
0.20
0.15
0.10
0.05
0.00
Se_T1T2
Se_T1T3
Se_T2T3
Sp_T1T2
Sp_T1T3
Sp_T2T3

Résultat de l’étude de simulation

- Distribution *a priori* Cauchy moins biaisée
- **Performances élevées chez les caprins et bovins**
  - Biais moyens < -2%
  - Probabilité de couverture > 80%
- Tendance à sous-estimer les Se et Sp chez les ovins (-3 à -7%)
- Mais probabilité de couverture acceptable (de 79 à 92%)



Discussion

- Impact potentiellement important du **prior sur les paramètres de dépendance** conditionnelle
- Dans notre cas
  - Possibilité d'identification du problème car « évident »
  - Amélioration de la robustesse du modèle grâce à la prise en compte de la structuration des données
- Pas trivial en l'absence de sous population identifiable…
  - Identification des différentes « solutions » possible?
  - Inclusion des « biologistes » pour choisir entre les différentes solutions possible?

Mélange de dirichlet?

Myriam Garrido

Pierre Latouche

Pierre Druilhet

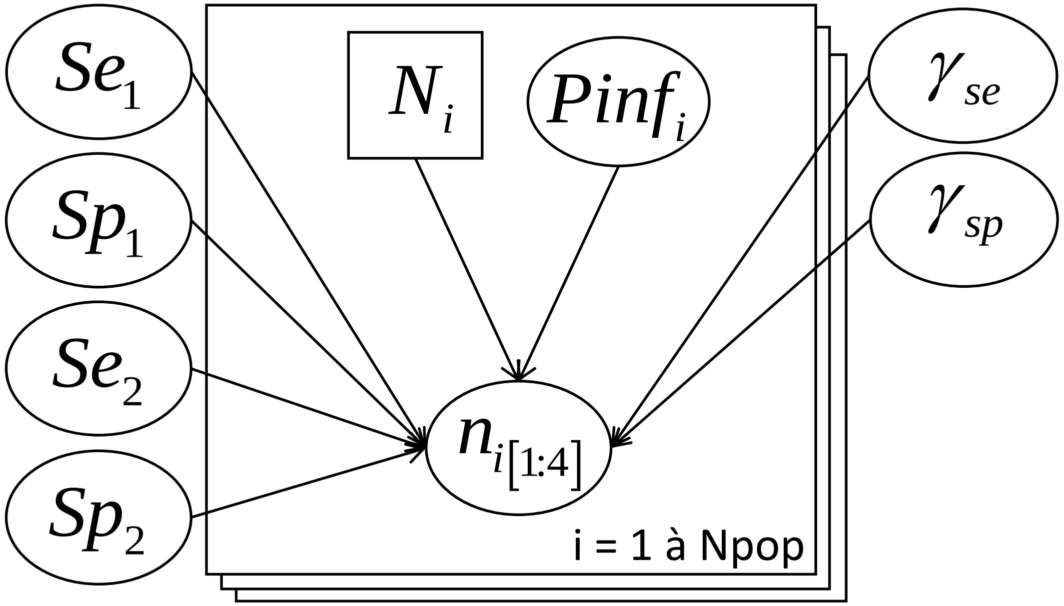
Prise en compte de la dépendance conditionnelle

- Ajout de **termes correcteurs** sur les probabilités

	Réactif 2 positif	Réactif 2 négatif
Réactif 1 positif	$p_{11} = \left(Se_1 \times Se_2 \boxed{+\gamma_{Se}}\right) \times P + \left((1-Sp_1)\times(1-Sp_2) \boxed{+\gamma_{Sp}}\right) \times (1-P)$	$p_{10} = \left(Se_1 \times (1-Se_2) \boxed{-\gamma_{Se}}\right) \times P + \left((1-Sp_1)\times Sp_2 \boxed{-\gamma_{Sp}}\right) \times (1-P)$
Réactif 1 négatif	$p_{01} = \left((1-Se_1)\times Se_2 \boxed{-\gamma_{Se}}\right) \times P + \left(Sp_1 \times (1-Sp_2) \boxed{-\gamma_{Sp}}\right) \times (1-P)$	$p_{00} = \left((1-Se_1)\times(1-Se_2) \boxed{+\gamma_{Se}}\right) \times P + \left(Sp_1 \times Sp_2 \boxed{+\gamma_{Sp}}\right) \times (1-P)$

- **Nouvelle paramétrisation du modèle proposée sous la forme d'un mélange de dirichlet**

Traditional LCM



$$n_{i[1:4]} \sim multinomial\left(N_i, p_{i[1:4]}\right)$$

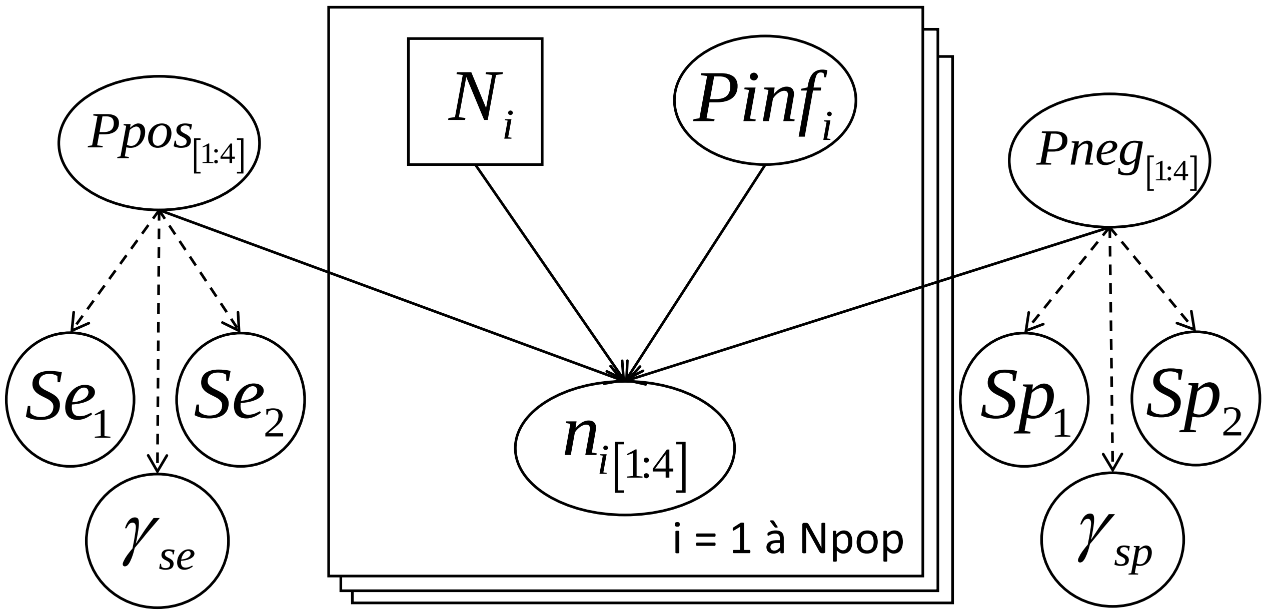
$$p_{i[1]} = Pinf_i \times \left(\left(1-Se_1\right)\times\left(1-Se_2\right)+\gamma_{se}\right)+\left(1-Pinf_i\right)\times\left(Sp_1\times Sp_2+\gamma_{sp}\right)$$

$$p_{i[2]} = Pinf_i \times \left(\left(1-Se_1\right)\times Se_2-\gamma_{se}\right)+\left(1-Pinf_i\right)\times\left(Sp_1\times\left(1-Sp_2\right)-\gamma_{sp}\right)$$

$$p_{i[3]} = Pinf_i \times \left(Se_1\times\left(1-Se_2\right)-\gamma_{se}\right)+\left(1-Pinf_i\right)\times\left(\left(1-Sp_1\right)\times Sp_2-\gamma_{sp}\right)$$

$$p_{i[4]} = Pinf_i \times \left(Se_1\times Se_2+\gamma_{se}\right)+\left(1-Pinf_i\right)\times\left(\left(1-Sp_1\right)\times\left(1-Sp_2\right)+\gamma_{sp}\right)$$

Dirichlet mixture LCM



$$n_{i[1:4]} \sim multinomial\left(N_i, p_{i[1:4]}\right)$$

$$p_{i[1:4]} = Pinf_i \times Ppos_{[1:4]} + \left(1-Pinf_i\right)\times Pneg_{[1:4]}$$

$$Se_1 = Ppos[3]+Ppos[4] \qquad Sp_1 = Pneg[1]+Pneg[2]$$

$$Se_2 = Ppos[2]+Ppos[4] \qquad Sp_2 = Pneg[1]+Pneg[3]$$

$$\gamma_{se} = Ppos[4]-Se_1\times Se_2 \qquad \gamma_{sp} = Pneg[1]-Sp_1\times Sp_2$$

Prior distribution

Traditional LCM:

$$Se_1 \sim beta(1,1) \qquad Sp_1 \sim beta(1,1)$$

$$Se_2 \sim beta(1,1) \qquad Sp_2 \sim beta(1,1) \qquad Pinf_i \sim beta(1,1)$$

$$\gamma_{se} \sim uniform(-1,1) \qquad \gamma_{sp} \sim uniform(-1,1)$$

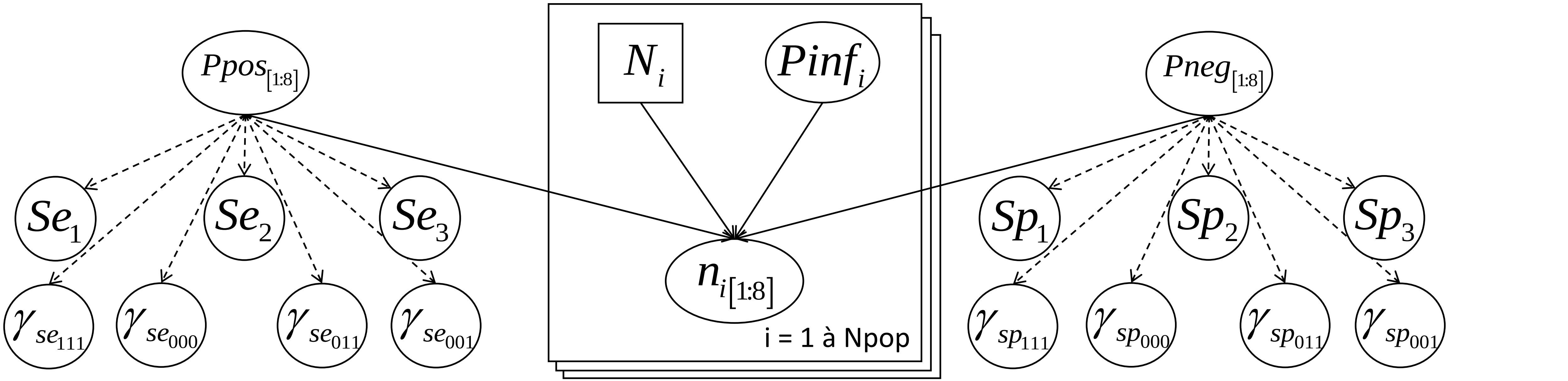
Dirichlet mixture LCM:

$$Pinf_i \sim beta(1,1)$$

$$Ppos_{[1:4]} \sim dirichlet\left(\alpha_{pos[1:4]}\right)$$

$$Pneg_{[1:4]} \sim dirichlet\left(\alpha_{neg[1:4]}\right)$$

Modèle sous forme dirichlet pour 3 tests



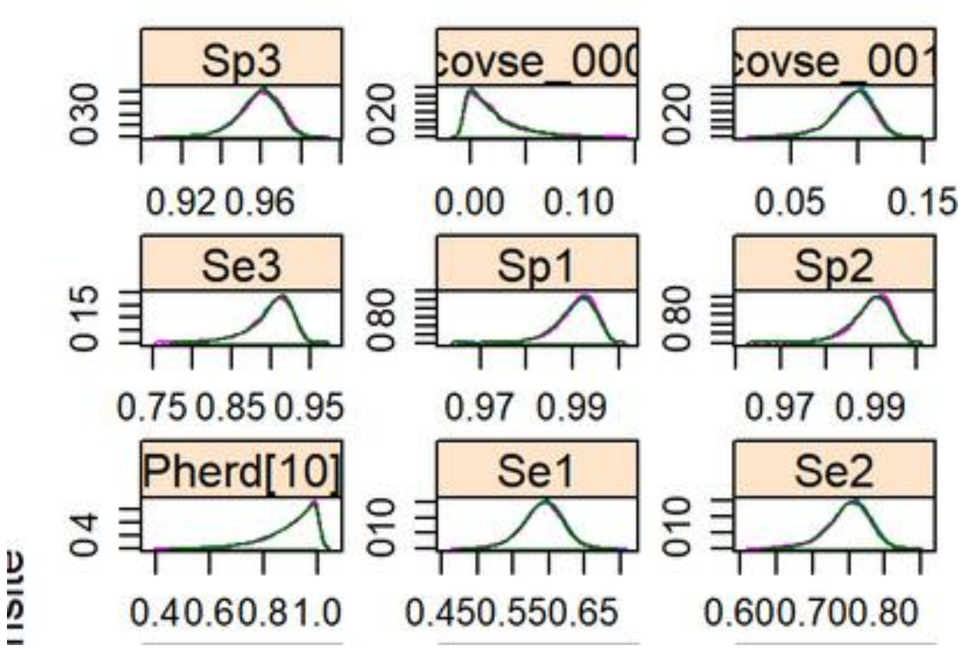
$$Se_1 = Ppos[5] + Ppos[6] + Ppos[7] + Ppos[8]$$
$$Se_2 = Ppos[3] + Ppos[4] + Ppos[7] + Ppos[8]$$
$$Se_3 = Ppos[2] + Ppos[4] + Ppos[6] + Ppos[8]$$
$$\gamma_{se111} = Ppos[8] - Se_1 \times Se_2 \times Se_3$$
$$\gamma_{se011} = Ppos[4] - (1 - Se_1) \times Se_2 \times Se_3$$
$$\gamma_{se001} = Ppos[2] - (1 - Se_1) \times (1 - Se_2) \times Se_3$$
$$\gamma_{se000} = Ppos[1] - (1 - Se_1) \times (1 - Se_2) \times (1 - Se_3)$$

$$n_{i[1:8]} \sim multinomial\left(N_i, p_{i[1:8]}\right)$$
$$p_{i[1:8]} = Pinf_i \times Ppos_{[1:8]} + (1 - Pinf_i) \times Pneg_{[1:8]}$$
$$Ppos_{[1:8]} \sim dirichlet\left(\alpha_{pos[1:8]}\right)$$
$$Pneg_{[1:8]} \sim dirichlet\left(\alpha_{neg[1:8]}\right)$$

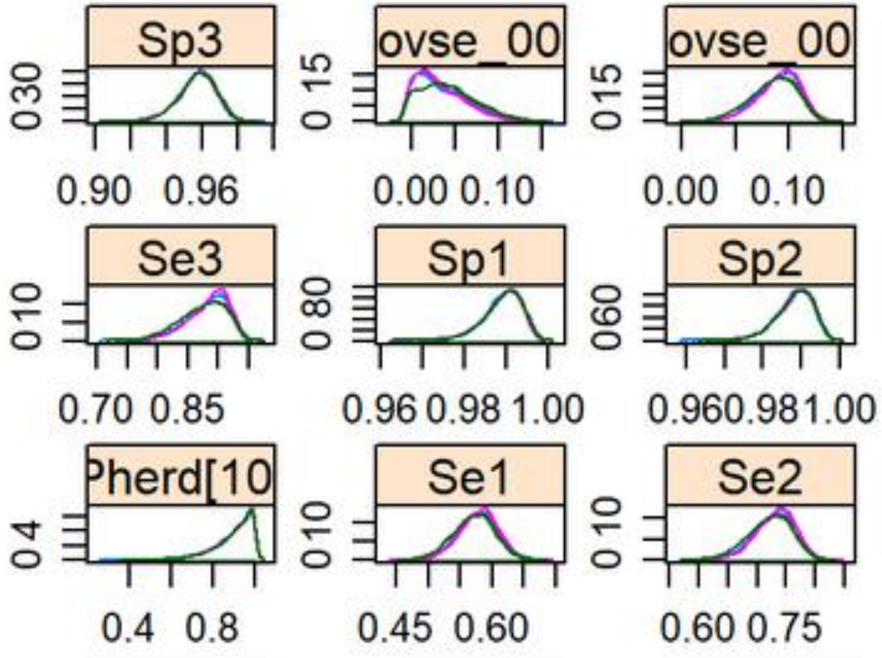
$$Sp_1 = Pneg[1] + Pneg[2] + Pneg[3] + Pneg[4]$$
$$Sp_2 = Pneg[1] + Pneg[2] + Pneg[5] + Pneg[6]$$
$$Sp_3 = Pneg[1] + Pneg[3] + Pneg[5] + Pneg[7]$$
$$\gamma_{sp111} = Pneg[8] - (1 - Sp_1) \times (1 - Sp_2) \times (1 - Sp_3)$$
$$\gamma_{se011} = Pneg[4] - Sp_1 \times (1 - Sp_2) \times (1 - Sp_3)$$
$$\gamma_{se001} = Pneg[2] - Sp_1 \times Sp_2 \times (1 - Sp_3)$$
$$\gamma_{se000} = Pneg[1] - Sp_1 \times Sp_2 \times (1 - Sp_3)$$

Application sur les données fièvre Q par élevage : 3 tests / 150 populations (1 par élevage)

Modèle « traditionnel »



Modèle « Dirichlet »



Temps de calcul = 20,76 minutes

Temps de calcul = 9,76 minutes

Application sans la structuration en élevage

- Modèle à trois tests et dix populations (1 par département)

$\Rightarrow$Estimations similaires à celles obtenues avec le prior uniforme

$\Rightarrow$Biais vers les dépendances conditionnelles fortes?

Perspectives

- Modèle qui tourne plus vite
- A l’air de favoriser les estimations à forte dépendance conditionnelle…
- Essayer d’inclure des distributions a priori informatives sur les deux Dirichlet
  - Actuellement $Dirichlet\left(\alpha,\left[0.25, 0.25, 0.25, 0.25\right]\right)$
  - Potentiellement

$$Dirichlet\left(\alpha,\begin{bmatrix} Se_{1Indep}\times Se_{2Indep}, \\ \left(1-Se_{1Indep}\right)\times Se_{2Indep}, \\ Se_{1Indep}\times\left(1-Se_{2Indep}\right), \\ \left(1-Se_{1Indep}\right)\times\left(1-Se_{2Indep}\right) \end{bmatrix}\right)$$

$$Se_{1Indep} \sim Beta\left(0.5, 0.5\right)$$
$$Se_{2Indep} \sim Beta\left(0.5, 0.5\right)$$
$$Sp_{1Indep} \sim Beta\left(0.5, 0.5\right)$$
$$Sp_{2Indep} \sim Beta\left(0.5, 0.5\right)$$

Estimation sur les données FQ par département

- 3 jeux de données environs 1500 ovins/bovins et caprins
- Testés avec 3 tests dans 10 départements

$\Rightarrow$ Modèle à 3 tests et 10 populations

- 3 modèles historiques avec des estimations incohérentes
- « vrai valeurs » estimées dans un modèle plus complexe

$\Rightarrow$Evaluer le comportement des modèles Dirichlet sur ces jeux de données

$\Rightarrow$Idéalement dans ce cas, peu biaisée et à large intervalle de crédibilité

Modèle à classes latentes

- Lclass : modèle à classe latente sans dépendance conditionnelle

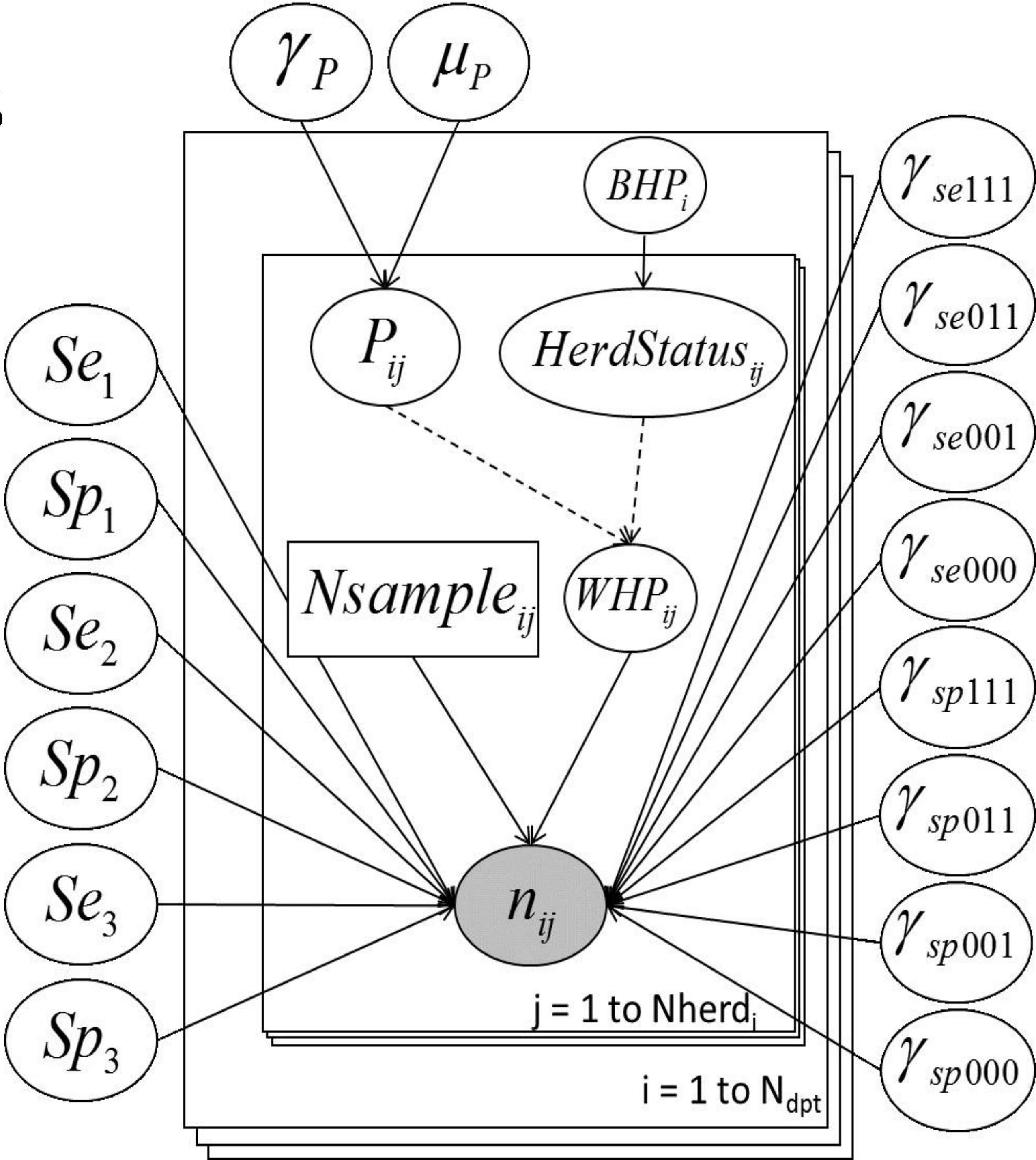
$$\gamma_{se\bullet\bullet\bullet} = 0 \qquad \gamma_{sp\bullet\bullet\bullet} = 0$$

- Unif : modèle à classe latente prior uniform[-0,5;0,5] sur la dep cond

$$\gamma_{se\bullet\bullet\bullet} \sim unif\left[0.5, 0.5\right] \qquad \gamma_{sp\bullet\bullet\bullet} \sim unif\left[0.5, 0.5\right]$$

- Cauchy : idem mais prior cauchy

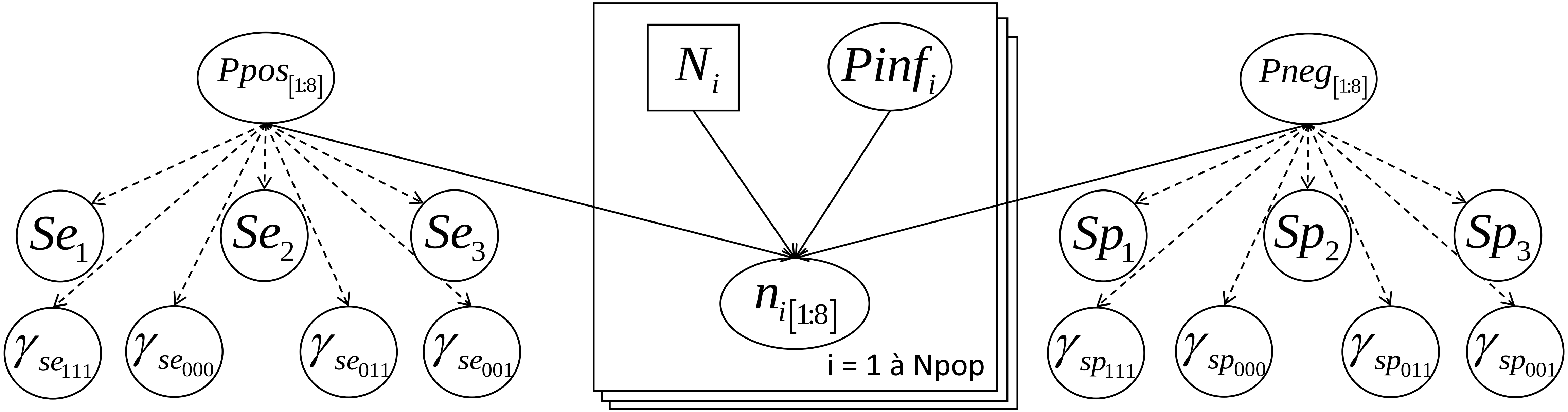
$$\gamma_{se\bullet\bullet\bullet} \sim cauchy\left[0, 0.039\right] \quad \gamma_{sp\bullet\bullet\bullet} \sim cauchy\left[0, 0.039\right]$$



7 Modèles testés

- Lclass : modèle à classe latente sans dépendance conditionnelle
- Unif : modèle à classe latente prior uniform[-0,5;0,5] sur la dep cond
- Cauchy : idem mais prior cauchy
- Dir : modèle à classe latente sous forme dirichlet
- Alpha 1 : modèle dirichlet avec hyperparamétrisation des Se et Sp et une valeur de alpha de 1
- Alpha8 : idem mais alpha 8
- Alpha 100 :idem mais alpha 100

Modèle sous forme dirichlet pour 3 tests sans hyperparametrisation

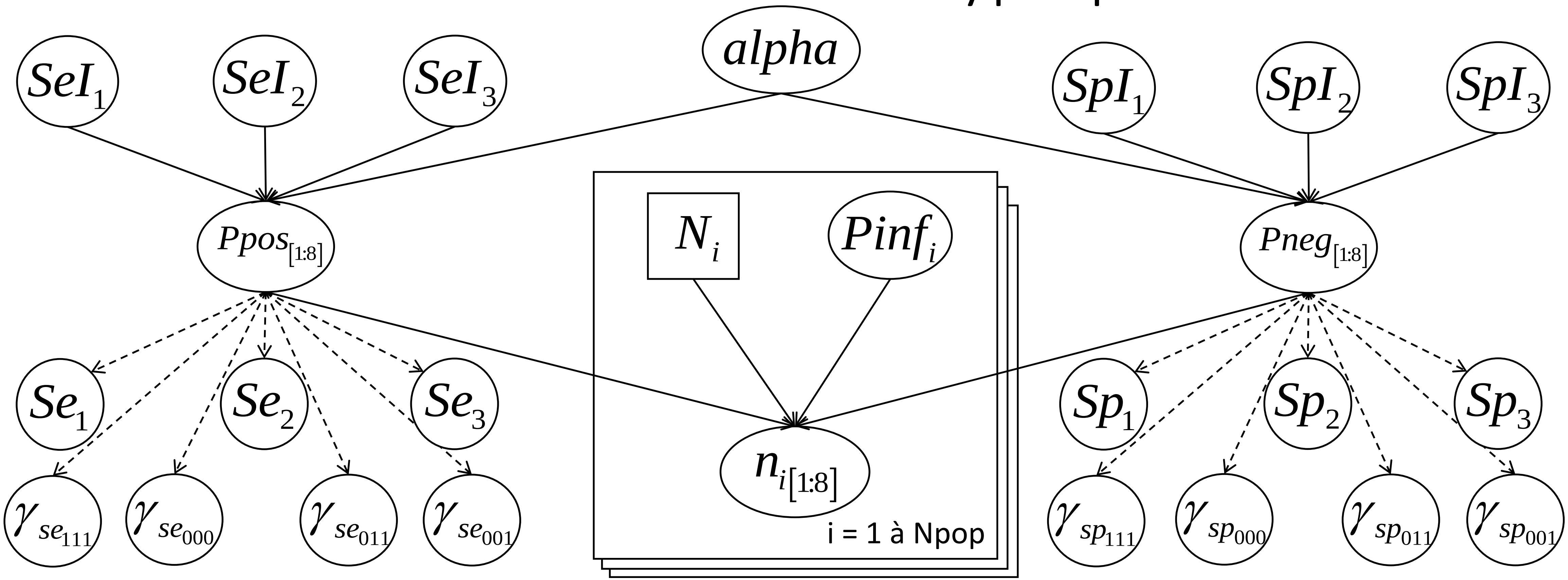


$$Se_1 = Ppos[5]+Ppos[6]+Ppos[7]+Ppos[8]$$
$$Se_2 = Ppos[3]+Ppos[4]+Ppos[7]+Ppos[8]$$
$$Se_3 = Ppos[2]+Ppos[4]+Ppos[6]+Ppos[8]$$
$$\gamma_{se111} = Ppos[8]-Se_1\times Se_2\times Se_3$$
$$\gamma_{se011} = Ppos[4]-(1-Se_1)\times Se_2\times Se_3$$
$$\gamma_{se001} = Ppos[2]-(1-Se_1)\times(1-Se_2)\times Se_3$$
$$\gamma_{se000} = Ppos[1]-(1-Se_1)\times(1-Se_2)\times(1-Se_3)$$

$$n_{i[1:8]} \sim multinomial\left(N_i, p_{i[1:8]}\right)$$
$$p_{i[1:8]} = Pinf_i\times Ppos_{[1:8]}+(1-Pinf_i)\times Pneg_{[1:8]}$$
$$Ppos_{[1:8]} \sim dirichlet\left(\alpha_{pos[1:8]}\right)$$
$$Pneg_{[1:8]} \sim dirichlet\left(\alpha_{neg[1:8]}\right)$$

$$Sp_1 = Pneg[1]+Pneg[2]+Pneg[3]+Pneg[4]$$
$$Sp_2 = Pneg[1]+Pneg[2]+Pneg[5]+Pneg[6]$$
$$Sp_3 = Pneg[1]+Pneg[3]+Pneg[5]+Pneg[7]$$
$$\gamma_{sp111} = Pneg[8]-(1-Sp_1)\times(1-Sp_2)\times(1-Sp_3)$$
$$\gamma_{se011} = Pneg[4]-Sp_1\times(1-Sp_2)\times(1-Sp_3)$$
$$\gamma_{se001} = Pneg[2]-Sp_1\times Sp_2\times(1-Sp_3)$$
$$\gamma_{se000} = Pneg[1]-Sp_1\times Sp_2\times(1-Sp_3)$$

Modèle sous forme dirichlet avec hyperparametrisation



$$Ppos_{[1:8]} \sim Dirichlet\left(\alpha, \begin{bmatrix} (1-SeI_1)\times(1-SeI_2)\times(1-SeI_3), (1-SeI_1)\times(1-SeI_2)\times SeI_3, \\ (1-SeI_1)\times SeI_2\times(1-SeI_3), (1-SeI_1)\times SeI_2\times SeI_3, \\ SeI_1\times(1-SeI_2)\times(1-SeI_3), SeI_1\times(1-SeI_2)\times SeI_3, \\ SeI_1\times SeI_2\times(1-SeI_3), SeI_1\times SeI_2\times SeI_3 \end{bmatrix}\right)$$

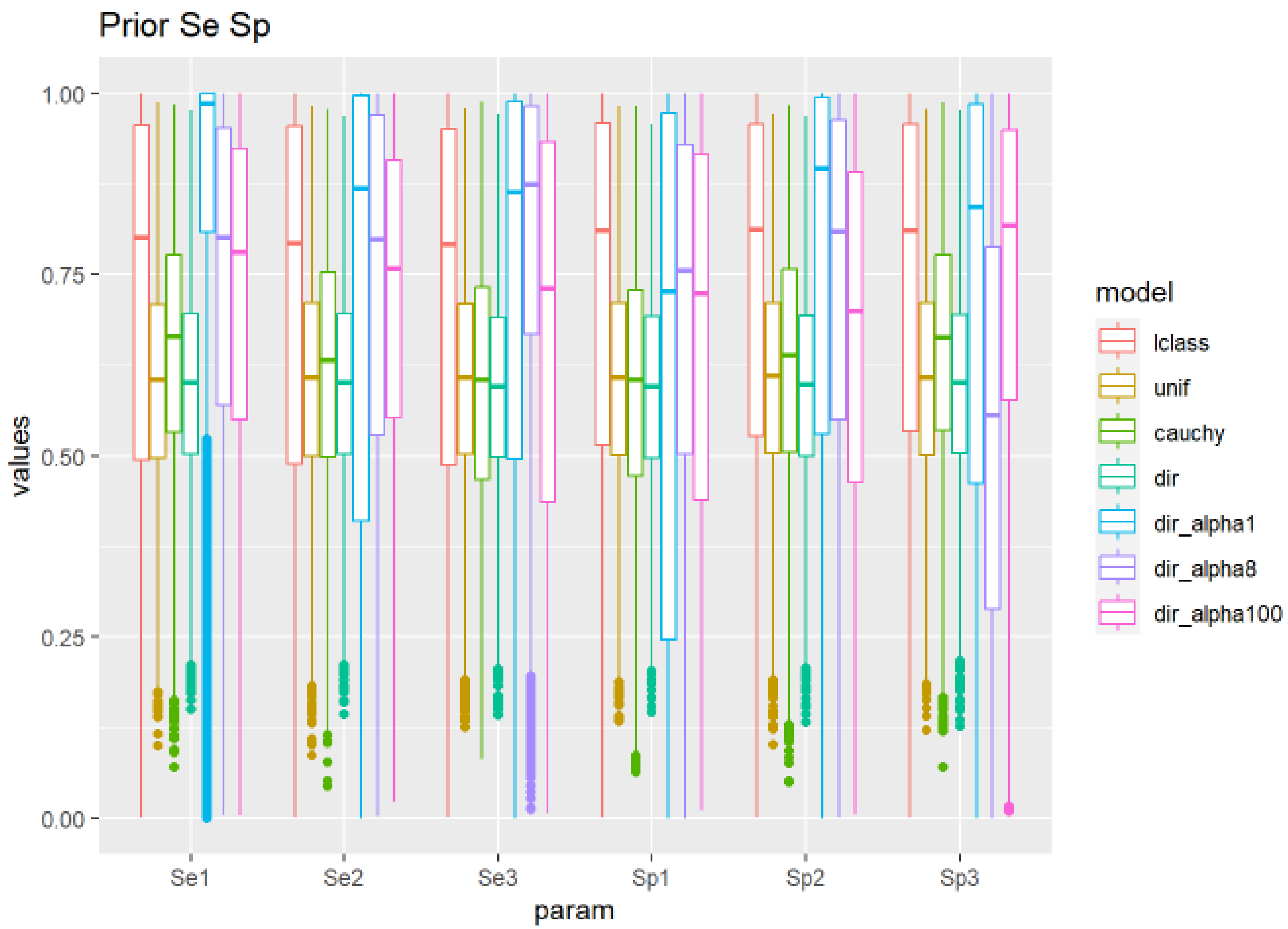
$$SeI_{\bullet} \sim Beta(0.5, 0.5)$$

$$Pneg_{[1:8]} \sim Dirichlet(\alpha, [\ldots])$$

$$SpI_{\bullet} \sim Beta(0.5, 0.5)$$

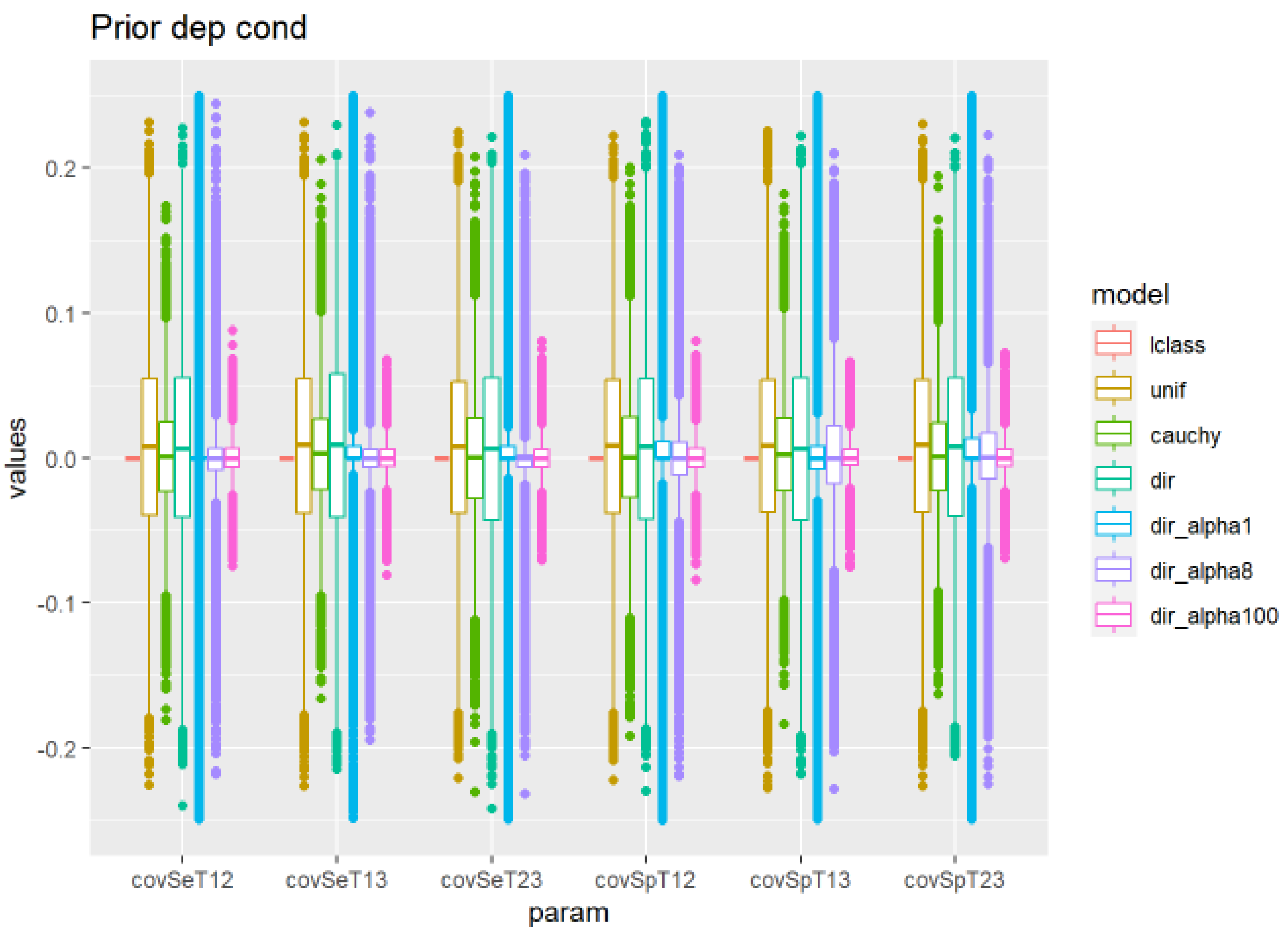
Distribution a priori Se-Sp

- Distribution a priori assez similaire quelques soit le modèle

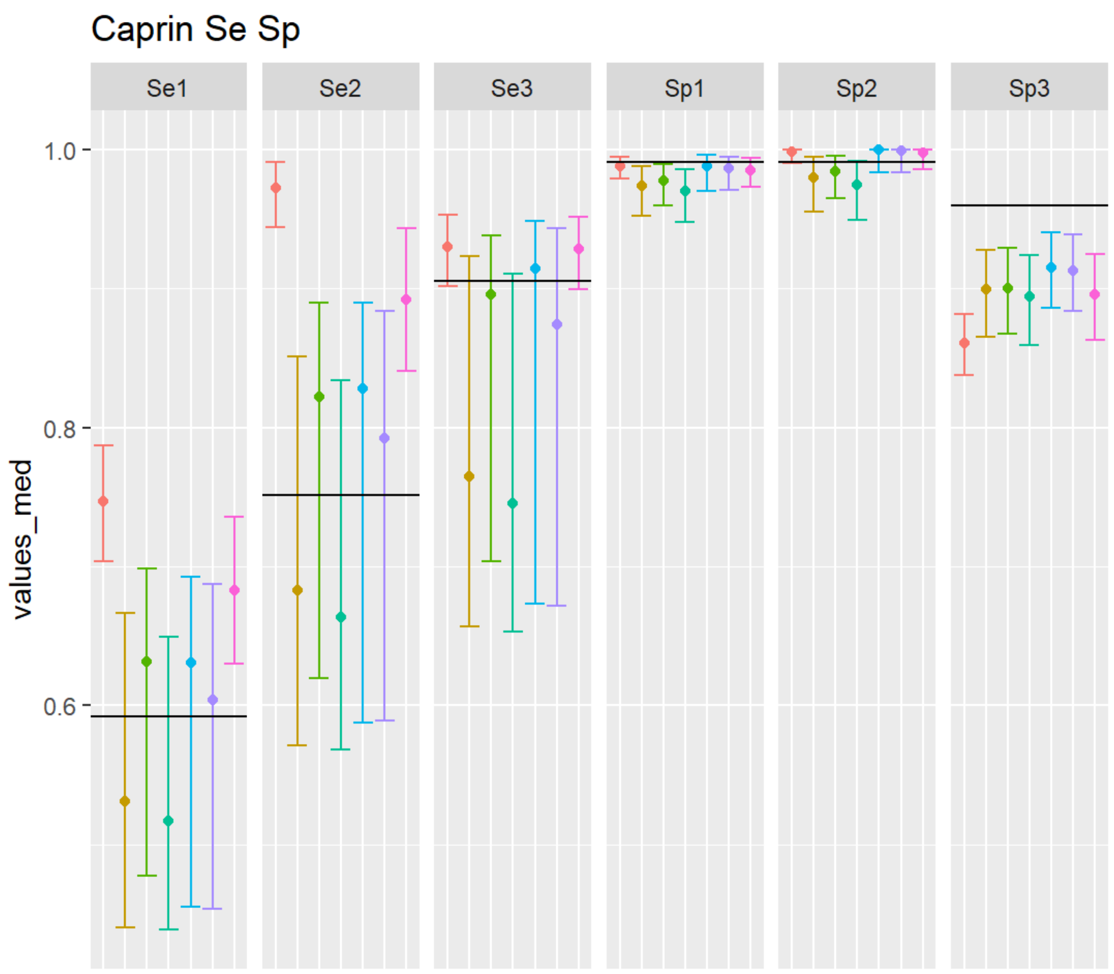


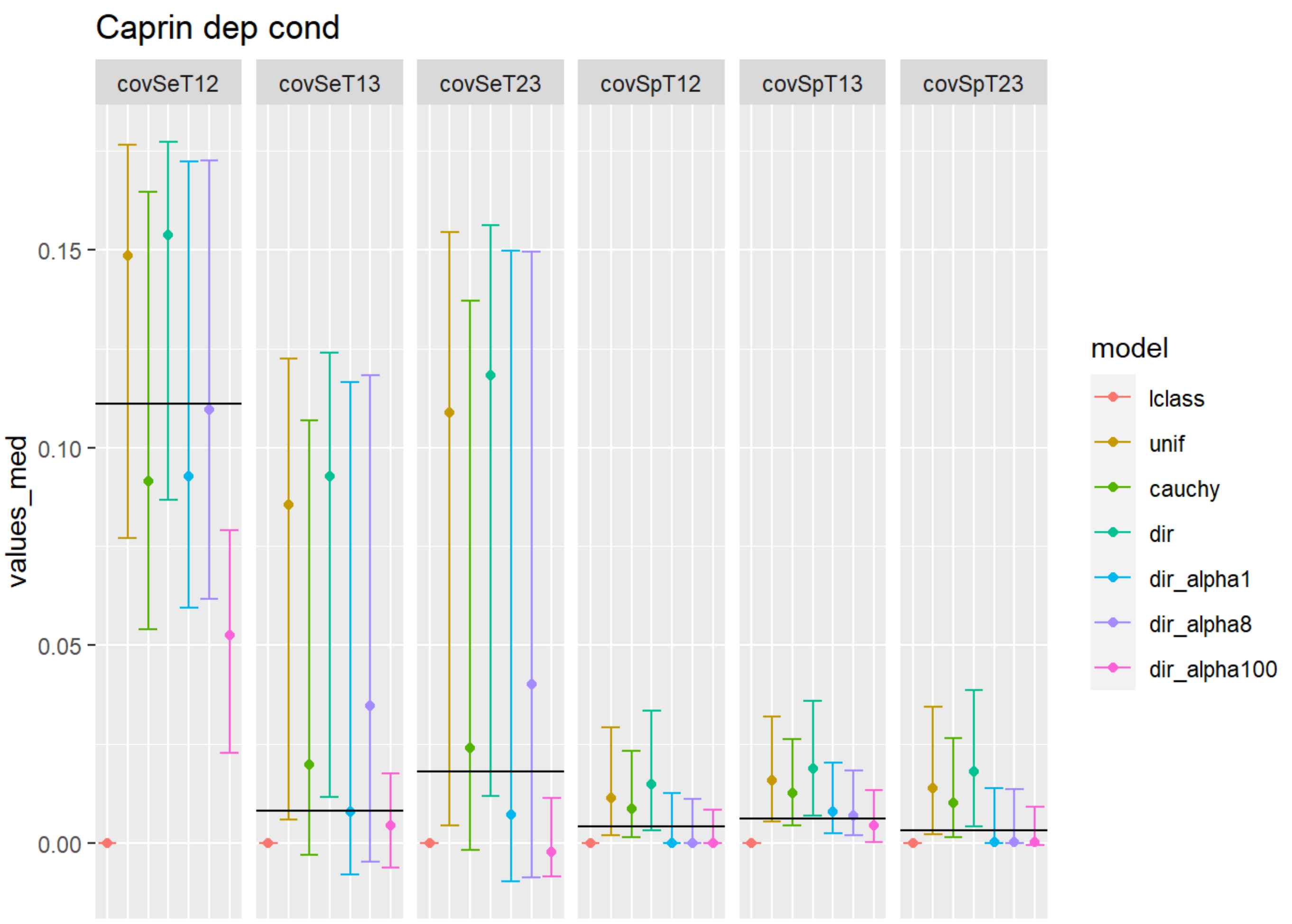
Distribution a priori dep-cond

- Dependance conditionnelle
- Lclass : nulle
- Unif : assez large
- Cauchy : plus étroite
- Dir : idem Unif
- Alpha 1 : centrée en 0 mais queue large
- Alpha8 : idem
- Alpha 100 : proche Lclass

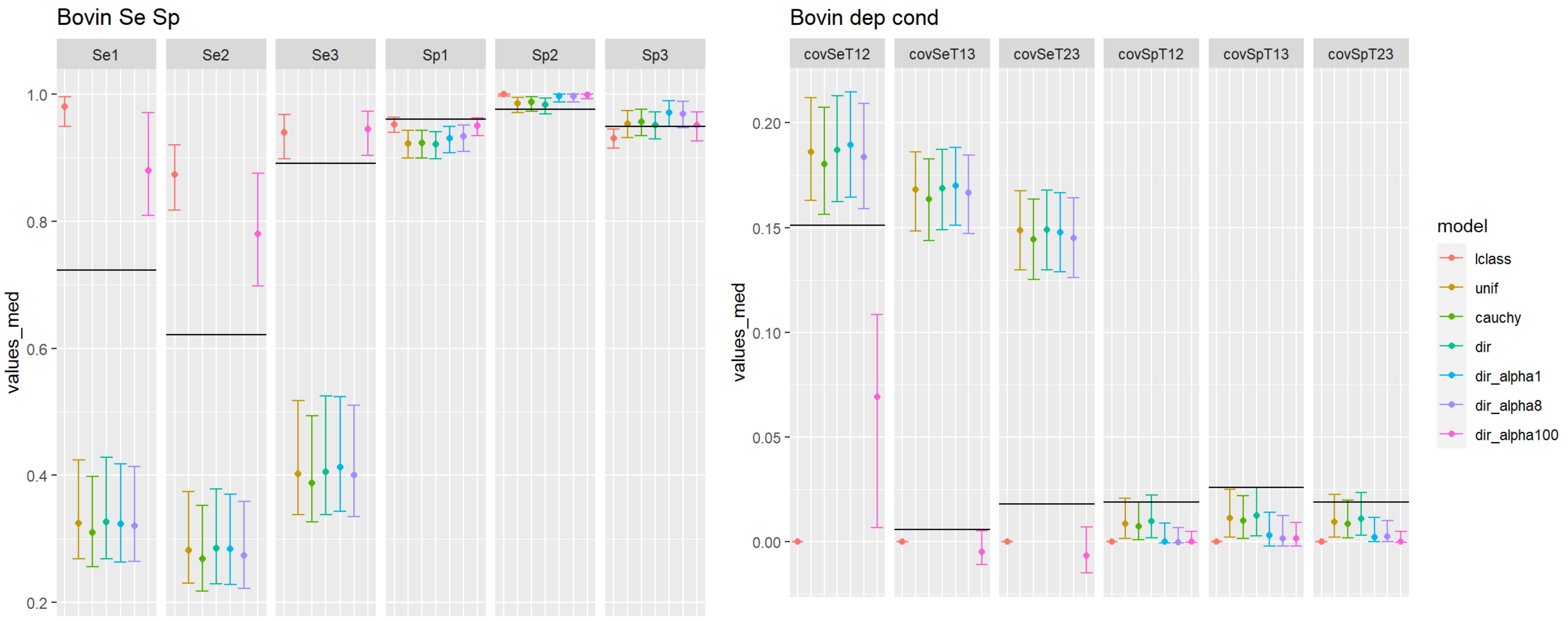


Estimation caprin

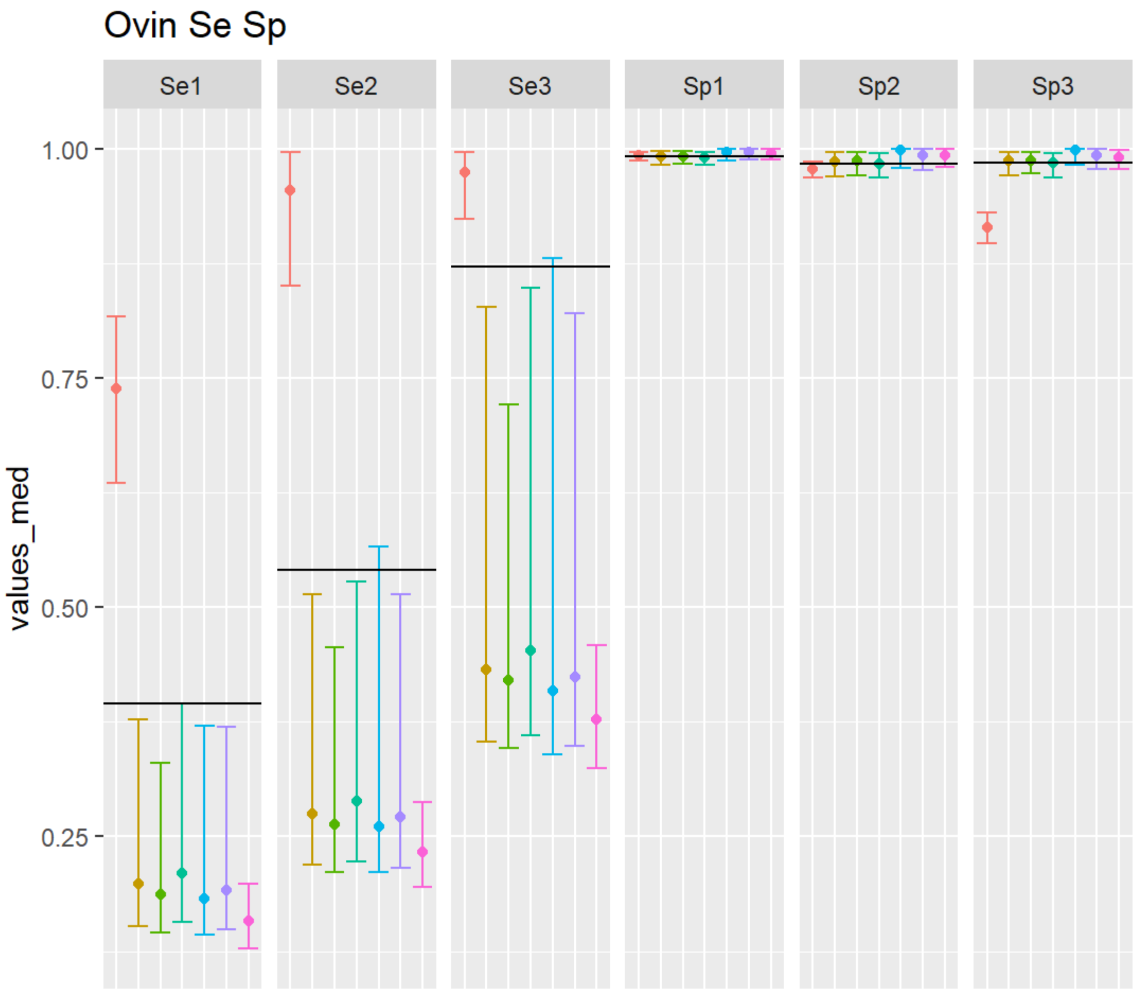


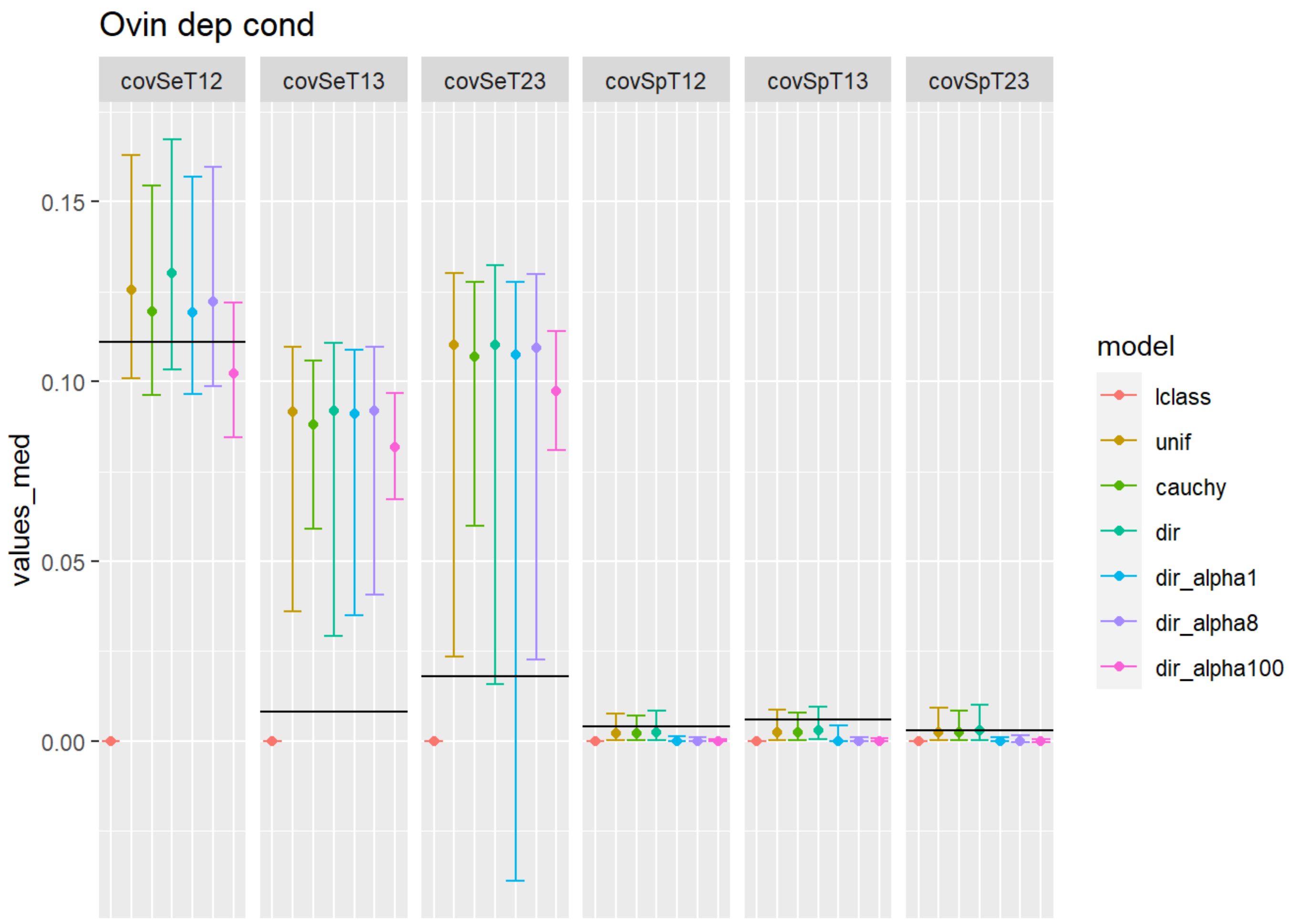


Estimation bovin



Estimation ovin



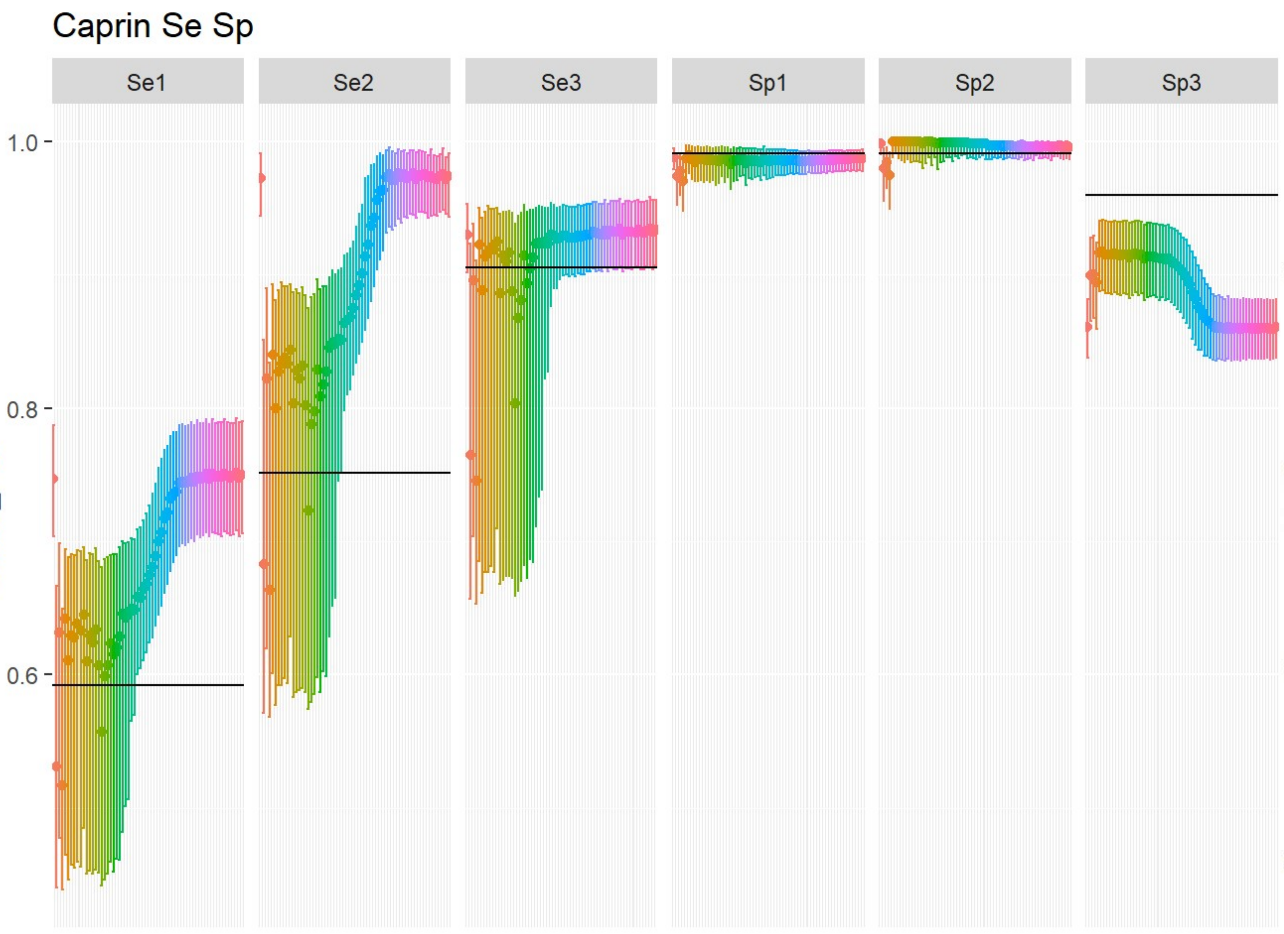


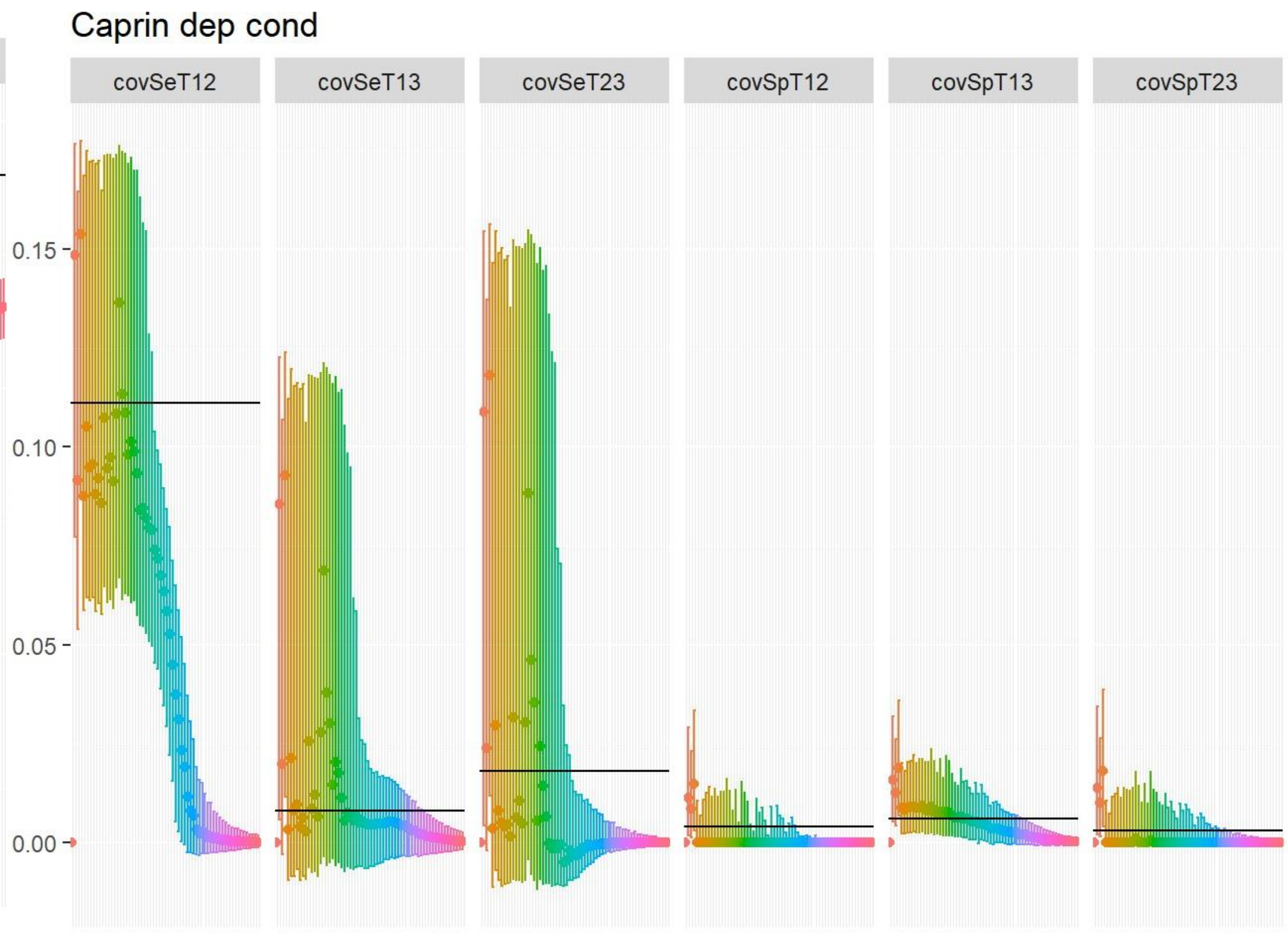
Bilan

- Dirichlet sans hyperparametrisation :
  - Comportement proche de modèle unif, forte dependance conditionnelles à posteriori => biais négatifs potentiellement important
- Dirichlet avec hyperarametrisation : alpha 1, alpha 8, alpha 100
  - Plus alpha est grand plus on est proche de Lclass (indépendance)

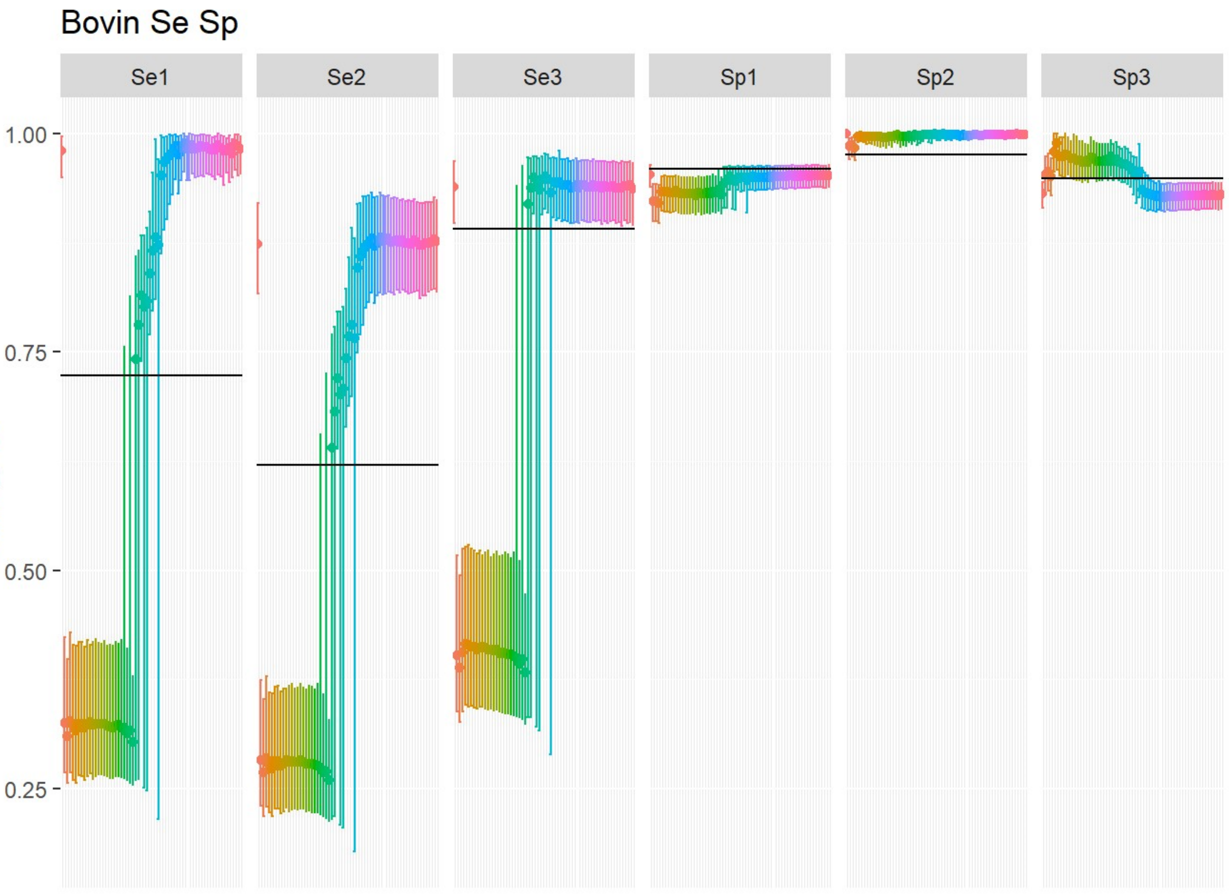
$\Rightarrow$Est-ce qu'en testant plus de valeur de alpha, on peut identifier les solutions attendues avec le jeu de données sans structuration à l'échelle de l'élevage?

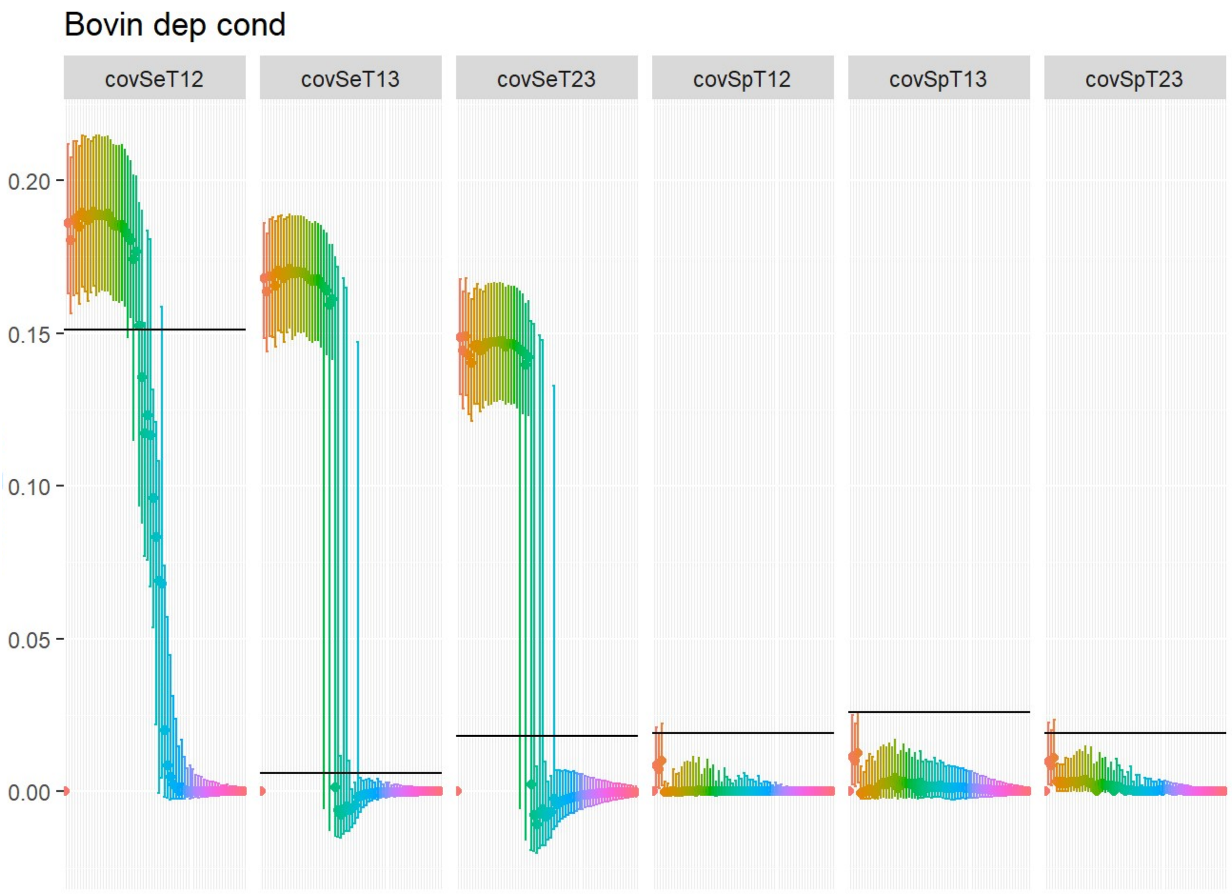
Estimation caprin



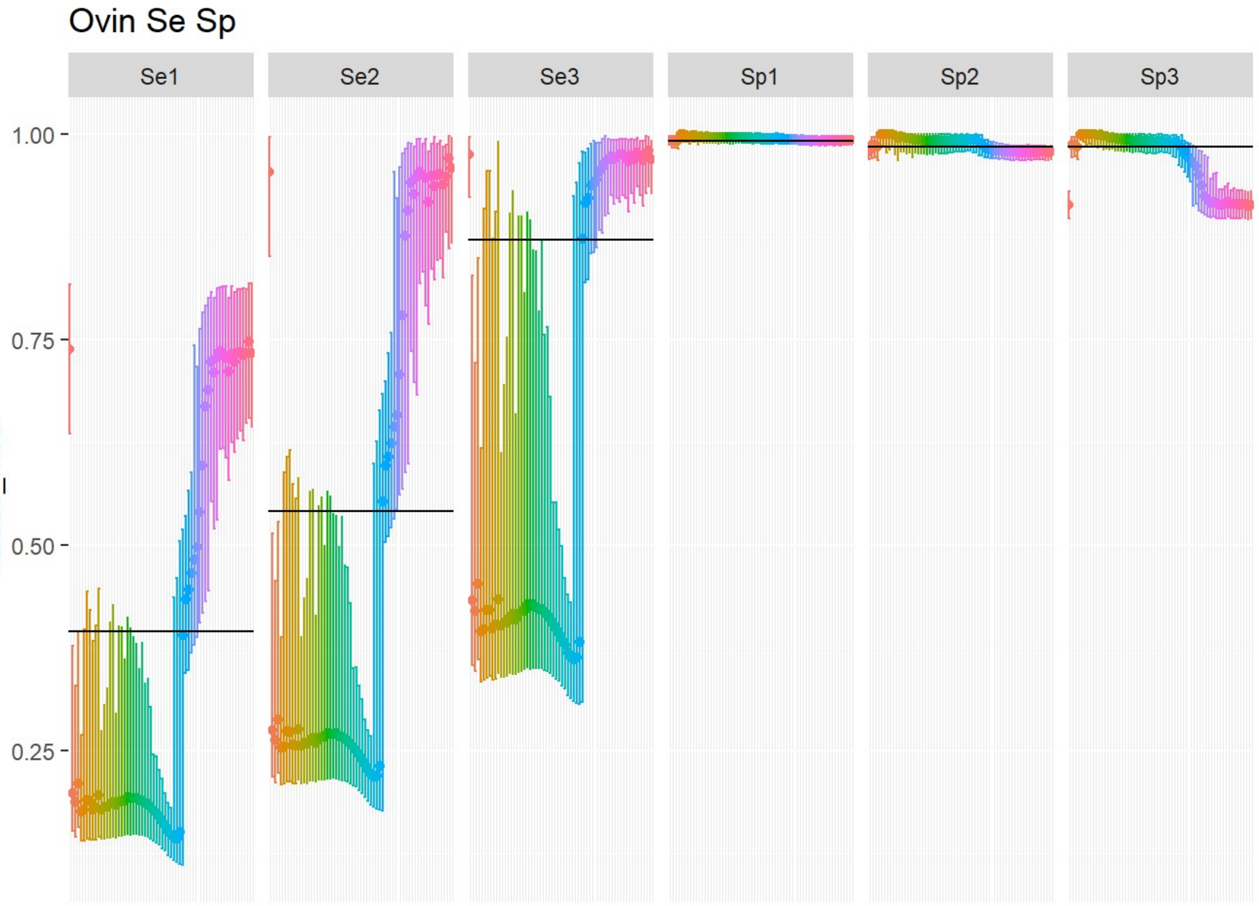


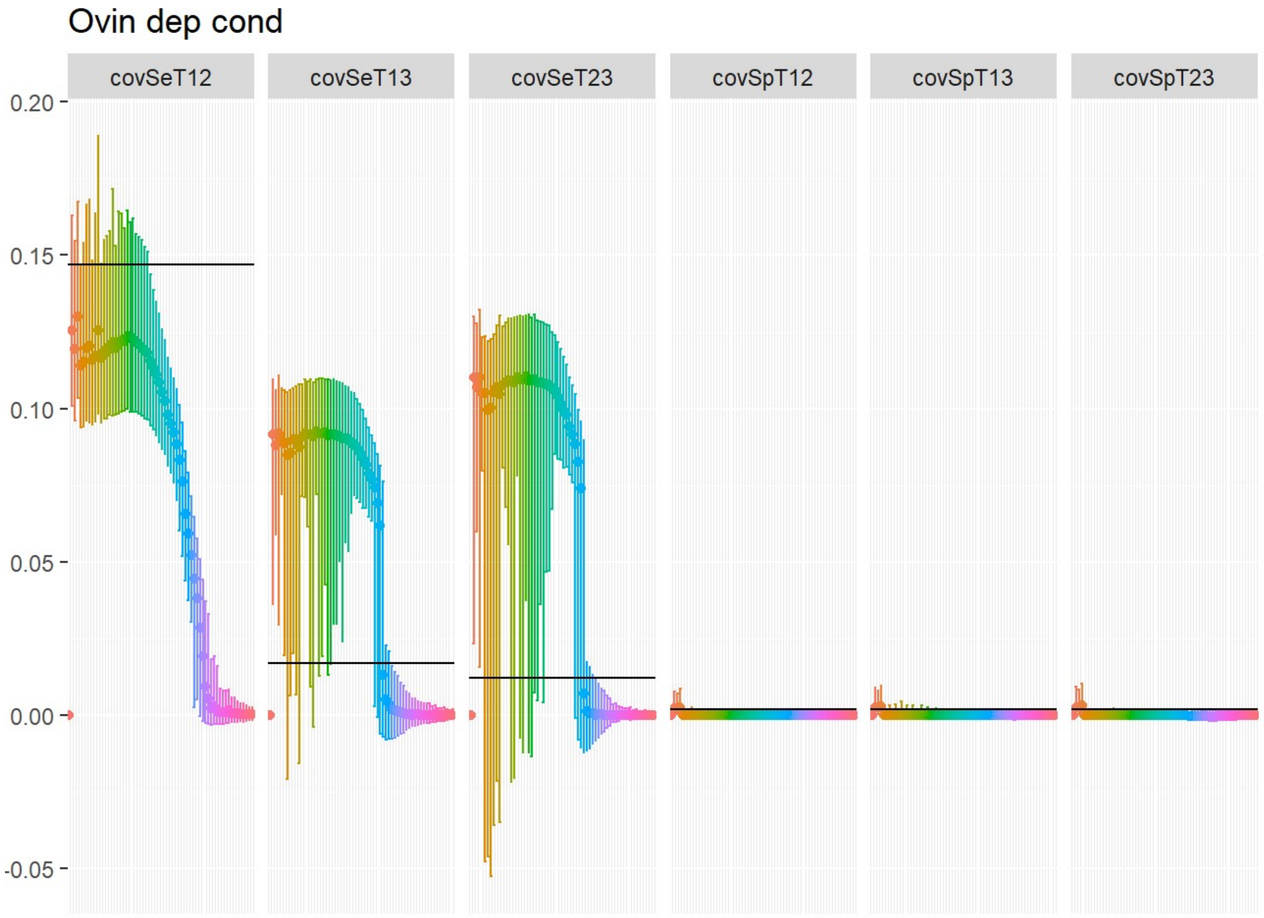
Estimation bovin



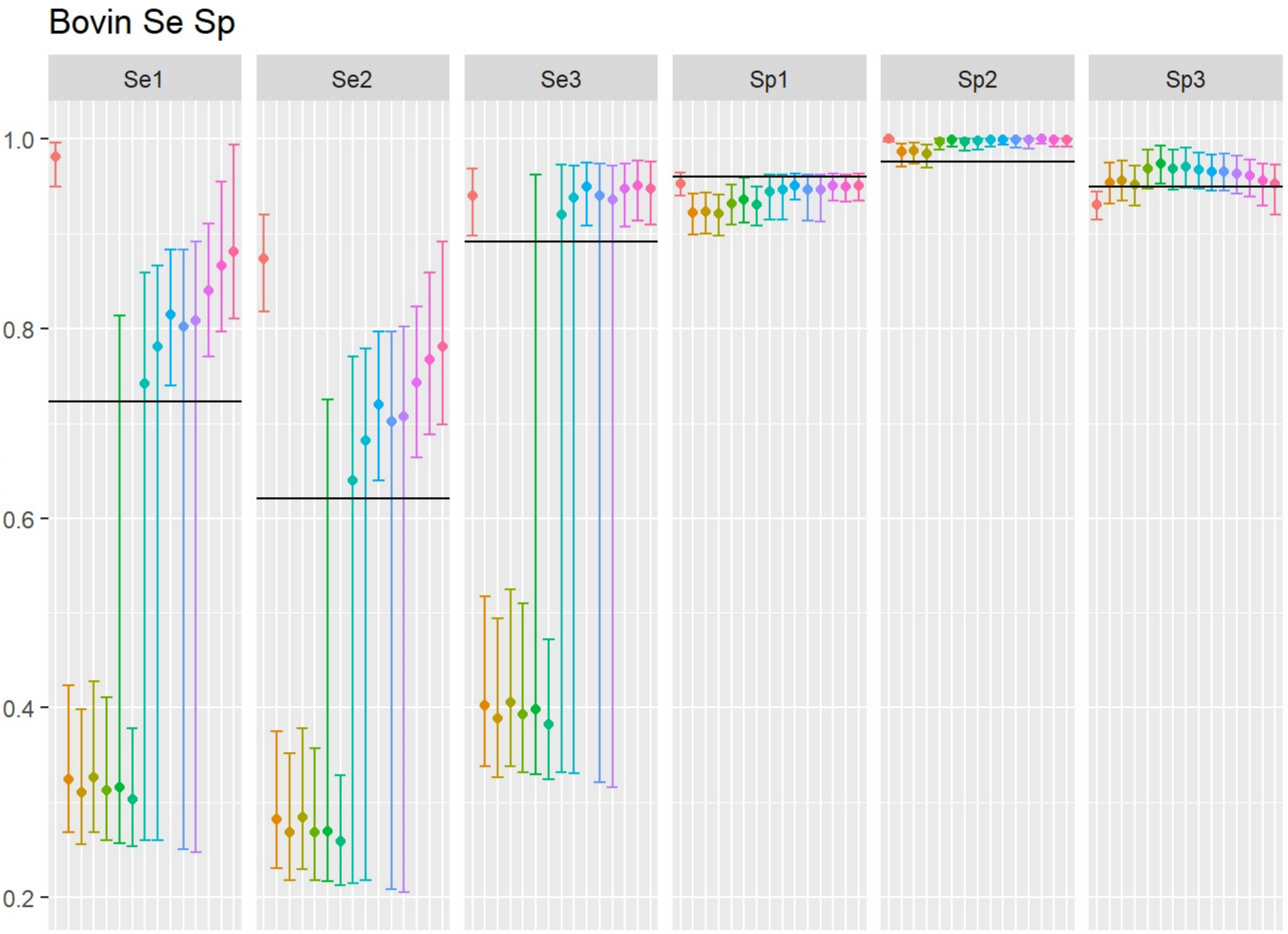


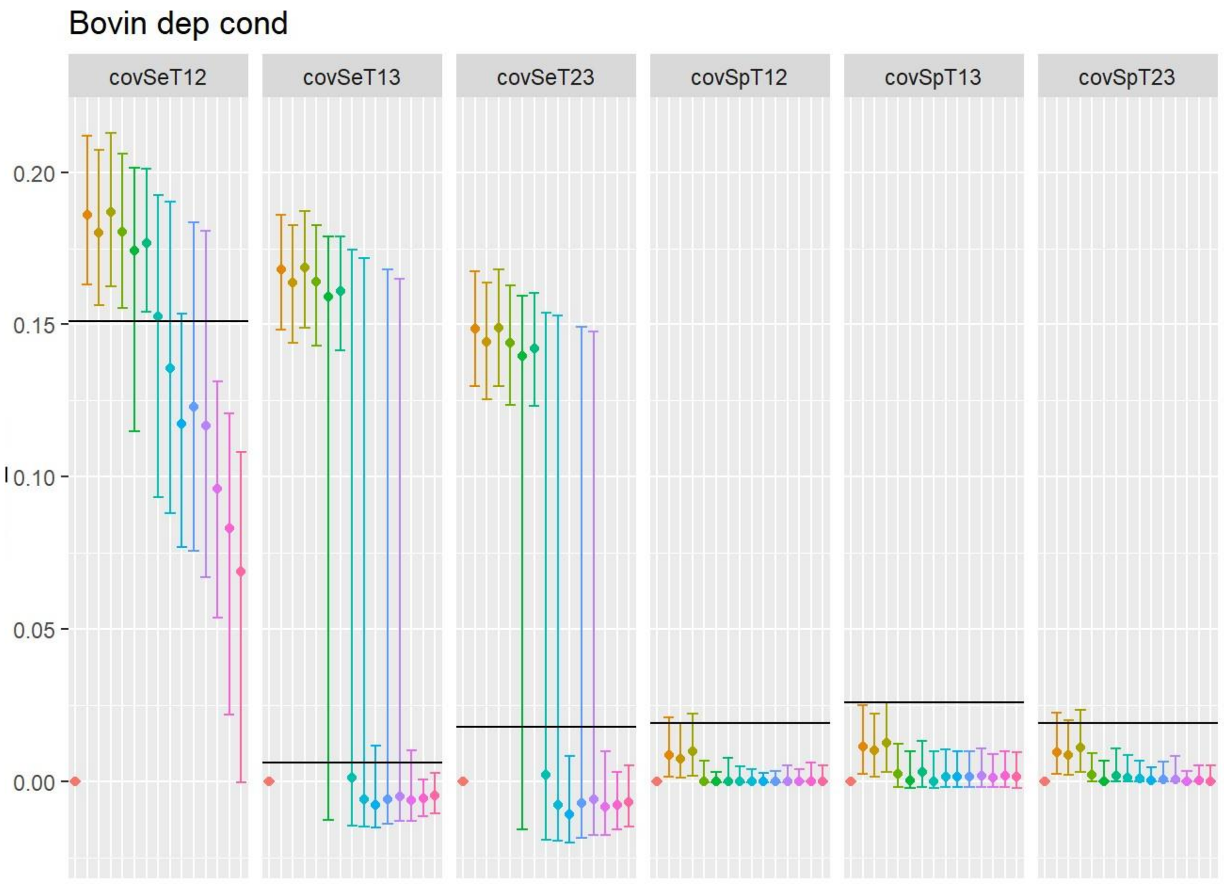
Estimation ovin



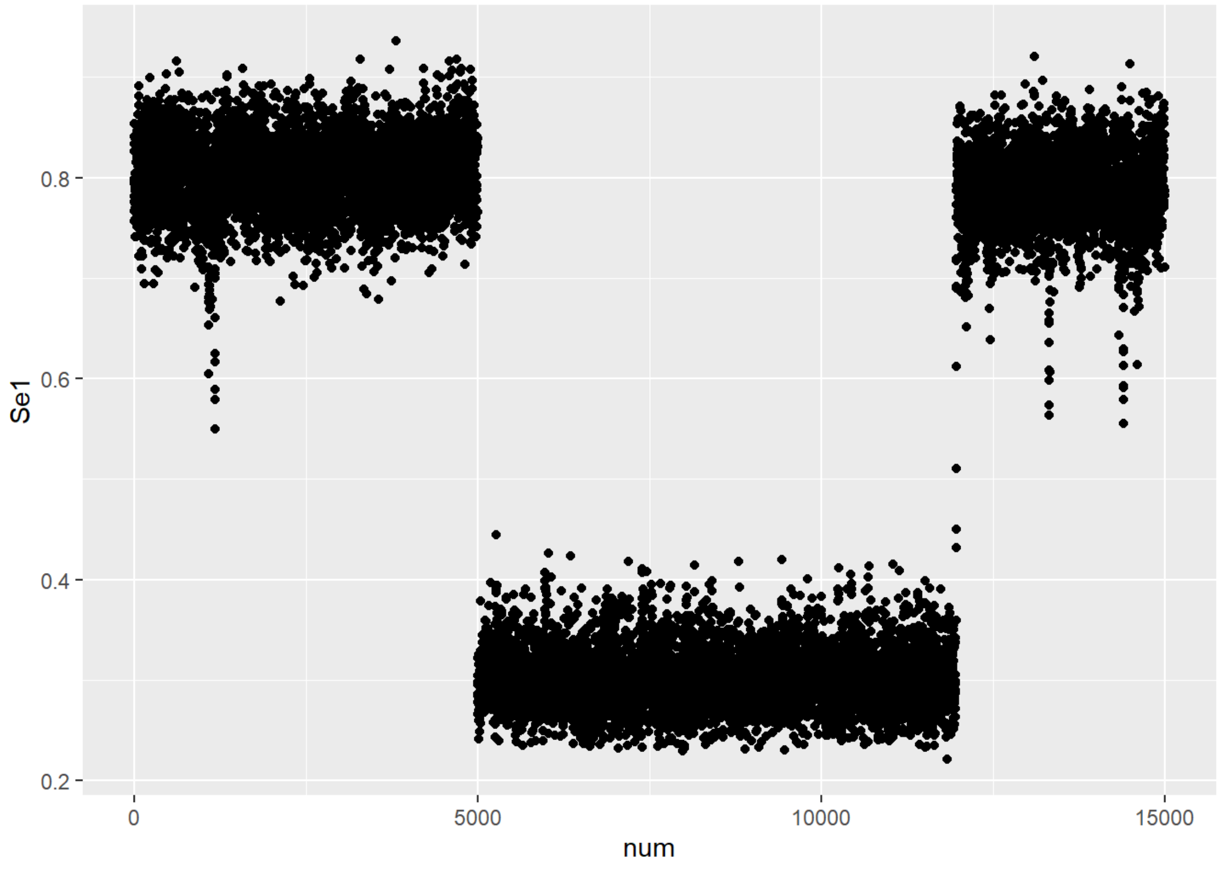


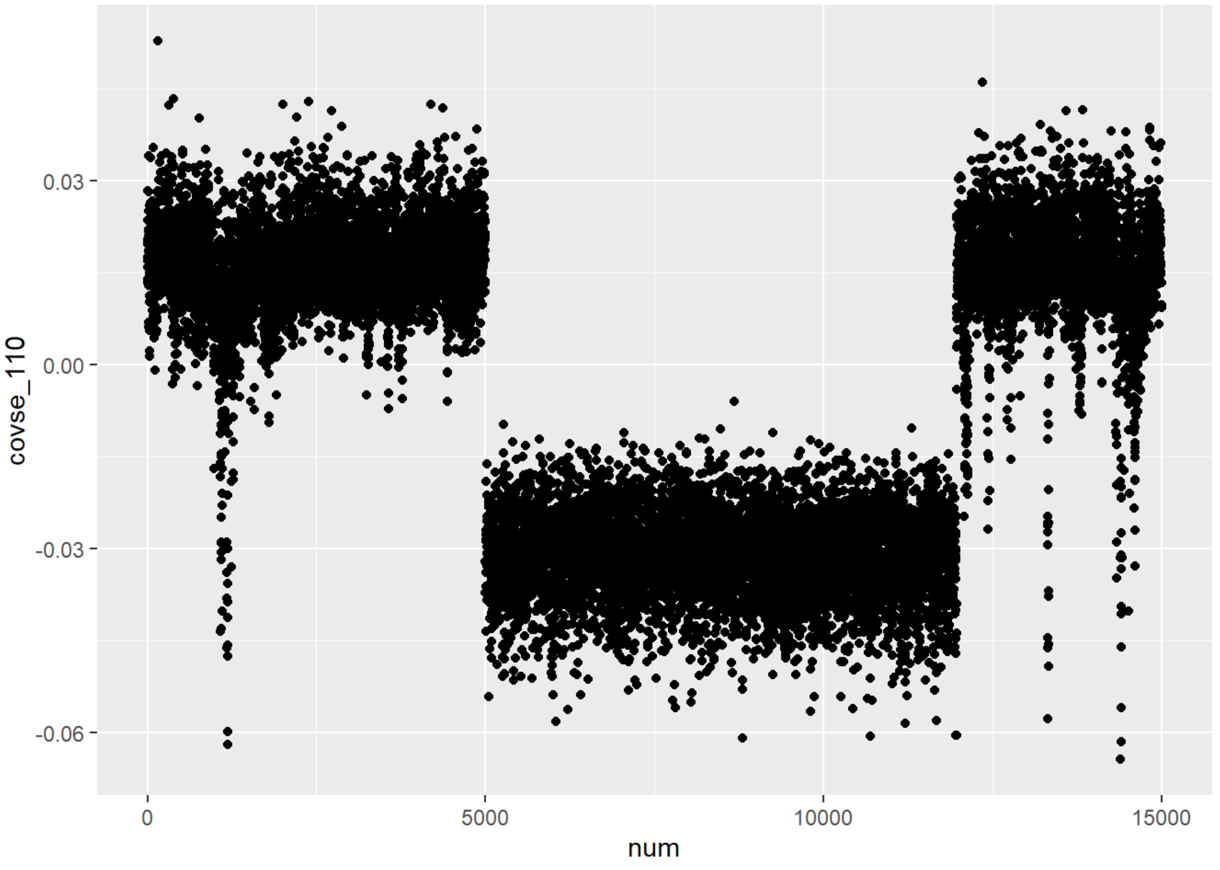
Estimation bovin





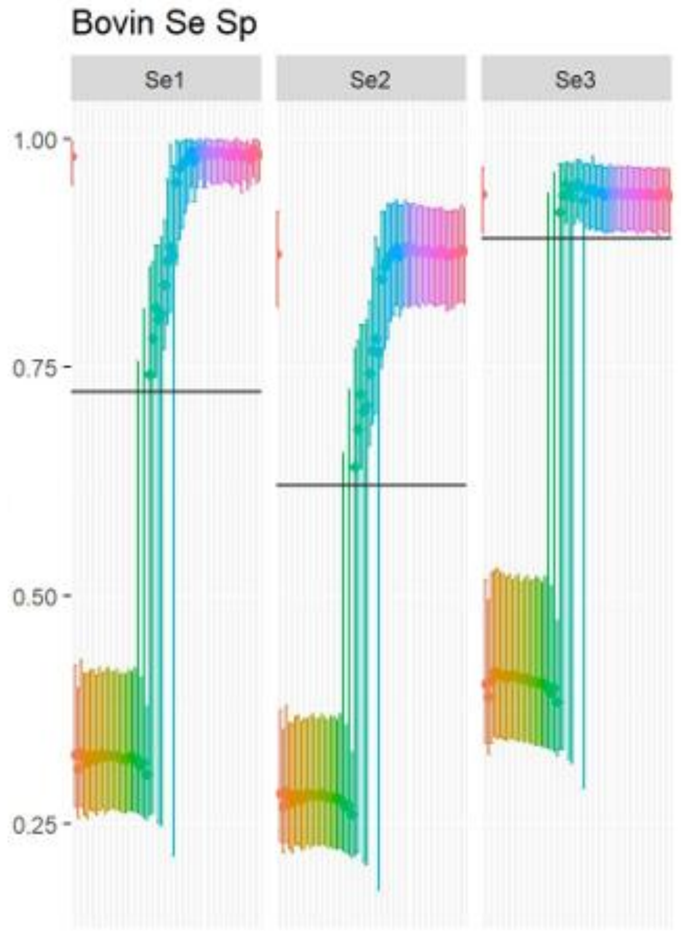
Trace, Se, CovSe, alpha = 22

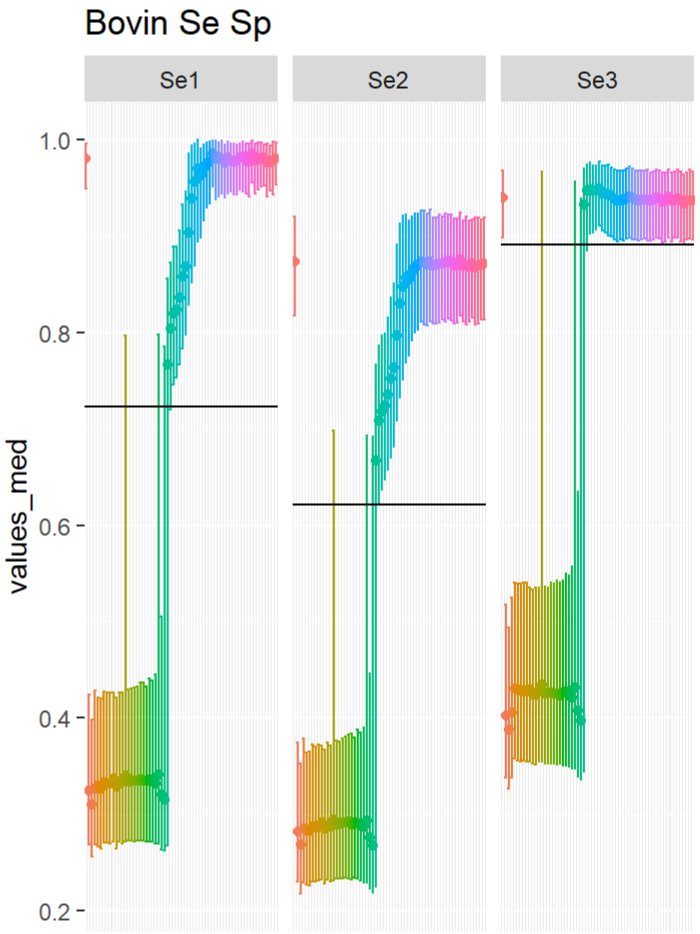


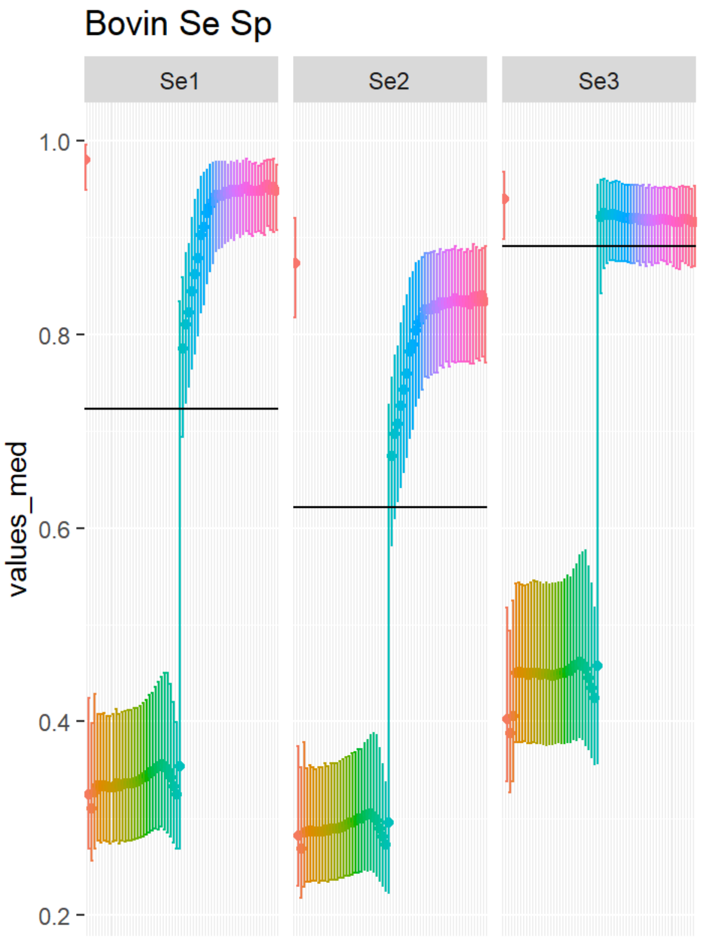


Distribution a priori sur les Se et Sp

$Beta(0.5, 0.5)$ $Beta(1,1)$ $Beta(5,5)$





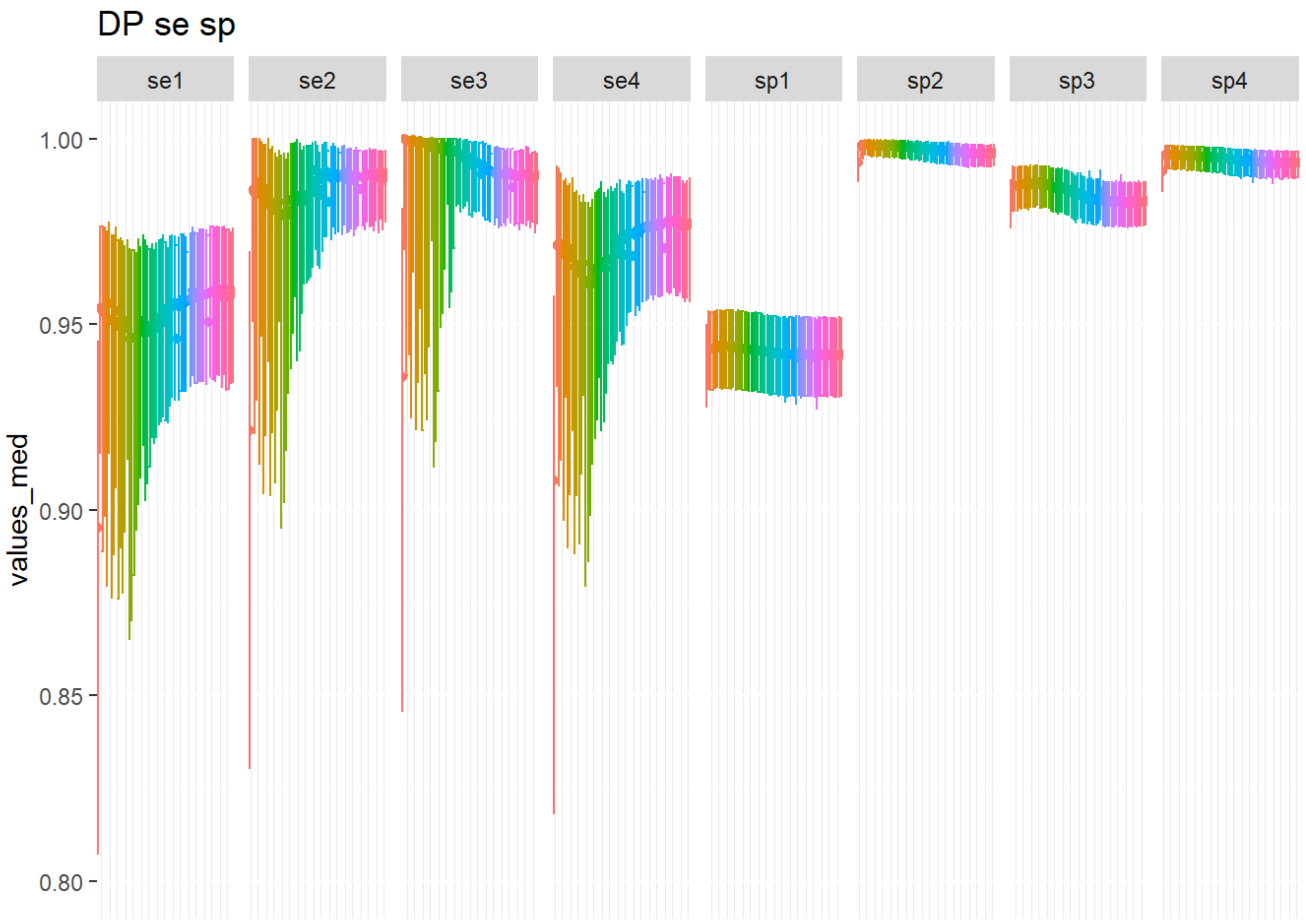


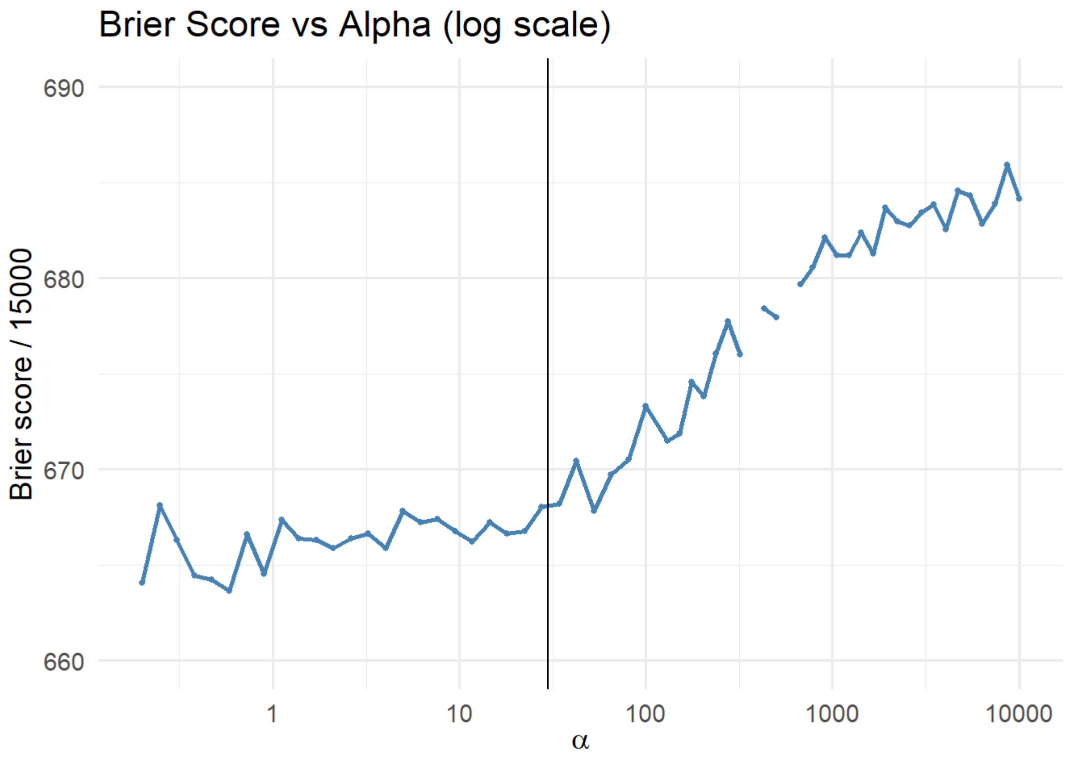
Bilan

- LCM paramétré en mélange de Dirichlet
  - Estimation similaire aux LCM classiques qui tendent à surestimer la dépendance conditionnelle
  - Plus facile à implémenter
  - Pas de contrainte sur les paramètres de dépendance conditionnelle
  - Plus rapide
- Avec hyper-paramétrisation
  - Exploration possible de solution avec moins de dépendance conditionnelle
  - Questionnement systématique de l’hypothèse d’indépendance conditionnelle?

Sélection de la valeur de alpha?

- Scores de prédiction:
  - Pseudo Score de brier pour données multinomiales?





Merci de votre attention