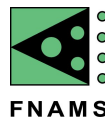


Étude exploratoire sur la cécidomyie des fleurs de la luzerne porte-graine



Clément CHEVALEYRE et Corentin BARBU

Coordinateur FNAMS : Benjamin COUSSY

Février 2026

Travail financé dans le cadre des actions spécifiques de la section fourragère de SEMAE

1. Introduction

Très peu d'études ont été menées sur la cécidomyie des fleurs de la luzerne (*Contarinia medicaginis*), il y a donc très peu de connaissances scientifiques sur ce bioagresseur. L'objectif de cette étude est de faire une première analyse des données récoltées en parcelles porte-graine par la FNAMS sur la cécidomyie des fleurs de la luzerne, en les mettant en regard des quelques données disponibles dans la littérature.

L'objectif final de la FNAMS sur la cécidomyie serait de prédire les vols, idéalement grâce à un OAD de prévision de la temporalité et de l'ampleur des vols. Ici nous étudions l'impact de différents facteurs sur la temporalité et l'ampleur des vols et identifions les points sur lesquels il faudrait renforcer les données pour parvenir à cet objectif et favoriser un traitement raisonné des cécidomyies plutôt que des traitements systématiques au stade sensible.

2. Matériel et méthode

2.1. Jeux de données

Jeu de données sur le suivi en parcelles de luzerne porte-graine sur la cécidomyie des fleurs de la luzerne

Base de données de relevés faits dans des parcelles d'expérimentations en luzerne porte-graine entre 2017 et 2019. Dans ces essais, l'objectif premier était d'évaluer la pertinence d'un piège à

phéromone de cécidomyie du framboisier pour avoir une idée des vols de cécidomyie des fleurs de luzerne (même genre de cécidomyie pour les deux : genre *Contarinia*).

On a donc des résultats de piège englué et aussi des comptages de galles de cécidomyie des fleurs sur les inflorescences de luzerne.

Le jeu de données contient également des informations sur :

- Age de la parcelle
- Dates des traitements contre le ravageur
- Variété de luzerne
- Stade de la culture lors des observations

Données fournies par l'INRAE

- **Données météorologiques** : issues de la base SAFRAN (Météo France), avec une maille de 8 km.

2.2. Procédure d'analyse

Pour approfondir nos connaissances sur la cécidomyie à partir du jeu de données disponible et des quelques sources bibliographiques à notre disposition, nous adoptons une démarche progressive en trois étapes : 1) Analyse descriptive des données et mise en relation avec les informations issues de la bibliographie ; 2) Modélisation linéaire de l'abondance du ravageur à partir des variables explicatives disponibles ; 3) Modélisation non linéaire et bayésienne de l'abondance du ravageur pour explorer des relations plus complexes et intégrer les *a priori* issus de la littérature.

A. Analyse descriptive de la donnée et mise en relation avec la bibliographie

L'objectif est de **décrire et visualiser simplement les données brutes** afin de mieux s'appropriier les informations disponibles. Cette étape permet d'identifier des tendances préliminaires ou de **formuler des hypothèses**, qui pourront ensuite être validées ou infirmées par des analyses statistiques ultérieures.

Le jeu de données comprend **122 observations**, réparties sur **15 communes**, dont 2 suivies sur 2 et 3 ans, ce qui représente un total de **17 communes-année**. Ce jeu de données, **relativement limité**, rend difficile une étude très approfondie.

En moyenne, chaque commune-année compte 6 à 7 observations. La majorité des parcelles bénéficient d'un **suivi hebdomadaire** (à l'exception de 4 d'entre elles) (Annexe 1). Le protocole d'échantillonnage repose sur un comptage direct. Il est donc nécessaire de **rapporter le nombre d'observations au nombre de jours de capture**. Les dates de relevé des pièges permettent par soustraction des dates de déduire le nombre de jours de pose. La date de pose initiale du piège n'étant pas connue, nous attribuons à la première observation 7 jours de pose, ce qui est habituellement le délai entre deux observations (Annexe 1). Il y a néanmoins une certaine variabilité de la durée de piégeage entre les points d'observation. Bien que les durées de piégeage correspondent habituellement à 7 jours, par exemple, pour la commune-année P3_2018, les deux derniers points d'observation couvrent chacun deux semaines de piégeage au lieu d'une seule d'où l'importance de prendre en compte les durées de piégeage pour bien interpréter les données.

L'analyse de l'histogramme des abondances de cécidomyies observées ne comprends pas d'excès de densité qui pourrait signer des **approximations dans le comptage** des cécidomyies par les

observateurs (par exemple, des arrondis à 5 ou 10 individus près) (Annexe 2). Si cela avait été le cas, il aurait fallu adapter la méthode d'analyse et simplifier les données pour renforcer la robustesse statistique. Certaines observations relèvent un nombre très élevé d'individus.

La figure 1 retrace l'évolution des abondances de cécidomyies au fil du temps, en lien avec les stades phénologiques de la luzerne et les interventions insecticides. On observe que les premières observations débutent généralement au début du stade sensible de la luzerne, marqué par l'apparition des boutons verts. Elles commencent pour la majorité en juin, parfois avec un certain retard par rapport à la date théorique du premier vol qui se situerait plutôt entre fin mai et début juin (Pierre E., 2018^[4]). De plus, la littérature définit un début de vol à partir de 230 degrés et un pic du premier vol à 250 degrés-jours (M. N. Berim, 2009^[2]), et nous supposons ici une base 10°C et commençant au 1er janvier car ces informations ne sont pas précisées dans la publication. Nous calculons que ces 230 degrés jours correspondent en général pour les observations à fin mai et, pour certaines localisations, début juin. Cela laisse penser que **le premier vol pourrait ne pas être capté par les relevés actuels** (voir le trait tireté sur les graphiques d'abondance, Figure 1).

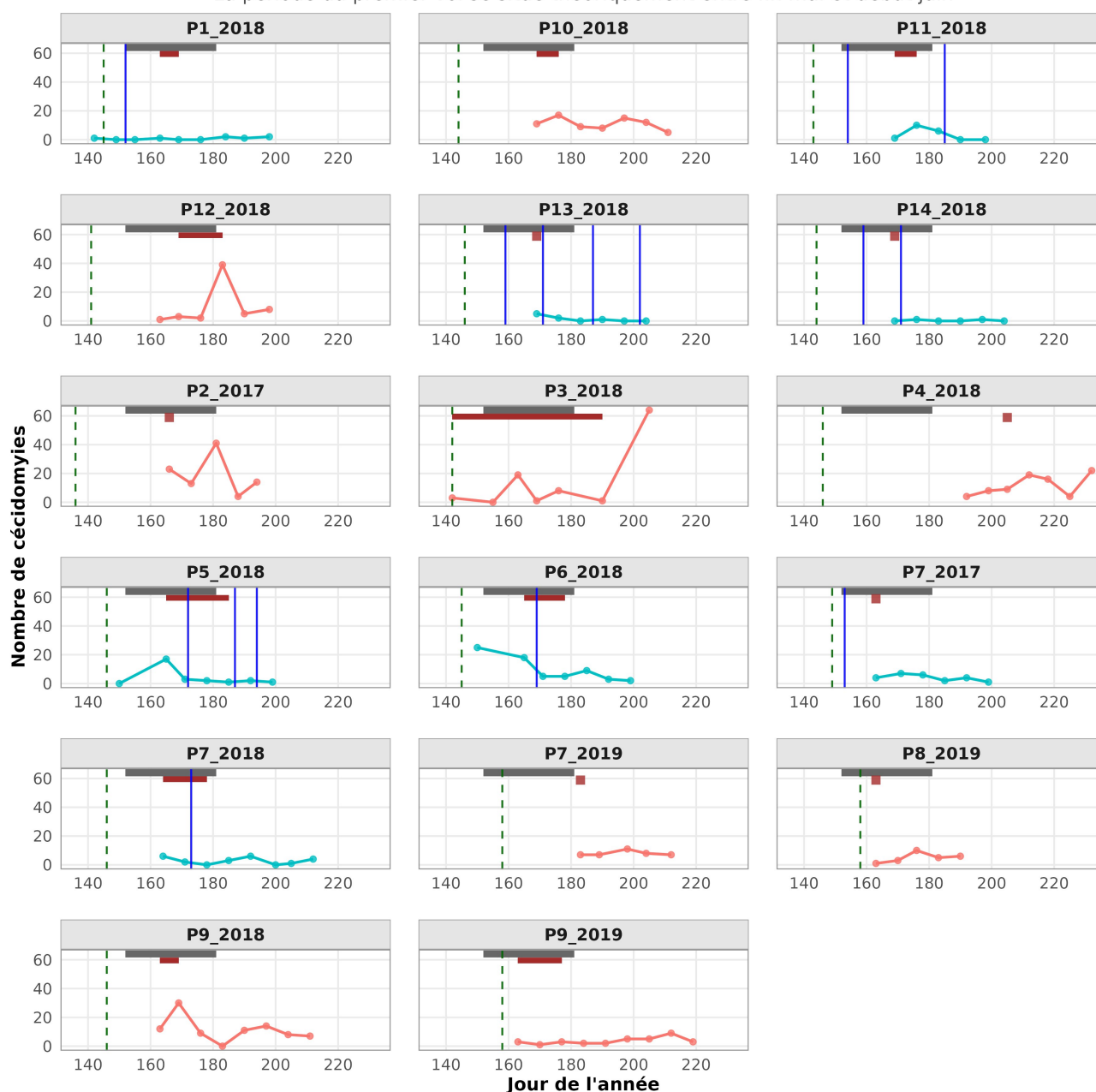
On **distingue plusieurs vagues successives** d'abondance de cécidomyies dans certaines communes-années. Cela pourrait correspondre à des générations successives après la première vague, sans que l'on puisse être sûr que la première vague observée correspondent bien au premier vol.

Les **traitements insecticides semblent efficaces** pour gérer le ravageur, car aucun pic d'abondance important n'est observé après une application. Nous revenons sur l'estimation de cette efficacité dans nos analyses statistiques.

Enfin, pour les deux parcelles **suivies sur plusieurs années**, aucune explosion d'abondance n'est constatée d'une année sur l'autre, même en l'absence de traitement insecticide (ex : P9_2018 et 2019). Cette **stabilité** est un première indice faible de l'existence de mécanismes de régulation biologique qui pourraient potentiellement se mettre en place dès la première année et évitant une augmentation d'année en année par effet d'inoculum croissant du ravageur d'une année sur l'autre.

Nombre de cécidomyies en fonction du jour de l'année par commune et année d'observation

La période du premier vol se situe théoriquement entre fin mai et début juin



Événement | Début du vol estimé (230 DJ) | Mois de juin | Stade sensible | Traitement

Conduite — Non traité — Traité

Figure 1. Évolution des abondances de cécidomyies par commune et par année, avec repères phénologiques et dates de traitement.

La figure intègre les dates de traitements, la période sensible (définie par la présence de boutons verts), ainsi que le mois de juin comme repère temporel par rapport au premier vol de cécidomyies, qui intervient généralement entre fin mai et début juin (source : Ephytia). Enfin, la date estimée du début de ce premier vol est calculée sur la base d'un seuil de 230 degrés-jours (Berim, 2009) en base 10, à partir du 1er janvier.

B. Modèles linéaires

Modélisation du logarithme de l'abondance moyenne des cécidomyies

Nous débutons l'analyse statistique avec des **modèles linéaires simples**, afin d'identifier les facteurs déterminants de l'abondance des cécidomyies à l'échelle des communes-année.

Les premiers modèles visent à étudier l'effet de **quelques variables explicatives disponibles** (type de conduite, nombre de traitements, variété, âge de la culture, année) sur le **logarithme de la moyenne des abondances observées par commune-année**, pondéré par le nombre de jours de piégeage. Le choix de modéliser la **moyenne** plutôt que l'ensemble des observations s'explique par le **manque d'indépendance** entre les relevés successifs au sein d'une même commune-année : chaque nouvelle observation dépend fortement de la précédente. En utilisant la moyenne, nous obtenons un jeu de données respectant l'hypothèse d'indépendance des observations. La **pondération par la durée de piégeage** est essentielle, car celle-ci peut varier. Cette correction évite de **sous-estimer ou surestimer** les abondances.

Le passage au **logarithme** permet de **réduire l'influence des valeurs extrêmes** qui ont sinon un poids très important en régression linéaire. L'ajout de **+1 aux observations** permet d'**inclure les valeurs nulles** et de conserver l'information biologique associée aux absences de cécidomyies.

Les données ont été analysées à l'aide d'un **modèle linéaire généralisé mixte (GLMM)** implémenté avec le package *glmmTMB*. Il permet d'inclure des **variables quantitatives** mais également des **variables qualitatives** comme effets fixes et/ou effets aléatoires. Le modèle estime alors un effet par modalité, comparé à une modalité de référence. Le modèle inclut comme **effets fixes** les variables explicatives d'intérêt ainsi que leurs interactions éventuelles. Ces effets expliquent la variation moyenne de la réponse dans la population. Des **effets aléatoires** sont introduits pour tenir compte de la non-indépendance des observations, comme par exemple des regroupements spatiaux.

Nous **testons des modèles intégrant différents groupes de variables**, chacun correspondant à des hypothèses spécifiques. L'objectif est de vérifier que les résultats de ces modèles **convergent vers les mêmes conclusions**, renforçant ainsi la **robustesse** des résultats aux différentes hypothèses de modélisation. Les résultats de cette étude sont présentés dans la **partie 4.1**.

C. Modèle non linéaire

Pour ce modèle, nous utilisons un modèle bayésien pour prendre en compte les informations de la bibliographie, avec leur incertitude. L'idée est de valider ou préciser ces informations à partir des données observées.

Ce modèle non linéaire prend en compte les conditions météorologiques et les pratiques culturales, ainsi que le cycle de reproduction des cécidomyies (très simplifié). Le premier vol et la reproduction sont fonction des conditions météorologiques (température et humidité relative). Nous prenons aussi en compte les traitements insecticides. Le modèle simule ainsi les abondances quotidiennes des cécidomyies. Un *a priori* est défini pour chacun des paramètres en se basant sur la bibliographie pour orienter l'ajustement du modèle vers des valeurs déjà connues ou supposées plausibles.

Le modèle ajuste les paramètres suivants, les *a priori* qui sont détaillés ci-dessous peuvent être visualisés sur les courbes de présentation des résultats (Figure 6) :

- **djv** : nombre de degrés-jour déterminant le pic du premier vol de l'insecte ; l'*a priori* étant de 250 DJ (M. N. Berim, 2009^[2]). La distribution normale d'écart type 25 degrés jours suivant un prior :

$$djv \sim \text{dnorm}(250, 25)$$

- **sv** : paramètre de variabilité du vol qui va permettre d'étaler plus ou moins le 1er vol du ravageur sur plusieurs jours, par défaut étalé sur 5 jours. Prior :

$$sv \sim \text{extraDistr}::\text{dbetapr}(5, 3)$$

- **Np_x** : taille de vol initial de la parcelle x. Chaque parcelle est soumise à des conditions de pratiques culturales et environnementales différentes (variété, paysage, âge,...). Elles disposent donc d'un pool potentiel d'insectes distinct (ce qui correspond à un effet parcelle). L'*a priori* est d'environ 20 individus mais suit une distribution log normale (donc à queue de distribution très longue) :

$$Np \sim \text{dlnorm}(\text{meanlog}=\log(20), \text{sdlog}=\log(2))$$

- **rb** : taux de reproduction de base. L'*a priori* sur la capacité à se reproduire est une multiplication par 5 pour chaque génération (Piart Jean, 1962^[3] ; Pierre E., 2018^[4]) ;

$$rb \sim \text{dlnorm}(\text{meanlog}=\log(5), \text{sdlog}=\log(1.5))$$

- **ng** : durée d'une génération. L'*a priori* choisi correspond à environ 25 jours, cette durée peut facilement varier de plus ou moins 5 jours (Bournoville et al, 1986^[1] ; Piart Jean, 1962^[3] ; Pierre E., 2018^[4]) ;

$$ng \sim \text{dnorm}(25,5)$$

- **nt** : paramètre quantifiant la rémanence (vitesse de décroissance de la quantité d'insecticide efficace). On considère ici que la quantité appliquée lors d'un traitement correspond à une efficacité de 100% le jour du traitement. La quantité diminue exponentiellement avec le nombre de jours (n_{jours}) avec $\exp(-n_{\text{jours}}/nt)$. Des traitements successifs peuvent induire une quantité supérieure à celle nécessaire pour induire une efficacité de 100% mais l'efficacité est bien sûr toujours limitée à 100% chaque jour. L'*a priori* utilisé correspond à une baisse d'environ 2/3 de la quantité de produit tous les 3 jours après le traitement ;

$$nt \sim \text{dlnorm}(\text{meanlog}=\log(3), \text{sdlog}=\log(1.5))$$

- **s** : paramètre rendu nécessaire par l'utilisation de la vraisemblance pour l'ajustement du modèle non-linéaire par MCMC. Intuitivement, il correspond à la variabilité des observations autour des tailles de population estimées. Mathématiquement, il correspond au paramètre de « taille » de la distribution binomiale négative utilisée pour calculer la vraisemblance des observations. Prior :

$$s \sim \text{dbetaapr}(\text{shape1}=5, \text{shape2}=3)$$

La **température** et l'**humidité relative** sont également utilisés pour influencer la capacité de reproduction au cours du temps. Le modèle de Brière est une fonction thermo-dépendante très utilisée en écologie pour décrire un taux biologique en fonction de la température :

$$r(T) = \begin{cases} aT(T-T_{min})((T_{max}-T))^m & \text{si } T_{min} < T < T_{max} \\ 0 & \text{sinon} \end{cases}$$

avec :

- o $r(T)$: taux biologique dépendant de la température
- o T : température (en °C)
- o T_{min} : température minimal biologique
- o T_{max} : température maximal biologique
- o a : constante d'échelle à estimer (>0)
- o m : constante d'échelle à estimer (>0)

Le taux de reproduction relatif en fonction de l'humidité relative est décrit par une fonction gaussienne :

$$h(RH) = \exp\left(-\frac{(RH - RH_{opt})^2}{2\sigma_{RH}^2}\right)$$

avec :

- o RH : humidité relative
- o $h(RH)$: taux biologique dépendant de l'humidité relative
- o RH_{opt} : humidité optimal
- o σ_{RH} : constante d'échelle (>0)
- plus σ est grand, plus la tolérance est large

Le taux de reproduction relatif qui en résulte est le suivant :

$$R(T, HR) = rb * r(T) * h(RH)$$

Pour limiter la complexité du modèle et réduire le nombre de paramètres à ajuster, nous avons choisi de fixer les paramètres des fonctions $r(T)$ et $h(RH)$. Ces paramètres sont déterminés à partir des informations disponibles dans la bibliographie, comme suit :

Paramètres de la fonction de température
(Bournoville et al, 1986^[1] ; M. N. Berim, 2009^[2])

$$\begin{aligned} T_{min} &= 10 \\ T_{max} &= 30 \\ a &= 1e-4 \\ m &= 1,2 \end{aligned}$$

Paramètres de la fonction d'humidité relative (M. N. Berim, 2009^[2] ; Pierre E., 2018^[4])

$$\begin{aligned} RH_{opt} &= 65 \\ \sigma_{RH} &= 10 \end{aligned}$$

La figure 2 illustre le taux de reproduction relatif, normalisé entre 0 et 1, sous la forme d'un graphique de contour. Ce graphique résulte de l'interaction entre les deux fonctions dépendant de la température et de l'humidité.

Reproduction relative de la cécidomyie

Modèle hypothétique basé sur température et humidité

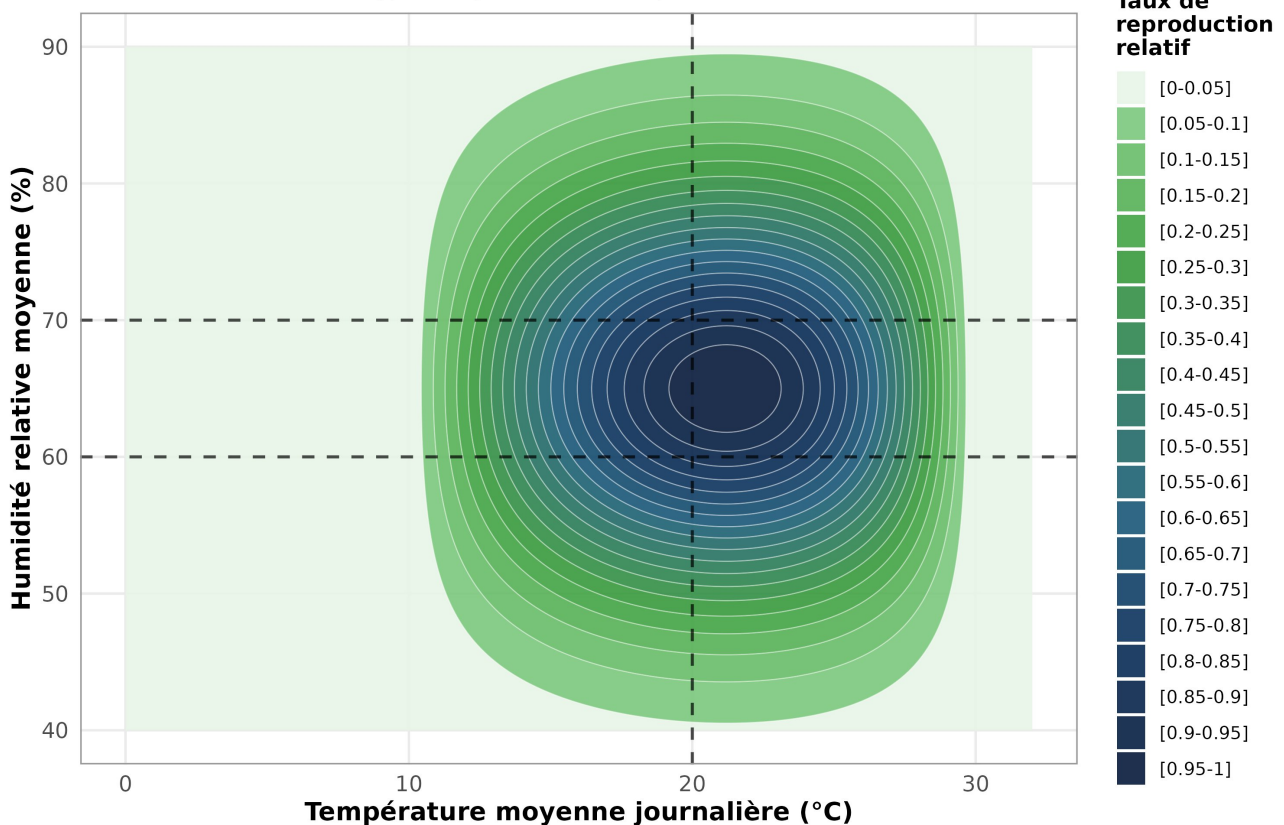


Figure 2 : Taux de reproduction relatif de la cécidomyie en fonction de la température et de l'humidité relative, normalisé entre 0 et 1.

L'ensemble des paramètres du modèle est ajusté sur le jeu de données d'abondance de cécidomyie au cours du temps dans les différentes communes-année par l'algorithme de Métropolis-Hastings, qui repose sur une *Monte Carlo Markov Chain* (MCMC). Cette approche statistique est utilisée pour estimer la distribution *a posteriori* des paramètres d'un modèle lorsque celle-ci ne peut pas être calculée analytiquement. Elle repose sur la construction d'une chaîne de Markov dont les états successifs représentent des valeurs possibles des paramètres. À chaque itération, une nouvelle valeur est proposée puis acceptée ou rejetée selon sa probabilité, déterminée à partir du rapport des vraisemblances entre la nouvelle et l'ancienne valeur, pondéré par leurs distributions *a priori*. Ce mécanisme permet d'explorer l'espace des paramètres en donnant plus de poids aux zones où la vraisemblance est élevée, c'est-à-dire où le modèle reproduit mieux les données. Après un certain nombre d'itérations, la chaîne converge vers la distribution cible, et les échantillons obtenus peuvent être utilisés pour estimer les paramètres, leurs incertitudes et les intervalles de crédibilité.

Plusieurs simulations ont été réalisées nous amenant notamment à une modification de la paramétrisation, afin de limiter les problèmes de convergence des chaînes résultant de la covariance entre les paramètres. Après avoir réalisé une ACP (Analyse en Composantes Principales) sur la chaîne de MCMC d'une des simulations, nous avons utilisé les composantes principales de l'ACP pour l'ajustement du modèle. Cela a permis de faciliter les transitions entre les différents modes de l'ajustement. On obtient ainsi des chaînes correctement convergées (Annexe 5.4).

3. Résultats de l'analyse statistique

3.1. Modélisation de l'abondance moyenne des cécidomyies

Pour structurer notre analyse nous définissons un « modèle central » qui nous paraît bien résumer les résultats obtenus avec les différentes variations testées du modèle et qui comprend l'ensemble des variables considérées. Le modèle central explique la moyenne des abondances observées rapportées au nombre de jour de piégeage par la variable binaire de traitement de la parcelle et par l'âge de la parcelle comme variable catégorielle. De plus, les effets variété, région et année sont pris en compte comme variables à effet aléatoire. On notera que vu le très petit nombre d'observations, on ne peut pas avoir d'effet aléatoire année emboîté dans un effet région ni des effets fixes à la fois pour les régions, les années et les variétés simultanément.

Le modèle central converge correctement (Figure 3). L'ajustement de ce modèle indique un **effet suppresseur fort et significatif des traitements** (-7.1, p-value=5.6x10⁻³) et un effet plutôt suppresseur de l'âge des parcelles, plus ou moins significatif suivant la spécification exacte des modèles. S'il y avait un effet âge de la parcelle, cet effet serait en tout cas nettement inférieur à l'effet traitement. Un modèle légèrement différent ajoutant un effet aléatoire parcelle estime un effet encore plus important du traitement (-9.5, p-value= 3.66x10⁻⁶, ensemble des résultats de ce modèle non détaillé ici).

```
Family: gaussian ( identity )
Formula:
meanValPond ~ conduite + age + (1 | Région) + (1 | an) + (1 | variete)
Data: cecidomyie_campagne

      AIC      BIC    logLik -2*log(L)  df.resid
  107.9    114.9    -45.0     89.9         7

Random effects:

Conditional model:
Groups   Name          Variance Std.Dev.
Région   (Intercept)  3.255e-06 0.001804
an       (Intercept)  2.518e-04 0.015867
variete  (Intercept)  2.341e+00 1.529968
Residual                    1.398e+01 3.738511
Number of obs: 16, groups: Région, 5; an, 3; variete, 10

Dispersion estimate for gaussian family (sigma^2): 14

Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    13.729      2.692    5.100 3.4e-07 ***
conduiteTraité -7.104      2.562   -2.772 0.00556 **
ageA2          -2.828      2.562   -1.103 0.26982
ageA3          -4.792      3.062   -1.565 0.11765
ageA4          -5.688      5.064   -1.123 0.26130
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3 Résultats du modèle linéaire initial central

Au lieu de seulement prendre en compte si les parcelles sont traitées ou pas, il est possible de prendre en compte le nombre de traitement. Les résultats obtenus sont similaires (non-détaillés) mais un peu plus faible, ce qui suggère un effet « moins que proportionnel » des traitements sur les abondances.

Prendre en compte uniquement l'effet traitement dans le modèle (et non l'âge) donne un résultat très similaire au modèle central (-7.104, p-value = 3.34x10⁻³).

Pour certains modèles, l'effet traitement peut ne plus être significatif, tout en restant négatif. C'est notamment le cas si l'on retire l'effet variété aléatoire et que l'on passe l'effet région en effet fixe. Il est assez attendu que certaines spécifications du modèle limitent la significativité de l'effet traitement, surtout avec le faible nombre d'observations. Nous estimons cependant qu'il y a bien un effet fort, robuste et significatif des traitements. L'effet âge de la parcelle n'est en revanche pas robuste.

Modéliser le log de la moyenne pondérée au lieu du log de la moyenne naturelle ne change pas la significativité du traitement mais change bien sûr la taille d'effet puisque la variable expliquée est très différente. La taille d'effet du traitement sur la moyenne des abondances est de l'ordre de deux fois plus importante pour la moyenne des moyennes d'abondances passées au log, ce qui suggère que les traitements réduisent particulièrement les hautes abondances.

L'ajout des variétés en effet fixe au modèle permet de comparer les abondances selon les variétés en prenant en compte les traitements et les effets aléatoires. Nous comparons ces effets fixes grâce au package emmean (Figure 4). On observe sur certaines variétés (Creno, Daphne, Melissa et Etincelle) nettement plus de cécidomyies que sur d'autres (Artemis et Europe), les autres variétés (Alexis, Galaxie, Sibemol et Timbale) présentent des abondances de cécidomyies intermédiaires ou très variables. Ces résultats sont bien sûr à confirmer étant donné le très petit nombre d'observations.

variete	emmean	SE	df	asympt.LCL	asympt.UCL
ALEXIS	3.283	2.58	Inf	-1.78	8.34
ARTEMIS	-0.325	2.12	Inf	-4.48	3.83
CRENO	9.070	1.44	Inf	6.24	11.90
DAPHNE	14.241	2.36	Inf	9.61	18.87
ETINCELLE	11.633	2.36	Inf	7.00	16.26
EUROPE	-1.258	2.51	Inf	-6.18	3.66
GALAXIE	3.592	3.43	Inf	-3.12	10.31
MELISSA	11.233	1.64	Inf	8.02	14.44
SIBEMOL	6.411	1.75	Inf	2.99	9.83
TIMBALE	5.498	3.39	Inf	-1.15	12.14

Results are averaged over the levels of: conduite, age
Confidence level used: 0.95

Figure 4 Comparaison des effets variétaux

3.2. Modélisation non-linéaire de la dynamique intra-annuelle de l'abondance des cécidomyies ajusté par MCMC

Les figures 6.1 et 6.2 représentent les distributions de densité de probabilité *a priori* des paramètres (en dehors des effets parcelle (Nps), présentés en annexe 6.3), utilisées comme base initiale pour l'ajustement du modèle (définie dans la partie 3.2.C). Les densités de probabilité *a posteriori*, quant à elles, correspondent à l'ensemble des valeurs possibles des paramètres explorées par la méthode MCMC. Plus un pic est resserré autour d'une valeur, plus la certitude sur cette valeur pour le paramètre est élevée.

L'ajustement du modèle de simulation de la dynamique annuelle des cécidomyies a permis de distinguer deux modes d'ajustement des paramètres. Les changements de mode résultent d'une **covariance entre plusieurs paramètres** et entraînent une modification de la durée d'une génération, de la capacité de reproduction de la cécidomyie, de la variabilité autour de la date du premier vol et du pool initial de ravageurs. L'évolution conjointe de tous ces paramètres permet d'atteindre des populations similaires aux observations, avec des probabilités comparables.

Deux durées de **génération** (ng) seraient possibles, avec des modes (maximums locaux de la densité de probabilité ou « hauts de pics ») à **19 jours ou 29 jours**. La durée la plus longue s'accompagne

notamment d'une **capacité de reproduction** (rb) plus importante pour la cécidomyie (environ 6,5 contre environ 8) et la **variabilité du premier vol** (sv) tend à se recentrer autour de la médiane. Le mode le plus probable reste cependant le mode 1 à 29 jours (72,6% de la densité). De plus, selon la bibliographie (Bournoville et al, 1986^[1] ; Piart Jean, 1962^[3] ; Pierre E., 2018^[4]), une durée entre générations de 29 jours serait plus probable, car 19 jours semblent trop courts pour couvrir l'ensemble des stades de développement de l'insecte. Dans la suite, nous nous concentrerons donc sur le mode à 29 jours.

Cette multimodalité ne semble cependant pas influencer le **nombre de degrés-jours** (djv) déterminant le pic du premier vol des cécidomyies. Celui-ci apparaît légèrement plus élevé que notre *a priori* initial (**à 275 DJ, intervalle de crédibilité à 95% 229 à 314 DJ**) au lieu de 250 DJ, avec un écart-type du pic de 11 jours [8-19], ce qui se traduit par un étalement estimé de 95% du premier vol de cécidomyies sur la moyenne des parcelles observées entre le 7 mai et le 27 juin avec un pic au 5 juin (Figure 5).

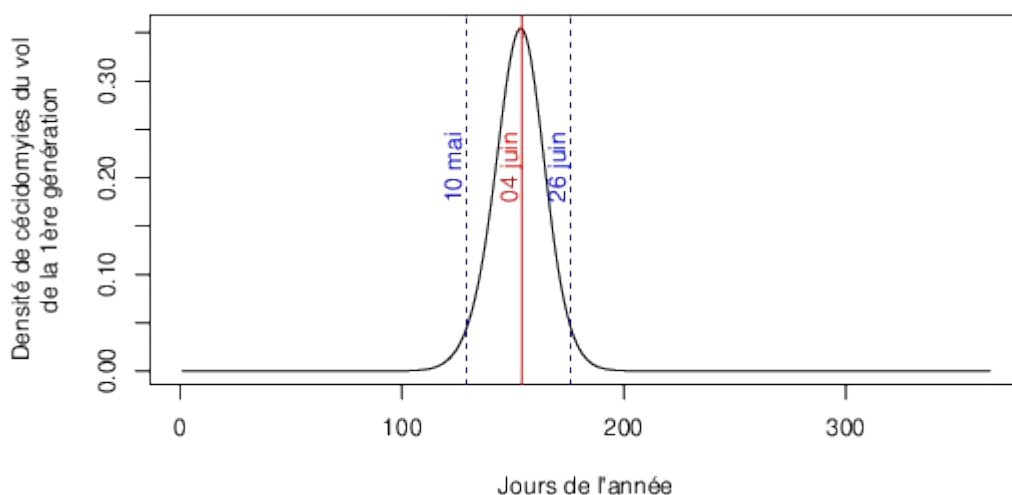


Figure 5 Distribution moyenne du vol de première génération de cécidomyies sur les parcelles observées

Enfin, la **rémanence des traitements** (nt) semble plus élevée que notre *a priori* initial. Le mode de la distribution de nt est d'un peu plus de 5 jours, ce qui correspond à l'élimination de 2/3 de la quantité de produit efficace tous les 5 jours.

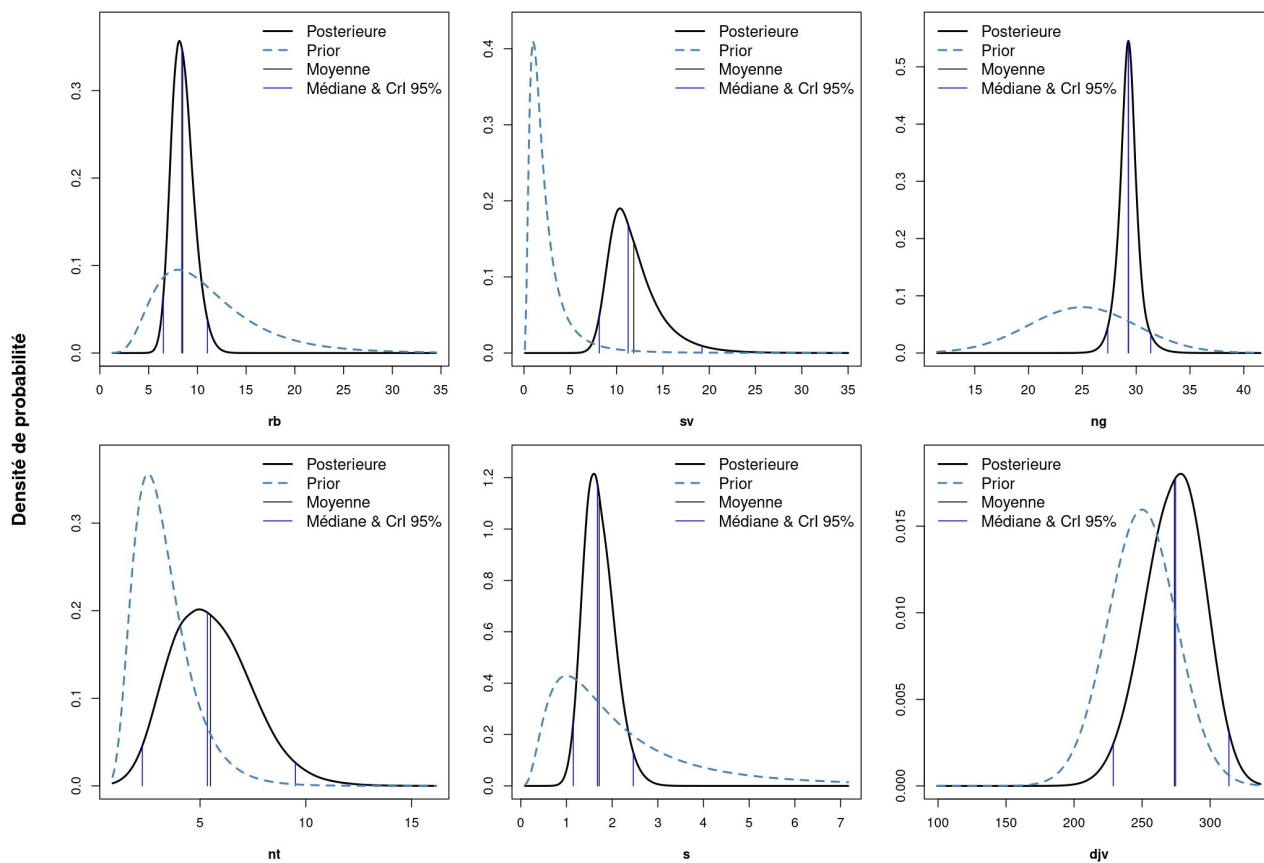


Figure 6.1 : Distribution des paramètres du modèle de cécidomyies *a priori* et *a posteriori* pour le 1er mode de l'ajustement. Les ordonnées correspondent aux densités de probabilités et les abscisses aux valeurs du paramètre : rb = capacité de reproduction, sv = variabilité du 1er vol, ng = durée de génération, djv = nombre de degrés-jour pour le 1er vol, nt = rémanence des traitements, s = variabilité autour de l'abondance simulée

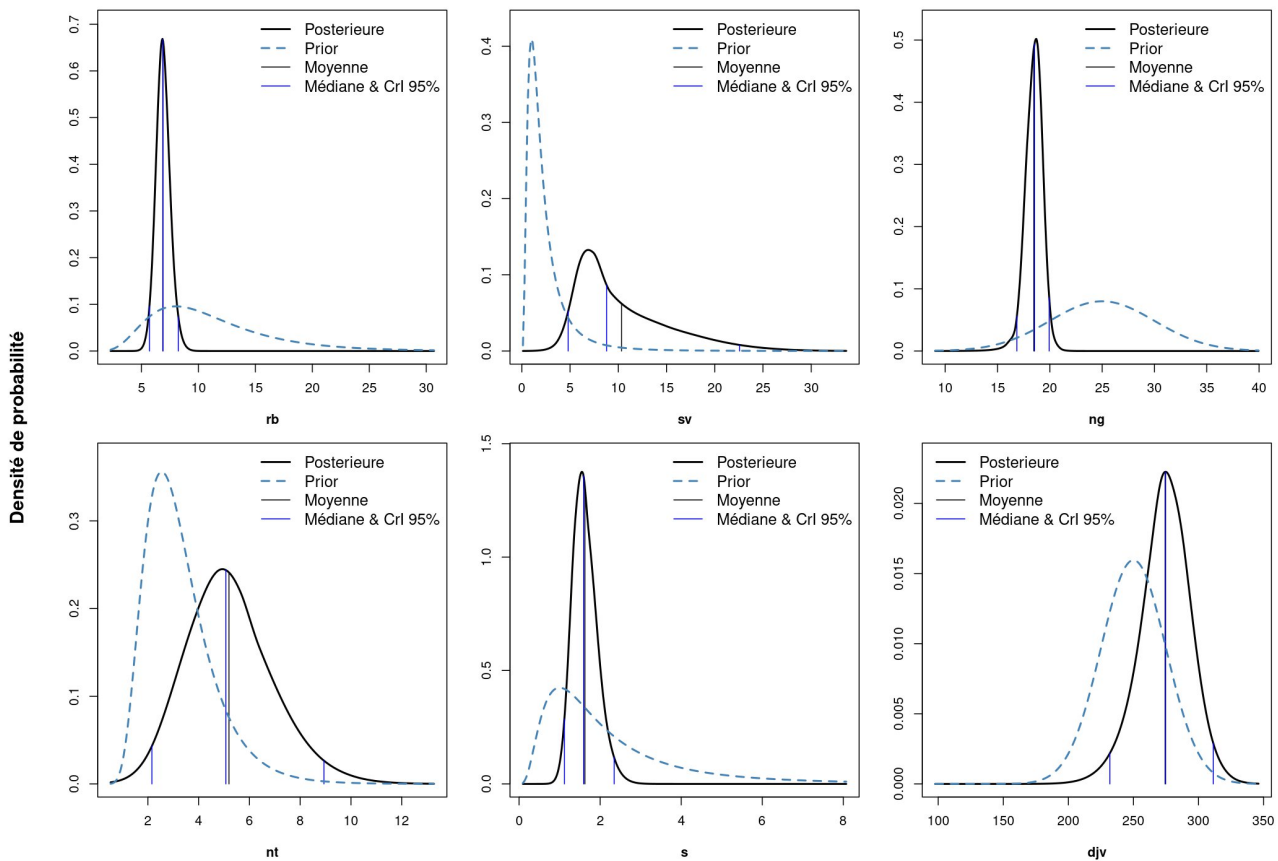


Figure 6.2 : Distribution des paramètres du modèle de cécidomyies a priori et a posteriori pour le 2nd mode de l'ajustement. Les ordonnées correspondent aux densités de probabilités et les abscisses aux valeurs du paramètre : rb = capacité de reproduction, sv = variabilité du 1er vol, ng = durée de génération, djv = nombre de degrés-jour pour le 1er vol, nt = rémanence des traitements, s = variabilité autour de l'abondance simulée.

La figure 7 superpose, pour les différentes communes-années, les observations et les simulations correspondant au nombre de cécidomyies capturées pendant la période de piégeage pour l'ajustement des paramètres sur le mode n°1, seul compatible avec les durées du stade de développement comme vu ci-dessus. Globalement, les simulations semblent bien correspondre aux observations pour plusieurs parcelles, même si certains pics apparaissent légèrement décalés. Ce décalage pourrait s'expliquer par le fait que la durée entre générations n'est pas fixe : selon les conditions météorologiques, elle pourrait être allongée ou raccourcie de quelques jours, en effet, nous prenons dans ce modèle l'effet de la météo sur la capacité de reproduction mais pas sur la durée du cycle.

L'observation des simulations avant les premières observations (Figure 7) permet également de voir que les abondances de cécidomyies augmentent avant les premières observations. Cela confirme que le premier vol de cécidomyies qui semble débiter aux alentours des semaines 19-20, soit mi-mai, n'est presque jamais observé.

On observe également une limite du modèle : il ne parvient pas à simuler la diminution de la population de cécidomyies à l'automne. L'impact des conditions environnementales ne permet donc pas de reproduire la saisonnalité de la population. Cela pourrait s'expliquer par des critères insuffisamment stricts sur les conditions de température et/ou d'hygrométrie favorables à la

reproduction du ravageur. D'autres paramètres non pris en compte, comme la photopériode, pourraient également jouer un rôle important, notamment sur l'entrée en diapause interrompant le développement.

Cependant, cette limite n'est pas problématique ici car on s'intéresse uniquement à la dynamique de la population au cours d'une seule campagne de la culture de la luzerne et aux dégâts causés par l'insecte tôt dans la saison (principalement pendant le stade « bouton vert », en début de période estivale). Si de nouvelles observations se faisaient sur plusieurs années, un suivi plus long chaque année serait intéressant pour quantifier l'importance des abondances d'une année sur l'infestation l'année suivante et potentiellement sur les déterminants de l'entrée en diapause.

Observations vs simulation par commune-année (à la semaine)

Chaque point représente l'effectif de cécidomyies capturées durant la semaine de piégeage (approximation pour les observations réel)

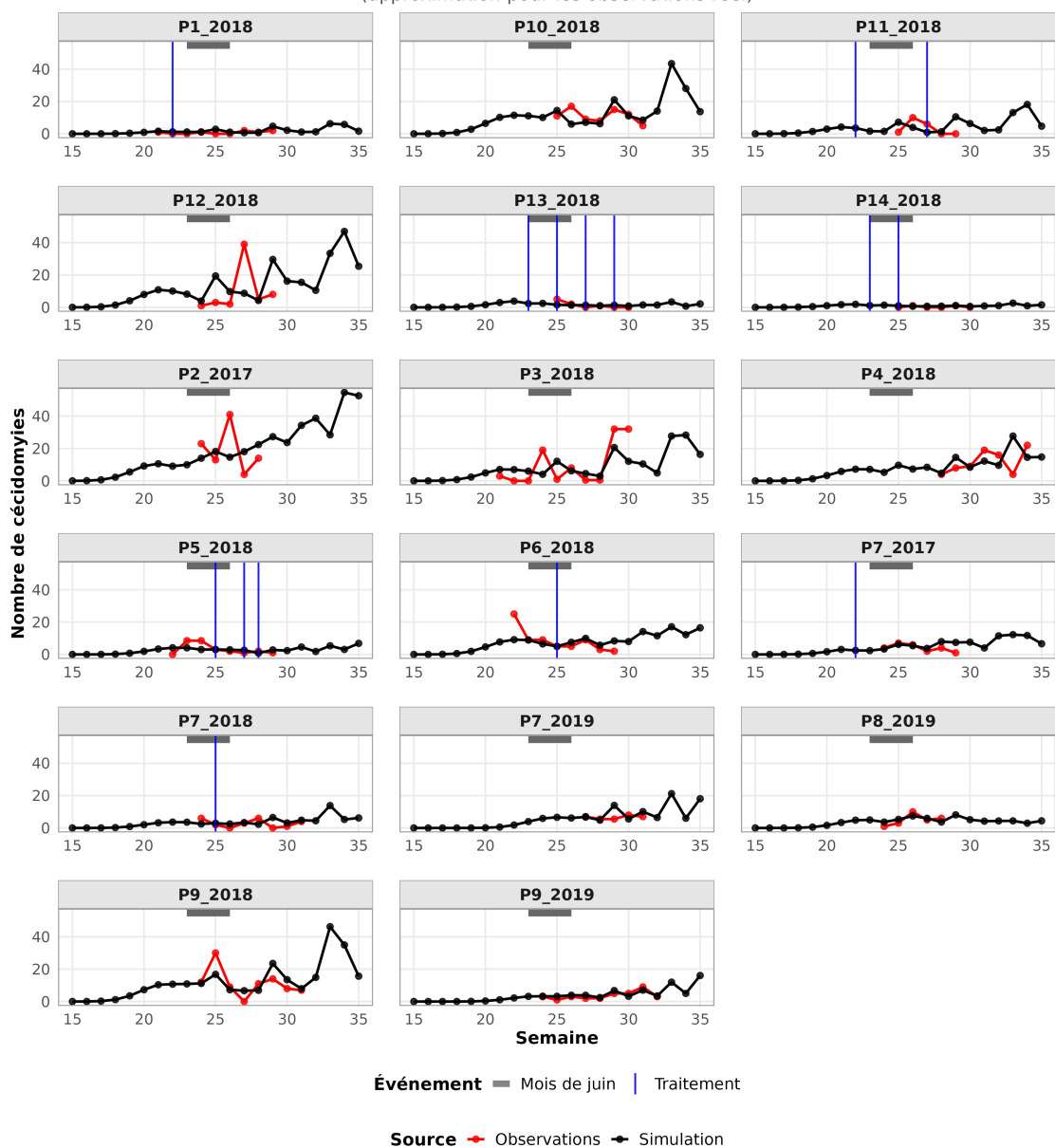


Figure 7 : Observations vs simulation pour l'ajustement sur le mode 1

Si l'ajustement général à l'échelle hebdomadaire laisse à désirer, l'ajustement à l'échelle annuelle (somme des observations annuelles) est bon (Figure 8). On présente ici les résultats pour le mode n°1, le R^2 est légèrement inférieur pour le mode n°2 non présenté, surtout à l'échelle annuelle (R^2 hebdo = 0.1353 au lieu de 0.1693 et R^2 annuel = 0.9262 au lieu de 0.8508). Un bon ajustement général annuel était cependant attendu puisqu'il y a dans le modèle une variable de niveau d'infestation générale par parcelle année.

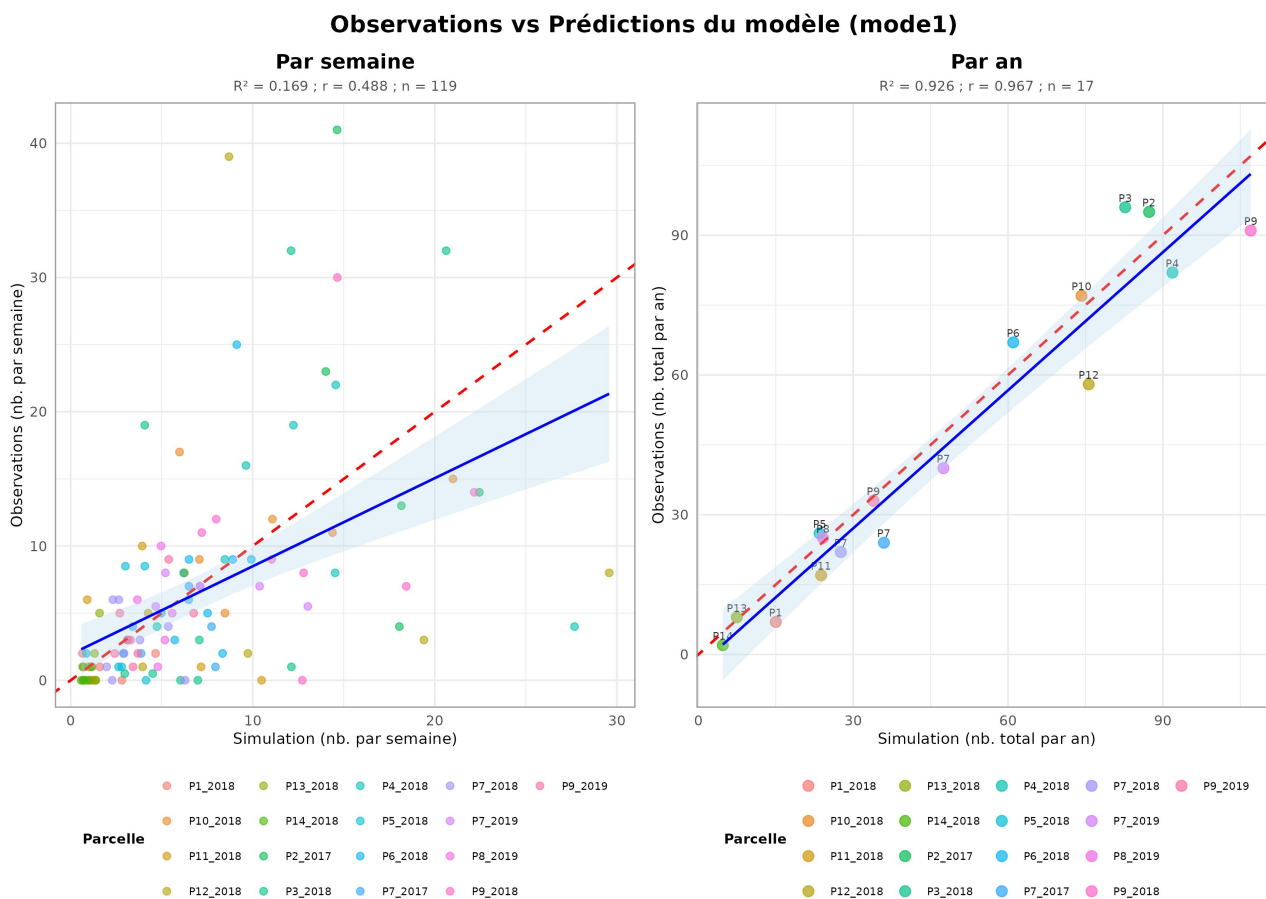


Figure 8 : valeurs prédites en fonction des valeurs observées à la parcelle pour les observations hebdomadaires (gauche) ou annuelles (droite) pour l'ajustement sur le mode 1. Les valeurs médianes d'ajustement sont utilisées pour chaque paramètre. La droite bleue indique la droite de régression et l'ombre bleue l'espace des positions alternatives de la droite de régression d'après les intervalles de confiance de ses paramètres. La droite tiretée rouge correspond à la bissectrice.

4. Conclusion

Les données disponibles, bien que limitées en volume et en durée d'observation, permettent — en s'appuyant sur la bibliographie — de dégager des tendances significatives concernant les vols de cécidomyies :

- Les traitements sont efficaces et ressortent dans tous les modèles testés (linéaires et non linéaires), il est de plus probable qu'ils aient une certaine rémanence dans l'environnement.

Notre estimé ici est une baisse de 2/3 de l'efficacité de l'insecticide tous les 5 jours après la date d'application.

- Les différentes variétés ont probablement des niveaux d'attractivité différents pour les cécidomyies.
- Les données sont cohérentes avec une date de pic de vol pouvant être déterminée approximativement à partir du nombre de degrés-jours, même si le nombre de degrés-jours estimé est légèrement supérieur à celui observé dans la littérature : 275 [235-315] au lieu de 230 degrés-jours au dessus de 10°C à partir du 1er janvier).
- La durée de génération habituelle semble être centrée autour de 29 jours.
- Il ne semble pas y avoir d'effet interannuel marqué sur une parcelle suivie deux années consécutives, aucun accroissement de la population n'a été observé. Cela pourrait être permis pas la mise en place de régulations par des auxiliaires des cultures.

Nos capacités d'analyses sont cependant limitées par les facteurs suivants :

- Pour confirmer les effets variétaux il faudrait augmenter le nombre de parcelles en croisant les variétés sur les différentes localisations.
- Les premières observations sont très probablement trop tardives (début juin) pour observer l'ensemble du premier pic de cécidomyies ce qui limite la fiabilité de nos résultats. Des observations dès début mai garantiraient de refléter l'ensemble du premier vol. Les dernières observations ne permettent pas non plus de caractériser la fin de cycle de la cécidomyie (avec moins d'impact pour l'analyse actuelle).
- Le nombre de communes-années suivies est limité dans ce jeu de données bien que l'approche bayésienne permette tout de même de préciser nos *a priori*.
- L'impact des conditions météorologiques sur la reproduction et plus généralement sur le cycle de vie ne peut pas être ajusté avec les données existantes (trop peu d'informations trop peu réparties dans le temps). En particulier, le modèle ne permet pas de simuler la saisonnalité de la présence de la cécidomyie.
- Au cours de ce contrat, nous nous sommes concentrés sur l'étude des adultes et nous n'avons pas pu étudier les données sur les galles.
- Les estimations réalisées au travers du modèle non-linéaire des observations hebdomadaires ne prennent pas en compte l'autocorrélation temporelle entre les observations pour l'estimation de la vraisemblance. Les intervalles de confiance sont donc assez peu fiables (tendance à être trop faibles) mais prendre en compte correctement cette autocorrélation est encore largement un problème ouvert à moins d'utiliser des méthodes calculatoirement très coûteuses telles que la vraisemblance synthétique qu'il aurait été difficile et peu rentable de mobiliser pour cette étude en raison du faible nombre d'observations et des durées limitées d'observation sur chaque point. Cette limite ne s'applique pas aux résultats des modèles à linéaires à l'échelle de la parcelle.

A condition que les limitations notées ci-dessus sur les données soient levées, il serait pertinent d'affiner le modèle en :

- Ajustant l'impact des conditions de développement (température et hygrométrie, voire photopériode) sur la capacité reproductive.
- Prenant en compte les conditions environnementales sur toute la durée de développement (et non uniquement le jour de ponte).
- Prenant en compte l'environnement paysager (présence de haies, prairies, cultures adjacentes,...).

- Ajustant l'attractivité de la parcelle en fonction de :
 - o la variété de la culture
 - o L'âge de la culture

Par rapport à l'objectif de création d'un OAD de prédiction de la temporalité et de l'ampleur des vols de cécidomyies, nous pouvons déjà dire que le premier vol de cécidomyie est probablement habituellement antérieur au premier stade de susceptibilité de la luzerne. Cependant, la prédiction de l'ampleur du vol (intensité de l'infestation) reste complexe. La variété de luzerne a sans doute un impact, à confirmer avec des données plus abondantes. L'âge de la parcelle n'a sans doute qu'un effet assez faible.

L'observation du premier vol pourrait permettre d'anticiper le deuxième vol, qui semble habituellement coïncider avec le début du stade sensible de la luzerne. Des données plus abondantes et sur une période couvrant plus largement la période de reproduction de la cécidomyie (mi-mai à fin-août) sur plusieurs années et plusieurs bassins de production décorrélés d'un point de vue météorologique pourrait permettre l'ajustement de l'impact des facteurs météorologiques sur le développement des cécidomyies, ce qui améliorerait sans doute la prédiction de l'ampleur des vols.

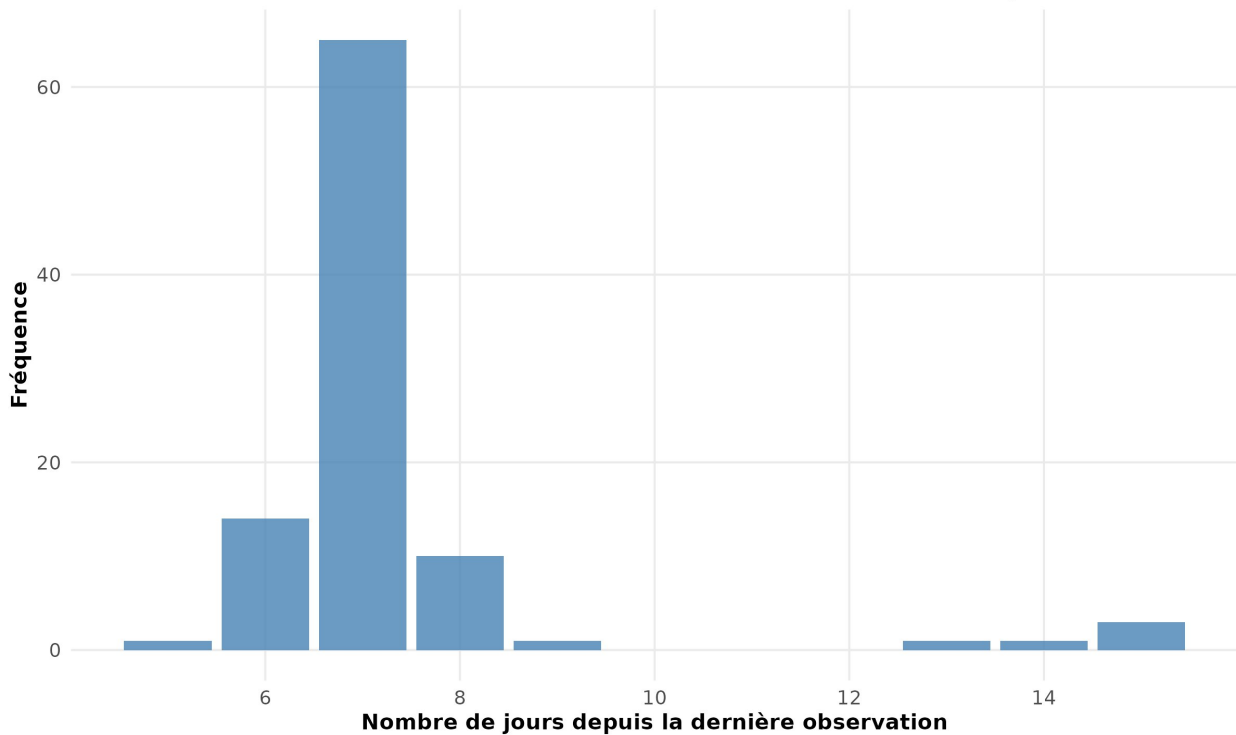
La géolocalisation des parcelles étudiées permettrait d'étudier l'influence du paysage pour affiner notamment l'ampleur prédite de l'infestation. A ce stade, le nombre de localisations était insuffisant pour étudier l'impact du paysage, même si les parcelles avaient été géolocalisées.

Pour la réalisation d'un OAD, **la prédiction de la temporalité paraît assez accessible si des données sont collectées dès début mai** sur un nombre de parcelle comparable à celui étudié ici. Pour pouvoir estimer l'ampleur des vols, il faut sans doute augmenter significativement la taille de l'échantillon et apportant un soin particulier aux variétés testées et à leur distribution au sein des sites d'échantillonnage. Le dimensionnement précis des efforts d'échantillonnage à réaliser en fonction de la précision souhaitée peut faire l'objet d'une prestation spécifique correspondant à quelques jours de travail.

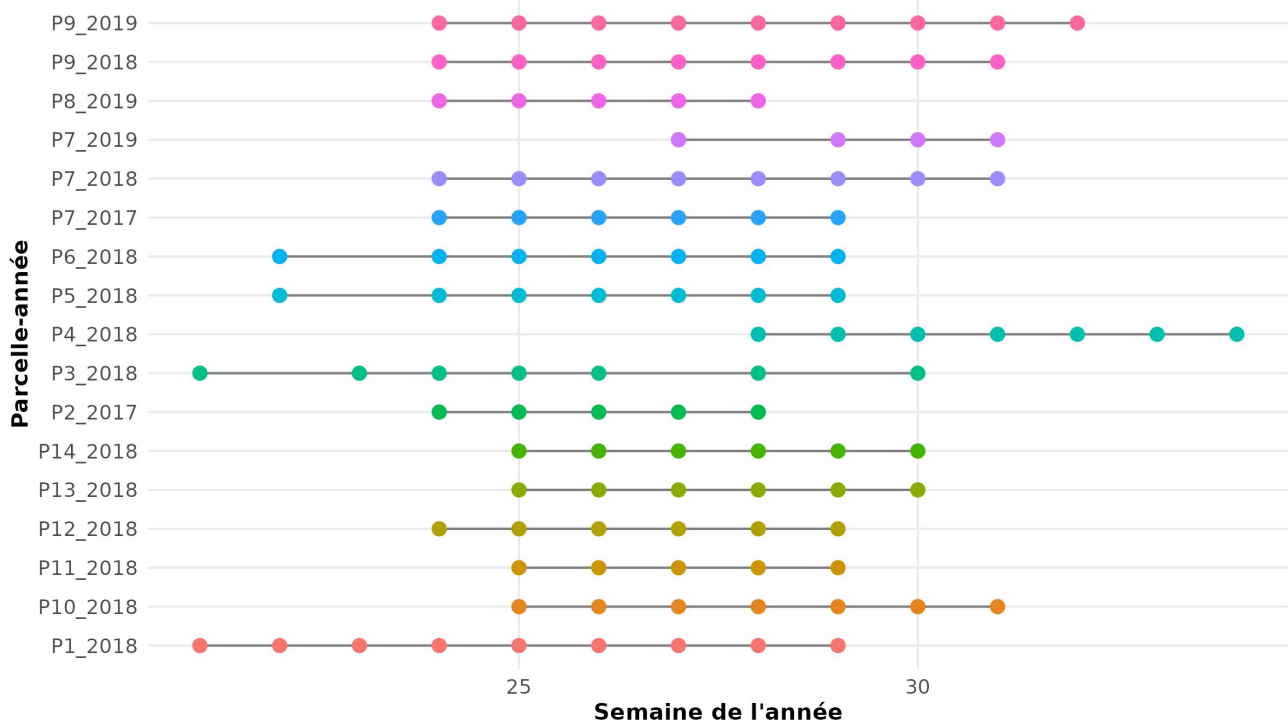
5. Annexes

5.1. Annexe 1 : Délai entre les observations

Délai entre les observations de suivi de la cécidomyie

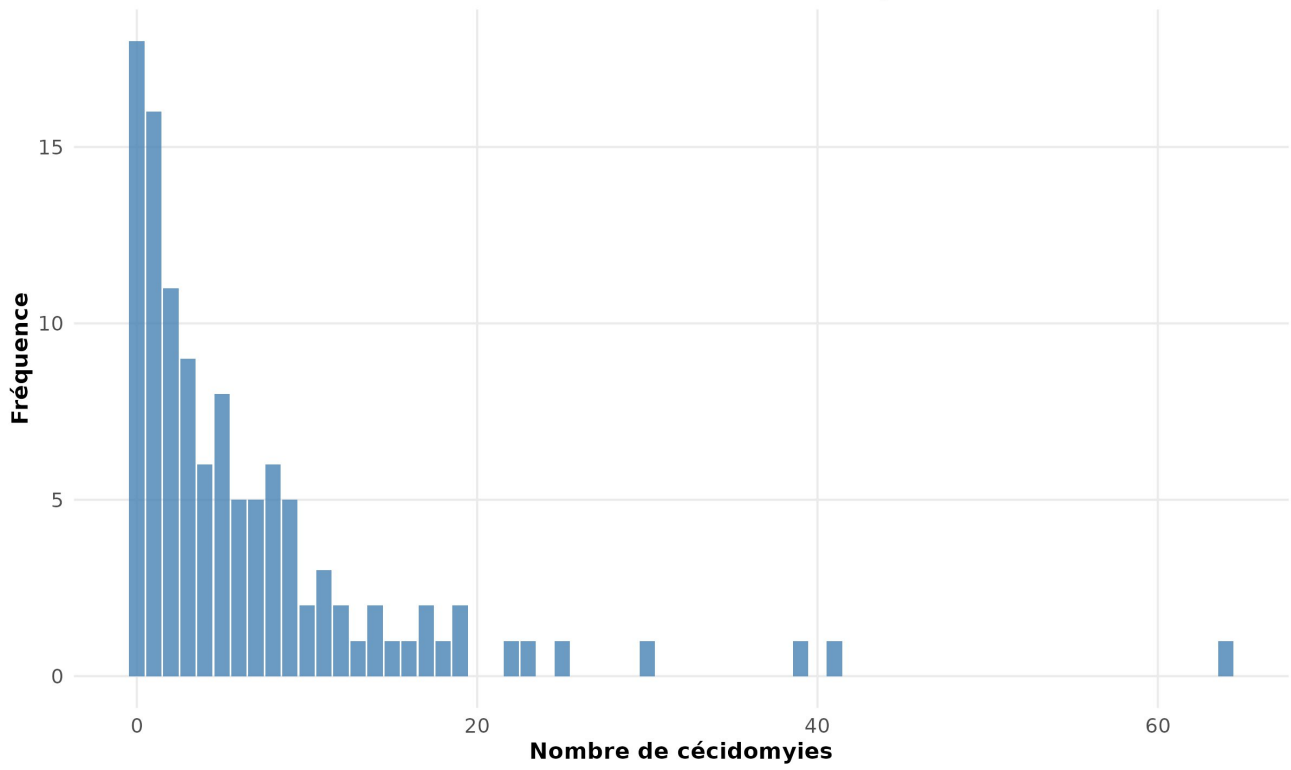


Semaines des observations de la cécidomyie de la luzerne par commune et année d'observation



5.2. Annexe 2 : Distribution des observations d'abondance

Distribution de l'abondance des cécidomyies observées

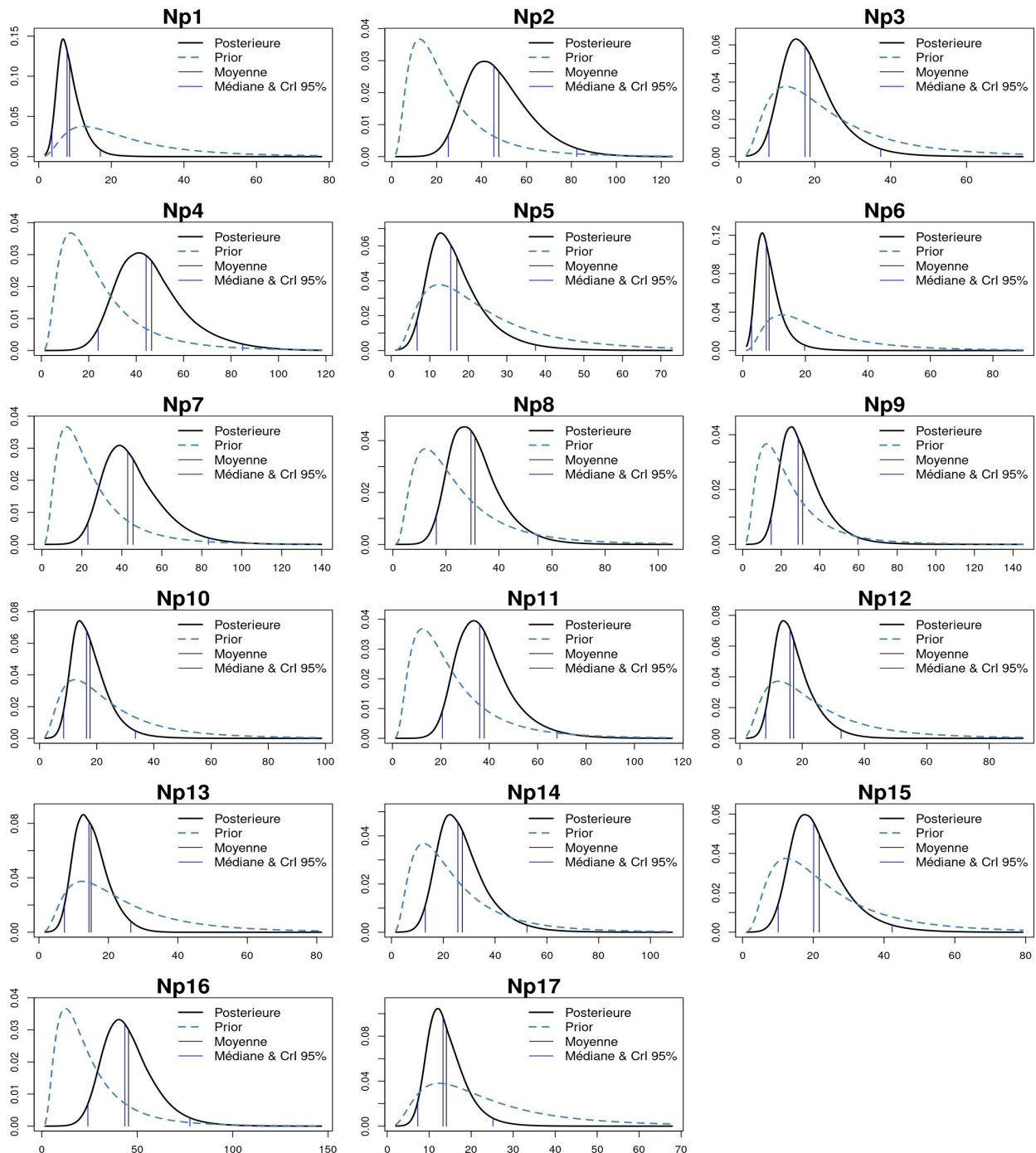


5.3. Annexe 3 : résultats d'ajustement complémentaires

5.3.1. Mode 1

Pour chaque graphique, les abscisses correspondent à la valeur de la variable et les ordonnées correspondent à la densité de probabilité de la valeur.

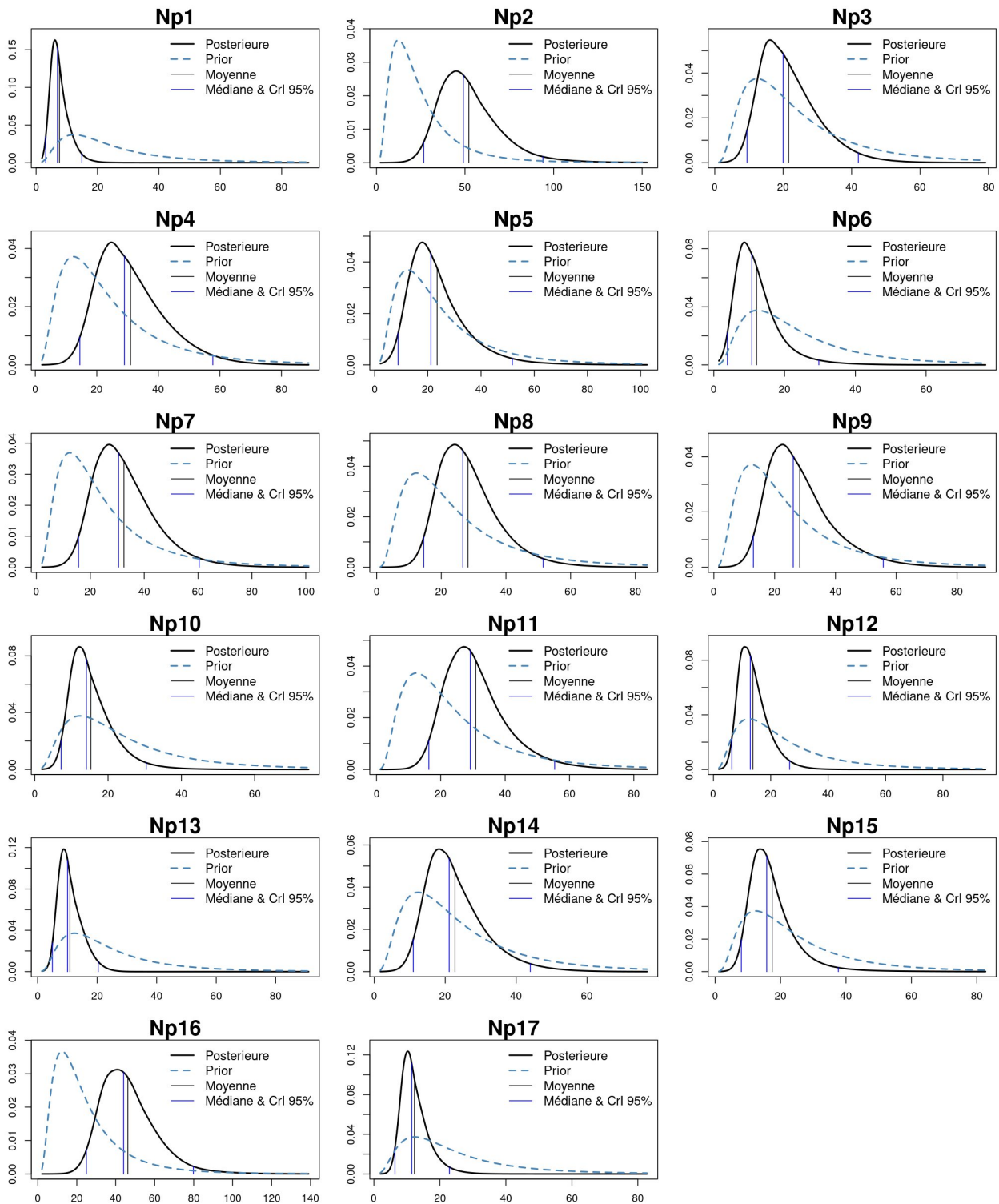
Prior vs Postérieure



5.3.2. Mode 2

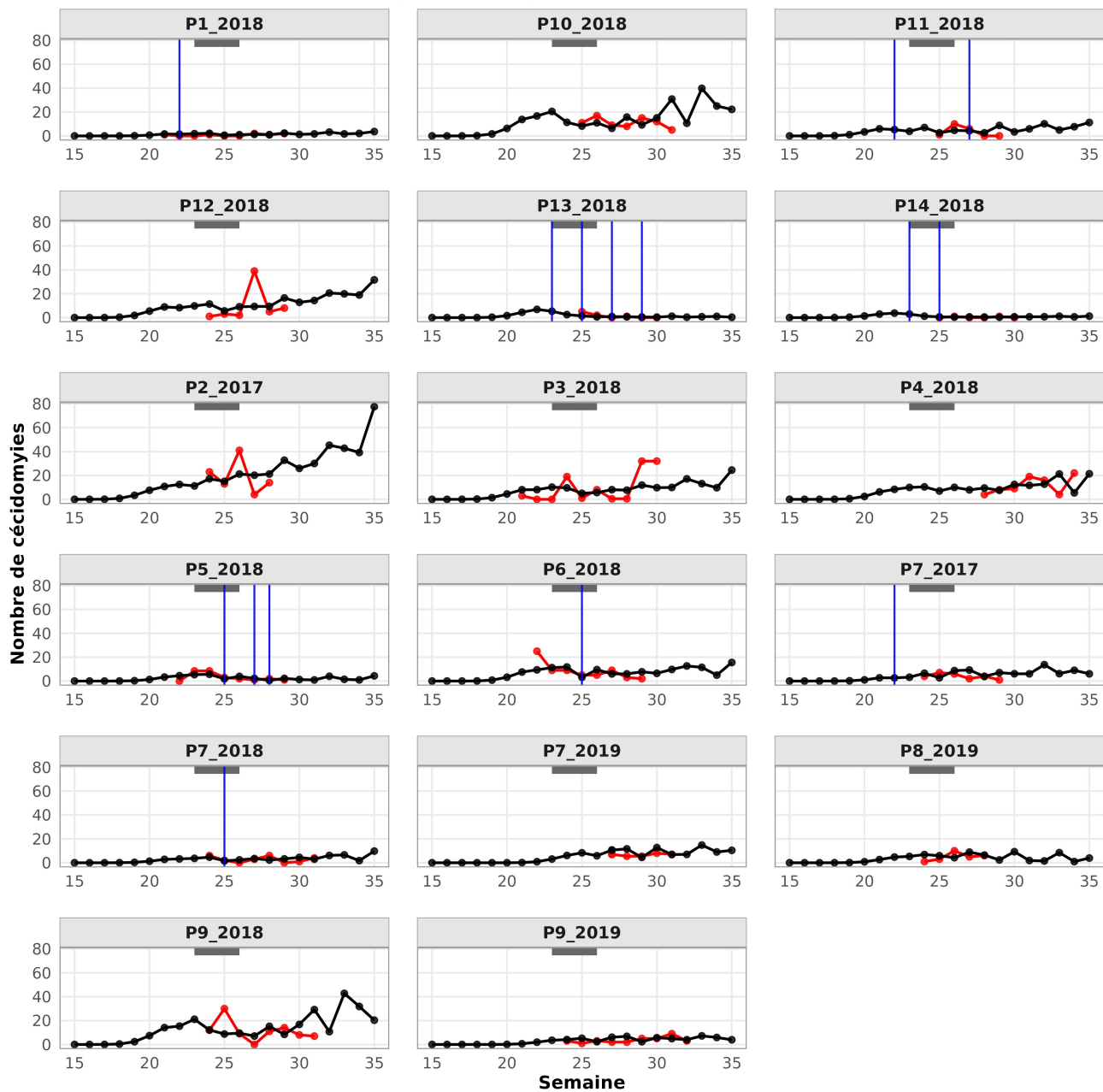
Pour chaque graphique, les abscisses correspondent à la valeur de la variable et les ordonnées correspondent à la densité de probabilité de la valeur.

Prior vs Postérieure



Observations vs simulation par commune-année (à la semaine)

Chaque point représente l'effectif de cécidomyies capturées durant la semaine de piégeage (approximation pour les observations réel)



Événement — Mois de juin | Traitement

Source — Observations — Simulation

5.4. Visualisation des chaînes d'estimation des paramètres

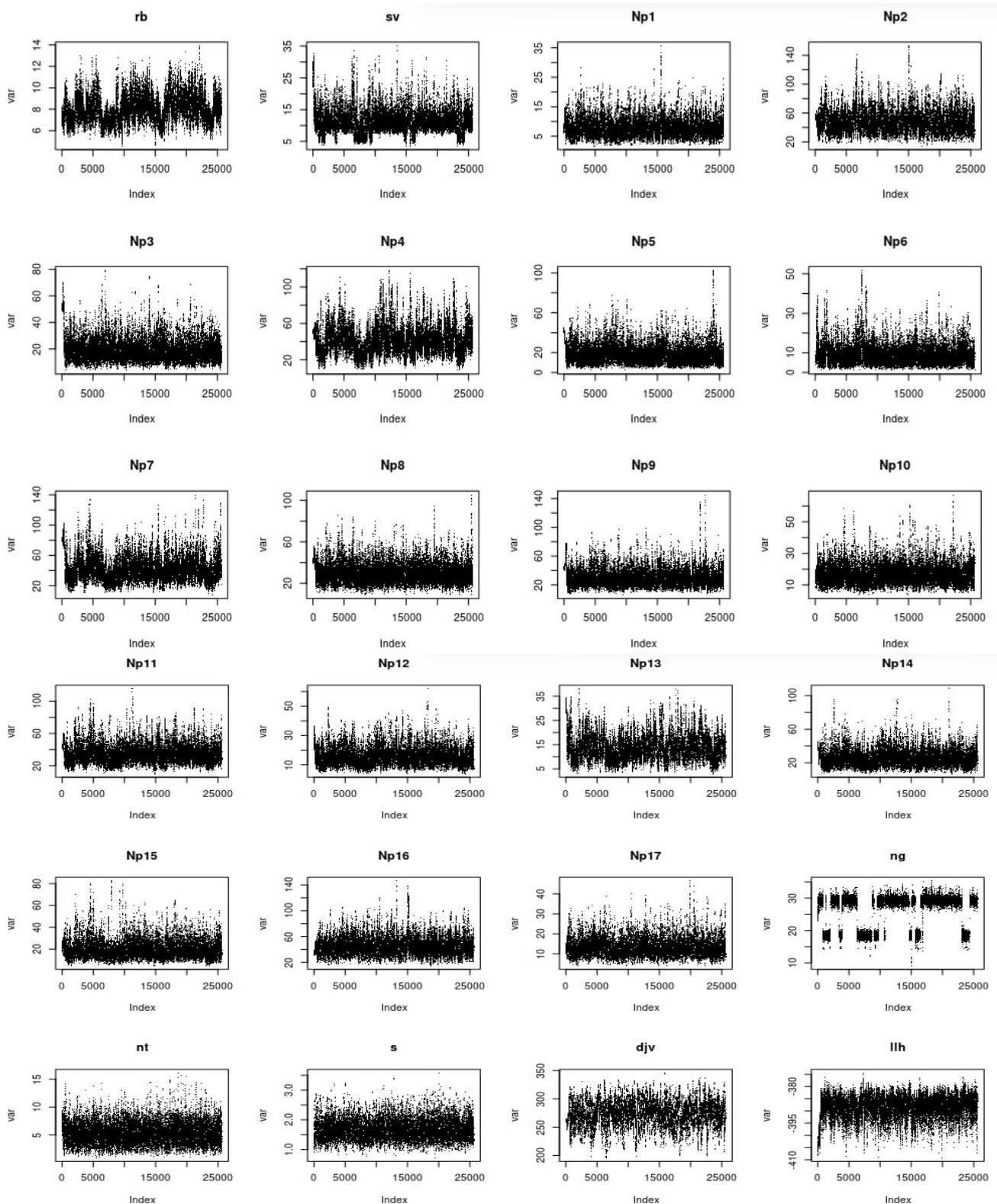


Figure supplémentaire : Visualisation des chaînes MCMC d'ajustement du modèle non linéaire. Les abscisses sont les itérations de la chaîne (un point par estimation), les y sont les valeurs estimées de chaque variable ou de la log-vraisemblance (llh). Les chaînes sont considérées comme convergentes si les moyennes et la variabilité sont stables au cours des itérations. Il peut néanmoins y avoir de

l'autocorrélation dans la chaîne (ex : rb), allant pour ng jusqu'à une « bimodalité stable » de ng. Pour estimer précisément la proportion des modalités haute (1) et basse (2) il faudrait plus d'itérations mais dans le rapport nous nous focalisons sur la modalité haute (1).

6. Références bibliographiques

[1] BOURNOVILLE, René, DELAUDE, Alain, et CANTOT, Pierre. L'organisation de la lutte contre la cécidomyie des fleurs de luzerne (*Contarinia medicaginis* Kieff.): contribution des travaux réalisés en conditions agronomiques. *Agronomie*, 1986, vol. 6, no 4, p. 393-400.

[2] M. N. Berim (2009) : *Contarinia medicaginis* Kieffer - Lucerne Flower Midge, Lucerne Blossom Gallfly. Project «Interactive Agricultural Ecological Atlas of Russia and Neighboring Countries. Economic Plants and their Diseases, Pests and Weeds ».

[3] Piart Jean. (1962). *Etudes biologiques sur la cécidomyie de la luzerne Contarinia medicaginis K. et ses parasites : résultats 1962*. sl : ORSTOM, 57 p. multigr.

[4] Pierre, E. *Contarinia medicaginis*. Dans : *Encyclopédie Hypp : Protection des plantes*. INRAE, 2018-07-04. Disponible sur : <https://ephytia.inra.fr/fr/C/11668/Hypp-encyclopedie-en-protection-des-plantes-Contarinia-medicaginis> (consulté le 29 janvier 2026).