

# Calibration et analyse de sensibilité du modèle

Le modèle possède cinq paramètres inconnus à calibrer, à savoir :

- $\gamma$  le paramètre relatifs aux individus exogènes ;
- $p_m$  le paramètre régulant les échanges intra-bloc ;
- $\mu_{ER}$  la probabilité de survie à l'enherbement ras ;
- $\mu_{EH}$  la probabilité de survie à l'enherbement haut ;
- $k$  le paramètre lié à la capacité d'accueil des inflorescences.

## 1 Qualité d'ajustement

Avant une quelconque estimation des paramètres, nous avons besoin de mesurer la qualité de nos estimations. Pour ce faire, on utilisera une fonction pour comparer le nombre de larves estimées avec le nombre de larves observés. Pour définir cette fonction, on pose :

- $m$ , le nombre de jours entre la première observation et la dernière ;
- $n$ , le nombre de relevés effectif ;
- $t$ , le nombre de jours passés depuis la première observation ;
- $t^j$ , le nombre de jours entre la première observation et le  $j^{\text{ème}}$  relevé.

(On a donc  $t^1 = 0$  et  $t^n = m$ .)

La fonction est alors définie par

$$f(y, \hat{y}) = \frac{\sqrt{\frac{1}{n} \sum_{j=2}^n (y_j^* - \hat{y}_j^*)^2}}{\max(y) - \min(y)},$$

où

$$y_j^* = y_{t^j} \quad \text{et} \quad \hat{y}_j^* = \frac{1}{t^j - t^{j-1}} \sum_{k=t^{j-1}}^{t^j} \hat{y}_k.$$

La figure 1 illustre le fonctionnement de la fonction. Appliquer la fonction à  $n - 1$  valeurs (correspondant aux relevés) plutôt qu'à chacun des  $m$  jours (correspondant à l'étendue des relevés) est motivé par le fait que les relevés ne furent pas fait à des intervalles de temps régulier. Ce faisant, on évite d'attribuer plus d'importance aux relevés qui ont eu un écart relativement important avec le relevé précédent.

Par la suite, l'objectif sera de minimiser cette fonction.

## 2 Analyse de sensibilité

Afin d'avoir une meilleure idée de l'influence des paramètres sur le modèle, on réalise une analyse de sensibilité du modèle. On utilisera la méthode Sobol, l'implémentation utilisée est décrite dans [Saltelli et al., 2008]. Il n'y a cependant pas de méthode permettant d'avoir une idée des intervalles dans lesquels évoluent les paramètres. Si pour les probabilités il paraît normal de les faire évoluer entre 0 et 1, les intervalles pour les paramètres  $\gamma$  et  $k$  sont plus délicats à choisir.

On choisira ici l'intervalle  $[0; 0.1]$  pour  $\gamma$  et l'intervalle  $[1; 70]$  pour  $k$ . Ces choix sont motivés par des résultats de calibration antérieurs qui proposaient des valeurs pour lesdits paramètres dans des intervalles strictement inclus dans ceux qui seront utilisés. Cela reflètera ainsi au mieux le comportement réel du modèle.

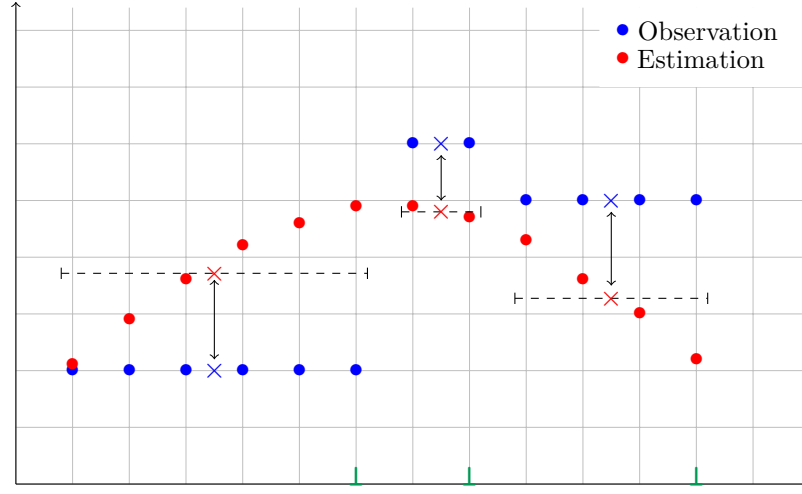


FIGURE 1 – Schéma illustrant le fonctionnement de la fonction objectif. À chaque relevé effectif (marqueurs verts), on fait correspondre la période correspondant à ce relevé (segments en pointillés). Et pour chacune de ces périodes, on calcule la moyenne des valeurs estimées (les croix rouges). On compare ensuite les moyennes ainsi calculées avec les valeurs observées associées (les croix bleues).

Les résultats sont visibles sur la figure 2. Il est intéressant de noter que les paramètres ont des effets différents d'un sous-bloc à l'autre, ce qui est rassurant. Le paramètre  $\gamma$  est important pour les trois sous-blocs, avec des interactions non-négligeables. Les probabilités de survie à la modalité de couverture du sol sont également important, variant selon les sous-blocs. Ainsi  $\mu_{ER}$  apporte beaucoup de variance pour les sous-blocs «ER» et «PS»; il n'apporte aucune variance pour le sous-bloc «EH». Le paramètre  $\mu_{EH}$  est important pour le sous-bloc «EH» et négligeable pour les deux autres. Le paramètre  $p_m$  apporte une variance comparable pour les trois sous-blocs. Enfin, le paramètre  $k$  n'a — semble-t-il — aucun impact sur le modèle.

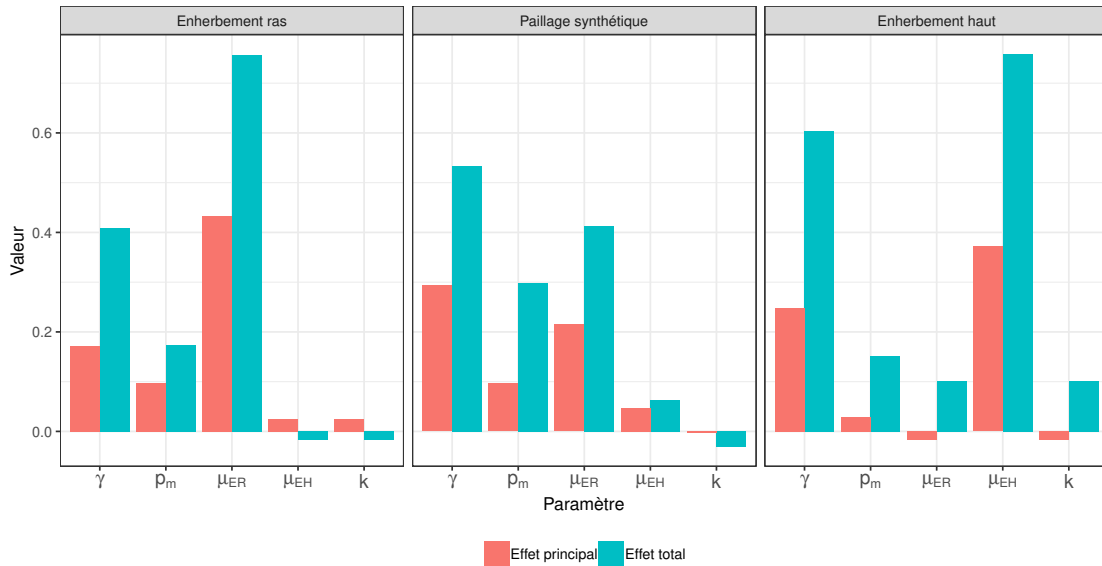


FIGURE 2 – Analyse de sensibilité du modèle. Méthode Sobol

### 3 Calibration

Une première calibration du modèle avec NSGA-II en prenant pour critère la qualité d'estimation sur chacun des trois sous blocs donne des résultats peu convenables. En effet, le paramètre  $\mu_{EH}$  est presque toujours fixé à 0 (ou très proche). Cela peut s'expliquer pour deux raisons. Premièrement,

ce paramètre n'est important que dans un seul des trois sous-blocs ; il n'est donc pas très important pour le modèle dans sa globalité. Deuxièmement, le modèle (dans sa version actuelle) n'a que le paramètre  $\gamma$  pour parvenir à une baisse en fin de saison. Il utilise donc en premier lieu le  $\gamma$  puis complète au besoin avec des individus émergents (qui eux se multiplient sans moyen de stopper la dynamique).

Pour pallier ce problème, l'idée est d'alors de rajouter un critère permettant la préférence des individus émergents sur les individus exogènes. Mais pour cela, il faut d'abord enlever un des trois critères déjà présent car NSGA-II devient mauvais au delà de trois critères [Ge et al., 2018].

À cette fin, on a lancé 5 répétitions de NSGA-II (*popsiz* = 100 et *generations* = 200). On a ainsi pu récupérer 500 solutions accompagnés de la valeur de la *NRMSE* produite sur chaque sous-bloc. Les corrélations trouvées sur ces valeurs sont les suivantes :

$$\rho(ER, PS) = -0.869 \quad \rho(ER, EH) = 0.761 \quad \rho(PS, EH) = -0.943$$

Les sous-blocs PS et EH sont très corrélés négativement. Baisser l'erreur de l'un entraîne une augmentation de l'autre. Comme le montre la figure 3 l'erreur associée à EH est plus grande que celle associée à PS. En retirant le critère PS, cela contribuera à baisser l'erreur sur l'enherbement haut de sorte que les erreurs entre PS et EH seront plus équilibrées. Cela libère également un critère permettant d'introduire notre nouveau critère.

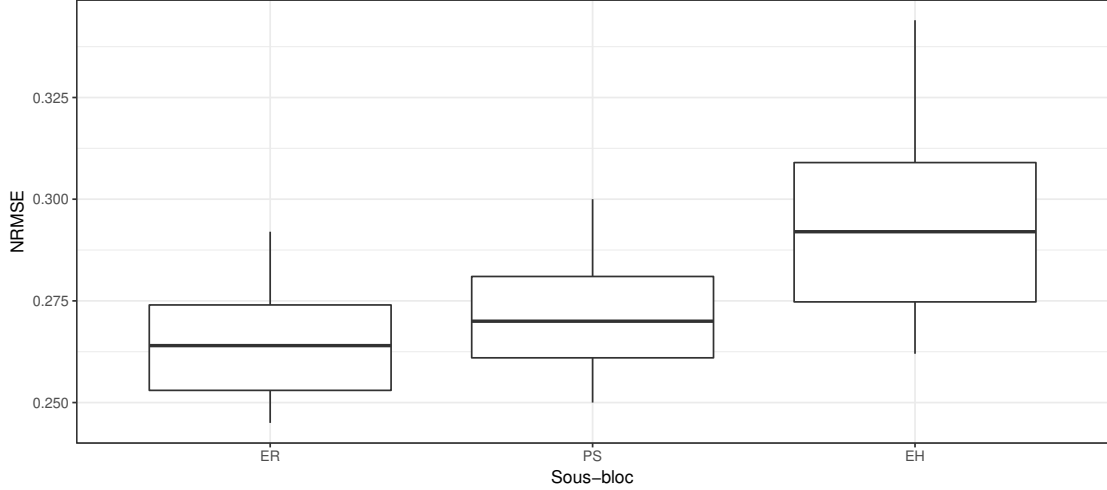


FIGURE 3 – Comparaison des valeurs de la NRMSE des solutions proposées par NSGA-II en fonction des trois sous-blocs.

Ce nouveau critère portant sur la préférence des émergents est défini par

$$\max \sum_{t,i} \frac{N_{t,i} - N_{t,i}^{\text{exo}}}{N_{t,i}}.$$

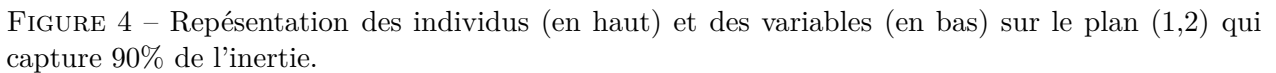
Nos trois critères sont donc maintenant : ER, EH et la maximisation des émergents.

Il ne reste plus maintenant qu'à lancer NSGA-II plusieurs fois et à en extraire le front de Pareto.

## 4 Résultats

Pour la calibration des paramètres en fonction de nos trois critères, on relance NSGA-II 15 fois avec une taille de population de 300. Il s'avère que parmi les 4500 solutions certaines sont dominées par d'autres. Au final, il nous reste 1622 solutions.

Choisir la solution qui minimise la distance à 0 n'est pas toujours la solution la plus pertinente vis à vis de notre problème. Il peut être intéressant de rechercher, d'explorer parmi les solutions proposées.

[illegible]

1. Une régression PLS2 serait même mieux dans la mesure où cela permettrait de mieux mettre en évidence l'impact

souvent fixé à 0 avant l'introduction du critère sur la maximisation des endogènes. Un phénomène comparable — mais de moindre envergure — est visible pour  $\mu_{ER}$  et le critère ER.

On observe une corrélation positive entre les critères ER et PS. Il y a aussi une corrélation positive entre les deux autre critères, mais il faut maximiser le critère de préférence des émergents et minimiser le critère EH : il y a donc un antagonisme entre l'amélioration de ces deux critères.

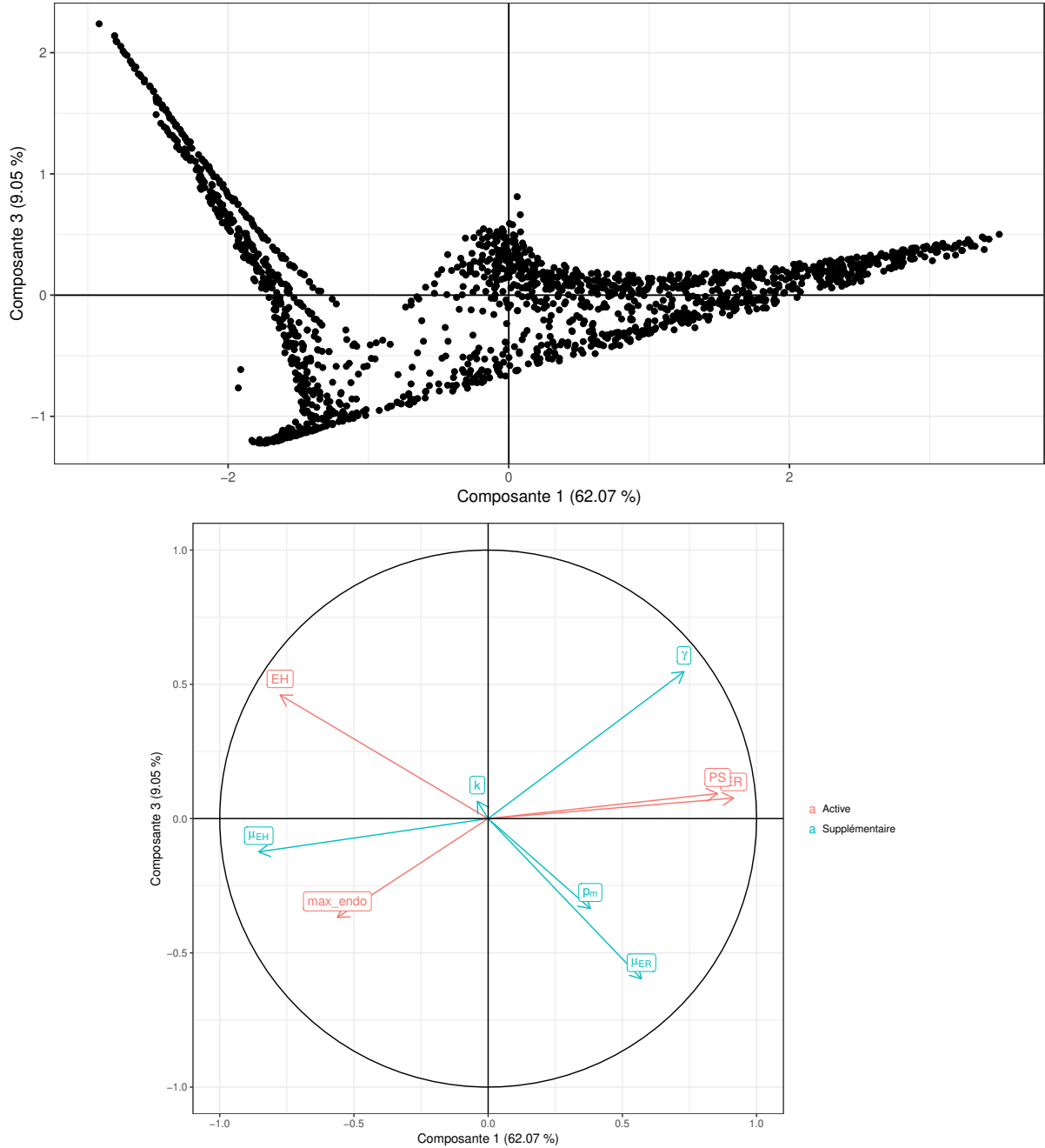


FIGURE 5 – Représentation des individus (en haut) et des variables (en bas) sur le plan (1,3) qui capture 71% de l'inertie.

Le plan (1,3) (figure 5) met en exergue des phénomènes moins évidents. Il semblerait que le paramètre  $\mu_{EH}$  soit corrélé négativement avec les critères PS et ER. Il en va de même pour le paramètre  $\mu_{ER}$  et le critère EH. Ainsi, donc maximiser les paramètres  $\mu_{ER}$  et  $\mu_{EH}$  pourrait avoir de bonnes conséquences sur l'estimation dans certains sous-blocs.

En résumé, le plan (1,2) — qui capture 90% de l'inertie — montre que :

- lorsque  $\gamma$  est élevé, le critère de préférence des émergents n'est pas très bien respecté et le critère EH est bien respecté ;

- lorsque  $\mu_{EH}$  est élevé, le critère EH n'est pas bien respecté et le critère de maximisation des émergents est bien respecté ;
  - lorsque  $\mu_{ER}$  est élevé, les critères ER et PS ne sont pas très bien respectés.
- Le plan (1,3) — qui capture 71% de l'inertie — montre que :
- lorsque  $\mu_{EH}$  est élevé, les critères ER et PS sont bien respectés ;
  - lorsque  $\mu_{ER}$  est élevé, le critère EH est bien respecté.

En conclusion, les critères EH et préférence des émergents sont incompatibles. Ensuite, il n'y a pas de paramètres qui ait une valeur qui soit bénéfique pour tous les aspects du modèle. Ainsi, il va donc falloir trouver un compromis qui permette de n'être pas trop mauvais sur tous les critères.

## 5 Exemples de solutions

Pour être bon sur le critère de préférence des émergents, on peut sélectionner une solution ayant un  $\gamma$  relativement faible (voir figure 6).

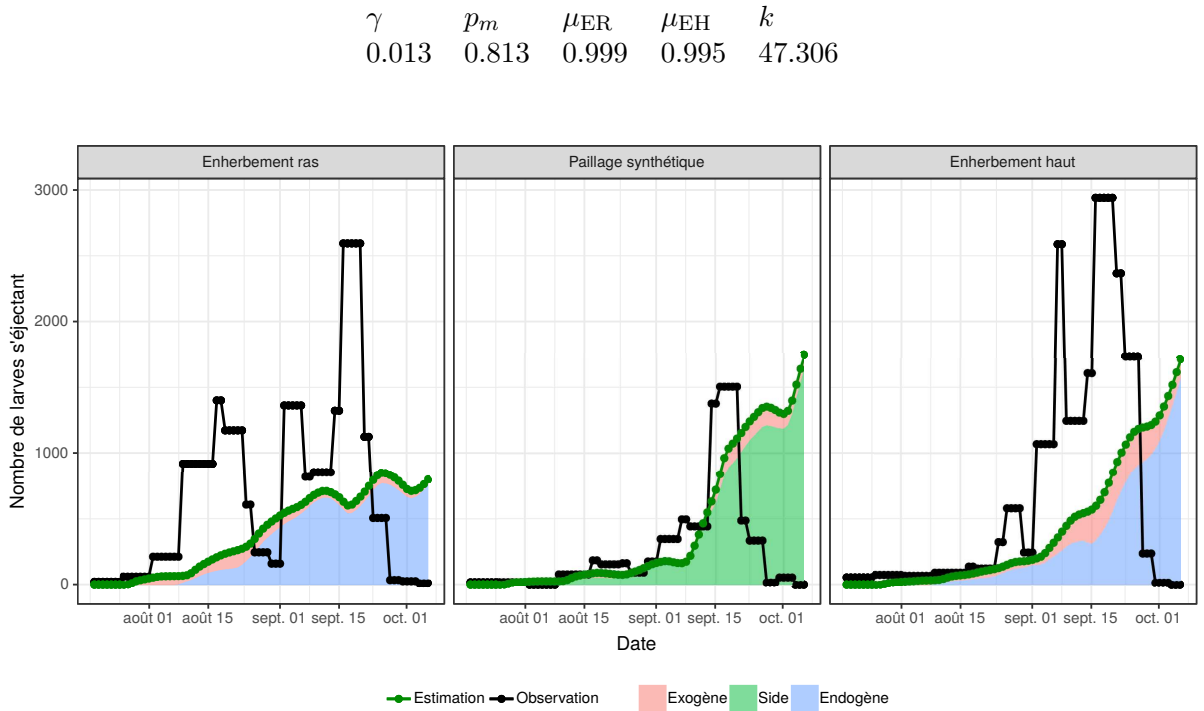


FIGURE 6 – Dynamiques estimées avec un jeu de paramètres vérifiant  $\gamma < q_{0.1}(\gamma)$ .

Pour être bon sur les critères ER et PS, on peut sélectionner une solution ayant un  $\mu_{ER}$  faible et un  $\mu_{EH}$  élevé (voir figure 7).

$\gamma$	$p_m$	$\mu_{ER}$	$\mu_{EH}$	$k$
0.046	0.790	0.497	0.971	12.543

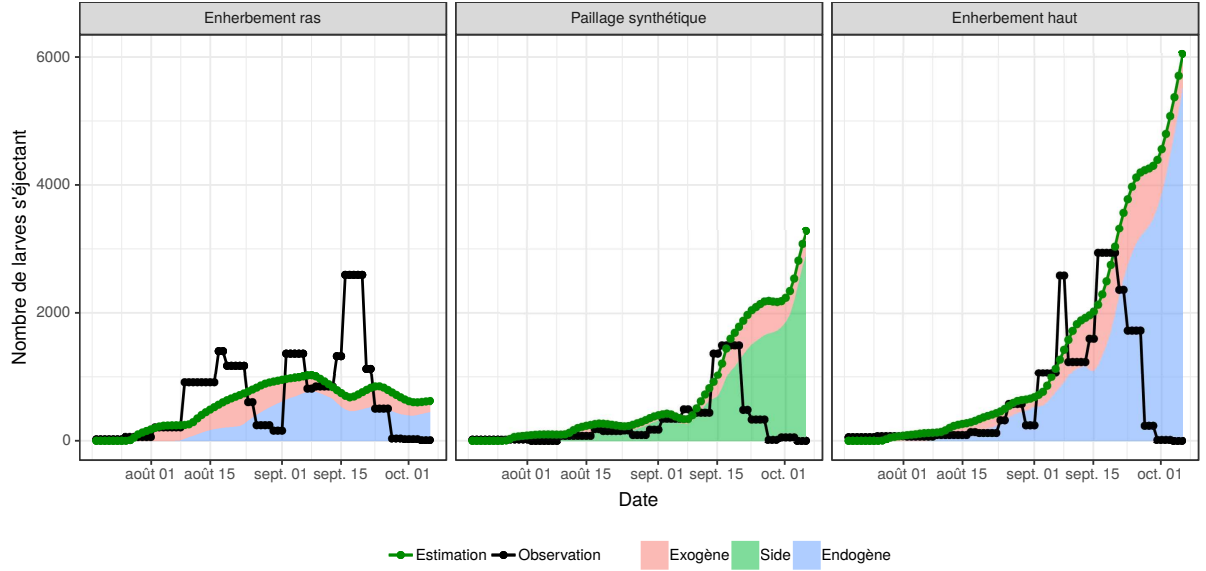


FIGURE 7 – Dynamiques estimées avec un jeu de paramètres vérifiant  $\mu_{ER} < q_{0.1}(\mu_{ER})$  et  $\mu_{EH} > q_{0.9}(\mu_{EH})$ .

On peut essayer d'être bon sur tous les critères en minimisant la norme 1 (voir figure 8).

$\gamma$	$p_m$	$\mu_{ER}$	$\mu_{EH}$	$k$
0.010	0.591	1.000	1.000	54.119

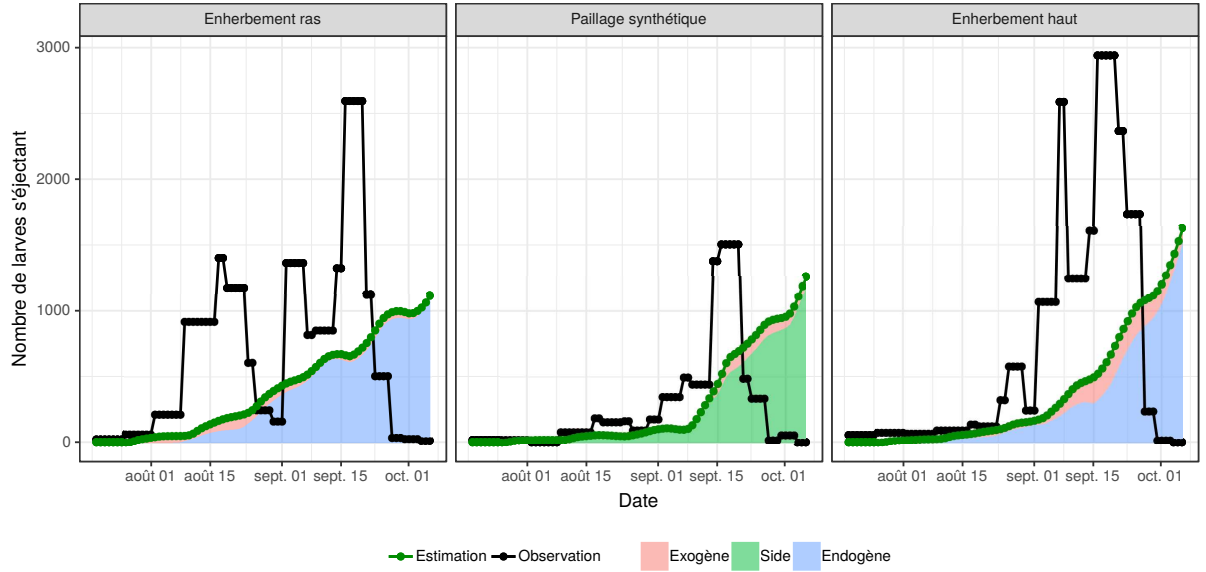


FIGURE 8 – Dynamiques estimées avec un jeu de paramètres mimisant la norme 1 des quatre critères.

## 6 Prochaines étapes

- Inclure la saisonnalité ;
- inclure le  $\mu_{global}$  dans la calibration.

## Références

- [Ge et al., 2018] Ge, H., Zhao, M., Sun, L., Wang, Z., Tan, G., Zhang, Q., and Chen, C. P. (2018). A many-objective evolutionary algorithm with two interacting processes : Cascade clustering and reference point incremental learning. *IEEE Transactions on Evolutionary Computation*.
- [Saltelli et al., 2008] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis : the primer*. John Wiley & Sons.