

Modèle conceptuel cécidomyies

Ce document s'intéresse à la possibilité d'utiliser le nombre de larves qu'il y a eu antérieurement (sur une plage de jours à définir) afin de prédire le nombre de larves au jour t .

1 Régression linéaire simple

Afin de rester un minimum cohérent avec la littérature, on testera la significativité du nombre de larves qu'il y a eu entre 7 et 17 jours auparavant. Pour ce faire, on effectue pour chacun des trois sous-blocs une régression linéaire simple défini par

$$L_t = \beta_0 + \sum_{j=7}^{17} \beta_j L_{t-j}.$$

Il faut noter que l'on utilisera uniquement les individus vérifiant $t > 17$ jours.

1.1 Enherbement ras

Le risque de première espèce du test de Student (pour la nullité des coefficients) sera fixé à $\alpha = 0.05$. Les résultats de la régression linéaire sont visibles dans la table 1. Il apparaît que seule l'ordonnée à l'origine est significative. Dès lors, il apparaît compliqué de prédire uniquement le nombre de larves à la date t avec les nombres de larves à des dates antérieures.

TABLE 1 – Coefficients trouvés par la régression linéaire simple pour l'enherbement ras, ainsi que leurs erreurs standards associés. Pour un risque de première espèce d' $\alpha = 0.05$, seule l'ordonnée à l'origine est significative

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|-----------|------------|---------|----------|
| (Intercept) | 1107.6499 | 184.8822 | 5.99 | 0.0000 |
| 'Lt-7' | 0.2387 | 0.3119 | 0.77 | 0.4476 |
| 'Lt-8' | -0.1055 | 0.4404 | -0.24 | 0.8117 |
| 'Lt-9' | -0.1190 | 0.4329 | -0.27 | 0.7845 |
| 'Lt-10' | -0.1849 | 0.4171 | -0.44 | 0.6593 |
| 'Lt-11' | -0.0115 | 0.4128 | -0.03 | 0.9779 |
| 'Lt-12' | -0.0087 | 0.4095 | -0.02 | 0.9831 |
| 'Lt-13' | -0.0624 | 0.4179 | -0.15 | 0.8819 |
| 'Lt-14' | 0.3418 | 0.4333 | 0.79 | 0.4339 |
| 'Lt-15' | -0.0895 | 0.4401 | -0.20 | 0.8396 |
| 'Lt-16' | 0.0927 | 0.5126 | 0.18 | 0.8572 |
| 'Lt-17' | -0.4645 | 0.4066 | -1.14 | 0.2587 |

La figure 1 montre ainsi la tendance du modèle à renvoyer pour tout les individus une valeur dans une plage de valeurs assez restreinte, ne reflétant pas la variabilité des observations. Ce faisant, le modèle proposé renvoie ainsi une valeur moyenne ne tenant finalement assez peu compte de la valeur au jour t . La figure 2 confirme cette tendance en montrant la piètre qualité de la prédiction.

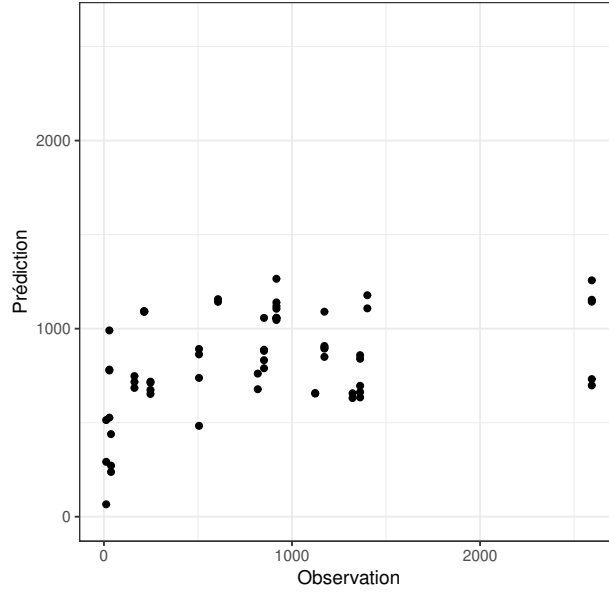


FIGURE 1 – Comparaison entre les valeurs observées et les valeurs prédites pour l’enherbement ras. Une bonne prédiction montrerait les points proches de la droite $y = x$. Ce n’est pas le cas ici.

1.2 Paillage synthétique

On réitère les calculs pour le paillage synthétique. Les résultats sont visibles dans la table 2, et comme pour l’enherbement ras seule l’ordonnée à l’origine est significative. Ainsi il n’est pas étonnant de trouver des résultats similaires concernant la prédiction comme l’illustrent les figures 3 et 4 : une prédiction mauvaise qui donne une valeur moyenne tout du long, manquant de réalisme.

TABLE 2 – Coefficients trouvés par la régression linéaire simple pour le paillage synthétique, ainsi que leurs erreurs standards associés. Pour un risque de première espèce d’ $\alpha = 0.05$, seule l’ordonnée à l’origine est significative

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 296.6741 | 78.6914 | 3.77 | 0.0004 |
| ‘Lt-7’ | 0.4112 | 0.3590 | 1.15 | 0.2575 |
| ‘Lt-8’ | 0.0161 | 0.5157 | 0.03 | 0.9753 |
| ‘Lt-9’ | -0.1417 | 0.5647 | -0.25 | 0.8029 |
| ‘Lt-10’ | -0.1033 | 0.5208 | -0.20 | 0.8435 |
| ‘Lt-11’ | 0.0148 | 0.4569 | 0.03 | 0.9743 |
| ‘Lt-12’ | 0.1155 | 0.4571 | 0.25 | 0.8015 |
| ‘Lt-13’ | -0.1878 | 0.4617 | -0.41 | 0.6858 |
| ‘Lt-14’ | 0.0936 | 0.5074 | 0.18 | 0.8544 |
| ‘Lt-15’ | 0.0908 | 0.5323 | 0.17 | 0.8652 |
| ‘Lt-16’ | -0.0222 | 0.7469 | -0.03 | 0.9764 |
| ‘Lt-17’ | -0.2156 | 0.6257 | -0.34 | 0.7319 |

1.3 Enherbement haut

Encore une fois, seule l’ordonnée à l’origine est significative (voir table 3). Cependant la figure 5 montre des prédictions plus variables, qui si elles ne sont pas parfaites (particulièrement vrai pour les valeurs observées très grandes ou très petites) demeurent bien plus acceptables que celles des deux autres sous-blocs. Cela se vérifie sur la figure 6 où l’on peut voir que la dynamique est plus ou moins captée (moins que plus, il est vrai).

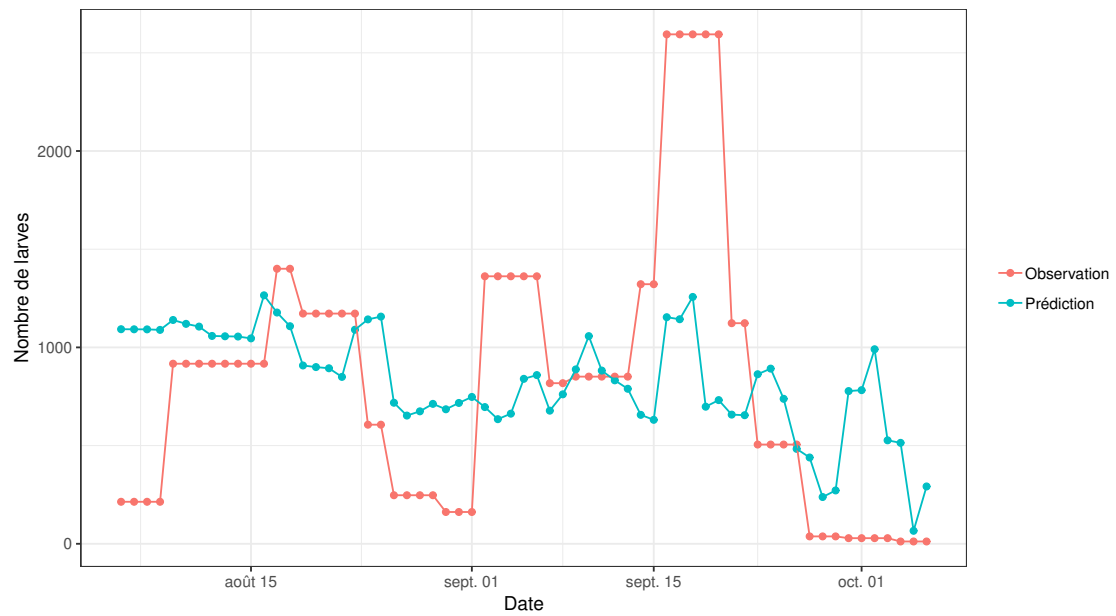


FIGURE 2 – Comparaison des nombres larves observées et des larves prédites à chaque date pour l'enherbement ras. La prédiction — plutôt mauvaise — renvoie une valeur assez proche tous les jours, ne captant pas la dynamique du phénomène observé.

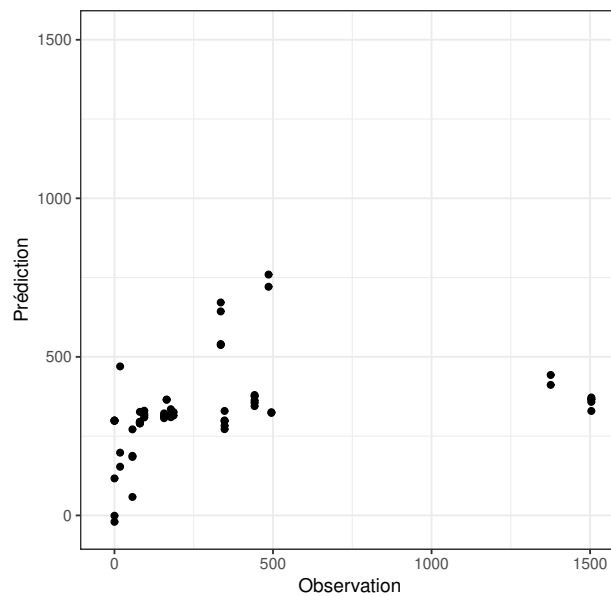


FIGURE 3 – Comparaison entre les valeurs observées et les valeurs prédites pour le paillage synthétique. Une bonne prédiction montrerait les points proches de la droite $y = x$. Ce n'est pas le cas ici.

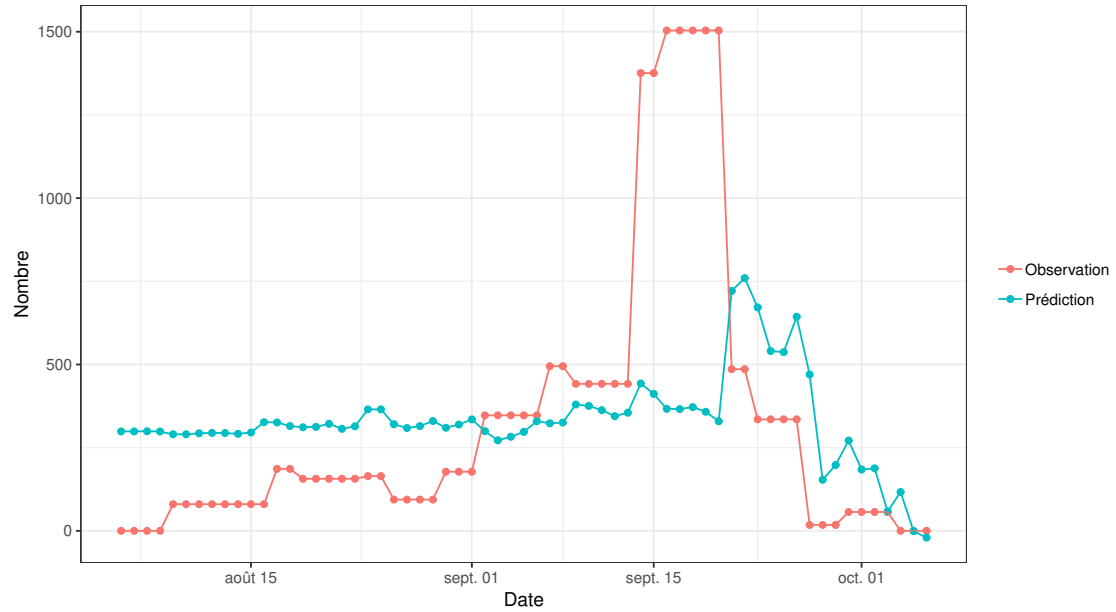


FIGURE 4 – Comparaison des nombres larves observées et des larves prédites à chaque date pour le paillage synthétique. La prédiction — plutôt mauvaise — renvoie une valeur assez proche tous les jours, ne captant pas la dynamique du phénomène observé.

TABLE 3 – Coefficients trouvés par la régression linéaire simple pour l’enherbement haut, ainsi que leurs erreurs standards associés. Pour un risque de première espèce d’ $\alpha = 0.05$, seule l’ordonnée à l’origine est significative

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 525.1541 | 159.7604 | 3.29 | 0.0018 |
| ‘Lt-7’ | 0.5592 | 0.3297 | 1.70 | 0.0960 |
| ‘Lt-8’ | -0.1018 | 0.4635 | -0.22 | 0.8270 |
| ‘Lt-9’ | 0.2186 | 0.4608 | 0.47 | 0.6372 |
| ‘Lt-10’ | 0.0512 | 0.4636 | 0.11 | 0.9125 |
| ‘Lt-11’ | -0.3174 | 0.4741 | -0.67 | 0.5062 |
| ‘Lt-12’ | 0.2680 | 0.4781 | 0.56 | 0.5776 |
| ‘Lt-13’ | -0.2279 | 0.4834 | -0.47 | 0.6392 |
| ‘Lt-14’ | 0.3811 | 0.5218 | 0.73 | 0.4685 |
| ‘Lt-15’ | -0.1342 | 0.5405 | -0.25 | 0.8048 |
| ‘Lt-16’ | -0.1415 | 0.5491 | -0.26 | 0.7977 |
| ‘Lt-17’ | -0.2839 | 0.4105 | -0.69 | 0.4923 |

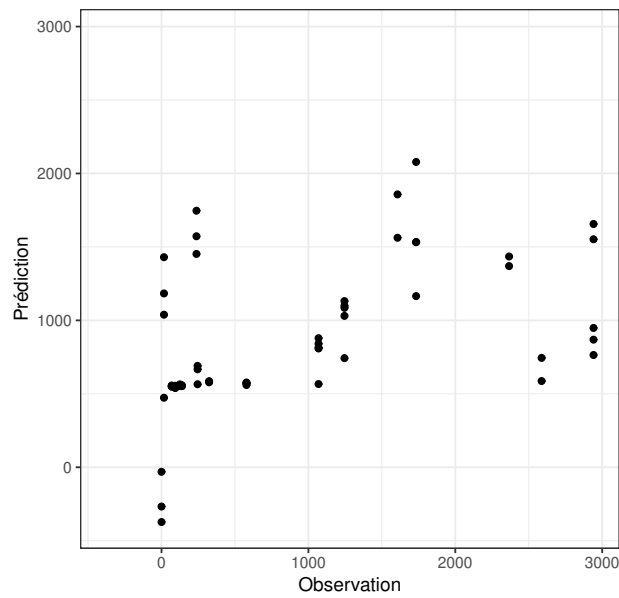


FIGURE 5 – Comparaison entre les valeurs observées et les valeurs prédites pour l'enherbement haut. Une bonne prédiction montrerait les points proches de la droite $y = x$. La prédiction n'est pas géniale ici.

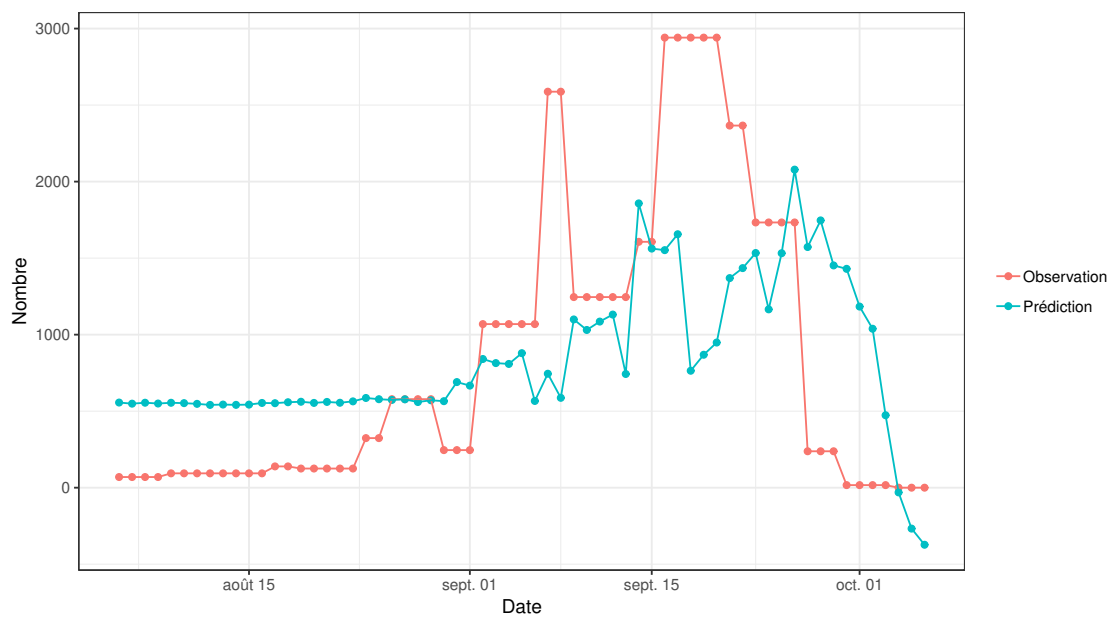


FIGURE 6 – Comparaison des nombres larves observées et des larves prédites à chaque date pour l'enherbement haut. La prédiction arrive à capter très grossièrement le phénomène observé.

1.4 Conclusion

Les résultats présentés ici indiquent clairement que le nombre de larves au jour t ne peut être modélisé par une combinaison linéaire des nombres de larves qu'il y a eu quelques jours plus tôt.

D'autres méthodes ont été testées, notamment un GLM Poisson et un GLM gaussien. Bien qu'il en ressortait que certaines variables étaient significatives, les prédictions étaient tout aussi mauvaises et ressemblaient fortement à celles de la régression linéaire simple.

Enfin un modèle linéaire mixte a été réalisé sur l'ensemble de la parcelle avec la modalité de couverture de sol comme effet fixe. Il en ressort qu'il y avait une différence significative entre les trois modalités de couverture du sol, ce qui est encourageant dans la mesure que cela implique que ce facteur est utile pour expliquer les différences entre les trois sous-blocs.

2 Corrélations

On s'intéresse maintenant à l'autocorrélation des larves et la corrélation entre les larves et les inflorescences. Le coefficient de corrélation de Spearman sera utilisé.

2.1 Autocorrélation

On calcule l'autocorrélation des larves avec un *lag* allant de 7 à 17 jours. Les résultats sont les suivants :

| | Enherbement ras | Paillage synthétique | Enherbement haut |
|-----------------------|-----------------|----------------------|------------------|
| $\rho(L_t, L_{t-7})$ | 0.292 | 0.613 | 0.574 |
| $\rho(L_t, L_{t-8})$ | 0.199 | 0.549 | 0.526 |
| $\rho(L_t, L_{t-9})$ | 0.080 | 0.462 | 0.475 |
| $\rho(L_t, L_{t-10})$ | 0.013 | 0.379 | 0.408 |
| $\rho(L_t, L_{t-11})$ | -0.012 | 0.338 | 0.381 |
| $\rho(L_t, L_{t-12})$ | -0.032 | 0.299 | 0.364 |
| $\rho(L_t, L_{t-13})$ | -0.083 | 0.260 | 0.342 |
| $\rho(L_t, L_{t-14})$ | -0.110 | 0.220 | 0.326 |
| $\rho(L_t, L_{t-15})$ | -0.153 | 0.186 | 0.307 |
| $\rho(L_t, L_{t-16})$ | -0.182 | 0.156 | 0.285 |
| $\rho(L_t, L_{t-17})$ | -0.239 | 0.127 | 0.271 |

L'autocorrélation est assez faible, en particulier pour l'enherbement ras. De plus cette autocorrélation n'a pas l'air de prendre en compte le cycle de développement des cécidomyies, elle est décroissante avec le temps alors que l'on s'attendrait à pic aux alentours de 12 – 14 jours (correspondant à un cycle complet) — exception faite pour l'enherbement ras, avec des corrélations qui remontent légèrement à la fin.

2.2 Corrélation entre les larves et les inflorescences

On s'intéresse maintenant à la corrélation entre les larves et les inflorescences. On fait toujours varier le *lag* entre 7 et 17 jours. Les résultats sont visibles ci-dessous :

| | Enherbement ras | Paillage synthétique | Enherbement haut |
|-----------------------|-----------------|----------------------|------------------|
| $\rho(L_t, I_{t-7})$ | 0.522 | 0.351 | 0.580 |
| $\rho(L_t, I_{t-8})$ | 0.422 | 0.298 | 0.556 |
| $\rho(L_t, I_{t-9})$ | 0.325 | 0.258 | 0.529 |
| $\rho(L_t, I_{t-10})$ | 0.244 | 0.243 | 0.494 |
| $\rho(L_t, I_{t-11})$ | 0.185 | 0.234 | 0.466 |
| $\rho(L_t, I_{t-12})$ | 0.123 | 0.217 | 0.442 |
| $\rho(L_t, I_{t-13})$ | 0.064 | 0.181 | 0.401 |
| $\rho(L_t, I_{t-14})$ | 0.015 | 0.127 | 0.357 |
| $\rho(L_t, I_{t-15})$ | -0.021 | 0.062 | 0.326 |
| $\rho(L_t, I_{t-16})$ | -0.020 | -0.001 | 0.294 |
| $\rho(L_t, I_{t-17})$ | 0.007 | -0.063 | 0.275 |

Un phénomène quelque peu semblable à l'autocorrélation des larves est présent. Une décroissance avec le temps d'une corrélation qui n'était pas bien forte au début. Cela reste néanmoins plus interprétable dans la mesure où la corrélation la plus importante coïncide avec la durée de larvation. Il convient cependant de rester prudent dans la mesure où l'on n'observe pas une franche diminution après la durée de larvation, ce qui semble plus indiquer que cela reste un phénomène lié au temps et indépendant du cycle de développement des cécidomyies.

2.3 Conclusion

Les corrélations calculées précédemment semblent en accord avec les résultats de la première partie. À savoir que les larves et les inflorescences ne peuvent (du moins linéairement) prédire le nombre de larves à venir. Cela a au moins le mérite de conforter l'approche choisie, à savoir un modèle décrivant les phénomènes observés. Un essai de modèle linéaire mixte tenant aussi compte des inflorescences a été réalisé, les résultats sont visibles dans l'annexe A.

3 Modèle conceptuel

On s'intéresse au modèle conceptuel

$$L_t = \alpha(t)R(\mu L_{t-12} + E(t-7)),$$

où R représente le coefficient de reproduction des femelles, μ la probabilité de survie des œufs, $E(t-7)$ les individus exogènes arrivant 7 jours avant, L_{t-12} le nombre de larves présentes un cycle auparavant et $\alpha(t)$ un coefficient journalier représentant l'ensemble des conditions environnementales. On a les contraintes suivantes sur les paramètres :

- $\forall t, \alpha(t) \in [0; 1]$,
- $R \in [0; 10]$,
- $\mu \in [0; 1]$,
- $\forall t, E(t-7) \in \mathbf{R}_+$.

Au vu des résultats précédents, on peut s'attendre à ce que l'optimisation ajuste sur les individus exogènes. Ce qui sera intéressant, c'est de voir dans quelle mesure les larves observées interviendront dans la prédiction.

3.1 Trois sous-modèles distincts

On commence par appliquer ce modèle indépendamment aux trois sous-blocs.

Sans vraiment trop de surprises, le modèle arrive à recréer les dynamiques observées comme le montrent les figures 7 et 8. Pour les coefficients R et μ , on trouve les résultats suivants :

| | Enherbement ras | Paillage synthétique | Enherbement haut |
|--------|-----------------|----------------------|------------------|
| R | 0.163 | 0.100 | 0.189 |
| μ | 0.896 | 0.201 | 0.868 |
| $R\mu$ | 0.146 | 0.020 | 0.164 |

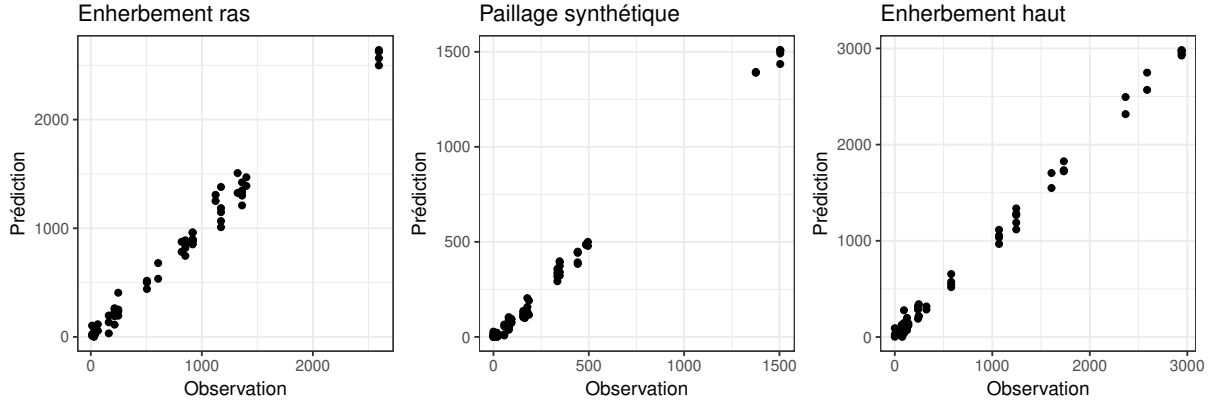


FIGURE 7 – Comparaison des valeurs prédites et observées par l’optimisation du modèle conceptuel par l’algorithme NSGA-II pour chacun des trois sous-blocs. Les points sont proches de la droite $y = x$, signe d’une prédiction convenable.

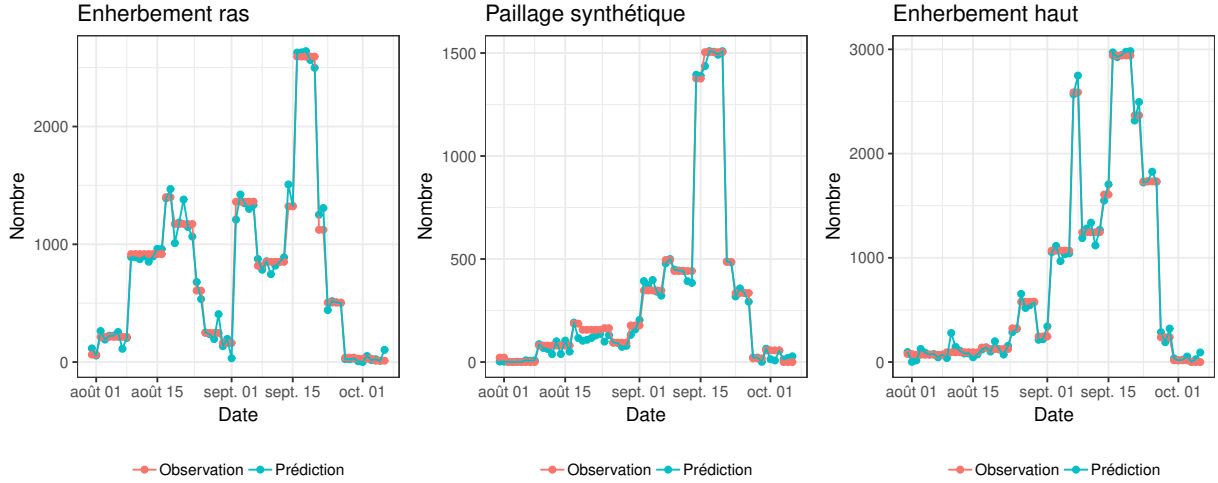


FIGURE 8 – Dynamiques observées et simulées pour chacun des trois sous-blocs par le modèle conceptuel et l’algorithme génétique NSGA-II.

Il demeure compliqué d’interpréter ces résultats dans la mesure où le modèle a deux paramètres libres chaque jour (et donc un nombre infini de façon de reproduire les dynamiques observées). Sans tirer de conclusion, on peut néanmoins apprécier le fait que $\mu^{ER} > \mu^{EH} > \mu^{PS}$, en accord avec nos hypothèses. Globalement, l’influence des larves antérieures est plutôt limitée.

Les coefficients $\alpha(t)$ ne semblent pas suivre un schéma commun aux trois sous blocs. Ils pourraient être interprétés comme l’effet du nombre d’inflorescences pour l’enherbement haut $\rho^{EH}(\alpha(t), I_t) = 0.72$ mais cette explication ne convient pas vraiment aux deux autres sous-blocs ($\rho^{ER}(\alpha(t), I_t) = 0.38$ et $\rho^{EH}(\alpha(t), I_t) = 0.51$). Les corrélations entre $\alpha(t)$ et I_{t-7} sont encore plus faibles.

3.2 Modèle global

On réitère ce qui vient d’être fait, à la différence que l’on impose que le R soit le même pour les trois sous-blocs. On obtient ainsi :

- $R = 0.171$
- $\mu^{ER} = 0.580$
- $\mu^{PS} = 0.287$
- $\mu^{EH} = 0.938$

L’algorithme converge plus lentement dans ce cas-ci. On peut noter que l’ordre des μ a changé.

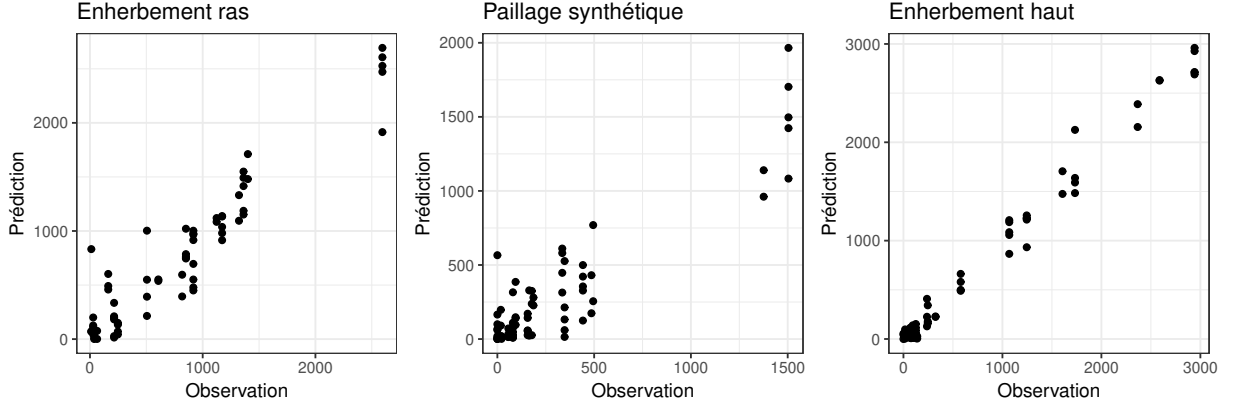


FIGURE 9 – Comparaison des valeurs prédites et observées par l’optimisation du modèle conceptuel par l’algorithme NSGA-II pour chacun des trois sous-blocs. Les points sont plus ou moins proches de la droite $y = x$ préfigurant une prédiction acceptable bien qu’améliorable.

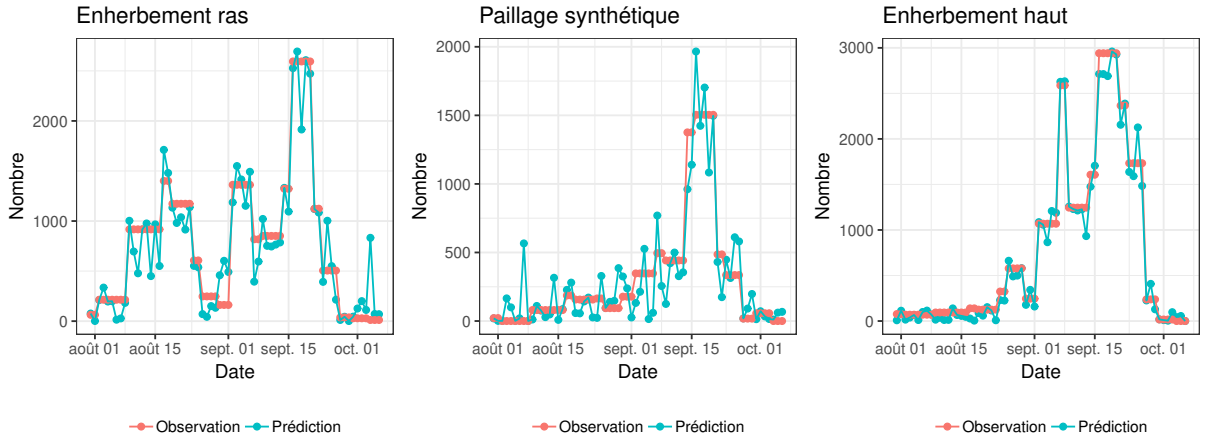


FIGURE 10 – Dynamiques observées et simulées pour chacun des trois sous-blocs par le modèle conceptuel et l’algorithme génétique NSGA-II.

Ce qui ne change pas en revanche, c’est la propension du modèle à ajuster sur les larves exogènes. Et une fois de plus les $\alpha(t)$ ne semblent pas suivre une logique précise.

3.3 Conclusion

Ce modèle n’est pas vraiment exploitable dans la mesure où les deux paramètres libres chaque jour permettent un nombre infini de façons de reproduire nos observations. On pourrait multiplier les tentatives en introduisant des contraintes mais les résultats précédent laissent présager que cela ne fonctionnerait pas.

D’une manière plus générale, les données que l’on possède sont trop peu nombreuses pour un modèle prédictif. Il faudrait bien plus d’individus et bien plus de variables explicatives pour s’essayer à une régression ou du *machine learning* qui soit exploitable quand le phénomène considéré est par nature très variable.

A Modèle linéaire mixte avec larves et inflorescences

On essaye ici de régresser L_t en fonction des larves tombés 7 à 17 jours auparavant, des inflorescences présentes de 0 à 17 jours auparavant et de la modalité de couverture du sol. Cette dernière variable est considéré comme un effet fixe. Comme le montre la table 4, les variables significatives sont peu nombreuses.

TABLE 4 – Coefficients trouvés par le modèle linéaire mixte, ainsi que leurs erreurs standards associés. Pour un risque de première espèce d' $\alpha = 0.05$, seules les inflorescences du jour et les modalités de couvertures des sols ER et PS sont significatives.

| | Estimate | Std. Error | t value | Pr(> t) |
|---------|-----------|------------|---------|----------|
| 'Lt-7' | 0.1436 | 0.1521 | 0.94 | 0.3465 |
| 'Lt-8' | -0.0009 | 0.1999 | -0.00 | 0.9966 |
| 'Lt-9' | -0.0501 | 0.2018 | -0.25 | 0.8041 |
| 'Lt-10' | -0.0284 | 0.2000 | -0.14 | 0.8872 |
| 'Lt-11' | -0.1284 | 0.1981 | -0.65 | 0.5179 |
| 'Lt-12' | 0.0040 | 0.1995 | 0.02 | 0.9841 |
| 'Lt-13' | -0.1130 | 0.2025 | -0.56 | 0.5776 |
| 'Lt-14' | 0.3194 | 0.2198 | 1.45 | 0.1483 |
| 'Lt-15' | -0.0690 | 0.2288 | -0.30 | 0.7635 |
| 'Lt-16' | -0.1480 | 0.2288 | -0.65 | 0.5185 |
| 'Lt-17' | -0.0934 | 0.1760 | -0.53 | 0.5964 |
| It | 0.7197 | 0.3304 | 2.18 | 0.0309 |
| 'It-1' | -0.6987 | 0.7118 | -0.98 | 0.3278 |
| 'It-2' | -0.0162 | 0.8284 | -0.02 | 0.9844 |
| 'It-3' | 0.0635 | 0.8387 | 0.08 | 0.9398 |
| 'It-4' | 0.1723 | 0.8653 | 0.20 | 0.8425 |
| 'It-5' | -0.1859 | 0.8702 | -0.21 | 0.8311 |
| 'It-6' | 0.3835 | 0.8804 | 0.44 | 0.6638 |
| 'It-7' | 0.4353 | 0.9197 | 0.47 | 0.6366 |
| 'It-8' | -0.3300 | 0.9323 | -0.35 | 0.7238 |
| 'It-9' | 0.8294 | 0.9548 | 0.87 | 0.3864 |
| 'It-10' | -0.9268 | 1.0591 | -0.88 | 0.3829 |
| 'It-11' | 0.8237 | 1.1512 | 0.72 | 0.4754 |
| 'It-12' | -0.8740 | 1.1979 | -0.73 | 0.4667 |
| 'It-13' | 0.6701 | 1.2190 | 0.55 | 0.5833 |
| 'It-14' | -1.6787 | 1.2321 | -1.36 | 0.1750 |
| 'It-15' | 1.1881 | 1.3621 | 0.87 | 0.3844 |
| 'It-16' | -1.1056 | 1.2267 | -0.90 | 0.3688 |
| 'It-17' | 0.9019 | 0.5901 | 1.53 | 0.1285 |
| SolER | 382.9572 | 90.2334 | 4.24 | 0.0000 |
| SolPS | -158.1030 | 77.0743 | -2.05 | 0.0419 |
| SolEH | -187.5178 | 113.9947 | -1.64 | 0.1020 |

Les figures 11 et 12 montrent les résultats de la prédiction. Les dynamiques du paillage synthétique et de l'enherbement haut sont bien reproduites. Ce n'est pas le cas de l'enherbement ras. C'est dû au fait que l'enherbement ras n'avait pas une dynamique de larves similaire à la dynamique d'inflorescences. En effet, on peut voir que le modèle se contente de *scaler* les inflorescences (le coefficient associé à I_t est de 0.7) pour s'approcher au mieux de la dynamique de larves. Ne dépendre uniquement des inflorescences du jour n'est probablement pas une bonne idée pour modéliser un phénomène complexe.

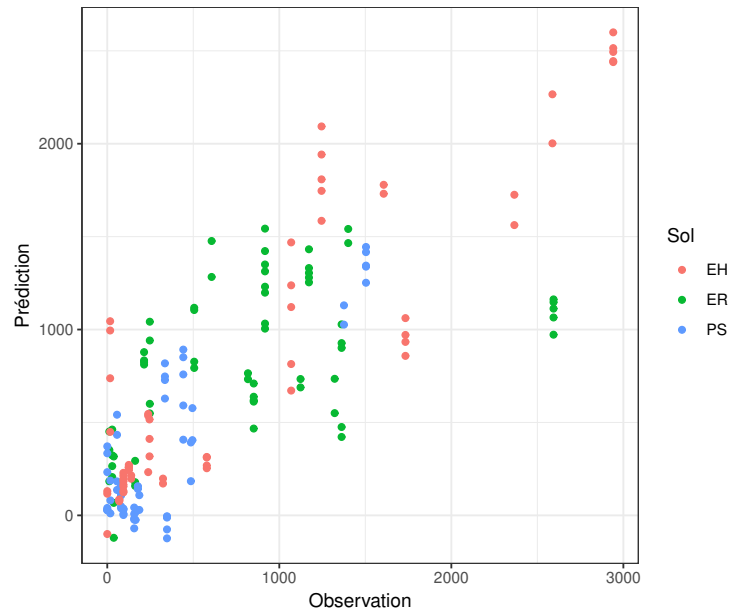


FIGURE 11 – Comparaison des valeurs observées et prédites par le modèle linéaire mixte pour chacun des trois sous-blocs.

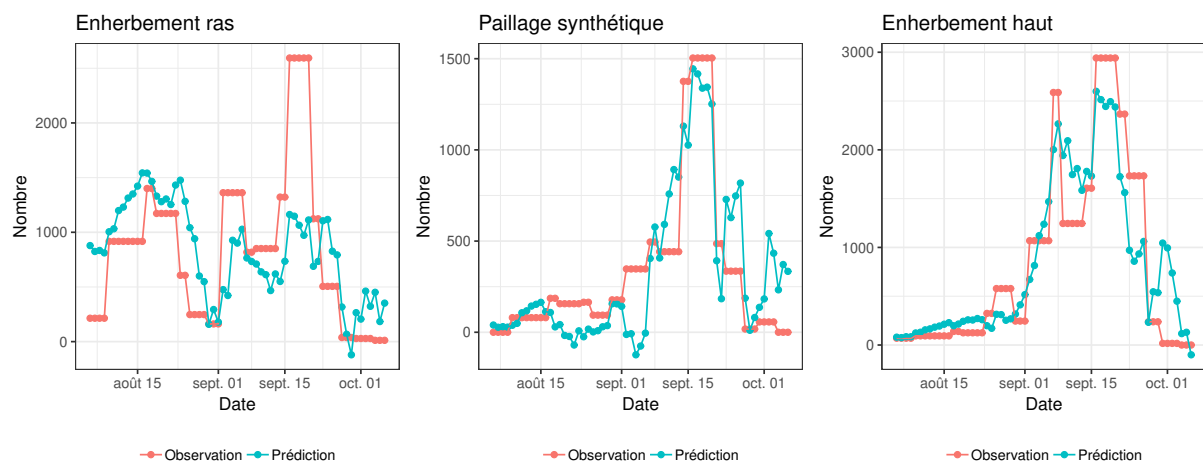


FIGURE 12 – Dynamiques observées et prédites par le modèle linéaire mixte.