

IMPLICATION DES OISEAUX DANS LE CYCLE ÉPIDÉMIOLOGIQUE DES *Borrelia*, AGENTS DE LA MALADIE DE LYME

Rapport de stage
M2 Recherche en Santé Publique
Parcours "biostatistiques"

Alex DROUIN

Encadrement :

MARSOT Maud - ANSES - LSA_n - EPI

HENRY Pierre-Yves - UMR 7179 CNRS MNHN - MECADEV et UMR 7204 CNRS
MNHN SU - CESCO

MOUTAILLER Sara - ANSES INRA ENVA Université Paris-Est - UMR BIPAR

Laboratoire d'accueil :

Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail
Laboratoire de Santé Animale de Maisons-Alfort et de Dozulé
Unité Épidémiologie

Période : février - juillet 2019

Remerciements

Je tiens à remercier tout particulièrement Maud MARSOT pour m'avoir à nouveau fait confiance pour ce projet passionnant, pour son encadrement, ses conseils et son soutien infaillible.

Je remercie également Pierre-Yves HENRY pour ses idées, ses conseils et pour ses précieux éclairages en écologie aviaire.

Je remercie Sara MOUTAILLER pour son aide et ses explications des techniques de laboratoire

Je remercie Clémence GALLON, Élodie DEVILLERS et l'ensemble de l'équipe de l'unité BIPAR pour leur patience et pour m'avoir formé aux techniques de laboratoire.

Je remercie l'intégralité de l'équipe EPI, Benoît, Gina, Laetitia, Noémie, Hélène, Simone, Gabriella, Cynthia, Erika et Viviane pour leur accueil et leur bonne humeur.

Je remercie Colin FONTAINE pour son aide méthodologique concernant les analyses de réseau

Je remercie Romain PROVOST et l'ensemble des bagueurs du CRBPO pour leur aide et leur participation à cette étude.

Enfin, je remercie Nadine et Pauline pour leur bienveillance et leur soutien tout au long de cette année.

Résumé

La borréliose de Lyme, affectant à la fois l'Homme, les bovins, les chevaux et les chiens, est une maladie causée par un ensemble de bactéries appartenant au genre *Borrelia*. Cette maladie est vectorielle et uniquement transmise par les tiques, principalement *Ixodes ricinus* en Europe. Le rôle des mammifères dans le cycle épidémiologique des *Borrelia* est bien étudié, mais il existe peu d'études concernant l'implication des oiseaux dans ce cycle, notamment pendant leur période de reproduction. Pourtant, cette période correspond au pic de densité et d'activité des tiques, mais également avec la période durant laquelle les oiseaux évoluent dans une zone géographique restreinte et la saison durant laquelle l'Homme réalise des activités de plein air, s'exposant aux piqûres de tiques et donc à l'infection par les *Borrelia*. Pour étudier le rôle des oiseaux communs dans le cycle épidémiologique des *Borrelia*, durant leur période de reproduction en France, nous nous sommes intéressés dans un premier volet aux déterminants responsables de la variabilité temporelle et interspécifiques de la charge en tique des oiseaux. Nous avons mis en place des modèles de comptage (Poisson, binomial négatif et Zero-Inflated Negative Binomial) et mis en évidence l'existence d'une variabilité temporelle. Par ailleurs, certaines espèces présentaient une charge en tiques plus élevée que la moyenne, comme le Rossignol philomèle, l'Accenteur mouchet, le Merle noir et la Grive musicienne, au contraire du Pic épeiche, de la Mésange nonnette, de la Mésange bleue, de la Mésange huppée, du Pouillot véloce et du Pouillot fitis. Dans cette étude, il est apparu que la variabilité de la charge en tique était liée à des traits d'histoire de vie aviaires, en augmentant avec le poids moyen de l'espèce et en diminuant avec la hauteur moyenne d'alimentation. Cette variabilité était aussi liée avec le sexe et l'âge, puisque les jeunes étaient plus porteurs de tiques que les mâles et les femelles adultes. Dans un second volet, nous avons réalisé des analyses sur un réseau 2-modes oiseaux-agents pathogènes, afin de mettre en place l'approche méthodologique qui sera utilisée dans la suite du projet. Nous avons étudié certaines métriques (degrés) et réalisé un partitionnement du réseau par maximum de modularité. Les résultats obtenus sont intéressants et prometteurs pour la suite de l'étude, lorsque ces méthodes seront appliquées sur les données réelles.

Table des matières

1	Introduction	1
1.1	Contexte	1
1.2	Objectifs du stage	6
2	Matériels et méthodes	7
2.1	Données	7
2.1.1	Premier volet : variabilité de la charge en tiques des oiseaux	7
2.1.2	Second volet : étude du réseau oiseaux-agents pathogènes	8
2.2	Méthodes statistiques utilisées	10
2.2.1	Premier volet : variabilité de la charge en tiques des oiseaux	10
2.2.2	Second volet : étude du réseau oiseaux-agents pathogènes	18
3	Résultats	24
3.1	Premier volet : Variabilité de la charge en tiques des oiseaux	24
3.2	Second volet : étude du réseau oiseaux-agents pathogènes	28
4	Discussion	32

1 Introduction

1.1 Contexte

La maladie de Lyme ou borréliose de Lyme est une maladie bactérienne causée par un ensemble de bactéries de la classe des spirochètes et du genre *Borrelia* (anciennement nommé "complexe *Borrelia burgdorferi sensu lato*") (Centre National de Référence des Borrelia, 2018). Le genre *Borrelia* compte plus de 20 espèces, dont certaines sont pathogènes pour l'Homme : *Borrelia burgdorferi* (majoritaire aux États-Unis), *B. afzelii* et *B. garinii* (majoritaires en Europe), *B. bavariensis* ou *B. spielmanii* (également présentes en Europe) (Rizzoli et al., 2011; Steere et al., 2016).

Cette maladie s'exprime cliniquement chez l'Homme, le Cheval, le Chien et les bovins (Boulouis et al., 2015). Chez l'Homme, elle se décompose en trois stades : le stade I (érythème migrant, syndrome grippal ou asymptomatique), le stade II (neuroborréliose, arthrite ou lymphocytome) (Rizzoli et al., 2011; Stanek et al., 2012; Bush & Vazquez-Pertejo, 2018) et le stade III, dont l'existence fait polémique au sein des autorités de santé françaises (Haute Autorité de Santé, 2018; Académie Nationale de Médecine, 2018). La diversité de la symptomatologie est liée à l'espèce impliquée dans l'infection. En effet, *B. burgdorferi* est le plus souvent associée à l'apparition d'arthrite et de neuroborréliose, *B. garinii* avec la neuroborréliose et *B. afzelii* avec l'acrodermatite chronique atrophiante (Rizzoli et al., 2011).

En Europe, la maladie est ubiquitaire, mais les incidences les plus élevées sont observées en Europe centrale et en Europe du Nord (Mead, 2015). En France, il y a eu 45 000 cas en 2017, selon Santé Publique France (Septfonds, 2019).

La borréliose de Lyme est une maladie vectorielle (c'est à dire dont l'agent pathogène est transmis par un arthropode hématophage), uniquement transmise par piqûre de tique. En Europe, le principal vecteur pour l'Homme est *Ixodes ricinus*, tique ubiquiste capable de se nourrir sur une très grande diversité d'hôtes vertébrés (mammifères, oiseaux,...). Le cycle de ce vecteur est triphasique, c'est à dire que chaque stase de la tique (larve, nymphe et adulte femelle) se nourrit sur un hôte différent, avant de se décrocher et de muer ou de pondre dans le cas des femelles adultes (Apanaskevich & Oliver, 2014). Chez les tiques, la transmission transtadiale de l'agent pathogène est possible, c'est à dire la transmission d'une stase à une autre, lors de la mue (comme montrée pour *B. garinii*, *B. valaisiana* et *B. burgdorferi* chez *I. ricinus* par Heylen et al. (2014b)). C'est pourquoi, lors de chaque repas sanguin, si la tique se nourrit sur un hôte infecté, elle sera susceptible de devenir elle-même infectée, de transmettre le spirochète à la stase suivante et enfin de transmettre l'agent pathogène à un nouvel hôte lors du repas sanguin suivant (Figure 1). En revanche, la transmission trans-ovarienne (c'est à dire d'une femelle adulte à ses oeufs puis aux larves) est considérée comme négligeable, bien que son existence soit possible. Il existe alors deux scénarios conduisant à la présence de *Borrelia* chez une larve : soit celle-ci a effectué son repas sanguin sur un hôte infecté, soit il y a eu transmission du spirochète entre une autre tique infectée et la larve se nourrissant à proximité l'une de l'autre, et sans bactériémie de l'hôte vertébré. Dans ce dernier cas, on parle alors de "co-feeding" (Heylen et al., 2013b; Coipan & Sprong, 2016; Palomar et al., 2017).

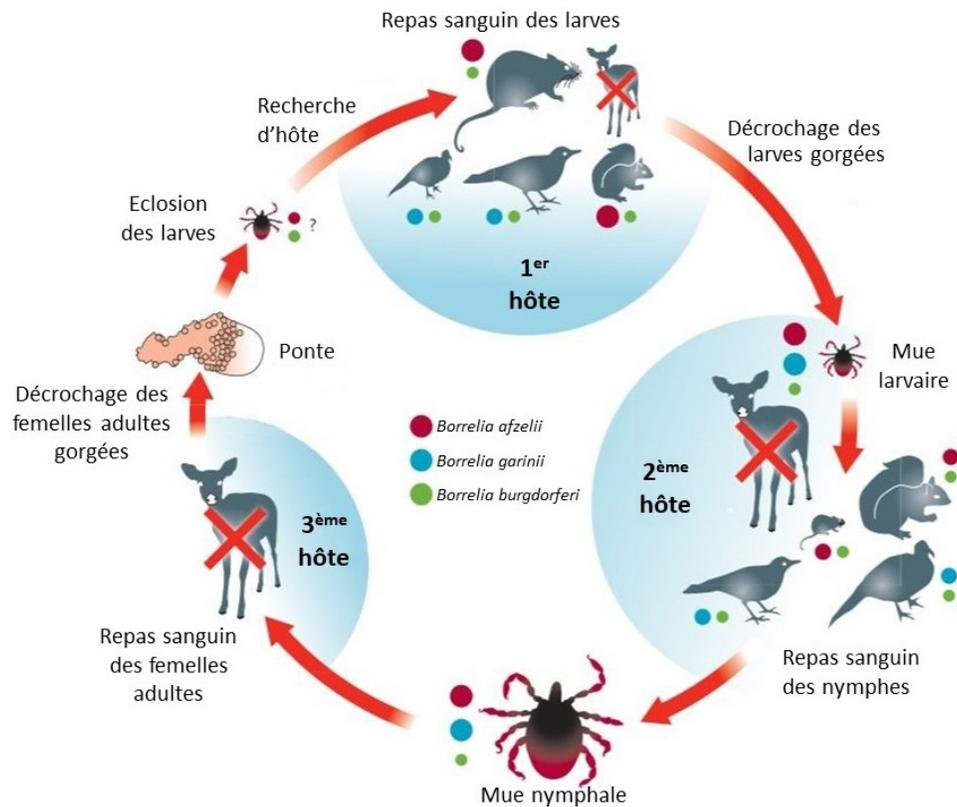


FIGURE 1 – Cycle européen des *Borreliella*, d'après Stanek et al. (2012)

Bien que certaines espèces de bactéries soient plus généralistes, comme *B. burgdorferi* qui est liée aux mammifères et aux oiseaux (Heylen et al., 2014b), d'autres espèces du genre *Borreliella* sont spécialisées dans les espèces vertébrées qu'elles infectent. En effet, certaines sont liées aux reptiles, comme *B. lusitaniae*, d'autres aux mammifères, comme *B. afzelii*, *B. bavariensis* ou *B. spielmanii*. D'autres encore infectent préférentiellement les oiseaux, comme *B. garinii*, *B. valaisiana* et *B. turdi* (Baranton & de Martino, 2009; Coipan & Sprong, 2016; Heylen et al., 2017). Les espèces de *Borreliella* des mammifères ne sont pas considérées comme susceptibles de pouvoir perdurer chez les oiseaux (Heylen et al., 2013b).

Alors que le cycle épidémiologique des *Borreliella* impliquant *I. ricinus* et les mammifères est bien documenté, ceux concernant les tiques spécialisées et leurs hôtes sont moins connus, comme les cycles enzootiques existant entre les tiques ornitophiles et les oiseaux (Heylen et al., 2013b).

L'existence d'un cycle épidémiologique entre les oiseaux et les tiques dépend de la compétence vectorielle de la tique et du rôle d'hôte réservoir de l'oiseau. La compétence vectorielle est l'aptitude d'un arthropode à ingérer un agent pathogène, en assurer la multiplication et/ou le développement et le transmettre à un hôte vertébré. Elle est propre à un couple vecteur-agent pathogène et dépend du degré de compatibilité entre ces deux acteurs. Un hôte réservoir est un animal jouant le rôle de réservoir pour un agent pathogène donné, c'est-à-dire

assurant sa conservation, en tant qu'espèce, et sa transmission aux sujets réceptifs. Cette espèce représente une source majeure d'agents pathogènes pour l'infection d'autres animaux (Dufour et al., 2018). La possibilité de transmission des spirochètes responsables de la maladie de Lyme au sein des cycles épidémiologiques dépend donc de l'espèce d'oiseau, de l'espèce de tique et de l'espèce de *Borrelia* elle-même.

Les trois espèces majoritaires de tiques que l'on retrouve sur les oiseaux en Europe sont *I. ricinus*, *I. frontalis* et *I. arboricola*. Cette dernière est une tique endophile, vivant dans les nids des oiseaux et s'y détachant après le repas sanguin. Elle partage plusieurs espèces hôtes avec *I. ricinus* (Heylen et al., 2014a,c). La biologie d'*I. frontalis* est peu connue, mais bien qu'on puisse l'observer dans le milieu extérieur, cette espèce est principalement retrouvée dans les nids d'oiseaux. *I. frontalis* réalise son repas sanguin sur de nombreuses espèces d'oiseaux et partage l'intégralité de ses hôtes avec *I. ricinus* (Hamer et al., 2011; Heylen et al., 2014a, 2017).

Le rôle des oiseaux en tant que réservoir des bactéries responsables de la maladie de Lyme est reconnu depuis de nombreuses années (Heylen et al., 2013b). En effet, le Merle noir (*Turdus merula*, Figure 2c) est connu comme hôte réservoir de certaines espèces de *Borrelia* (Heylen et al., 2017), et notamment de *B. turdi* et *B. valaisiana* (Palomar et al., 2017). La Mésange charbonnière (*Parus major*, Figure 2a) pourrait également avoir ce rôle (Heylen et al., 2013b) et cela a été prouvé pour *B. turdi* (Norte et al., 2015), bien que la pathogénicité de cette espèce n'ait pas été montrée pour l'Homme. Enfin, le Rouge-gorge familier (*Erithacus rubecula*, Figure 2d) serait un réservoir probable de *Borrelia* (Norte et al., 2015). Concernant les tiques, *I. arboricola* et *I. frontalis* sont des vecteurs compétents pour certaines espèces de *Borrelia* (Heylen et al., 2013b). Il a été montré que *I. frontalis* serait un vecteur compétent de au moins *B. turdi* chez *T. merula* et que *I. ricinus* serait un vecteur compétent de *B. valaisiana* et *B. turdi* chez cette même espèce d'oiseau (Heylen et al., 2017). Il a également été mis en évidence que *B. garinii* est plus associée à *P. major* que *B. valaisiana*, alors qu'on observe l'effet inverse chez les oiseaux de la famille des Turdidae (comme le Merle noir *T. merula* ou la Grive musicienne *T. philomelos*) (Heylen et al., 2014b). Ces associations différentielles entre espèces d'oiseaux et de *Borrelia* pourrait être liées à des différences concernant le système immunitaire des oiseaux (Heylen et al., 2017).

Pour certaines espèces de *Borrelia* (*B. garinii*, *B. valaisiana*, *B. turdi*), il existerait donc un cycle épidémiologique, évoluant entre les oiseaux et les tiques ornithophiles. La transmission et l'amplification des souches de *Borrelia* au sein de ces cycles dépend d'interactions uniques entre l'espèce hôte et probablement de facteurs relatifs au vecteur (Heylen et al., 2014b). *I. ricinus* étant ubiquiste et partageant plusieurs espèces hôtes avec les tiques ornithophiles, elle pourrait être capable de faire le lien entre les cycles épidémiologiques ornithologiques et d'autres espèces de mammifères, dont l'Homme (Heylen et al., 2017).

En marge des possibles interactions existant entre les tiques, les oiseaux et les *Borrelia*, la majorité des études concernant la contribution des oiseaux à l'évolution de la maladie de Lyme portent sur les périodes de migration et sur la capacité des oiseaux à disperser les tiques et les *Borrelia* qui les infectent (Olsén et al., 1995; Poupon et al., 2006). En revanche, peu d'études ont porté sur la période de reproduction des oiseaux et la participation de ceux-ci

aux dynamiques locales de populations de tiques (par la quantité de tiques qui les infeste, on parle de "charge en tiques") et de *Borreliella* associées à celles-ci, et aux éventuels facteurs de variation. Pourtant, cette période, s'étendant environ de avril à juin, coïncide avec la période de forte activité des tiques, et particulièrement *I. ricinus*. Durant cette période, les oiseaux sont sédentaires et contribuent dans une plus grande mesure à la circulation locale de *Borreliella* (Marsot et al., 2012; Heylen et al., 2013a). Enfin, la période de reproduction des oiseaux correspond au printemps, saison durant laquelle l'homme réalise des activités de plein air, ce qui l'expose aux piqûres de tiques et donc à l'infection par les *Borreliella*.



(a) *Parus major*



(b) *Cyanistes caeruleus*



(c) *Turdus merula*



(d) *Erithacus rubecula*

FIGURE 2 – Principales espèces d'oiseaux infestées par les tiques (Muséum National d'Histoire Naturel, 2019)

Certaines études ont permis de mettre en évidence *T. merula* comme étant une des espèces les plus infestées par les tiques (Doby, 1999; Norte et al., 2015). D'autres ont montré que *P. major* a une infestation en tiques plus importante que *Cyanistes caeruleus* (Mésange bleue, Figure 2b) (Heylen et al., 2013b), notamment en *I. ricinus*, mais qu'il n'y a pas de différence pour *I. arboricola*. Cette différence pourrait être liée aux différences de cycles biologiques des tiques et aux traits d'histoire de vie des oiseaux. En effet, *I. arboricola* est nidicole et infesterait ces deux espèces d'oiseaux de la même manière, puisqu'elles utilisent toutes deux des nids situés dans des cavités. Au contraire, la différence d'infestation en

I. ricinus est certainement liée au fait que *P. major* passe plus de temps à chercher de la nourriture au sol, dans le milieu de vie de *I. ricinus*, que *C. caeruleus* (Heylen et al., 2014c). Il a ainsi été montré que la charge en tiques varie en fonction de l'utilisation verticale de l'espace. Les oiseaux qui se nourrissent dans les étages inférieurs de la végétation, comme *T. merula*, *T. philomelos* et *E. rubecula* seraient plus infestés de tiques que d'autres espèces qui se nourrissent en dehors de la zone d'habitat de *I. ricinus*, comme *C. caeruleus* ou *Sitta europaea* (Sittelle torchepot). Parallèlement aux variations interspécifiques, la charge en tiques des oiseaux pourrait dépendre de facteurs climatiques, de la densité en tiques dans l'environnement, ou de caractéristiques propres à l'oiseau comme le sexe, l'âge, le poids ou l'immunité (Marsot et al., 2012; Heylen et al., 2013a). La densité en tiques, correspondant à une évaluation de la présence des tiques à l'affût sur la végétation, est variable au cours des années (Ruyts et al., 2018), et ce phénomène est principalement lié aux variations climatiques (Heylen et al., 2013a). On peut donc penser que la charge en tiques des oiseaux pourrait également présenter des variations temporelles, bien que cela n'ait pour l'instant pas été démontré en France (Marsot et al., 2012).

Concernant la prévalence en *Borrelia* chez les tiques portées par les oiseaux, une étude belge a mis en évidence *P. major* comme étant plus souvent porteur de tiques infectées que *C. caeruleus*, bien que la proportion de tiques infectées ne présentait pas de différence significative entre ces deux espèces (Heylen et al., 2013b). Il n'existe à notre connaissance pas d'autres études européennes ayant étudié la variabilité de prévalence en *Borrelia* chez les tiques portées par les oiseaux. De plus, l'étude de Marsot et al. (2012) constitue la seule étude menée en France concernant la variabilité de la charge en tiques des oiseaux durant la période de reproduction. Pourtant, il existe en France une grande diversité d'espèces aviaires, dont seules quelques unes ont été étudiées comme réservoirs potentiels de l'agent pathogène responsable de la maladie de Lyme.

À la lumière de ce contexte, nous envisageons deux questions scientifiques portant sur l'implication des oiseaux dans le cycle épidémiologique des *Borrelia*, en France et durant la période de reproduction des espèces aviaires les plus communes :

- Quelles caractéristiques individuelles et spécifiques des oiseaux peuvent expliquer la variabilité annuelle et interspécifique de leur charge en tiques ?
- Certaines espèces de *Borrelia* sont-elles inféodées à des espèces aviaires particulières ?

Pour y répondre, nous étudierons dans un premier volet la variabilité de la charge en tiques des espèces communes d'oiseaux en France, en utilisant une base de données du CRBPO (Centre de Recherches sur la Biologie des Populations d'Oiseaux, du Muséum National d'Histoire Naturelle) en ajustant le modèle de comptage le plus adapté pour expliquer les données et mettre en évidence d'éventuels facteurs de variabilité. Dans un second volet, nous nous intéresserons au réseau oiseaux-*Borrelia* afin de déterminer s'il existe des espèces de *Borrelia* inféodées à certaines espèces aviaires, en étudiant le partitionnement du réseau par maximum de modularité. Nous allons pour cela réaliser l'analyse des prélèvements de tiques collectées sur des oiseaux au printemps 2018 par les bagueurs du CRBPO.

1.2 Objectifs du stage

Ce stage s'inscrit dans le cadre du projet OUTLYER (OiseaU Tiques LYmE Risque), consistant à déterminer le rôle des oiseaux en reproduction dans la dynamique locale des populations de tiques et de *Borrelia*, et d'évaluer la contribution relative des différentes communautés aviaires au risque associé à la maladie de Lyme pour l'Homme en France. Ce projet est coordonné par Maud MARSOT (Unité EPI, Laboratoire de Santé Animale, ANSES), Pierre-Yves HENRY (UMR 7179 CNRS MNHN - MECADEV et UMR 7204 CNRS MNHN/SU - CESCO) et Sara MOUTAILLER (Équipe Vectotiq, UMR BIPAR, ANSES/INRA/ENVA/Université Paris-Est) et donnera lieu à un projet de thèse débutant en 2019. L'objectif de ce stage est de réaliser une première exploration des méthodes statistiques permettant d'analyser les données obtenues, ainsi que d'obtenir des résultats préliminaires permettant d'orienter la suite du projet. L'ensemble des analyses statistiques de ce travail ont été implémentées sous R (version 3.5.2.).

2 Matériels et méthodes

2.1 Données

2.1.1 Premier volet : variabilité de la charge en tiques des oiseaux

Afin de répondre à la question portant sur les déterminants de la variabilité de la charge en tiques, nous disposons de la base de données du CRBPO. Celle-ci contient les données de capture d’oiseaux depuis les années 1990, captures réalisés par les bagueurs bénévoles contribuant aux différents programmes de capture. Nous nous sommes intéressés aux oiseaux étudiés dans le cadre du programme STOC-Capture (Suivi Temporel des Oiseaux Communs par Capture), c’est-à-dire durant la saison de reproduction (avril-juin). Pour chaque oiseau capturé, diverses informations sont relevées, dont :

- la date de capture ;
- l’action réalisée (bagueage ou contrôle de l’oiseau, reprise si l’oiseau est retrouvé mort) ;
- le numéro de bague de l’individu ;
- son espèce ;
- son sexe ;
- son âge ;
- des informations relatives au bagueurs et à la station de capture ;
- le nombre éventuel de tiques comptées autour du bec (zone sur laquelle on retrouve 95% des tiques sur les oiseaux)

Dans le cadre de cette étude, et afin d’assurer une standardisation de la méthode de comptage du nombre de tiques sur les oiseaux, nous avons limité notre analyse à la station STOC 204, située dans la forêt de Sénart (Draveil, 91). Ce choix permettait également d’assurer un nombre conséquent de comptages ainsi qu’un suivi temporel sur plus de 10 ans (2007-2018). Pour chaque espèce d’oiseau, des variables relatives aux traits d’histoire de vie ont été créées et ajoutées à la base de données principale :

- le poids moyen de l’espèce, d’après Storchová & Hořák (2018) ;
- la hauteur moyenne d’alimentation, calculée d’après la proportion de temps passé à s’alimenter dans les différentes strates de la végétation (Wilman et al., 2014). Soit P_G la proportion de temps passé à s’alimenter au niveau du sol, P_U dans le sous-étage de la végétation ($< 2\text{m}$), P_M à mi-hauteur ($> 2\text{m}$ mais sous la canopée) et P_C celle passée à s’alimenter au niveau de la canopée ou juste au dessus. On calcule alors la hauteur moyenne d’alimentation H (en cm) par : $H = 5 \times P_G + 100 \times P_U + 400 \times (P_M + P_C)$;
- une recodification du sexe et de l’âge en variable catégorielle prenant en compte les deux effets ensembles. En effet, les jeunes oiseaux (avant leur mue post-juvénile) ne présentant généralement pas de dimorphisme sexuel, ni phénotypique ni comportemental, il apparaît judicieux de considérer trois classes : les jeunes oiseaux / les femelles adultes / les mâles adultes.

Enfin, dans les cas pour lesquels un oiseau avait été capturé plusieurs fois dans la même saison, nous avons gardé uniquement le premier comptage de cette saison. En effet, étant donné que certaines tiques ont été retirées de l'oiseau lors de la première capture, le nombre de tiques présentes lors des suivantes est dépendant de la première.

En s'intéressant à l'histogramme du nombre moyen de tiques par année et par espèce (Figure 3), on peut faire l'hypothèse qu'il existe une variabilité entre espèces au sein d'une même année, ainsi qu'une variabilité globale entre les années.

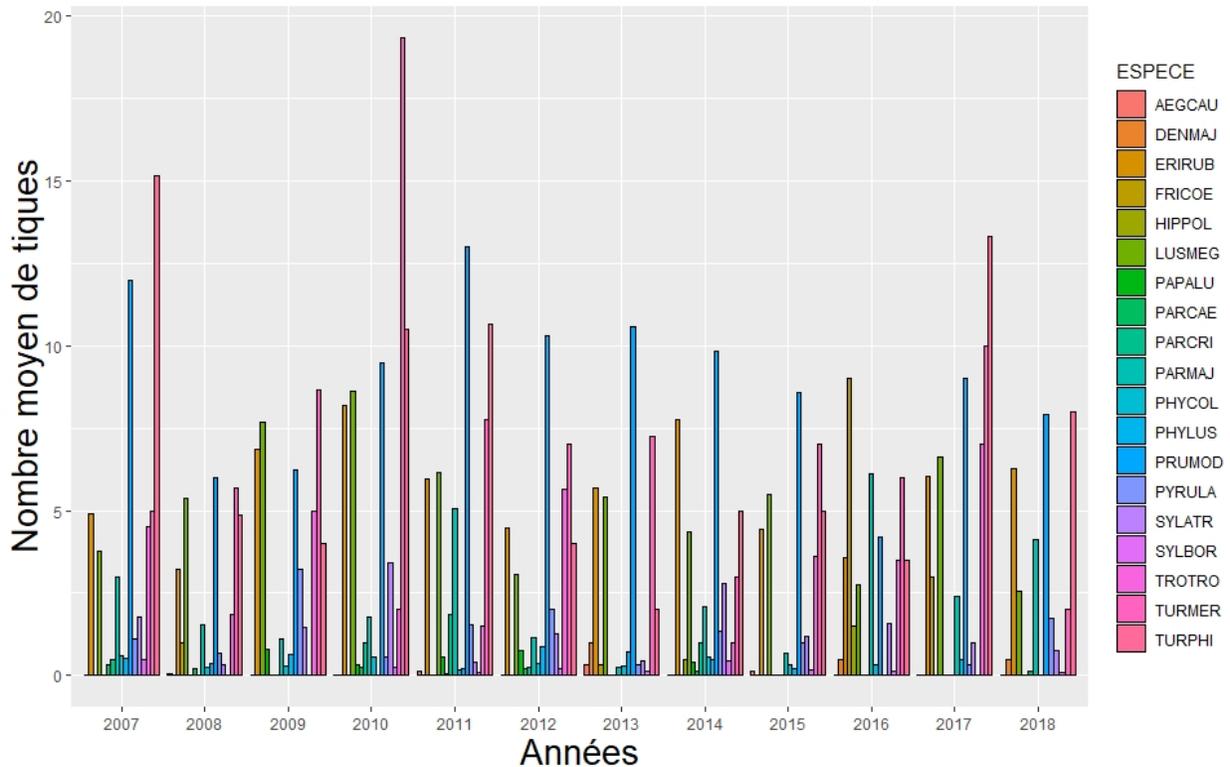


FIGURE 3 – Nombre moyen de tiques par espèce d'oiseau et par année - Station STOC 204

AEGCAU : *Aegithalos caudatus*, DENMAJ : *Dendrocopos major*, ERIRUB : *Erithacus rubecula*, FRICOE : *Fringilla coelebs*, HIPPOL : *Hippolais polyglotta*, LUSMEG : *Luscinia megarhynchos*, PAPALU : *Parus palustris*, PARCAE : *Cyanistes caeruleus*, PARCRI : *Parus cristatus*, PARMAJ : *Parus major*, PHYCOL : *Phylloscopus collybita*, PHYLUS : *Phylloscopus trochilus*, PRUMOD : *Prunella modularis*, PYRULA : *Pyrrhula pyrrhula*, SYLATR : *Sylvia atricapilla*, SYLBOR : *Sylvia borin*, TROTRO : *Troglodytes troglodytes*, TURMER : *Turdus merula*, TURPHI : *Turdus philomelos*

2.1.2 Second volet : étude du réseau oiseaux-agents pathogènes

Les données utilisées dans la seconde partie de cette étude ont également été collectées dans le cadre du programme STOC-Capture. En effet, il a été demandé aux bagueurs participant à ce programme de prélever des tiques sur les oiseaux capturés lors des sessions de la

saison de reproduction de 2018. Chaque bagueur volontaire devait envoyer un maximum de trente prélèvements par station de capture dont il a la charge, avec au plus cinq prélèvements par espèce d'oiseau différente et sans dépasser trente tiques par prélèvement. Ces tiques devaient être prélevées sur l'intégralité du corps de l'oiseau. Les tiques étaient conservées dans l'alcool à 70° et acheminées par voie postale jusqu'à l'ANSES (site de Maisons-Alfort) pour être analysées. Chaque prélèvement était identifié par la date de capture et le numéro de bague de l'oiseau, permettant par la suite de récupérer les caractéristiques de celui-ci dans la base de données du CRBPO.

- Constitution des échantillons et extraction d'ADN

Pour chaque prélèvement provenant d'un oiseau, les tiques étaient triées et regroupées par stase (larve/nympe/adulte). Pour un prélèvement d'oiseau il pouvait donc y avoir jusqu'à trois échantillons différents.

Pour chaque échantillon, les tiques étaient broyées par des billes de 2,8 mm de diamètre, à l'aide d'un ribolyseur, dans un tampon de lyse additionné de Protéinase K. Puis les étapes d'extraction d'ADN étaient réalisées en suivant les recommandations du kit *Nucleospin Tissue* (Macherey-Nagel).

- Pré-amplification de l'ADN

Cette étape permet d'augmenter le signal de l'ADN des agents pathogènes éventuels par rapport à celui de la tique. Elle était réalisée en suivant les recommandations du fabricant du kit *Perfecta Preamp supermix* (Quanta). La PCR était exécutée en ajoutant 2,5 μL de TaqMan PreAmp Master Mix (2X) et 1,2 μL d'un mélange d'amorces ciblant les agents pathogènes à 1,3 μL d'échantillon. Puis ce mélange était passé au thermocycleur pour un cycle à 95°C pendant 10 minutes et 14 cycles à 95°C pendant 15 secondes et 4 minutes à 60°C. Puis les ADN pré-amplifiés étaient dilués au 1/10.

- PCR micro-fluidique

L'analyse a été réalisée en utilisant la méthode développée par Michelet et al. (2014), permettant la détection de 35 agents pathogènes (*Anaplasma bovis*, *A. centrale*, *A. marginale*, *A. ovis*, *A. phagocytophilum*, *A. platys*, *Bartonella henselae*, *Babesia bovis*, *B. caballi*, *B. canis*, *B. divergens*, *B. microti*, *B. ovis*, *B. venatorum* sp.EU1, *B. vulpes*, *Borrelia afzelii*, *B. bavariensis*, *B. bissettii*, *B. burgdorferi*, *B. garinii*, *B. lusitaniae*, *B. mayonii*, *B. miyamotoi*, *B. spielmannii*, *B. valaisiana*, *Coxiella burnetii*, *Ehrlichia canis*, *Francisella tularensis*, *Hepatozoon* spp., *Neoehrlichia mikurensis*, *Rickettsia aeschlimannii*, *R. helvetica*, *R. massiliae*, *R. slovacica* et *R. felis*). Elle permet également de savoir si l'échantillon provient d'un pool de tiques de l'espèce *I. ricinus* ou d'une autre espèce de tique. Elle est fondée sur l'utilisation de sondes TaqMan (FAM, BHQ1) en suivant le protocole du kit *Perfecta qPCR Tough Mix* (Quanta). Une fois les échantillons et les sondes déposés sur une puce, la PCR est réalisée par l'appareil BioMark™ (Fluidigm, USA). Le cycle de PCR est composé d'un premier chauffage à 95°C pendant 5 minutes et de 45 cycles à 95°C pendant 10 secondes, 60°C pendant 15 secondes et 40°C pendant 10 secondes. Les données (luminescence) sont acquises par l'appareil, en temps réel, et analysées par son logiciel.

- Confirmation des agents pathogènes

Connaissant l'existence de croisements potentiels entre les sondes à ADN correspondant à plusieurs espèces de *Borrelia*, certains échantillons ont été confirmés par Nested-PCR, en amplifiant un gène autre que celui amplifié lors de la PCR micro-fluidique. Ces séquences amplifiées ont ensuite été séquencées afin de connaître l'espèce de *Borrelia* impliquée (Devos et al., 2019).

En plus des données déjà présentes dans la base de données du CRBPO, ces différentes analyses ont permis d'obtenir les informations suivantes pour chaque échantillon :

- l'espèce à laquelle appartient l'oiseau ;
- la stase des tiques ;
- l'espèce de tique impliquée (*I. ricinus* versus autre espèce). Étant donné que *I. frontalis* est principalement active en hiver, et que la collecte a été réalisée durant le printemps 2018, nous ferons l'hypothèse dans le cadre de ce travail que si les tiques impliquées ne sont pas des *I. ricinus* alors ce sont des *I. arboricola* ;
- la présence ou l'absence d'agents pathogènes cités ci-dessus.

N'ayant à l'heure actuelle pas l'information de l'espèce de *Borrelia* au sein des pools de larves pour la totalité des oiseaux, il n'a pas été possible de travailler sur un réseau oiseaux-*Borrelia*. Nous avons donc choisi de travailler sur un jeu de données d'entraînement correspondant au réseau oiseaux-agents pathogènes afin de mettre en place les méthodes statistiques qui seront utilisées dans la suite du projet OUTLYER. Dans ce réseau d'entraînement, nous avons donc conservé les informations au niveau espèce de *Borrelia* des oiseaux pour lesquels l'information était disponible, mais nous avons considéré comme négatifs les oiseaux positifs en *Borrelia* spp. sans information sur l'espèce. Nous avons également pris en compte la présence ou l'absence d'autres pathogènes (*C. burnetti*, *Hepatozoon* spp., *F. tularensis*, *Bor. miyamotoi*, *Bor. afzelii*, *Ehrlichia* spp., *Bab. venorum*, *A. phagocitophilum*, *Bartonella* spp., *Bor. burgdorferi*, *Bor. lusitaniae*, *N. mikurensis*, *Bor. valaisiana*), n'ayant *a priori* pas de lien entre eux, mais permettant de réaliser un entraînement sur les méthodes de réseau. Il faut donc noter que les résultats présentés pour cette partie ne sont pas des conclusions biologiques mais des pistes de réflexion pour la future mise en place de ce type d'analyse sur les données réelles.

2.2 Méthodes statistiques utilisées

2.2.1 Premier volet : variabilité de la charge en tiques des oiseaux

La variable à expliquer étant le nombre de tiques observées sur les oiseaux, on est donc en présence de données de comptage, dont l'histogramme est représenté en Figure 4. A la lecture de ce graphique, il paraît nécessaire de s'interroger sur une éventuelle surdispersion des données, ainsi qu'un probable excès de zéros. Une première approche de données de comptage a été réalisée via une régression de Poisson. Étant donnée la possible surdispersion des données, nous avons ensuite mis en œuvre un modèle binomial négatif, approche classique dans

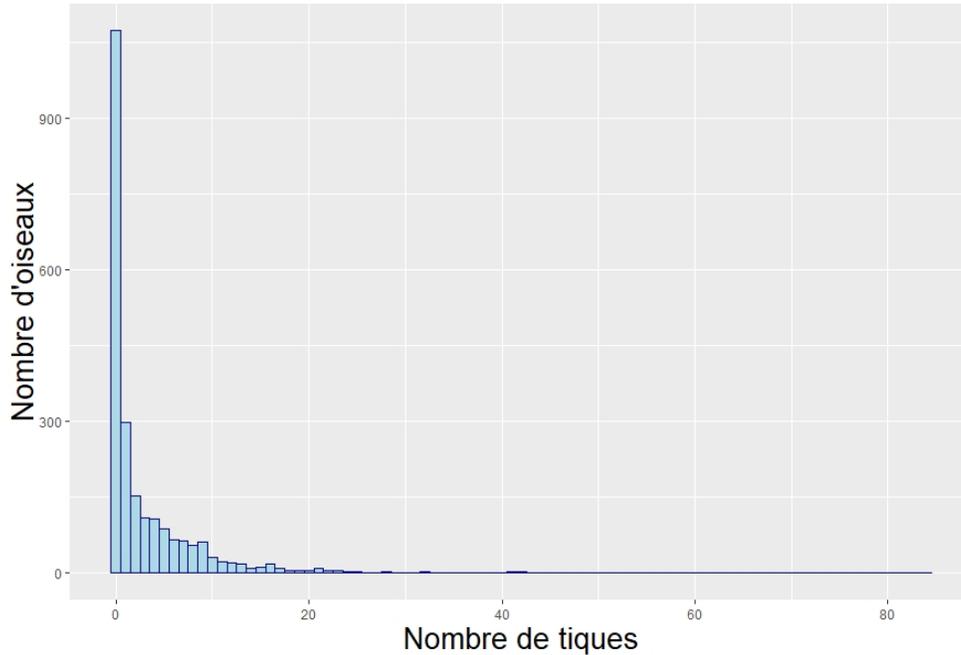


FIGURE 4 – Distribution du nombre de tiques par oiseau - Station STOC 204

l'étude de données de charge parasitaire (Lloyd-Smith, 2007), qui serait probablement mieux adapté à nos données. Enfin, du fait du potentiel excès de zéros, un modèle de régression à inflation de zéros a été implémenté. Ce type de modèle semblerait *a priori* correspondre à nos données puisqu'il permettrait de modéliser deux processus biologiques : d'une part la rencontre ou non des parasites, et d'autre part le nombre de tiques s'accrochant sur l'oiseau.

- **Modèles linéaires généralisés (GLM)**

D'après McCullagh & Nelder (1989); Jais (2018); Guilloux (2018); Chesneau (2018)

Principe et écriture du modèle

Ces modèles sont des extensions des modèles de régression classiques, spécifiés par trois composantes :

- Une composante aléatoire

Elle se réfère à la variable réponse Y . On note (y_1, \dots, y_n) les n réalisations indépendantes de Y . On suppose que chaque y_i provient d'une distribution dont la fonction de densité appartient à la famille exponentielle linéaire, de la forme :

$$f(y_i; \theta_i, \phi) = \exp \left(\frac{y_i \theta_i - v(\theta_i)}{u(\phi)} + w(y_i, \phi) \right) \quad (1)$$

θ_i est le paramètre naturel de la distribution exponentielle.

La fonction $u(\phi)$ est le plus souvent de la forme $u(\phi) = \frac{\phi}{\omega_i}$ ou ω_i est une pondération connue des observations, que l'on fixera ici à 1, et ϕ est appelé le paramètre de dispersion.

On peut alors mettre l'expression de la structure exponentielle sous la forme canonique en posant :

$$Q(\theta) = \frac{\theta}{\phi} \quad (2)$$

$$a(\theta) = \exp\left(-\frac{v(\theta)}{\phi}\right) \quad (3)$$

$$b(y) = \exp(w(y, \phi)) \quad (4)$$

On obtient alors :

$$f(y_i, \theta_i) = a(\theta_i)b(y_i)\exp(y_i Q(\theta_i)) \quad (5)$$

- Une composante systématique

Soit $x_{i1}, x_{i2}, \dots, x_{ip}$ les valeurs des p covariables observées à la i ème observation. La composante systématique relie le paramètre η_i aux covariables par une combinaison linéaire :

$$\eta_i = \sum_{k=1}^n x_{ik}\beta_k, \quad i = 1, \dots, n \quad (6)$$

Pour simplifier, on écrira sous forme matricielle :

$$\eta_i = x'_i\beta \quad (7)$$

- Une fonction de lien

Cette fonction relie l'espérance de y_i et la composante systématique. En notant $\mu_i = \mathbb{E}(Y_i)$ pour $i = 1, \dots, n$, on a :

$$g(\mu_i) = \eta_i \quad (8)$$

avec g une fonction inversible nommée fonction de lien. La fonction de lien qui associe l'espérance μ_i au paramètre naturel est appelée fonction de lien canonique, et on a alors :

$$g(\mu_i) = \theta_i = x'_i\beta \quad (9)$$

• Régression de Poisson

Pour la régression de Poisson, si l'on considère n variables indépendantes Y_i de loi de Poisson de paramètre $\lambda_i = \mathbb{E}(Y_i)$, ces variables admettent pour densité :

$$f(y_i, \lambda_i) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} = \exp(-\lambda_i) \frac{1}{y_i!} \exp(y_i \ln(\lambda_i)) \quad (10)$$

On reconnaît une structure exponentielle mise sous la forme canonique, de paramètre naturel $\theta_i = \ln(\lambda_i)$ et de fonction de lien canonique la fonction logarithme.

On peut alors écrire le modèle sous la forme :

$$\ln(\lambda_i) = x_i' \beta \quad (11)$$

Estimation des paramètres

L'estimation des paramètres du modèle est une estimation par maximum de vraisemblance. Pour un GLM, la log-vraisemblance s'écrit :

$$\mathcal{L}(\beta) = \sum_{i=1}^n \ln f(y_i; \theta_i, \phi) = \sum_{i=1}^n l(y_i; \theta_i, \phi) \quad (12)$$

Dans le cas de la régression de Poisson on a :

$$\mathcal{L}(\beta) = \sum_{i=1}^n (y_i x_i' \beta - \exp(x_i' \beta) - \ln(y_i!)) \quad (13)$$

En dérivant la log-vraisemblance par rapport à β , on obtient :

$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta} = \sum_{i=1}^n (y_i - \exp(x_i' \beta)) X_i \quad (14)$$

En égalisant cette dérivée à zéro, ces équations sont des fonctions non linéaires de β , et le calcul l'estimation de $\hat{\beta}$ fait donc appel à des processus itératifs de type Newton-Raphson. L'algorithme par défaut utilisé dans R par la fonction `glm()` du package `stats` (R Core Team, 2018) est l'algorithme IRLS (Iteratively Reweighted Least Squares).

Ajustement

Le diagnostic de nos modèles a été réalisé par la construction de rootogrammes. Ce sont des graphiques adaptés pour les données de comptage, permettant de comparer visuellement les valeurs prédites par le modèle aux données observées. L'écart entre ces valeurs s'étudie au niveau de l'axe des abscisses. Sur R, nous utiliserons la commande `rootogram()` du package `countreg` (Kleiber & Zeileis, 2016).

Pour étudier l'ajustement des GLM, on utilise principalement la mesure de la déviance (ou déviance résiduelle), appelé en anglais "Deviance goodness of fit test". On considère un modèle saturé, c'est à dire comportant autant de covariables que d'observations, et on définit la déviance comme :

$$D = -2 \ln \left(\frac{\mathcal{L}_n(\hat{\beta})}{\mathcal{L}_{sat}} \right) \quad (15)$$

Avec \mathcal{L}_{sat} la log-vraisemblance du modèle saturé et $\mathcal{L}_n(\hat{\beta})$ celle du modèle estimé. D suit une loi du χ^2 à $n - p$ degrés de liberté, ce qui permet de construire un test d'acceptation du modèle.

On définit également le rapport de vraisemblance (ou différence des déviances) pour deux modèles emboîtés (ou LRT pour Likelihood ratio test en anglais), permettant d'évaluer l'apport de paramètres supplémentaires dans l'ajustement du modèle. Pour deux modèles emboîtés à p_1 et p_2 paramètres, on a :

$$D_2 - D_1 = 2(\mathcal{L}_1 - \mathcal{L}_{sat}) - 2(\mathcal{L}_2 - \mathcal{L}_{sat}) = 2(\mathcal{L}_1 - \mathcal{L}_2) \quad (16)$$

Cette différence suit une loi du χ^2 à $(p_2 - p_1)$ degrés de liberté pour les lois à un paramètre, et une loi de Fisher pour les lois à deux paramètres.

Sur **R**, le calcul du LRT sera réalisé par la commande `lrtest()` du package `lmtest` (Zeileis & Hothorn, 2002).

Un mauvais ajustement d'une régression de Poisson peut traduire une surdispersion des données, c'est à dire que la variance est plus grande que la moyenne, alors qu'elles devraient être égales. Il existe des tests conçus spécifiquement pour tester l'équi-dispersion dans une régression de Poisson, comme celui construit par Cameron & Trivedi (1990), que nous n'utiliserons pas ici. En revanche, nous allons ajuster un modèle prenant en compte la surdispersion, ici le modèle binomial négatif, et comparer celui-ci à la régression de Poisson à l'aide d'un LRT.

• Modèle binomial négatif

D'après Ver Hoef & Boveng (2007); Hilbe (2011)

Principe et écriture du modèle

La loi binomiale négative est une loi comptant le nombre de tirages nécessaires pour obtenir r succès, dans une succession d'expériences de Bernoulli de probabilité p . Soit X le nombre de tirages, on a alors la fonction de masse :

$$\begin{aligned} \mathbb{P}(X = x) &= \binom{x+r-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots \\ \text{ou } \mathbb{P}(X = x) &= \binom{x+r-1}{r-1} p^r (1-p)^x, \quad x = 0, 1, 2, \dots \end{aligned} \quad (17)$$

On peut généraliser cette loi à tout réel r positif, en exprimant le coefficient binomial à l'aide d'une fonction gamma d'Euler :

$$\binom{y}{k} = \frac{\Gamma(y+1)}{\Gamma(k+1)\Gamma(y-k+1)} \quad (18)$$

L'espérance et la variance de cette loi sont données par :

$$\begin{aligned} \mathbb{E}(X) &= r \frac{1-p}{p} = \mu \\ \mathbb{V}(X) &= r \frac{1-p}{p^2} = \mu + \frac{1}{r} \mu^2 \end{aligned} \quad (19)$$

La loi binomiale peut également être considérée comme une généralisation de la loi de Poisson, c'est à dire comme une mixture Poisson-Gamma. On considère alors cette loi comme une loi de Poisson de paramètre $\lambda\theta$ dont θ suit une loi Gamma de paramètres $\nu > 0$ (paramètre d'échelle) et $\delta > 0$ (paramètre de forme) et de moyenne 1 (avec donc $\nu = \frac{1}{\delta}$). On peut écrire la distribution conditionnelle suivante :

$$\mathbb{P}(Y = y_i | \Theta = \theta_i) = \frac{(\lambda_i \theta_i)^{y_i} \exp(-\lambda_i \theta_i)}{y_i!} \quad (20)$$

Avec θ_i admettant la densité de probabilité :

$$f(\theta_i; \nu) = \frac{\nu^\nu}{\Gamma(\nu)} \theta_i^{\nu-1} \exp(-\nu \theta_i), \quad \theta_i > 0 \quad (21)$$

On a donc une densité jointe définie par :

$$\mathbb{P}(Y = y_i | \Theta = \theta_i) f_\Theta(\theta_i) = \frac{(\lambda_i \theta_i)^{y_i}}{y_i!} \exp(-\lambda_i \theta_i) \frac{\nu^\nu}{\Gamma(\nu)} \theta_i^{\nu-1} \exp(-\nu \theta_i), \quad \theta_i > 0 \quad (22)$$

On intègre cette densité par rapport à θ_i pour obtenir la loi de Y :

$$\begin{aligned} \mathbb{P}(Y = y_i; \nu) &= \int_0^{+\infty} \mathbb{P}(Y = y | \Theta = \theta_i) f_\Theta(\theta_i) d\theta_i \\ &= \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1)\Gamma(\nu)} \left(\frac{\nu}{\lambda_i + \nu} \right)^\nu \left(\frac{\lambda_i}{\lambda_i + \nu} \right)^{y_i} \\ &= \binom{y_i + \nu - 1}{y_i} \left(\frac{\nu}{\lambda_i + \nu} \right)^\nu \left(\frac{\lambda_i}{\lambda_i + \nu} \right)^{y_i} \\ &= \binom{y_i + \frac{1}{\alpha} - 1}{y_i} \left(\frac{1}{1 + \lambda_i \alpha} \right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \lambda_i \alpha} \right)^{y_i}, \quad \text{avec } \alpha = \frac{1}{\nu} \end{aligned} \quad (23)$$

On reconnaît la forme de la fonction de masse de la loi binomiale négative explicitée dans l'équation (17), avec $r = \alpha$ et $p = \frac{1}{1 + \lambda_i \alpha}$. On peut réécrire cette probabilité sous la forme :

$$\mathbb{P}(Y = y_i; \alpha) = \exp \left(y \ln \left(\frac{\lambda_i \alpha}{1 + \lambda_i \alpha} \right) - \frac{1}{\alpha} \ln(1 + \lambda_i \alpha) + \ln \left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \right) \right) \quad (24)$$

On reconnaît une loi de famille exponentielle, dont le lien canonique est :

$$g(\lambda_i) = \ln \left(\frac{\lambda_i/\alpha}{1 + \lambda_i/\alpha} \right) \quad (25)$$

On peut alors écrire le modèle sous la forme :

$$\ln \left(\frac{\lambda_i/\alpha}{1 + \lambda_i/\alpha} \right) = x'_i \beta \quad (26)$$

On notera enfin qu'on peut montrer qu'une loi binomiale négative de paramètres (r, p) converge vers une loi de Poisson lorsque r tend vers l'infini.

Pour une loi binomiale de paramètre r et p , l'espérance et la variance sont exprimées par :

$$\begin{aligned} \mathbb{E}(Y) &= r \frac{1-p}{p} = \lambda_i \\ \mathbb{V}(Y) &= r \frac{1-p}{p^2} = \lambda_i + \frac{1}{r} \lambda_i^2 \end{aligned} \quad (27)$$

On observe que la variance est fonction quadratique de l'espérance, et est supérieure à celle-ci. Donc la loi binomiale permet bien de prendre en compte une surdispersion d'une loi de Poisson.

Estimation des paramètres

Les paramètres du modèle binomial négatif sont également estimés par maximum de vraisemblance. Ici, la log-vraisemblance s'écrit :

$$\mathcal{L}(\lambda; y, \alpha) = \sum_{i=1}^n \left(y_i \ln \left(\frac{\alpha \lambda_i}{1 + \alpha \lambda_i} \right) - \frac{1}{\alpha} \ln(1 + \alpha \lambda_i) + \ln \left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1) \Gamma(1/\alpha)} \right) \right) \quad (28)$$

En paramétrisant en fonction de β , on peut montrer que l'on a :

$$\mathcal{L}(\beta; y, \alpha) = \sum_{i=1}^n \left(y_i \ln \left(\frac{\alpha \exp(x'_i \beta)}{1 + \alpha \exp(x'_i \beta)} \right) - \frac{1}{\alpha} \ln(1 + \alpha \exp(x'_i \beta)) + \ln \left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1) \Gamma(1/\alpha)} \right) \right) \quad (29)$$

Sur \mathbf{R} , nous utiliserons la fonction `glm.nb()` du package `MASS` (Venables & Ripley, 2002). Le calcul de $\hat{\beta}$ est également réalisé par l'algorithme IRLS.

Ajustement

Le modèle de Poisson étant emboîté dans le modèle binomial négatif, il sera possible de comparer les deux modèles par un LRT, afin de mettre en évidence le modèle le mieux ajusté à nos données.

- **Modèle Zero-Inflated Negative Binomial (ZINB)**

D'après Hilbe (2011); UCLA (2014); Dupuy (2017)

Principe et écriture du modèle

Les modèles de régression à inflation de zéros (ou "zero-inflated models") permettent de modéliser des données de comptage en prenant en compte une surdispersion liée à un excès de zéros. Ces modèles postulent de l'existence de deux types de zéros : les zéros aléatoires et les zéros structurels. Les premiers sont liés à l'échantillonnage, au modèle de comptage. Ce sont des individus à risque pour l'effet étudié, et dont la variable de comptage peut prendre les valeurs 0, 1, 2, Les zéros structurels, au contraire proviennent d'individus qui ne sont pas à risque pour l'effet étudié, et dont la variable de comptage ne peut prendre que la valeur 0. Le modèle ZINB contient donc deux composantes : un modèle binaire, le plus souvent logistique, modélisant l'appartenance aux groupe non exposé avec une probabilité π_i et un modèle de comptage, ici le modèle binomial négatif. Soit Z_i les covariables associées à la partie logistique du modèle et X_i les covariables associées au modèle de comptage. On a alors :

$$P(Y_i = y_i | X_i, Z_i) = \begin{cases} \pi_i + (1 - \pi_i) \left(\frac{1}{1 + \alpha \lambda_i} \right)^\alpha & \text{si } y_i = 0 \\ (1 - \pi_i) \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(1/\alpha) y_i!} \left(\frac{\alpha \lambda_i}{1 + \alpha \lambda_i} \right)^{y_i} \left(\frac{1}{1 + \alpha \lambda_i} \right)^{1/\alpha} & \text{si } y_i > 0 \end{cases} \quad (30)$$

On a également :

$$\begin{aligned} \mathbb{E}(Y_i | X_i, Z_i) &= (1 - \pi_i) \lambda_i \\ \mathbb{V}(Y_i | X_i, Z_i) &= (1 - \pi_i) \lambda_i (1 + (\alpha + \pi_i) \lambda_i) \end{aligned} \quad (31)$$

La probabilité π_i et le paramètre λ_i sont modélisés par :

$$\begin{aligned} \text{logit}(\pi_i) &= z_i' \gamma \\ \ln(\lambda_i) &= x_i' \beta \end{aligned} \quad (32)$$

Estimation des paramètres

A nouveau, les paramètres du modèle sont estimés par maximum de vraisemblance. La log-vraisemblance du modèle s'écrit :

si $y = 0$:

$$\mathcal{L}(\beta, \gamma; y, \alpha) = \sum_{i=1}^n \left(\ln \left(\frac{1}{1 + \exp(-z_i' \gamma)} \right) + \frac{1}{1 + \exp(-z_i' \gamma)} \left(\frac{1}{1 + \alpha \exp(-x_i' \beta)} \right)^{1/\alpha} \right)$$

si $y \neq 0$:

$$\mathcal{L}(\beta, \gamma; y, \alpha) = \sum_{i=1}^n \left(y_i \ln \left(\frac{\alpha \exp(x'_i \beta)}{1 + \alpha \exp(x'_i \beta)} \right) + \ln \left(\frac{1}{1 + \exp(-z'_i \gamma)} \right) + \ln \left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \right) + \frac{1}{\alpha} \ln \left(\frac{1}{1 + \alpha \exp(x'_i \beta)} \right) \right) \quad (33)$$

Ajustement

Dans **R**, nous utiliserons la fonction `zeroinfl()` du package `countreg` (Zeileis et al., 2008). Les coefficients du modèle sont calculés par l'algorithme BFGS. Le modèle binomial négatif étant emboîté dans le modèle ZINB, il est possible de comparer l'ajustement de ces deux modèles par un LRT.

Afin d'analyser nos données de comptage de tiques, les variables explicatives espèce d'oiseaux, sexe/âge, année d'observation, hauteur du nid et poids moyen de l'espèce ont été introduites dans les modèles. Les variables de traits d'histoire de vie (hauteur d'alimentation, poids moyen de l'espèce) étant colinéaires à la variable espèce d'oiseaux, nous avons du réaliser deux modèles à chaque fois :

- un modèle du nombre de tiques en fonction de l'espèce d'oiseau, du sexe/âge et de l'année d'observation (que nous appellerons ci-après modèle "espèces") ;
- un modèle du nombre de tiques en fonction des traits d'histoire de vie aviaire (la hauteur d'alimentation et le poids moyen de l'espèce), du sexe/âge et de l'année d'observation (appelé ci-après modèle "THV").

2.2.2 Second volet : étude du réseau oiseaux-agents pathogènes

D'après Wasserman & Faust (1994); Beauquitte (2013); Kolaczyk & Csárdi (2014); Sueur (2015); Soulie Fogelman & Viennet (2016); Tackx (2018)

Pour ce second volet, nous avons étudié des données de présence ou d'absence d'agents pathogènes pour chaque échantillon de tiques prélevées par individu d'oiseau. Afin d'étudier la structure des communautés aviaires et des agents pathogènes qui leur sont inféodées, et pour mettre en évidence des éventuelles relations particulières entre les espèces, nous avons décidé de mettre en place l'analyse d'un réseau 2-modes pondéré. Pour cela, nous avons calculé la prévalence des agents pathogènes par espèce aviaire et utilisé ces données comme pondération.

• Principe des réseaux sociaux

Définitions

L'analyse de données de réseaux sociaux consiste à mesurer les relations au sein d'un ou de plusieurs groupes d'acteurs. Elle permet d'étudier la structure sociale d'un groupe,

à différentes échelles (groupe, dyade, individu). La table de contingence, ou "sociomatrice" est la base de la représentation et de l'analyse d'un réseau social. Cette matrice contient des lignes et des colonnes définissant les individus. Les données contenues dans cette matrice décrivent les relations entre ces acteurs.

Les réseaux sont communément représentés sous forme de graphes, dans lesquels les nœuds (aussi appelés "sommets") représentent les acteurs, et les arcs (appelées "liens" , "connexions" ou encore "arrêtes") représentent les liens entre les différents acteurs. Ces graphes peuvent être dirigés, dans le cas d'un lien orienté depuis un émetteur vers un récepteur, et la matrice de contingence sera alors asymétrique. Dans d'autres situations, les graphes sont non dirigés et représentent uniquement les connexions entre individus, de manière symétrique, ce qui est le cas dans le cadre de notre étude.

Un réseau peut être non pondéré, s'il décrit uniquement la présence ou l'absence de lien entre les acteurs, mais il peut également être pondéré, si l'on attribue un poids aux liens entre ces acteurs. Sur le graphe, on représentera alors cette pondération par l'ajout de la valeur chiffrée du poids au niveau du lien entre deux nœuds, ou par d'autres représentations graphiques (épaisseur du trait, couleur, etc.).

Enfin, les acteurs d'un réseau peuvent appartenir à un seul ou à plusieurs groupes. Lorsque ceux-ci peuvent être séparés en deux groupes distincts, comme c'est le cas dans le cadre de notre étude, on parle de réseau 2-modes, (ou "two-mode network" en anglais). Ces types de réseaux sont représentés par des graphes dits "bipartis" ("bipartite graphs" en anglais).

Notations mathématiques

Un graphe \mathcal{G} possède donc deux ensembles d'informations : un ensemble \mathcal{N} de g nœuds, tel que $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$, et un ensemble \mathcal{L} de m arcs tel que $\mathcal{L} = \{l_1, l_2, \dots, l_m\}$. On peut alors noter le graphe $\mathcal{G}(\mathcal{N}, \mathcal{L})$. Chaque arc est un couple non ordonné de nœuds : $l_k = (n_i, n_j)$. Le graphe étant non orienté, la ligne entre les nœuds n_i et n_j est la même que celle entre les nœuds n_j et n_i , et on a donc : $l_k = (n_i, n_j) = (n_j, n_i)$.

Dans le cas d'un graphe \mathcal{G} simple, non pondéré, la matrice d'adjacence $A_{\mathcal{G}} = [a_{ij}]$ est une matrice carrée $g \times g$ définie par :

$$a_{ij} = \begin{cases} 1 & \text{si } (n_i, n_j) \in \mathcal{L} \\ 0 & \text{sinon} \end{cases} \quad (34)$$

Pour un graphe pondéré, la matrice d'adjacence $A_{\mathcal{G}}$ est définie par :

$$a_{ij} = \begin{cases} w_{ij} & \text{si } (n_i, n_j) \in \mathcal{L} \\ 0 & \text{sinon} \end{cases} \quad (35)$$

avec w_{ij} le poids du lien (n_i, n_j) .

Dans le cadre des réseaux 2-modes, avec deux groupes distincts de nœuds \mathcal{N}_1 et \mathcal{N}_2 , de tailles respectives g et h , la matrice d'adjacence peut s'écrire sous la forme d'une matrice d'affiliation $g \times h$, telle que $A = [a_{ij}]$, définie comme la matrice $A_{\mathcal{G}}$ précédente. On peut également décrire ce réseau par une sociomatrice A' de taille $(g + h) \times (g + h)$, tel que :

$$A' = \begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix} \quad (36)$$

avec A la matrice d'affiliation et A^t sa transposée. En haut à gauche, la sous-matrice de taille $g \times g$ est nulle, indiquant qu'il n'y a pas de liens entre les acteurs du groupe \mathcal{N}_1 . De même, la sous matrice de taille $h \times h$, du coin en bas à droite est nulle, indiquant une absence de connexions au sein des acteurs du groupe \mathcal{N}_2 . Dans \mathbf{R} , nous avons tracé le graphe biparti à partir de la matrice d'affiliation par la fonction `plotweb()` du package `bipartite` (Dormann et al., 2008).

On dit que deux nœuds n_i et n_j sont adjacents si le lien $l_k = (n_i, n_j)$ appartient à \mathcal{L} . Un nœud est incident avec un lien (ou inversement) s'il appartient au couple définissant ce lien. Par exemple, n_1 et n_2 sont incident avec le lien $l_1 = (n_1, n_2)$.

Dans un graphe \mathcal{G} , le voisinage d'un sommet n_i , noté $N(n_i)$, est l'ensemble des voisins de n_i , c'est-à-dire reliés à n_i par un arc. On définit mathématiquement ce voisinage par :

$$N(n_i) = \{n_j \in \mathcal{N} | (n_i, n_j) \in \mathcal{L}\} \quad (37)$$

Pour décrire la structure de ces réseaux, il est possible d'utiliser de nombreuses métriques, aussi bien à l'échelle du nœud ou de la connexion que du réseau dans son ensemble. Nous décrirons ci-dessous les indices utilisés dans le cadre de notre étude.

• Métriques utilisées dans le cadre de l'analyse de réseaux

Degrés

Le degré d'un nœud, noté $d(n_i)$ est égal au nombre de liens incidents avec lui, ou encore au nombre de nœuds adjacents avec celui-ci. Dans le cadre de réseaux en écologie, le degré permet de mesurer le généralisme d'une espèce dans un réseau social donné. On le définit mathématiquement par :

$$d(n_i) = |N(n_i)|, \quad (38)$$

avec $N(n_i)$ le voisinage de n_i .

Dans le cas d'un réseau pondéré, on applique la formule suivante :

$$d(n_i) = \sum_{n_j \in N(n_i)} w_{ij} \quad (39)$$

Enfin, dans un réseau 2-modes de sous-groupes \mathcal{N}_1 , et \mathcal{N}_2 , on définit le degré moyen de chacun des groupes par :

$$\bar{d}(\mathcal{N}_k) = \frac{1}{g_k} \sum_{n_{ik} \in \mathcal{N}_k} d(n_{ik}) = \frac{m_k}{g_k}, \quad k = 1, 2 \quad (40)$$

avec g_k le nombre de nœuds du groupe k , m_k le nombre de liens incidents avec les nœuds du groupe k et n_{i_k} le nœud i du groupe k .

Dans **R**, les degrés non-pondérés ont été calculés à l'aide de la fonction `specieslevel()` du package `bipartite`. Les autres degrés ont été calculés manuellement.

Détection de communautés et modularité

Un réseau peut être divisé en sous-groupes, appelés "communautés" ou "modules", au sein desquels les nœuds sont très connectés, mais ayant peu de liens avec les autres sous-groupes. Dans le cadre de réseau appliqués à l'écologie, cela signifie que les espèces au sein des modules interagissent plus entre elles qu'elles n'interagissent avec les espèces externes. Il peut alors être intéressant de partitionner les réseaux pour étudier ces communautés.

Afin d'évaluer une partition donnée d'un réseau, Newman & Girvan (2004) introduisent un critère de qualité appelé "modularité". Elle permet de mesurer l'écart entre le nombre de liens dans un module à ce que serait ce nombre dans un réseau aléatoire, dans lequel les liens seraient distribués aléatoirement, et est définie pour un réseau uniparti par :

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - P_{ij}) \delta(c_i, c_j) \quad (41)$$

avec A_{ij} le coefficient de la i -ème ligne et j -ème colonne de la sociomatrice, m le nombre total de liens dans le réseau, c_i et c_j les communautés auxquelles appartiennent les sommets i et j , $\delta(c_i, c_j)$, le delta de Kronecker (valant 1 si $c_i = c_j$ et 0 sinon) et P_{ij} la matrice du réseau aléatoire décrivant la probabilité d'existence de l'arête i, j dans ce réseau.

Cette définition a été adaptée pour les réseaux bipartis par Barber (2007). Soit la sociomatrice A , d'un réseau 2-modes de sous-groupes \mathcal{N}_1 , et \mathcal{N}_2 , respectivement de taille g et h notée :

$$A = \begin{pmatrix} 0_{g \times g} & \tilde{A}_{g \times h} \\ \tilde{A}_{h \times g}^t & 0_{h \times h} \end{pmatrix} \quad (42)$$

La modularité s'écrit alors :

$$Q_B = \frac{1}{m} \sum_{i=1}^g \sum_{j=1}^h (\tilde{A}_{ij} - P_{ij}) \delta(c_i, c_j) = \frac{1}{m} \sum_{i=1}^g \sum_{j=1}^h \left(\tilde{A}_{ij} - \frac{k_i d_j}{m} \right) \delta(c_i, c_j) \quad (43)$$

avec \tilde{A} la matrice d'affiliation du réseau. Ici, on remplace P_{ij} par sa valeur calculée à l'aide de m , k_i le degré du nœud i et d_j le degré du nœud j .

Enfin, Dormann & Strauss (2014) ont adapté cette définition pour les réseaux bipartis pondérés :

$$Q_W = \frac{1}{M} \sum_{i=1}^g \sum_{j=1}^h (\tilde{W}_{ij} - \tilde{E}_{ij}) \delta(c_i, c_j) = \frac{1}{M} \sum_{i=1}^g \sum_{j=1}^h \left(\tilde{W}_{ij} - \frac{y_i k_j}{M} \right) \delta(c_i, c_j) \quad (44)$$

avec \tilde{W}_{ij} la matrice d'affiliation pondérée du réseau, M la somme des liens pondérés, \tilde{E} la matrice des liens attendues sous le modèle aléatoire, y_i la somme marginale de la ligne i et k_j la somme marginale de la colonne j . On peut réécrire cette formule sous forme matricielle, pour une partition de F communautés :

$$Q_W = \frac{1}{M} \text{tr}(M_1(\tilde{W} - \tilde{E})M_2) \quad (45)$$

avec M_1 la matrice $F \times g$ des étiquettes (ou "label matrix") du sous-groupe \mathcal{N}_1 , et M_2 celle du sous-groupe \mathcal{N}_2 . La matrice M_1 comporte une colonne par nœud et une ligne par étiquette, avec pour chaque colonne une unique valeur 1 sur la ligne correspondant à l'étiquette de ce nœud. La matrice M_2 est construite de manière similaire, avec les nœuds en ligne et les étiquettes en colonne.

Certaines méthodes classiques de détection de communautés au sein de réseaux utilisent la maximisation de la modularité. C'est le cas de celui utilisé durant notre étude, l'algorithme DIRTLPawb+ conçu par Beckett (2016) et implémenté dans le package `bipartite` sur R dans la fonction `computeModules()` (Dormann et al., 2008).

Cet algorithme est fondé sur l'algorithme LPawb+ (pour "Label Propagation Algorithm weighted bipartite"). L'idée du LPawb+ est d'assigner à chaque nœud de l'un des sous-groupes une étiquette unique (initialisation), puis d'attribuer de nouvelles étiquettes au second groupe en maximisant la modularité, de procéder de même sur le premier groupe, puis recommencer sur le second groupe, et ce jusqu'à atteindre un maximum de modularité. Ensuite, on essaie de fusionner les communautés trouvées de la sorte, et on garde les groupes ainsi si cela augmente la modularité. Le LPawb+ étant sensible aux valeurs d'initialisation, l'algorithme DIRTLPawb+ l'améliore en répétant un grand nombre de fois le LPawb+, en modifiant les valeurs de l'initialisation et en conservant la partition maximisant la modularité.

Graphique z_c et rôle des nœuds

Afin de déterminer la relation qu'entretient un espèce (un nœud) n_i avec les autres, au sein de son module s et avec les autres modules t . On peut calculer le degré intra-module standardisé z et la connectivité entre les modules c :

$$z = \frac{d(n_i)_s - \bar{d}_s}{SD_{ds}}$$

$$c = 1 - \sum_{t=1}^F \left(\frac{d(n_i)_t}{d(n_i)} \right)^2 \quad (46)$$

avec $d(n_i)_s$ le degré intra-module de n_i , \bar{d}_s le degré moyen de toutes les espèces du module s et SD_{ds} la variance, F le nombre total de modules t dans le réseau, $d(n_i)_t$ le nombre de lien du nœud n_i avec les espèces du module t et $d(n_i)$ le degré de n_i .

En positionnant les espèces sur un graphique z_c et en utilisant les valeurs seuils définies par Guimerà & Amaral (2005); Olesen et al. (2007), on peut définir 4 rôles pour les nœuds :

- $z \leq 2, 5; c \leq 0, 62$: les espèces périphériques, ayant peu de liens à l'intérieur de leurs modules, et presque jamais en dehors ;
- $z \leq 2, 5; c > 0, 62$: les espèces connectrices, importantes pour la connectivité entre les modules,
- $z > 2, 5; c \leq 0, 62$: les espèces "module hub", permettant la cohérence au sein de leur propre module
- $z > 2, 5; c > 0, 62$: les espèces "network hub", à la fois connectrices entre les modules et participant à la cohérence de leurs modules.

3 Résultats

3.1 Premier volet : Variabilité de la charge en tiques des oiseaux

Dans la base de données de la station STOC 204 pour laquelle nous avons des comptages de tiques sur les oiseaux entre 2007 et 2018, les oiseaux portent en moyenne 3 tiques (pour une médiane à 1 tique), les valeurs s'étalant de 0 à 84 tiques.

Afin de décrire nos données de comptage, nous avons commencé par ajuster un modèle de Poisson (modèle "espèces").

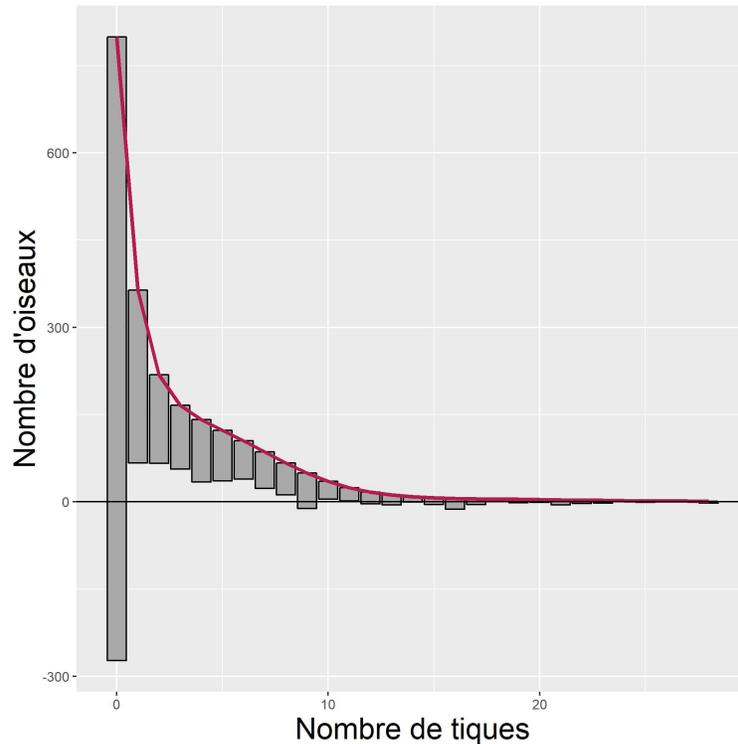


FIGURE 5 – Rootogramme du modèle de Poisson (modèle "espèces")

Le rootogramme (Figure 5) représente les données prédites par la courbe rouge et les données observées par l'histogramme. On remarque un écart important entre les prédictions et les observations, notamment en zéro. En effet, les faibles valeurs de comptage sont sur-estimées par le modèle, au contraire des zéros qui sont fortement sous-estimés. Le modèle de Poisson semble donc peu ajusté pour nos données. Cette observation graphique est confortée par le calcul du deviance goodness of fit test qui, étant très significatif ($p \ll 0,05$), confirme l'inadéquation du modèle.

Pour le modèle de Poisson "THV", les résultats observés sont similaires.

On émet l'hypothèse que ce mauvais ajustement du modèle est dû à une surdispersion de nos données par rapport à une loi de Poisson, et nous avons donc implémenté un modèle binomial négatif pour tester cette hypothèse.

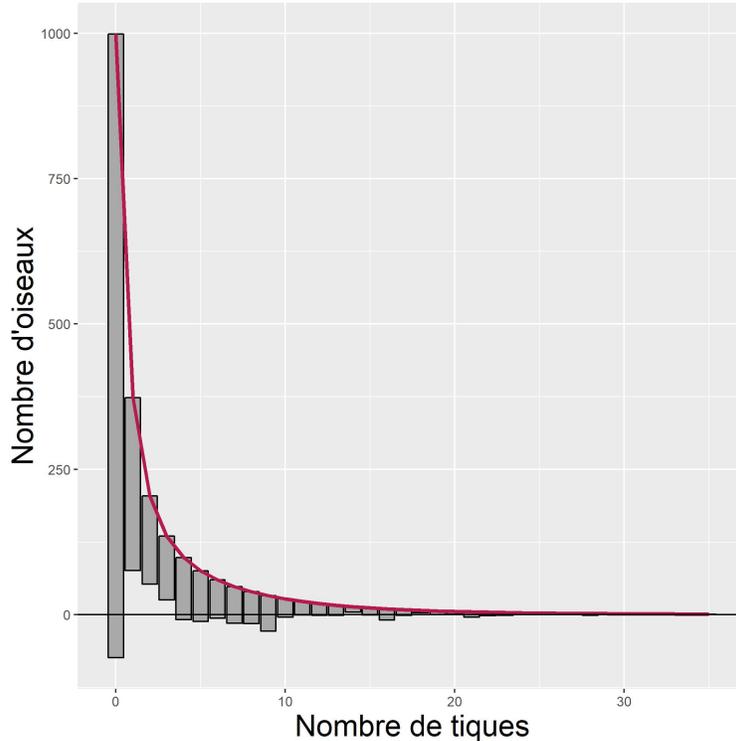


FIGURE 6 – Rootogramme du modèle binomial négatif (modèle "espèces")

On observe sur le rootogramme du modèle binomial négatif (modèle "espèces", Figure 6), qu'il y a moins d'écart entre les valeurs prédites et les valeurs observées, et notamment que les zéros sont moins sous estimés par cette approche. En revanche, ce modèle semble encore inadapté pour nos données.

Le calcul du LRT entre ce modèle et le modèle de Poisson est significatif ($p < 2,2 \times 10^{-16}$), et on en conclut donc que le modèle binomial négatif est significativement mieux ajusté que celui de Poisson.

Le deviance goodness of fit test n'est pas significatif ($p = 0,81$), ce modèle est donc convenablement adapté à nos données. En revanche, après lecture du rootogramme, il semble judicieux d'implémenter un modèle ZINB afin de prendre en compte la surdispersion restante de nos données, et notamment la sous-estimation des zéros. Par ailleurs, un LRT entre le modèle ZINB et le modèle binomial négatif nous permettra de conclure quant à l'utilité de cette approche.

Les conclusions concernant le modèle binomial négatif "THV" sont similaires à celles du modèle "espèces".

La construction du rootogramme (Figure 7) du modèle ZINB montre un plus faible écart entre les valeurs prédites et les données observée. On remarque également que les zéros sont presque parfaitement prédits par ce modèle, et il semble donc mieux ajusté à nos données.

Le LRT entre le modèle binomial négatif et celui ZINB est fortement significatif ($p < 2.2 \times 10^{-16}$), ce dernier est donc significativement mieux ajusté à nos données, c'est

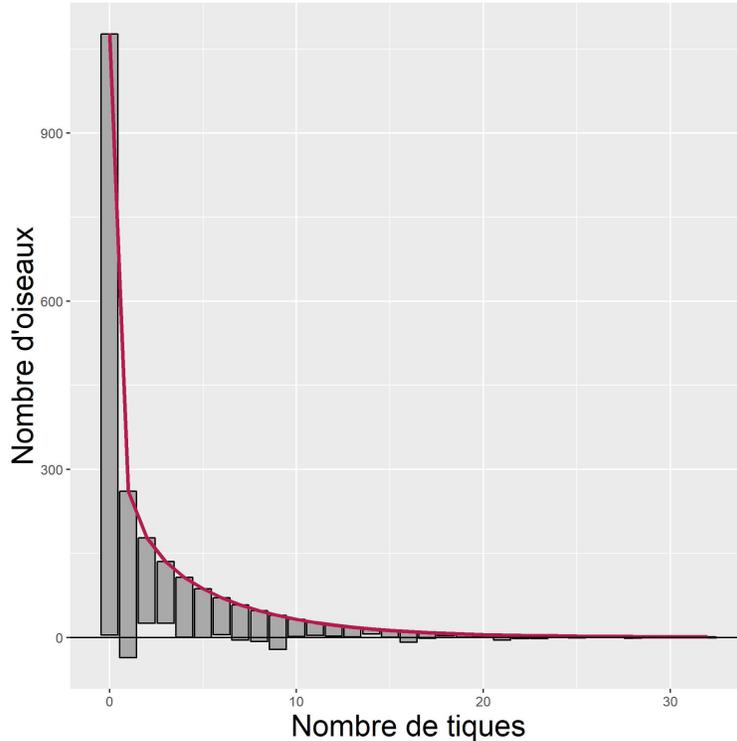


FIGURE 7 – Rootogramme du modèle ZINB (modèle "espèces")

donc le modèle que nous allons garder. Les conclusions du modèle "THV" sont équivalentes à celles présentées ici.

Les tableaux 1 et 2 donnent respectivement les odds-ratio significatifs des modèles "espèce" et "THV". Les modalités de référence utilisées pour les variables catégorielles sont :

- la Mésange charbonnière pour la variable "espèce", correspondant à un niveau moyen de charge en tiques ;
- les jeunes (oiseaux nés au cours du printemps de l'année considérée) pour la variable "sexe/âge"
- l'année 2007, première année de comptage de tiques sur la station STOC 204 pour la variable "année".

Concernant les composantes logistiques du modèle "espèce", on observe que les mésanges à longue queue ont plus de probabilité d'appartenir au groupe des oiseaux n'ayant jamais de tiques que les mésanges charbonnières, au contraire des rossignols philomèles. Les mâles adultes sont plus souvent dans ce groupe que les oiseaux de moins de un an.

La composante binomiale négative du même modèle montre une variabilité de la moyenne du nombre de tiques par espèce, puisque certaines espèces ont une moyenne plus faible que la Mésange charbonnière, comme le Pic épeiche, la Mésange bleue ou la Mésange huppée, tandis que d'autres ont une moyenne plus élevée comme le Rossignol philomèle, l'Accenteur mouchet, le Merle noir et la Grive musicienne. Les adultes, mâles et femelles ont un niveau

Tableau 1 – Coefficients significatifs du modèle ZINB (Modèle "espèce")

Composante logistique		Composante binomiale négative	
Mésange à longue queue	16,76 [2,25 ; 124,5]	Pic épeiche	0,07 [0,01 ; 0,77]
Rossignol philomèle	0,11 [0,05 ; 0,28]	Rossignol philomèle	2,00 [1,55 ; 2,58]
Mâle	4,36 [2,56 ; 7,44]	Mésange nonnette	0,09 [0,04 ; 0,21]
		Mésange bleue	0,03 [0,02 ; 0,05]
		Mésange huppée	0,35 [0,16 ; 0,78]
		Pouillot véloce	0,17 [0,11 ; 0,28]
		Pouillot fitis	0,19 [0,09 ; 0,39]
		Accenteur mouchet	3,31 [2,60 ; 4,22]
		Fauvette à tête noire	0,64 [0,50 ; 0,81]
		Fauvette des jardins	0,07 [0,04 ; 0,11]
		Merle noir	3,08 [2,35 ; 4,02]
		Grive musicienne	2,90 [2,15 ; 3,93]
		Femelle	0,44 [0,37 ; 0,53]
		Mâle	0,37 [0,32 ; 0,43]
		2008	0,67 [0,52 ; 0,87]
		2010	1,29 [1,02 ; 1,63]
		2015	0,76 [0,58 ; 0,99]
		2016	0,73 [0,56 ; 0,95]

Tableau 2 – Coefficients significatifs du modèle ZINB (Modèle "THV")

Composante logistique		Composante binomiale négative	
Hauteur moyenne d'alimentation	1,91 [1,64 ; 2,22]	Hauteur moyenne d'alimentation	0,70 [0,67 ; 0,73]
Poids moyen	0,77 [0,73 ; 0,81]	Poids moyen	1,01 [1,01 ; 1,01]
Mâle	3,48 [2,21 ; 5,47]	Femelle	0,40 [0,33 ; 0,48]
2008	2,57 [1,01 ; 6,53]	Mâle	0,41 [0,35 ; 0,48]
2011	4,26 [1,78 ; 10,20]	2008	0,72 [0,54 ; 0,96]
		2010	1,35 [1,04 ; 1,76]
		2011	1,32 [1,01 ; 1,74]
		2015	0,69 [0,52 ; 0,92]

moyen de charge en tiques inférieur aux jeunes de moins de un an. Enfin, il existe une variabilité annuelle puisque les niveaux moyens de charge en tiques observés en 2008, 2015 et 2016 étaient inférieurs à ceux observés en 2007, au contraire de l'année 2010.

Les coefficients du modèle "THV" permettent d'apporter des résultats complémentaires. Ainsi, on observe que l'augmentation de la hauteur d'alimentation augmente la probabilité que l'oiseau n'ait jamais de tiques et diminue le nombre de tiques portées par celui-ci. Le poids a un effet inverse : une augmentation du poids moyen de l'espèce diminue les chances

que l'oiseau n'ait jamais de tiques et augmente le nombre moyen de tiques observées sur l'oiseau. Les coefficients de ce modèle permettent également de confirmer l'effet du sexe/âge et la variabilité annuelle du précédent modèle.

3.2 Second volet : étude du réseau oiseaux-agents pathogènes¹

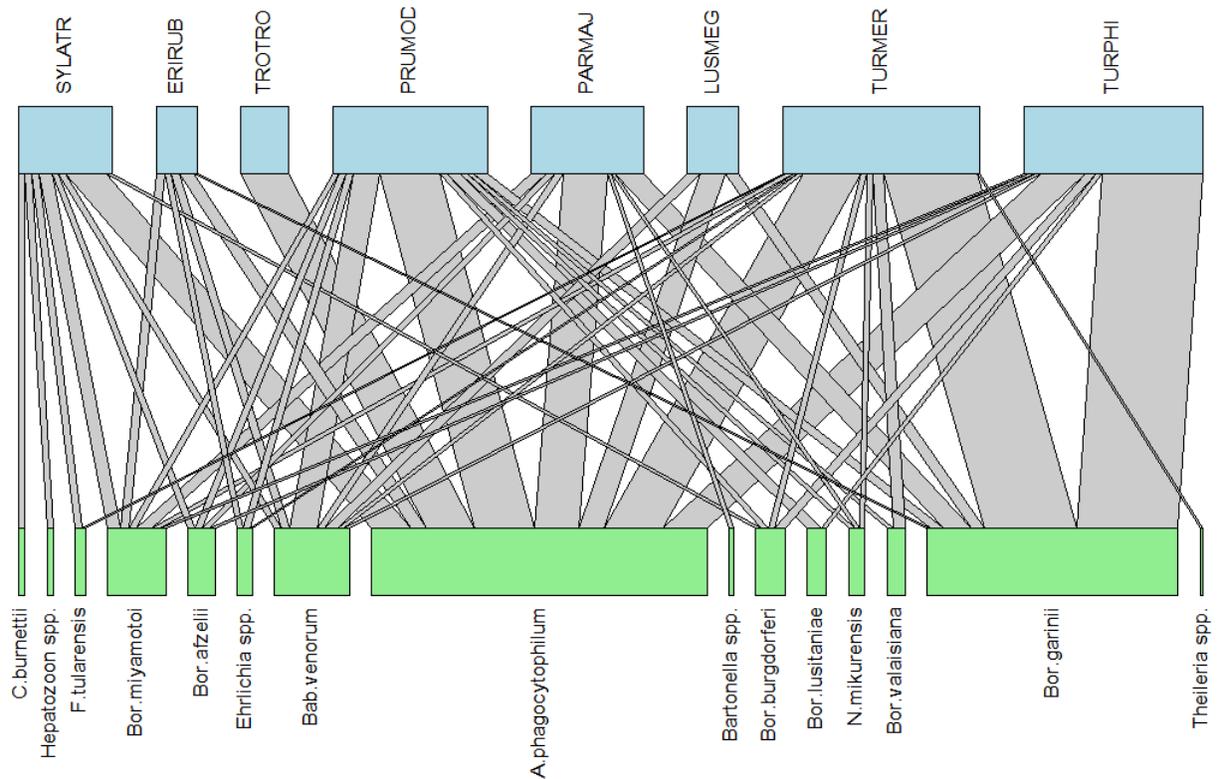


FIGURE 8 – Graphe du réseau oiseaux-agents pathogènes

En admettant que le co-feeding entre une nymphe infectée et une larve soit négligeable, tout comme la transmission transovarienne des *Borreliella*, on fait l'hypothèse que les larves de tiques à l'affut sont naïves et que si un pool de tiques analysé au laboratoire est positif en *Borreliella*, alors cela signifie que l'oiseau sur lequel elles ont été prélevées est lui-même porteur de bactéries du genre *Borreliella*. Nous avons conservé uniquement les agents pathogènes pour lesquels la transmission transovarienne est impossible ou négligeable.

Sur les 811 échantillons analysés au laboratoire, nous avons sélectionné uniquement ceux provenant de larves de tiques et ayant au moins un résultat positif pour les agents patho-

1. Pour toutes les figures de cette partie, on utilisera les notations : ERIRUB : *Erithacus rubecula* (Rouge-gorge familier), LUSMEG : *Luscinia megarhynchos* (Rossignol philomèle), PARMAJ : *Parus major* (Mésange charbonnière), PRUMOC : *Prunella modularis* (Accenteur mouchet), SYLATR : *Sylvia atricapilla* (Fauvette à tête noire), TROTRO : *Troglodytes troglodytes* (Troglodyte mignon), TURMER : *Turdus merula* (Merle noir), TURPHI : *Turdus philomelos* (Grive musicienne)

gènes étudiés. Parmi ces échantillons, nous avons considéré uniquement les espèces aviaires ayant au moins 10 individus échantillonnés dans la base de données. Enfin, pour les oiseaux positifs en *Borrelia*, nous avons fait le choix de conserver uniquement l'information au niveau espèce de *Borrelia* quand elle était disponible, et de considérer les oiseaux positifs en *Borrelia* mais sans information au niveau de l'espèce comme négatifs (20,2 % des échantillons sélectionnés avaient un résultat positif pour *Borrelia* spp., sans qu'une espèce soit identifiée). Au final, cette seconde partie de l'étude a porté sur 265 oiseaux.

Après avoir construit la matrice d'affiliation du réseau, en utilisant les prévalences de pathogènes observées par espèces d'oiseaux pour pondérer les connexions, nous avons pu construire le réseau présenté en figure 8.

Les degrés calculés pour les différents nœuds du réseau sont présentés dans les tableaux 3 et 4, ainsi que les degrés moyens pour chaque sous-groupe d'acteurs. On observe que la Mésange charbonnière, l'Accenteur mouchet, la Fauvette à tête noire, le Merle noir et la Grive musicienne avaient tous un degré non pondéré supérieur à la moyenne, et donc que ces espèces avaient un plus grand nombre de connexions avec des agents pathogènes que les autres oiseaux. En revanche, en s'intéressant au degré pondéré, en prenant en compte le poids de ces connexions, la Fauvette à tête noire n'était plus au dessus de la moyenne. Elle entretenait donc plus de connexions que la moyenne, mais avec une faible intensité de ces relations. L'Accenteur mouchet, le Merle noir et la Grive musicienne apparaissaient donc comme les oiseaux les plus centraux dans ce réseau.

Parmi les agents pathogènes, *A. phagocytophilum*, *Bab. venorum*, *Bor. afzelii*, *Bor. burgdorferi*, *Bor. garinii*, *Bor. miyamotoi* avaient un degré non pondéré supérieur à la moyenne, et étaient donc observés chez un plus grand nombre d'espèces que les autres agents pathogènes. En s'intéressant au degré pondéré, seules *A. phagocytophilum*, *Bab. venorum*, *Bor. garinii* et *Bor. miyamotoi* avaient une valeur supérieure ou égale à la moyenne, et étaient ainsi plus centrales que les autres agents pathogènes dans le réseau.

En appliquant l'algorithme DIRTLPawb+ sur notre réseau, nous avons obtenu une partition maximisant la modularité, présentée en Figure 9. On remarque que les deux turdidés du réseau (le Merle noir et la Grive musicienne) formaient un sous-graphe avec les deux espèces de *Borrelia* inféodées aux oiseaux (*Bor. garinii* et *Bor. valaisiana*) ainsi que avec *Bor. burgdorferi* et *Theileria* spp. Les deux turdidés avaient donc plus de connexions avec ces agents pathogènes que avec les autres agents présents dans le réseau. Concernant les autres *Borrelia*, nous avons retrouvé *Bor. lusitaniae* dans un cluster avec *Bab. venorum*, *Ehrlichia* spp., *N. mikurensis*, le Rouge-gorge familier et l'Accenteur mouchet, tandis que *Bor. afzelii* était retrouvé dans une communauté comprenant *Bartonella* spp., *Bor. miyamotoi*, *C. burnettii*, *F. tularensis*, *Hepatozoon* spp., la Mésange charbonnière et la Fauvette à tête noire.

Enfin, après avoir calculé les indices c et z pour chaque espèce, on peut les placer sur un graphique zc (Figure 10), et leur attribuer un rôle dans le réseau, en accord avec les valeurs seuils proposées par Guimera & Amaral (2005); Olesen et al. (2007). *A. phagocytophilum* ne figure pas sur ce graphe puisque cette espèce étant le seul agent pathogène dans son mo-

Tableau 3 – Degrés des oiseaux

	Degré non-pondéré	Degré pondéré
Rouge-gorge familier	5	0,24
Rossignol philomèle	3	0,31
Mésange charbonnière	7	0,67
Accenteur mouchet	10	0,92
Fauvette à tête noire	8	0,56
Troglodyte mignon	1	0,29
Merle noir	10	1,17
Grive musicienne	7	1,07
Degré moyen	6,38	0,65

Tableau 4 – Degrés des agents pathogènes

	Degré non-pondéré	Degré pondéré
<i>A. phagocytophilum</i>	8	2
<i>Bab. venorum</i>	7	0,44
Bartonella spp.	1	0,02
<i>Bor. afzelii</i>	4	0,16
<i>Bor. burgdorferi</i>	4	0,18
<i>Bor. garinii</i>	6	1,49
<i>Bor. lusitaniae</i>	2	0,11
<i>Bor. miyamotoi</i>	6	0,35
<i>Bor. valaisiana</i>	2	0,11
<i>C. burnettii</i>	1	0,04
Ehrlichia spp.	3	0,09
<i>F. tularensis</i>	2	0,06
Hepatozoon spp.	1	0,04
<i>N. mikurensis</i>	3	0,1
Theileria spp.	1	0,02
Degré moyen	3,4	0,35

dule, il n'était pas possible de calculer d'écart type (équation 46) . Nous avons constaté qu'il n'existait pas, dans notre réseau, d'espèces jouant les rôles de "Module hub" ou "Network hub", c'est à dire des espèces permettant la cohérence au sein de leurs propres modules. En revanche, nous avons observé des espèces connectrices, permettant la connectivité entre les différents modules, et des espèces périphériques, ayant peu de liens avec les autres espèces, que ce soit à l'intérieur ou à l'extérieur de leurs modules, comme *Bab. venorum*, *Bor. burgdorferi*, *N. mikurensis* ou la Grive musicienne (Figure 10).

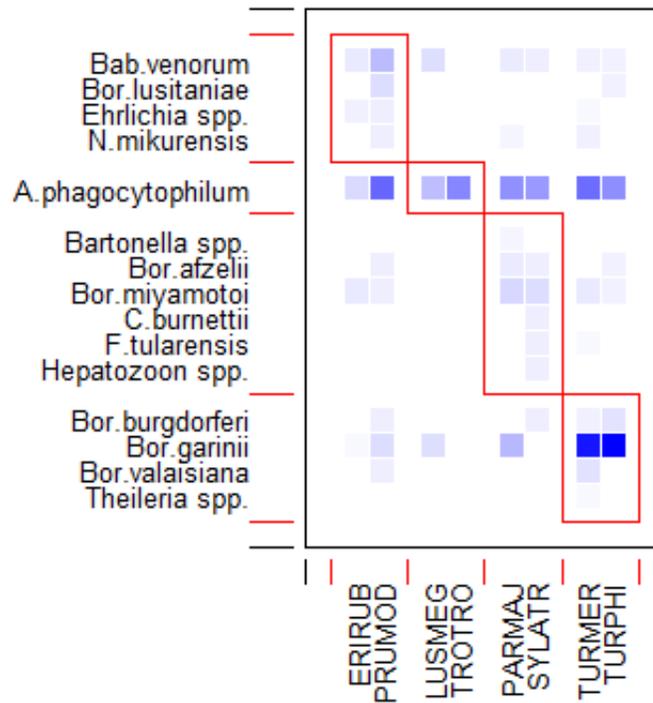


FIGURE 9 – Partition du réseau oiseaux-agents pathogènes

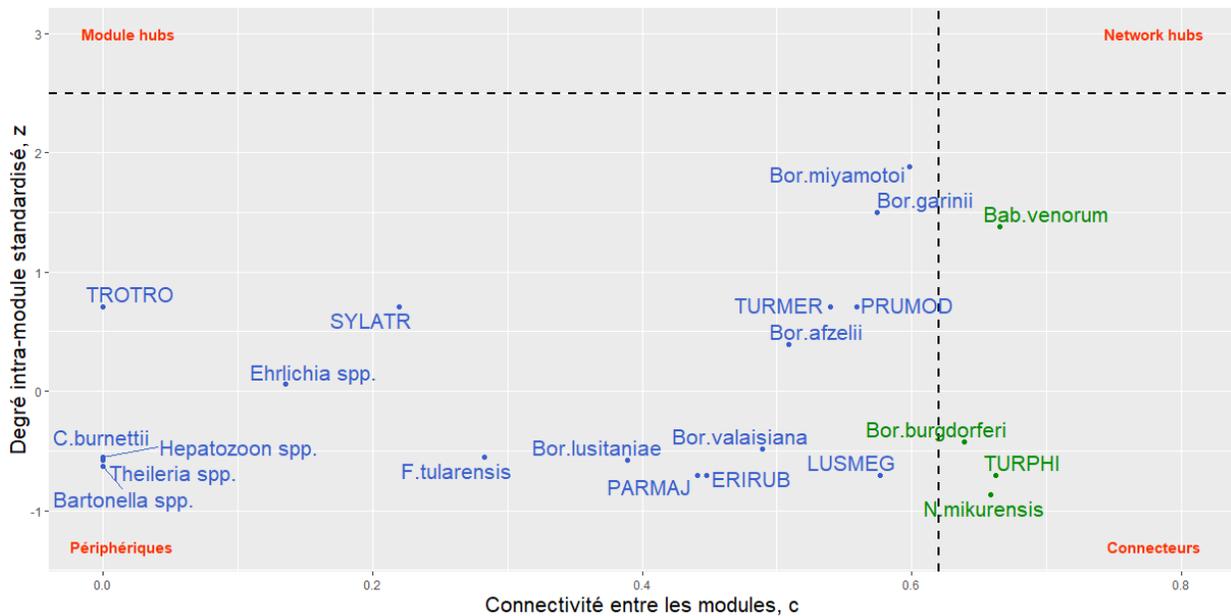


FIGURE 10 – Graphe zc des acteurs du réseau oiseaux-agents pathogènes

4 Discussion

Cette étude est une première étape du projet OUTLYER, qui vise à étudier la contribution des espèces communes d'oiseaux en période de reproduction en France dans la dynamique des populations de tiques et la circulation des *Borrelia* en France. A notre connaissance, il s'agit de la première étude portant sur ce sujet en France. Elle a permis de répondre à notre première question et confirmer qu'il existe réellement une variabilité interspécifique et temporelle de la charge en tique des oiseaux commune en France, et que celle-ci est liée aux traits d'histoire de vie des oiseaux (poids moyen de l'espèce, hauteur moyenne d'alimentation) ainsi qu'à des caractéristiques individuelles (sexe et âge). D'autre part, l'absence de la totalité des données du niveau espèce des *Borrelia* ne nous permet pas de savoir si certaines espèces de *Borrelia* sont inféodées à des espèces aviaires particulières, mais cette étude a permis de mettre en place les méthodes statistiques d'analyse de réseau, de montrer qu'elles sont applicables et qu'elles apportent des conclusions potentiellement intéressantes pour la suite du projet.

Le premier volet de l'étude nous a permis de montrer qu'il existe une variabilité annuelle de la charge en tiques, puisque certaines années sont marquées par une augmentation de celle-ci par rapport à 2007 (comme l'année 2010), et d'autres par une diminution (comme 2008 ou 2015). Cette variabilité pourrait être expliquée par des variations de densité en tiques à l'affût sur la végétation (Marsot et al., 2012), des variations météorologiques annuelles (Heylen et al., 2014c) (taux d'humidité, gelées ou neige durant l'hiver), ainsi que par les fluctuations de populations de la faune sauvage, notamment les rongeurs et les chevreuils, espèces sur lesquelles se nourrissent les différentes stases de tiques. Nous avons pu mettre en évidence une variabilité interspécifique de la charge en tiques, puisque certaines espèces aviaires telles que le Merle noir, la Grive musicienne ou l'Accenteur mouchet sont porteurs d'un plus grand nombre de tiques, tandis que d'autres espèces comme la Mésange bleue ou la Fauvette à tête noire ont une charge en tiques significativement inférieure à la moyenne. Ces résultats sont conformes à la littérature puisque le Rossignol philomèle et le Merle noir ont été caractérisés comme plus infestés par les tiques que les autres espèces au Portugal (Norte et al., 2015), tout comme les turdidés en France (Doby, 1999).

Ces résultats peuvent être expliqués en partie par les traits d'histoire de vie et par des caractéristiques individuelles propres aux oiseaux, que nous avons étudiés dans le modèle "THV". En effet, nous avons observé que plus la hauteur moyenne d'alimentation augmente, plus la charge en tiques de l'oiseau diminue. En comparant ce résultat avec les valeurs de hauteur d'alimentation par espèce (voir tableau 5 en annexe), on remarque que les espèces ayant une charge en tiques plus élevée que la moyenne sont celles qui se nourrissent le plus près du sol, et à l'inverse celles ayant une charge en tiques moins importante se nourrissent en hauteur. L'effet de l'utilisation verticale de l'espace a déjà été mise en évidence sur les oiseaux communs en France et pendant la période de reproduction (Marsot et al., 2012) et en Suède sur les oiseaux migrateurs (Comstedt et al., 2006). Le poids de l'oiseau, considéré ici comme une caractéristique propre à l'espèce et non individuelle, est mis en évidence comme un facteur de risque d'augmentation de la charge en tiques, puisque plus celui-ci augmente, plus

la cette charge est élevée. Cela pourrait s'expliquer par une plus grande surface corporelle et donc une plus grande zone de contact entre l'oiseau et la végétation. Cela tendrait à accroître la probabilité de rencontre entre les tiques et leurs hôtes, mais aussi à augmenter le nombre de tiques qui pourraient s'y accrocher. En revanche, l'effet mis en évidence dans cette étude est faible (OR = 1,01), et cela se traduit par le fait que certaines espèces comme le Pic épeiche ont une masse corporelle importante mais une faible charge en tiques. Le lien entre l'augmentation du poids et l'intensité de la charge en tiques a été montré précédemment en France (Marsot et al., 2012), mais pas en Hongrie (Hornok et al., 2016). Enfin, en considérant que les jeunes ne présentent pas de dimorphisme sexuel, ni phénotypique ni comportemental, nous avons mis en évidence un effet individuel de l'âge sur l'intensité de la charge en tiques, puisque les adultes mâles et femelles ont moins de tiques que les jeunes oiseaux dans notre étude. Cet effet est cohérent avec les résultats montrés par Heylen et al. (2013a) pour les femelles mais pas pour les mâles. En effet, les auteurs de cette étude ont séparé les femelles des mâles chez les jeunes et ont montré en Belgique que la charge en tiques de la Mésange charbonnière diminue avec l'âge chez les femelles mais augmente chez les mâles. Par ailleurs, les effets du sexe et de l'âge sur la charge en tiques n'avaient pas été mis en évidence par Marsot et al. (2012), et nous n'avons donc pas d'hypothèse biologique pouvant expliquer nos résultats.

Concernant la méthodologie, nous avons mis en évidence que le modèle ZINB est significativement mieux ajusté à nos données que le modèle de Poisson et que le modèle binomial négatif. Cela s'explique par le fait que nos données présentent une surdispersion par rapport à une loi de Poisson, et qu'il existe un excès de zéro. Biologiquement, ce modèle est cohérent, puisqu'il permet de considérer deux processus : la probabilité de rencontre entre les oiseaux et les tiques d'une part, et le nombre de tiques s'accrochant à l'oiseau d'autre part. En revanche, ce modèle présente des limites pour la taille de nos données, puisqu'il y a un grand nombre de coefficients à calculer, et que certains n'ont pas pu être estimés dans le modèle logistique. Une approche par modèle de quasi-Poisson aurait pu être envisagée comme second modèle, pour prendre en compte la surdispersion de nos données, mais celui-ci étant estimé par quasi-vraisemblance, il apparaissait plus compliqué de le comparer à nos modèles, notamment dans la perspective d'implémenter un modèle ZINB par la suite. Il aurait également pu être intéressant d'envisager un modèle ZINB à effets mixtes afin d'étudier les variabilités intra- et interspécifiques de charge en tiques, mais le nombre de paramètres à estimer aurait probablement été trop élevé comparé à la taille de nos données. Par ailleurs, le calcul des quatre types différents de coefficients (modèle logistique vs. modèle de comptage avec effets aléatoires vs. effets fixes) aurait posé des problèmes d'interprétation biologique. Enfin, la comparaison des modèles "THV" et "espèces" par analyse de déviance aurait été intéressante afin de calculer le pourcentage de déviance du modèle "espèce" expliqué par les traits d'histoire de vie ajoutés au second modèle, mais le concept de déviance n'est pas défini pour les modèles de régression à inflation de zéros.

Pour le second volet, nous n'avons pas pu mettre en place l'analyse du réseau oiseaux-*Borrelia*, et nous avons donc travaillé sur un réseau d'entraînement oiseaux-agents patho-

gènes. Nous avons pu montrer que cette méthode est aisée à mettre en place car bien implémentée sous R au sein du package `bipartite` et que ces méthode permettraient de mettre en évidence des résultats intéressants lorsqu'elles seront appliquées aux données réelles. Les conclusions de ce volet de l'étude ne doivent donc pas être considérés comme des résultats biologiques mais comme des pistes de réflexion pour la future application de ces méthodes sur les données complètes. Dans le réseau d'entraînement, nous avons montré que les espèces aviaires les plus centrales étaient la Mésange charbonnière, l'Accenteur mouchet, le Merle noir et la Grive musicienne. Nous avons pu observer que les espèces de *Borrelia* connues comme associées aux oiseaux se positionnaient dans les même clusters que les turdidés (Merle noir et Grive musicienne). Ces résultats sont cohérents avec ceux de Norte et al. (2015); Palomar et al. (2017), qui montrent que les turdidés sont les plus porteurs de *Borrelia*. On peut faire l'hypothèse que ce lien entre les turdidés et les *Borrelia* sera de nouveau mis en évidence lors de l'application de l'analyse de réseaux aux données réelles. En revanche, *Parus major* est souvent cité comme infecté par différentes espèces de *Borrelia* (Heylen et al., 2013b) mais nous n'avons pas mis en exergue cette relation dans le partitionnement de notre réseau d'entraînement. D'une part, on peut expliquer cette différence par le fait que notre étude compare plus d'espèces que les études précédentes et plus d'agents pathogènes, ce qui permet de prendre en compte les relations entre espèces différentes des autres études et donc de mettre en avant d'autre relations que celles oiseaux-*Borrelia*. D'autre part, on se pose la question de la validité biologique du partitionnement réalisé par l'algorithme de clustering non supervisé. En effet, il paraît contre-intuitif que le Rouge-gorge familier, le Rossignol philomèle et les turdidés appartiennent à des clusters différents alors que ces espèces sont proches phylogénétiquement et partagent des traits d'histoire de vie communs. On aurait également pu penser que le Troglodyte mignon et l'Accenteur mouchet formeraient un cluster avec le Rouge-gorge familier et le Rossignol philomèle, puisqu'ils partagent les mêmes niches écologiques. Ces résultats, différents de ceux attendus, peuvent être dus au nombre de données disponibles par espèces d'oiseaux (seulement 13 individus pour le Troglodyte mignon et 14 pour le Rossignol philomèle), et il sera intéressant de voir si cette séparation des espèces persiste lors de l'application du partitionnement sur réseau des données réelles. La collecte de tiques commencée en 2019 permettra également d'augmenter la taille du jeu de données, et donc d'avoir une plus grande confiance quant au partitionnement proposé par l'algorithme DIRTLPawb+. Enfin, le graphe z_c de ce réseau d'entraînement n'a pas permis de montrer l'existence d'espèces "hub" permettant la cohérence au sein des modules, mais uniquement des espèces périphériques avec peu de liens et certaines espèces connectrices entre les modules, comme la Grive musicienne. En appliquant cette méthode sur le réseau oiseaux-*Borrelia*, nous pourrions potentiellement montrer l'existence de certaines espèces ayant un rôle dans la cohérence des modules.

Nous avons choisi de réaliser ici une analyse de réseau afin de mettre en évidence des structures particulières de communauté, et d'étudier les relations entre oiseaux et pathogènes, et cette méthode paraît satisfaisante pour savoir s'il existe certaines espèces de *Borrelia* inféodées à des espèces aviaires particulières. Néanmoins, une approche de type GLM aurait également pu être possible afin d'étudier la variabilité interspécifique de prévalence en *Bor-*

reliella. La méthode de partitionnement du réseau par maximisation de la modularité, bien que classique dans l'analyse de réseaux 2-modes, est une méthode de classification non supervisée et elle ne prend en compte que la structure du réseau et pas les similarités entre les nœuds. Nous aurions pu utiliser d'autres méthodes de partitionnement, comme des méthodes de clustering hiérarchique permettant de regrouper les nœuds entre eux en fonction du calcul d'un indice de similarité (comme l'indice de Jaccard, fondé sur le nombre de voisins communs entre les nœuds). Par ailleurs, une approche des analyses par des méthodes de statistiques bayésiennes permettrait d'estimer le risque *a posteriori* lié à chaque espèce aviaire, en partant des observations *a priori* des prévalences. Concernant les méthodes de laboratoire, les sondes à ADN utilisées lors de la PCR micro-fluidique de cette étude peuvent présenter des réactions croisées entre espèces de *Borreliella*, ce qui nous a conduit à devoir confirmer des échantillons et à ne pas avoir toutes les informations au niveau espèces pour celles-ci. Les méthodes utilisées dans cette étude n'ont également pas permis d'identifier les espèces de tiques prélevées sur les oiseaux, puisque les sondes à ADN utilisées dans la PCR micro-fluidique ont mis en évidence la présence d'*I. ricinus*, mais pas d'*I. frontalis* ni d'*I. arboricola*, espèces de tiques ornitophiles que l'on pourrait éventuellement avoir dans nos échantillons. L'obtention de cette information permettrait d'inclure un niveau "tiques" dans le réseau, et de voir si celles-ci structurent les communautés d'oiseaux et de *Borreliella*.

Notre étude permet donc d'obtenir des premiers résultats et de mettre en place une méthodologie, qui pourront permettre au terme du projet OUTLYER d'étudier le système Habitats-Oiseaux-Tiques-*Borreliella* et de quantifier l'importance relative des oiseaux en reproduction pour le risque acarologique, c'est à dire à la production de nymphes infectées à l'affût dans la végétation. En effet, la caractérisation des relations spécifiques entre les espèces d'oiseaux et de *Borreliella*, couplée aux données portant sur l'intensité de la charge en tiques de ces oiseaux et à celles de l'abondance de leurs populations permettra de développer un indice de risque à l'échelle de la communauté d'oiseaux. Pour la suite de ce projet, il pourrait être intéressant d'inclure une dimension spatiale dans les analyses. En effet, pour les données de comptage, nous n'avons utilisé que les données de la station STOC 204, mais il serait pertinent de considérer d'autres stations afin de pouvoir éventuellement mettre en évidence un effet spatial ou environnemental sur l'intensité de la charge en tiques des oiseaux, si les données le permettent. Cette approche spatiale pourrait être envisagée dans l'analyse de réseau, soit en réalisant un réseau par type d'habitat et en comparant les métriques et partitions obtenues, soit en utilisant d'autres méthodes incluant des données phylogénétiques et environnementales. La transformation du réseau en dendrogrammes permettrait ainsi de déterminer si les agents pathogènes partagent préférentiellement des hôtes ayant une proximité phylogénétique ou évoluant dans les mêmes environnements, comme réalisé par Estrada-Peña et al. (2015).

Références

- Académie Nationale de Médecine (2018). Communiqué de presse du 2 juillet 2018 : "Mise au point de la Haute Autorité de Santé (HAS) à propos de la maladie de Lyme : Réactions et déception de l'Académie nationale de médecine". <http://www.academie-medecine.fr/communiquede-presse-du-2-juillet-2018-mise-au-point-de-la-haute-autorite-de-sante-has-a-propos-de-la-maladie-de-lyme-reactions-et-deception-de-lacademie-nationale-de-medecine/>.
- Apanaskevich, D. A., & Oliver, J. H. J. (2014). Life cycles and natural history of ticks. In *Biology of Ticks*, vol. Volume 1, (pp. 59–73). Oxford : Oxford University Press, 2nd ed.
- Baranton, G., & de Martino, S. (2009). *Borrelia burgdorferi sensu lato Diversity and Its Influence on Pathogenicity in Humans*. In *Lyme Borreliosis : Biological and Clinical Aspects*, (pp. 1–17). Basel; New York : Karger. OCLC : 433173332.
- Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, 76(6), 066102.
- Beauguitte, L. (2013). L'analyse des graphes bipartis. *halshs-00794976*, (p. 11).
- Beckett, S. J. (2016). Improved community detection in weighted bipartite networks. *Royal Society Open Science*, 3(1), 140536.
- Boulouis, H.-J., Lagrée, A.-C., Dugat, T., & Haddad, N. (2015). Les animaux vertébrés et les maladies dues à des bactéries vectorisées par les tiques. *Revue Francophone des Laboratoires*, 2015(472), 77–87.
- Bush, L. M., & Vazquez-Pertejo, M. T. (2018). Tick borne illness—Lyme disease. *Disease-a-Month*, 64(5), 195–212.
- Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, 46(3), 347–364.
- Centre National de Référence des *Borrelia* (2018). La bactérie responsable de la Borréliose de Lyme. *Les hôpitaux universitaires de Strasbourg*.
- Chesneau, C. (2018). Modèles de régression. *Université de Caen*.
- Coipan, E. C., & Sprong, H. (2016). Ecology of *Borrelia burgdorferi sensu lato*. In *Ecology and Prevention of Lyme Borreliosis*, vol. 4 of *Ecology and Control of Vector-Borne Diseases*, (pp. 41–62). Wageningen Academic Publishers.
- Comstedt, P., Bergström, S., Olsen, B., Garpmo, U., Marjavaara, L., Mejlön, H., Barbour, A., & Bunikis, J. (2006). Migratory passerine birds as reservoirs of Lyme borreliosis in Europe. *Emerging Infectious Diseases*, 12(7), 1087–1095.

- Devos, J., Galon, C., & Moutailler, S. (2019). Etude épidémiologique des agents infectieux transmis par les tiques. *Bulletin des GTV*, (93), 59–63.
- Doby, J. M. (1999). Contribution to the knowledge of Ixodes (Trichotoixodes) pari Leach (= *I. frontalis* (Panzer)) (Acari : Ixodidae), a tick specific to birds. *Acarologia*, 39(4), 315–325.
- Dormann, C. F., Gruber, B., & Fruend, J. (2008). Introducing the bipartite Package : Analysing Ecological Networks. *R News*, 8(2), 8–11.
- Dormann, C. F., & Strauss, R. (2014). A method for detecting modules in quantitative bipartite networks. *Methods in Ecology and Evolution*, 5(1), 90–98.
- Dufour, B., Brucker, G., Bénet, J.-J., Moutou, F., & Toma, B. (2018). Terminologie en épidémiologie animale. <http://aeema.vet-alfort.fr/index.php/component/glossary/Glossaire-1/>.
- Dupuy, J.-F. (2017). Models for overdispersed and zero-inflated count data - Application to health economics. *ISUP - filière Biostatistique*, (p. 95).
- Estrada-Peña, A., de la Fuente, J., Ostfeld, R. S., & Cabezas-Cruz, A. (2015). Interactions between tick and transmitted pathogens evolved to minimise competition through nested and coherent networks. *Scientific Reports*, 5, 10361.
- Guiloux, A. (2018). Le modèle linéaire généralisé. *Université d'Évry - Paris Saclay*.
- Guimerà, R., & Amaral, L. A. N. (2005). Cartography of complex networks : Modules and universal roles. *Journal of statistical mechanics (Online)*, 2005(P02001), P02001–P02001–13.
- Hamer, S. A., Hickling, G. J., Sidge, J. L., Rosen, M. E., Walker, E. D., & Tsao, J. I. (2011). Diverse *Borrelia burgdorferi* Strains in a Bird-Tick Cryptic Cycle. *Appl. Environ. Microbiol.*, 77(6), 1999–2007.
- Haute Autorité de Santé (2018). Borréliose de Lyme et autres maladies vectorielles à tiques (MVT). Tech. rep., HAS.
- Heylen, D., Adriaensen, F., Van Dongen, S., Sprong, H., & Matthysen, E. (2013a). Ecological factors that determine Ixodes ricinus tick burdens in the great tit (*Parus major*), an avian reservoir of *Borrelia burgdorferi* s.l. *International Journal for Parasitology*, 43(8), 603–611.
- Heylen, D., De Coninck, E., Jansen, F., & Madder, M. (2014a). Differential diagnosis of three common Ixodes spp. ticks infesting songbirds of Western Europe : *Ixodes arboricola*, *I. frontalis* and *I. ricinus*. *Ticks and Tick-borne Diseases*, 5(6), 693–700.
- Heylen, D., Krawczyk, A., de Carvalho, I. L., Nuncio, M. S., Sprong, H., & Norte, A. C. (2017). Bridging of cryptic *Borrelia* cycles in European songbirds. *Environmental Microbiology*, 19(5), 1857–1867.

- Heylen, D., Matthysen, E., Fonville, M., & Sprong, H. (2014b). Songbirds as general transmitters but selective amplifiers of *Borrelia burgdorferi* sensu lato genotypes in Ixodes ricinus ticks. *Environmental Microbiology*, *16*(9), 2859–2868.
- Heylen, D., Tijssse, E., Fonville, M., Matthysen, E., & Sprong, H. (2013b). Transmission dynamics of *Borrelia burgdorferi* s.l. in a bird tick community. *Environmental Microbiology*, *15*(2), 663–673.
- Heylen, D. J. A., Oosten, A. R. V., Devriendt, N., Elst, J., Bruyn, L. D., & Matthysen, E. (2014c). Seasonal feeding activity of the tree-hole tick, *Ixodes arboricola*. *Parasitology*, *141*(8), 1044–1051.
- Hilbe, J. (2011). *Negative Binomial Regression*. Cambridge ; New York : Cambridge University Press.
- Hornok, S., Flaisz, B., Takács, N., Kontschán, J., Csörgő, T., Csipak, Á., Jaksa, B. R., & Kováts, D. (2016). Bird ticks in Hungary reflect western, southern, eastern flyway connections and two genetic lineages of *Ixodes frontalis* and *Haemaphysalis concinna*. *Parasites & Vectors*, *9*.
- Jais, J.-P. (2018). Les Modèles Linéaires Généralisés et leurs extensions. *Université Paris 5*.
- Kleiber, C., & Zeileis, A. (2016). Visualizing Count Data Regressions Using Rootograms. *The American Statistician*, *70*(3), 296–303.
- Kolaczyk, E. D., & Csárdi, G. (2014). *Statistical Analysis of Network Data with R*. Use R! New York : Springer-Verlag.
- Lloyd-Smith, J. O. (2007). Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases. *PLOS ONE*, *2*(2), e180.
- Marsot, M., Henry, P. Y., Vourc'h, G., Gasqui, P., Ferquel, E., Laignel, J., Grysan, M., & Chapuis, J. L. (2012). Which forest bird species are the main hosts of the tick, *Ixodes ricinus*, the vector of *Borrelia burgdorferi* sensu lato, during the breeding season? *International Journal for Parasitology*, *42*(8), 781–788.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Boca Raton, Fla. : Chapman & Hall/CRC. OCLC : 45722035.
- Mead, P. S. (2015). Epidemiology of Lyme disease. *Infectious Disease Clinics of North America*, *29*(2), 187–210.
- Michelet, L., Delannoy, S., Devillers, E., Umhang, G., Aspan, A., Juremalm, M., Chirico, J., van der Wal, F. J., Sprong, H., Boye Pihl, T. P., Klitgaard, K., Bødker, R., Fach, P., & Moutailler, S. (2014). High-throughput screening of tick-borne pathogens in Europe. *Frontiers in Cellular and Infection Microbiology*, *4*.

- Muséum National d'Histoire Naturel (2019). Inventaire Natinal du Patrimoine Naturel. <https://inpn.mnhn.fr/accueil/index>.
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.
- Norte, A. C., da Silva, L. P., Tenreiro, P. J. Q., Felgueiras, M. S., Araújo, P. M., Lopes, P. B., Matos, C., Rosa, A., Ferreira, P. J. S. G., Encarnação, P., Rocha, A., Escudero, R., Anda, P., Nuncio, M. S., & Lopes de Carvalho, I. (2015). Patterns of tick infestation and their *Borrelia burgdorferi* s.l. infection in wild birds in Portugal. *Ticks and Tick-borne Diseases*, *6*(6), 743–750.
- Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, *104*(50), 19891–19896.
- Olsén, B., Jaenson, T. G., & Bergström, S. (1995). Prevalence of *Borrelia burgdorferi* sensu lato-infected ticks on migrating birds. *Applied and Environmental Microbiology*, *61*(8), 3082–3087.
- Palomar, A. M., Portillo, A., Santibáñez, P., Mazuelas, D., Roncero, L., Gutiérrez, Ó., & Oteo, J. A. (2017). Presence of *Borrelia turdi* and *Borrelia valaisiana* (Spirochaetales : Spirochaetaceae) in Ticks Removed From Birds in the North of Spain, 2009–2011. *Journal of Medical Entomology*, *54*(1), 243–246.
- Poupon, M.-A., Lommano, E., Humair, P.-F., Douet, V., Rais, O., Schaad, M., Jenni, L., & Gern, L. (2006). Prevalence of *Borrelia burgdorferi* sensu lato in ticks collected from migratory birds in Switzerland. *Applied and Environmental Microbiology*, *72*(1), 976–979.
- R Core Team (2018). *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing.
- Rizzoli, A., Hauffe, H. C., Carpi, G., Vourc'h, G. I., Neteler, M., & Rosà, R. (2011). Lyme borreliosis in Europe. *Eurosurveillance*, *16*, dsf.
- Ruyts, S. C., Tack, W., Ampoorter, E., Coipan, E. C., Matthysen, E., Heylen, D., Sprong, H., & Verheyen, K. (2018). Year-to-year variation in the density of *Ixodes ricinus* ticks and the prevalence of the rodent-associated human pathogens *Borrelia afzelii* and *B. miyamotoi* in different forest types. *Ticks and Tick-borne Diseases*, *9*(2), 141–145.
- Septfons, A. (2019). Commission des affaires sociales : Compte rendu de la semaine du 25 mars 2019. <http://www.senat.fr/compte-rendu-commissions/20190325/soc.html#toc2>.
- Soulie Fogelman, F., & Viennet, E. (2016). L'analyse des réseaux sociaux. *1024 – Bulletin de la société informatique de France*, (pp. 1–19).

- Stanek, G., Wormser, G. P., Gray, J., & Strle, F. (2012). Lyme borreliosis. *The Lancet*, 379(9814), 461–473.
- Steere, A. C., Strle, F., Wormser, G. P., Hu, L. T., Branda, J. A., Hovius, J. W. R., Li, X., & Mead, P. S. (2016). Lyme borreliosis. *Nature Reviews. Disease Primers*, 2.
- Storchová, L., & Hořák, D. (2018). Life-history characteristics of European birds. *Global Ecology and Biogeography*, 27(4), 400–406.
- Sueur, C. (2015). *Analyse des réseaux sociaux appliquée à l'éthologie et l'écologie. Modélisations, Simulations, Systèmes complexes*. Paris : Editions Matériologiques. OCLC : 966294557.
- Tackx, R. (2018). Analyse de la structure communautaire des réseaux bipartis. *Thèse de doctorat - Université Pierre et Marie Curie*, (p. 118).
- UCLA, S. C. G. (2014). Zero-Inflated Negative Binomial Regression | R Data Analysis Examples. <https://stats.idre.ucla.edu/r/dae/zinb/>.
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York : Springer, fourth ed.
- Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-poisson vs. negative binomial regression : How should we model overdispersed count data? *Ecology*, 88(11), 2766–2772.
- Wasserman, S., & Faust, K. L. M. (1994). *Social Network Analysis : Methods and Applications*. Cambridge : Cambridge University Press. OCLC : 782060651.
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0 : Species-level foraging attributes of the world's birds and mammals. *Ecology*, 95(7), 2027–2027.
- Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News*, 2(3), 7–10.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8), 1–25.

Annexes

Tableau 5 – Tableau des valeurs utilisées pour les traits d’histoire de vie

Nom scientifique	Nom vernaculaire	Hauteur moyenne d'alimentation (m)	Poids moyen (g)
<i>Prunella modularis</i>	Accenteur mouchet	0,05	21,25
<i>Turdus philomelos</i>	Grive musicienne	0,05	76
<i>Luscinia megarhynchos</i>	Rossignol philomèle	0,34	20,15
<i>Erithacus rubecula</i>	Rougegorge familier	0,52	17,55
<i>Troglodytes troglodytes</i>	Troglodyte mignon	0,52	8,9
<i>Turdus merula</i>	Merle noir	1,03	97
<i>Fringilla coelebs</i>	Pinson des arbres	1,52	22,15
<i>Sylvia borin</i>	Fauvette des jardins	2,2	18,7
<i>Phylloscopus collybita</i>	Pouillot véloce	2,26	7,7
<i>Aegithalos caudatus</i>	Mésange à longue queue	2,4	8,5
<i>Pyrrhula pyrrhula</i>	Bouvreuil pivoine	2,5	31
<i>Cyanistes caeruleus</i>	Mésange bleue	2,7	11,5
<i>Hypolaïs polyglotta</i>	Hypolaïs polyglotte	2,8	11,7
<i>Dendrocopos major</i>	Pic épeiche	3	74
<i>Lophophanes cristatus</i>	Mésange huppée	3,31	11,35
<i>Poecile palustris</i>	Mésange nonnette	3,4	10,5
<i>Parus major</i>	Mésange charbonnière	3,4	18,25
<i>Sylvia atricapilla</i>	Fauvette à tête noire	4	18,6
<i>Phylloscopus trochilus</i>	Pouillot fitis	4	9,1