



HAL
open science

Modélisation spatio-temporelle de la dynamique d'émergence de lépidoptères univoltins à partir de données opportunistes

Clara Navarro

► **To cite this version:**

Clara Navarro. Modélisation spatio-temporelle de la dynamique d'émergence de lépidoptères univoltins à partir de données opportunistes. Sciences de l'environnement. 2021. hal-04694077

HAL Id: hal-04694077

<https://hal.inrae.fr/hal-04694077v1>

Submitted on 11 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Master Biodiversité Écologie Évolution
Parcours Modélisation en écologie

**Modélisation spatio-temporelle de la
dynamique d'émergence de lépidoptères
univoltins à partir de données
opportunistes**

Clara Navarro

Étude réalisée à l'institut national de recherche pour l'agriculture,
l'alimentation et l'environnement, du 6 avril au 11 juin 2021
Sous la direction de Julien Papaïx et Emily Walker

En collaboration avec Yoann Bunz et Damien Combrisson du Parc national
des Écrins et Christophe Botella du laboratoire d'écologie alpine

Soutenu à Rennes le 24 juin 2021
Enseignant référent : Andreas Prinzing

Résumé

Les gestionnaires des espaces naturels souhaitent exploiter leurs bases de données opportunistes pour évaluer les dynamiques de biodiversité au sein des espaces qu'ils gèrent. Cependant, l'analyse de données opportunistes nécessite une attention particulière puisque l'effort d'échantillonnage est inconnu et hétérogène dans l'espace. De plus, les modèles de distribution d'espèce basés sur la niche ne prennent que rarement la dimension temporelle et n'explicitent pas les processus impliqués dans la dynamique écologique étudiée. En se basant sur les données opportunistes d'occurrence de papillons provenant du parc national des Écrins, nous avons développé un modèle spatio-temporel décrivant la dynamique d'émergence du stade adulte le long de gradients environnementaux. Le modèle permet de séparer les variations spatiales de l'effort d'échantillonnage de la distribution spatiale des espèces considérées ainsi que d'estimer des paramètres clés de la dynamique d'émergence. Les estimations des dates de première émergence sont relativement précoces dans l'année, mais leurs tendances sont cohérentes. Les variables environnementales (altitude et forêt) ont un effet négatif sur l'abondance relative des espèces. L'intégration de données climatiques permettrait une meilleure compréhension de la variabilité spatio-temporelle des dates d'émergence ainsi que l'étude des capacités des espèces à s'adapter au changement climatique par la modification de leur phénologie et de leur aire de répartition.

Mots clés— données opportunistes, processus ponctuel, échantillonnage préférentiel, modèle spatio-temporel, émergence

Spatio-temporal modelling of univoltine lepidoptera dynamic from opportunistic data

Abstract

Managers of natural areas wish to use their opportunistic databases to assess the dynamics of biodiversity in the areas they manage. However, the analysis of opportunistic data requires special attention since the sampling effort is unknown and spatially heterogeneous. Moreover, species distribution models based on the niche rarely take into account the temporal dimension and do not explain the processes involved in the ecological dynamics studied. Based on opportunistic butterfly occurrence data from the Écrins National Park, we developed a spatio-temporal model describing the emergence dynamics of the adult stage along environmental gradients. The model allows us to separate spatial variations in sampling effort from the spatial distribution of the species considered and to estimate key parameters of the emergence dynamics. The estimates of first emergence dates are relatively early in the year, but their trends are consistent. Environmental variables (altitude and forest) have a negative effect on the relative abundance of species. The integration of climatic data would allow a better understanding of the spatio-temporal variability of emergence dates and the study of the capacity of species to adapt to climate change by modifying their phenology and range.

Keywords— opportunistic data, point process, preferential sampling, spatio-temporal model, emergence

Remerciements

Je remercie vivement Julien Papaïx et Emily Walker, mes encadrants, à la fois pour leur partage d'expertise et le temps consacré à répondre à toutes mes interrogations, mais également pour la confiance qu'ils m'ont accordé tout au long de ce stage.

Merci à l'ensemble des membres du réseau scientifique CiSStat, nos rencontres ont participé à faire de mon stage une expérience enrichissante et pleine d'intérêt. Je remercie en particulier Christophe Botella pour ses conseils et son recul scientifique.

Je tiens également à remercier Yoann Bunz et Damien Combrisson, chargés de mission dans le parc national des Écrins, pour nos interactions très intéressantes, et la collaboration sans laquelle ce stage n'aurait pas vu le jour.

Un grand merci à toute l'équipe BioSP pour leur accueil et leur bonne humeur tout au long de ce stage.

Merci également à Denis Poinot, responsable des stages, pour ses mails récapitulant les objectifs de chaque semaine, et à Andréas Prinzing, mon référent universitaire, d'avoir veillé au bon déroulement de celui-ci.

Ce stage a été financé par le réseau CiSStats soutenu par le département MATHNUM de l'INRAE.

Table des matières

1	Introduction	1
2	Matériels et Méthodes	3
2.1	Données	3
2.1.1	Zone d'étude	3
2.1.2	Suivi opportuniste des papillons	3
2.1.3	Données environnementales	3
2.2	Modèle spatio-temporel	3
2.2.1	Modèle mécaniste pour la dynamique d'émergence des papillons	4
2.2.2	Modélisation de l'abondance	5
2.2.3	Processus d'observation	5
2.2.4	Estimation des paramètres	5
3	Résultats	7
3.1	Estimation de l'effort d'échantillonnage	7
3.2	Dynamique d'émergence des papillons	8
3.3	Date de début d'éclosion par année	9
3.4	Durée de la période d'éclosion	10
3.5	Effet des variables environnementales sur la distribution des espèces	10
4	Discussion	12
4.1	Modélisation de l'effort d'échantillonnage	12
4.1.1	La détectabilité des espèces	12
4.2	Modèle mécaniste pour la dynamique d'émergence des papillons	13
4.2.1	La dynamique d'abondance	13
4.2.2	La date de première émergence et la durée de la période d'éclosion	13
4.2.3	Rôle de la température dans l'émergence des papillons	14
4.3	Variables environnementales pour la répartition des espèces	14
4.4	Apport du modèle pour la gestion et la conservation des espèces	15
	Références	16

1 Introduction

L'évaluation des dynamiques de biodiversité permet d'identifier les besoins et les méthodes nécessaires pour la conservation et la gestion des espaces naturels protégés. En particulier, les mesures de conservation prises par les gestionnaires d'espaces naturels reposent souvent sur l'étude de la répartition des espèces (Knight *et al.*, 2008; Sutherland *et al.*, 2004; Van Strien *et al.*, 2013). Les modèles de distributions d'espèces (SDM), établissent un lien entre la présence d'une espèce et des prédicteurs environnementaux (Elith *et al.*, 2006; Norris, 2004; Thomas *et al.*, 2004). Cette information sur l'endroit où se trouve une espèce, et la variation de son abondance au cours du temps permet la planification et l'évaluation des stratégies de gestion (Ferrier, 2002). Ceci nécessite toutefois, des données d'une grande couverture spatiale et temporelle (Rushton *et al.*, 2004).

La montée en puissance des bases de données de science participative a apporté de nouvelles possibilités pour la conservation et les recherches en écologie au cours des dernières décennies (Dickinson *et al.*, 2010; Isaac *et al.*, 2014; Powney & Isaac, 2015). Alimentées de manière opportuniste par des naturalistes professionnels et amateurs, les sciences participatives produisent une grande quantité de données (Callaghan & Gawlik, 2015; Cima *et al.*, 2020). Ces données signalent souvent la présence des espèces en un point de l'espace et du temps, sans aucune information sur les absences (Liu *et al.*, 2016; Pearce & Boyce, 2006). Les données de présence seule sont dites opportunistes, lorsqu'il n'y a pas de protocole guidant le processus d'échantillonnage (Kery *et al.*, 2010). Elles présentent des avantages : elles sont peu coûteuses, faciles à collecter, et offrent un suivi spatio-temporel souvent inégalable par des données issues de suivis protocolés (Schmeller *et al.*, 2009). Pour répondre aux objectifs des politiques publiques environnementales, les gestionnaires souhaitent utiliser ces données opportunistes pour la prise de décisions d'actions de conservation à grande échelle géographique et temporelle (Horns *et al.*, 2018).

La principale difficulté dans l'analyse d'un tel type de données provient d'une caractérisation partielle voire nulle de la distribution des observateurs et donc de l'effort d'échantillonnage (Dobson *et al.*, 2020; Henckel *et al.*, 2020). L'effort d'échantillonnage mesure le nombre et la durée des visites durant lesquelles les observateurs peuvent signaler l'occurrence d'un individu à un point et un temps donné (Dickinson *et al.*, 2010). Ainsi, un nombre élevé d'individus observés appartenant à une espèce peut être lié à une forte abondance de celle-ci ou à la présence de nombreux observateurs (Giraud *et al.*, 2016). Une corrélation entre les facteurs environnementaux et la variation spatiale de l'effort d'échantillonnage peut donc entraîner des biais dans les modèles de distribution d'espèces (Beck *et al.*, 2014; Boakes *et al.*, 2010). Toute approche de modélisation doit utiliser ces données avec précaution pour éliminer le biais d'observation causé par un effort d'échantillonnage hétérogène (Phillips *et al.*, 2009).

Plusieurs techniques de modélisation ont été proposées pour tenir compte de l'incertitude du processus d'échantillonnage dans les SDM à partir d'occurrences opportunistes (De Solan *et al.*, 2019; Phillips *et al.*, 2009; Warton *et al.*, 2013). Parmi elles, la modélisation par des processus ponctuels est particulièrement bien adaptée à l'analyse de données de présence seule (Aarts *et al.*, 2012; Renner *et al.*, 2015; Warton & Shepherd, 2010). L'intensité du processus modélisé, correspond à un nombre moyen d'individus présents en un point de l'espace et du temps (Aarts *et al.*, 2012). L'intensité n'est pas une probabilité mais une mesure d'abondance, ce qui constitue un avantage pour l'interprétation et la mise en œuvre du modèle (Dorazio, 2012). Botella *et al.* (2021), proposent un nouveau SDM multi-espèces pour des données opportunistes de présence seule, ne nécessitant pas de connaissance préalable sur la

variation spatiale de l'effort d'échantillonnage. La plupart des approches, y compris celle de Botella *et al.* (2021) décrivent la distribution des espèces de façon spatiale mais pas spatio-temporelle (Elith & Leathwick, 2009). Dans le but de planifier des stratégies de gestion et d'évaluer la réalisation des objectifs de conservation dans le temps, il est nécessaire de concevoir des modèles incluant la variation temporelle de l'abondance des espèces (Hodgson *et al.*, 2011; Soulsby & Thomas, 2012).

L'objectif de cette étude est de concevoir une estimation conjointe de l'effort d'échantillonnage et de la dynamique spatio-temporelle de la distribution d'espèces à partir de données opportunistes. La méthode utilisée pour dissocier l'abondance de l'effort d'échantillonnage, repose sur la modélisation de l'occurrence de plusieurs espèces en tant que processus ponctuels de Poisson inhomogène dont l'intensité est le produit de l'effort d'échantillonnage et de la distribution des espèces d'intérêt le long de gradients environnementaux. La dynamique temporelle au cours d'une année donnée est modélisée de façon mécaniste. L'équation utilisée est conçue pour le suivi de l'abondance des populations d'insectes. En plus de l'abondance totale, elle dépend de trois paramètres : la date de première émergence, la durée de la période d'éclosion et l'espérance de vie. Nous appliquons le modèle à un jeu de données opportunistes d'occurrence de lépidoptères univoltins provenant du parc national des Écrins. Plus particulièrement, nous nous intéressons à la variation du nombre d'individus adultes observés sur un site au cours de la période de vol et à l'évolution au cours des années des dates de première émergence du stade adulte.

2 Matériels et Méthodes

2.1 Données

2.1.1 Zone d'étude

La zone d'étude correspond à l'ensemble du parc national des Écrins (44°51'21"N, 6°15'49"E). Le parc s'étend sur 925 km², couvrant une partie du massif des Écrins dans les Alpes. Le domaine couvrant le parc national des Écrins a été découpé en un ensemble de 196 cellules dans lesquelles seront réalisées les comptages d'occurrence (voir partie 2.2.3).

2.1.2 Suivi opportuniste des papillons

Les données utilisées ont été récoltées de manière opportuniste par des agents du parc. Il s'agit de données d'occurrences géolocalisées pour plusieurs espèces de lépidoptères. Les occurrences des espèces ont été relevées de 1957 à aujourd'hui. Nous restreignons l'étude à la période entre 2009 et 2020, correspondant à l'arrivée du numérique pour les récoltes opportunistes. Le jeu de données contient 50 439 lignes, correspondant au nombre total d'occurrences relevées. Ce sont des données de présence seule, qui consistent en un ensemble de localisation $x = \{x_1, x_2, x_3, \dots, x_m\}$ où une espèce a été observée. Les dates de collecte ont été converties en nombre de jours après le 31 décembre (jours juliens). Au total, 1411 espèces sont observées avec en moyenne 117 observations par espèce pour les plus abondantes (980 espèces sont vues moins de 10 fois). Pour 5 espèces informatives quant aux besoins et aux conséquences des mesures de gestion pour les papillons univoltins diurnes patrimoniaux, nous modélisons la dynamique d'émergence. Les espèces d'intérêt sont les suivantes : le Citron (*Gonepteryx rhamni*), l'Aurore (*Anthocharis cardamines*), le Gazé (*Aporia crataegi*), le Demi-deuil (*Melanargia galathea*) et l'Apollon (*Parnassius apollo*).

2.1.3 Données environnementales

L'altitude et l'occupation du sol ont été utilisées comme variables environnementales pour modéliser l'abondance relative des espèces. L'altitude moyenne de chaque cellule du domaine spatial a été extraite à partir du modèle numérique de terrain (MNT) du parc national des Écrins. Les types d'occupation du sol ont été récupérés à partir de la carte d'occupation des sols du CES OSO et sont représentés par des proportions d'occupation au sein d'une cellule donnée. Seule la classe d'habitat correspondant aux couverts forestiers a été retenue dans le modèle, car elle influence fortement l'abondance des espèces de papillons choisies (Swaay *et al.*, 2006).

2.2 Modèle spatio-temporel

Le modèle estime conjointement l'effort d'échantillonnage et la dynamique spatio-temporelle de l'émergence de papillons. La zone d'étude est divisée en cellules spatio-temporelles dans lesquelles le nombre d'occurrences à un endroit et pour un pas de temps considéré est rapporté. Ce nombre d'occurrences suit une loi de poisson inhomogène dépendant d'une fonction d'intensité θ , variant dans l'espace et dans le temps. L'intensité θ est modélisée comme le produit de la dynamique d'abondance à un temps donné

et de l'effort d'échantillonnage. Nous avons utilisé une méthode bayésienne pour l'estimation des paramètres d'intérêts (figure 2). Un schéma simplifié de la structure du modèle est représenté ci-dessous (figure 1).

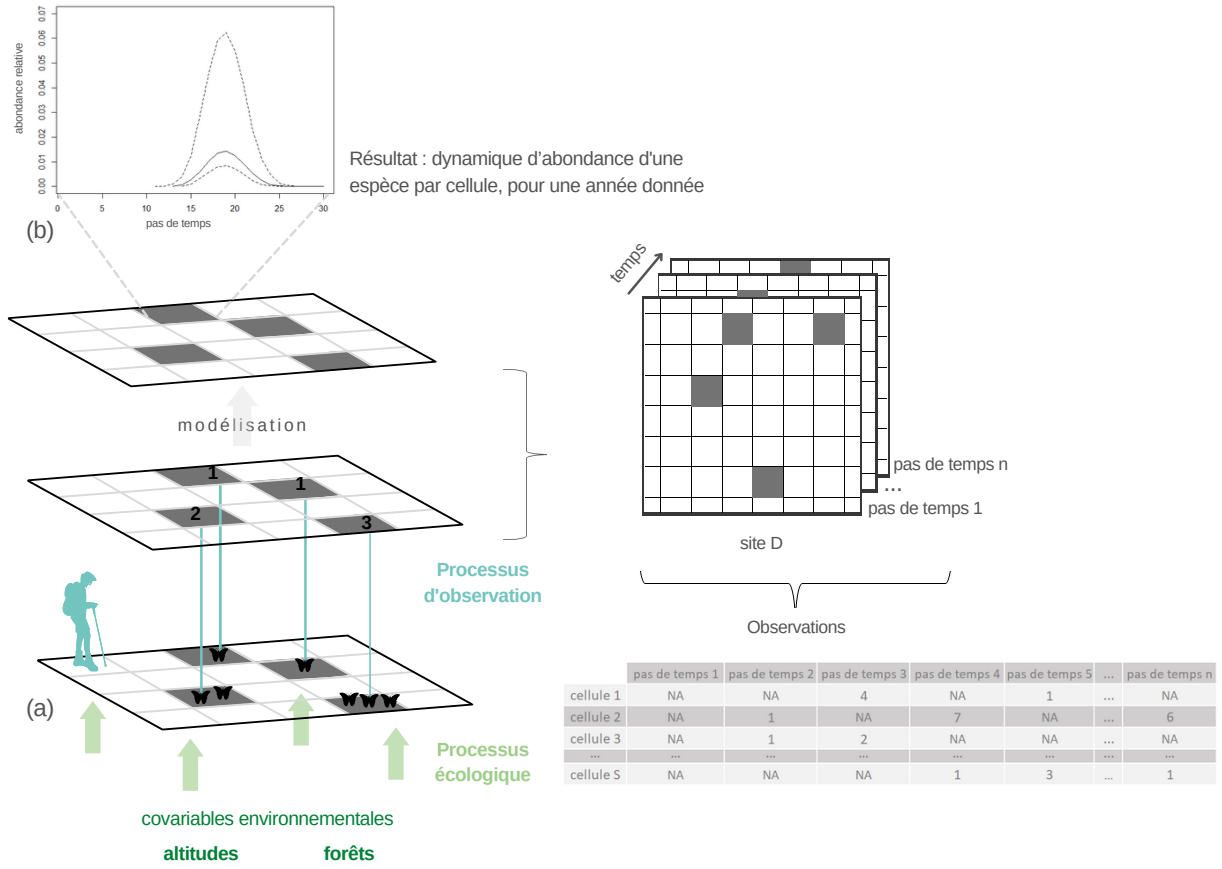


Figure 1 – Structure et sorties du modèle. (a) Distribution spatiale d'espèces en fonction de covariables environnementales. (b) Modélisation de l'abondance d'adultes en vol dans la cellule z , pour plusieurs fenêtres temporelles dans une année y .

2.2.1 Modèle mécaniste pour la dynamique d'émergence des papillons

Les variations du nombre d'adultes en vol $n_{i,z,y}(t)$ pour l'espèce i , sur le site z , au cours de l'année y peuvent être modélisées par l'équation différentielle suivante (Soulsby & Thomas, 2012) :

$$\frac{dn_{i,z,y}(t)}{dt} = \frac{N_{i,z,y}}{T_{E_{i,y}}} E\left(\frac{t'_{i,y}}{T_{E_{i,y}}}\right) - \frac{N_{i,z,y}}{T_i} \quad (1)$$

où $t'_{i,y} = t - t_{0_{i,y}}$ est le temps écoulé depuis le début de l'éclosion et $E(t'_{i,y}/T_{E_{i,y}})/T_{E_{i,y}}$ est une fonction d'éclosion dont l'intégrale sur la période d'éclosion $0 \leq t'_{i,y} \leq T_{E_{i,y}}$ vaut 1. L'équation (1) est fonction de quatre paramètres : $N_{i,z,y} = \int_t n_{i,z,y,t} dt$ est la population totale détectable, $t_{0_{i,y}}$ la date du début de l'éclosion, $T_{E_{i,y}}$ la durée de la période d'éclosion et T_i l'espérance de vie moyenne. Cette équation admet la solution analytique suivante :

$$\begin{cases} n_{i,z,y}(t) = 0 & \text{pour } t'_{i,y} < 0 \\ n_{i,z,y}(t) = \frac{3N_{i,z,y}}{4(a^2+9)} [a \sin^3 X - 3 \sin^2 X \cos X + \frac{6}{a^2+1} (a \sin X - \cos X + e^{-aX})] & \text{pour } 0 \leq t'_{i,y} \leq T_E \\ n_{i,z,y}(t) = \frac{9N_{i,z,y} e^{-aX} (1+e^{-a\pi})}{2(a^2+9)(a^2+1)} & \text{pour } t'_{i,y} > T_{E_{i,y}} \end{cases}$$

où $X = \pi t' / T_{E_{i,y}}$ et $a = T_{E_{i,y}} / \pi T_i$. L'abondance d'adultes en vol dans la cellule spatio-temporelle $(z, \delta t)$, l'année y , $\lambda_{i,t,z,y}$, est donc :

$$\lambda_{i,t,z,y} = \int_{\delta t} n_{i,z,y}(\tau) d\tau \quad (2)$$

2.2.2 Modélisation de l'abondance

L'abondance des espèces $N_{i,z,y}$ représente une distribution spatiale d'espèces le long de gradients environnementaux. Elle s'écrit comme une fonction log-linéaire des variables environnementales :

$$N_{i,z,y} = \exp(\beta_{0_{i,y}} + \beta_{1_i} * altitude(z) + \beta_{2_i} * foret(z)) \quad (3)$$

où les paramètres β dépendent de l'espèce, et de l'année pour l'ordonnée à l'origine. Les variables environnementales *altitude* et *foret* dépendent du site z . Il est impossible d'estimer l'abondance absolue des espèces à partir de données de présence uniquement. Nous estimons par conséquent une abondance relative dans l'espace.

2.2.3 Processus d'observation

L'intégralité du parc national des Écrins est définie comme un domaine spatial D bidimensionnel. Ce domaine est découpé en cellules en intersectant D avec une grille régulière de pas 4000 m, résultant en un ensemble de 196 cellules (logiciel R studio version 4.0.4, package sf, Pebesma (2018)). Pour chaque cellule z , le nombre d'occurrences de l'espèce d'intérêt i au pas de temps t , l'année y est modélisée comme un processus ponctuel de Poisson inhomogène :

$$Y_{(i,t,z,y)} \sim \text{Poisson}(\lambda_{i,t,z,y} s_z) \quad (4)$$

où s_z est l'effort d'échantillonnage de la cellule z et $\lambda_{i,t,z,y}$ l'abondance d'adultes en vol sur le site z , au pas de temps t , l'année y . Cette abondance est donnée par l'équation (2). L'effort d'échantillonnage correspond à la probabilité que la présence d'au moins une espèce soit reportée sachant qu'elle est réellement présente sur le site. D'après les experts du parc, l'effort d'échantillonnage est hétérogène dans l'espace, mais constant dans le temps, sur une année. Pour une première approximation, nous le considérons constant entre années également.

2.2.4 Estimation des paramètres

L'inférence des paramètres a été réalisée dans un cadre bayésien à l'aide du package Nimble du logiciel R (de Valpine *et al.*, 2021). Les paramètres t_0 , T_E et T suivent *a priori* des lois log-normales dont les

paramètres sont 1, 3, 2 pour les moyennes et 4, 5, 4 pour les précisions, respectivement. Les paramètres β_0 , β_1 et β_2 suivent *a priori* des lois normales centrées de variance 1000. Toutes les lois *a priori* choisies sont non informatives. Nous avons lancé 3 chaînes de Markov Monte Carlo (MCMC) de 30 000 itérations avec une période de chauffe de 20 000 itérations. Une itération sur 20 a été conservée sur les 10 000 itérations restantes pour construire les distributions *a posteriori* des paramètres. Ce paramétrage amène à un bon mélange et une bonne convergence des chaînes comme attesté par les valeurs du diagnostic de Gelman et Rubin (\hat{R}) proches de 1 (Brooks & Gelman, 1998; Gelman & Rubin, 1992).

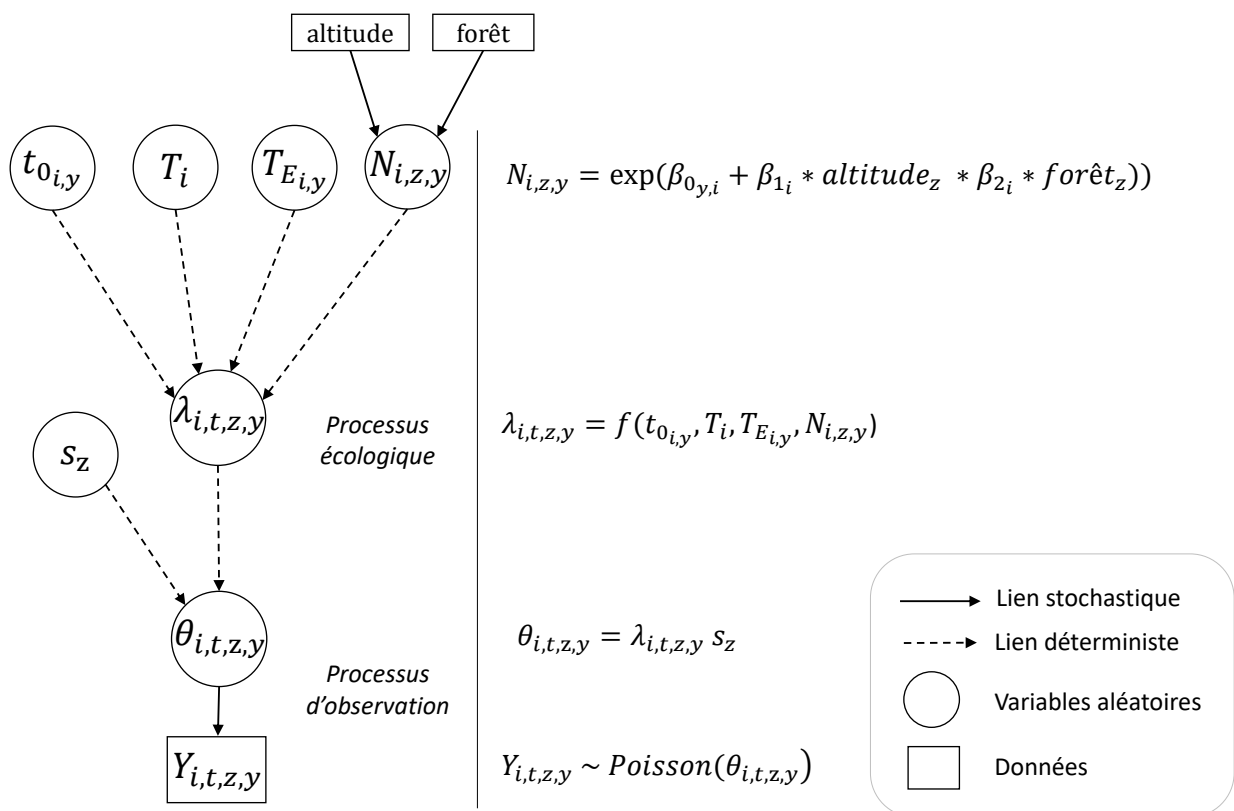


Figure 2 – Réseau bayésien correspondant au modèle

3 Résultats

3.1 Estimation de l'effort d'échantillonnage

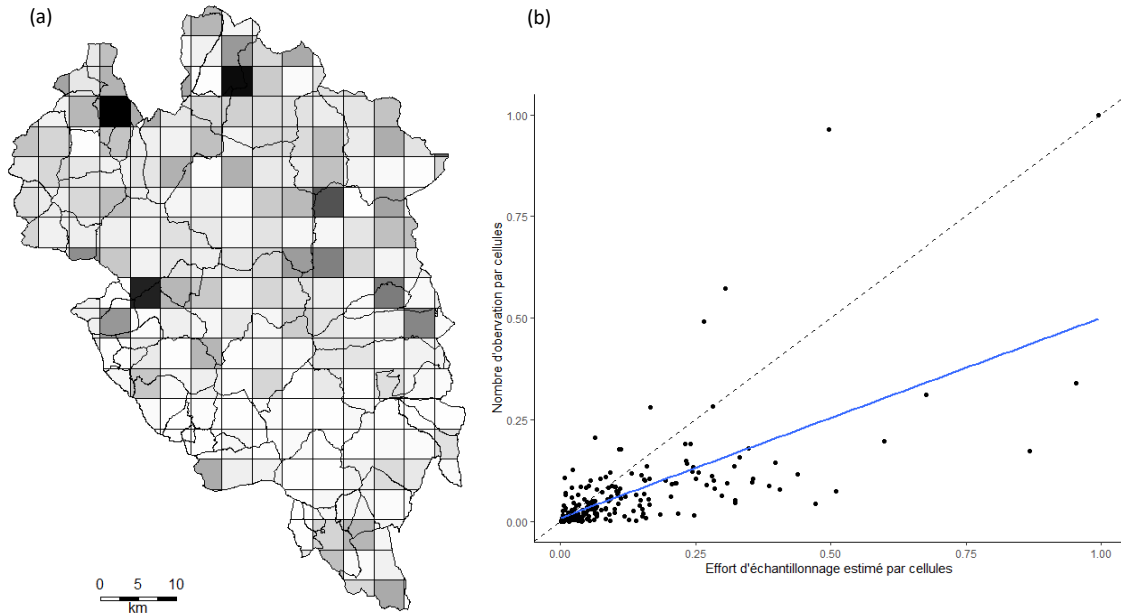


Figure 3 – (a) Carte de la répartition potentielle de l'effort d'échantillonnage dans le parc national des Écrins. Le gradient clair à foncé indique une probabilité d'observer une des espèces de papillon faible à élevé. (b) Nombre d'observations réels par cellules en fonction de l'effort d'échantillonnage estimé. La droite de régression est représentée en bleue et la première bissectrice en pointillés. Le nombre d'observations a été ramené sur une échelle comprise entre 0 et 1 pour tracer la première bissectrice.

L'estimation de l'effort d'échantillonnage est hétérogène dans l'espace avec un gradient nord-sud perceptible (figure 3a). Le résultat du test correspondant à la droite de régression figure 3b est significatif ($r = 0,42$; 194ddl; $P < 2,2e-16$). L'effort estimé augmente lorsque le nombre d'observations réel dans une cellule donnée est élevé. Cependant, le modèle estime un effort d'échantillonnage réel plus élevé dans certaines cellules ayant un faible nombre d'observations.

3.2 Dynamique d'émergence des papillons

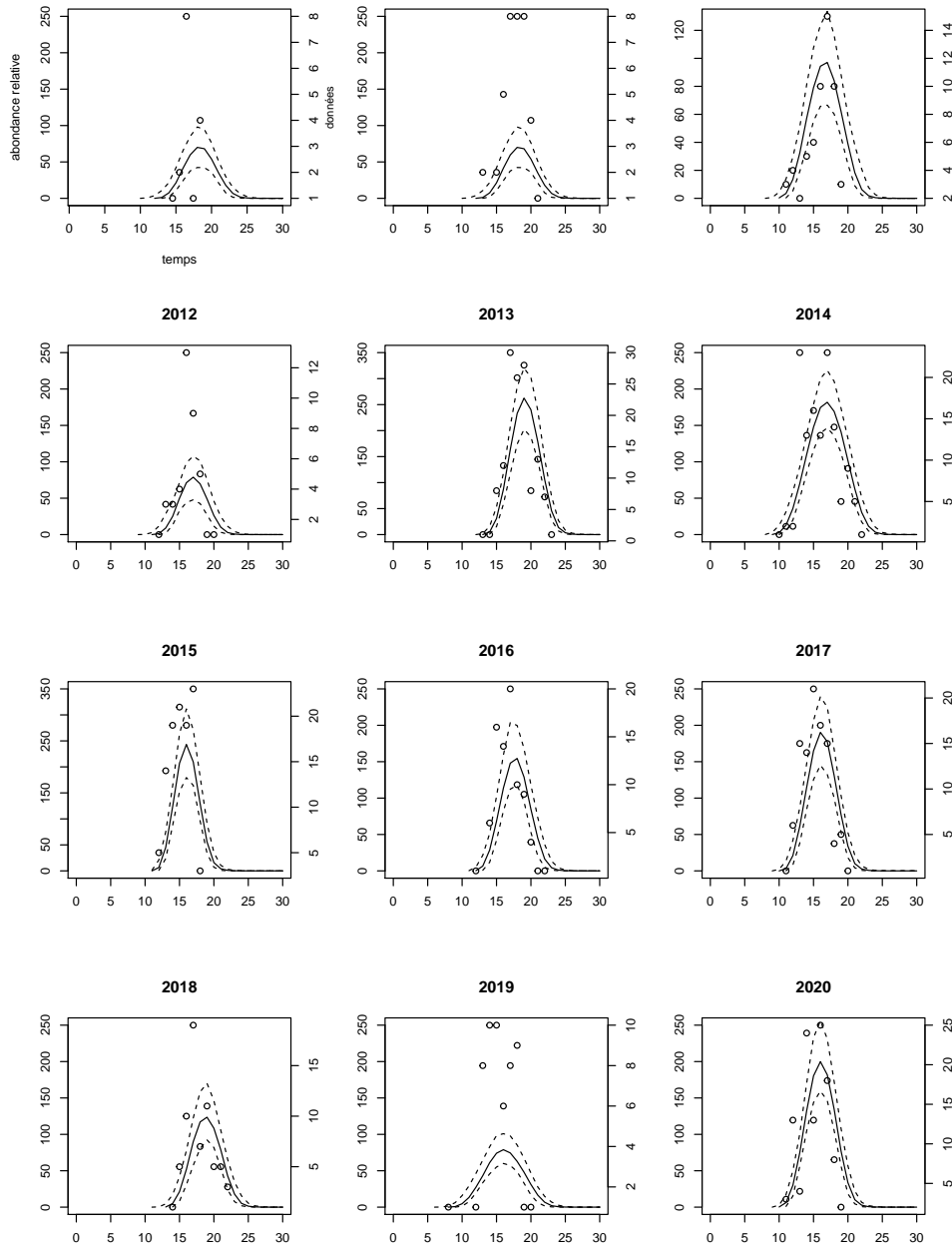


Figure 4 – Estimations de la variation d'abondance de l'Apollon (*Parnassius apollo*) en fonction du temps, sur l'ensemble du parc des Écrins et pour 12 années. Une année est divisée en 30 pas de temps, chacun correspondant à environ 13 jours. Les courbes pleines représentent les médianes et les courbes en pointillés les quantiles a posteriori à 97,5% et 2,5%. Les points sont un ensemble d'observations comptabilisé par pas de temps, pour chaque année.

La figure 4 montre les dynamiques d'émergence sur l'ensemble du parc pour chaque année, ainsi que les comptages d'occurrences pour l'espèce *Parnassius apollo* (Apollon). Globalement les dynamiques estimées d'émergence semblent cohérentes. Les variations d'abondance entre années sont aussi cohérentes avec les observations.

3.3 Date de début d'éclosion par année

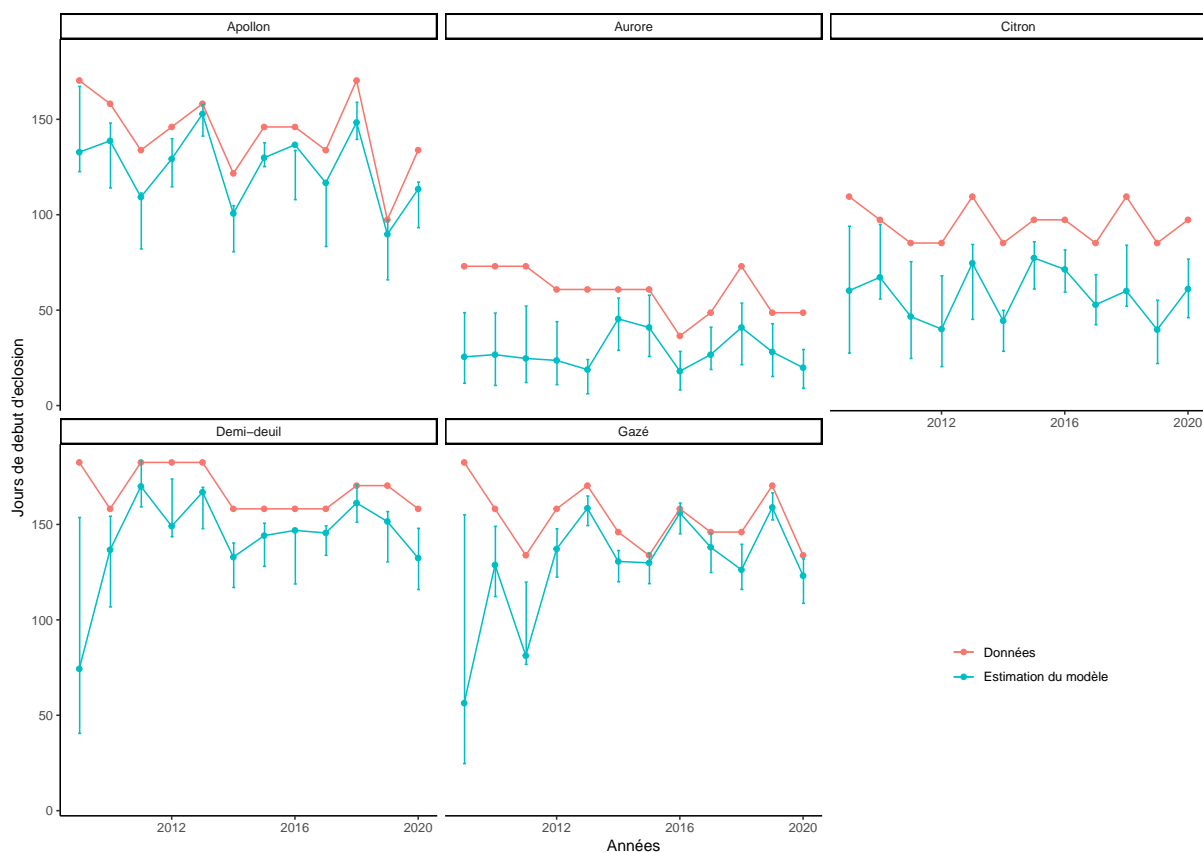


Figure 5 – Estimation des jours de début d'éclosion t_0 en fonction des années, pour 5 espèces de lépidoptères dans le parc national des Écrins. En bleu, sont représentés les t_0 estimés et en rouge les premiers jours d'observation d'un papillon par année extraits des données. Les barres d'erreurs indiquent l'incertitude de l'estimation avec les quantiles *a posteriori* à 97,5% et 2,5%.

Pour les 5 espèces, le modèle anticipe les dates de première émergence t_0 par rapport aux premiers jours d'observation d'un papillon dans l'année extraits directement à partir des données (figure 5). Pour le Gazé (*A. crataegi*), l'Apollon (*P. apollo*) et le Demi-deuil (*M. galathea*), la majorité des t_0 estimés ont lieu entre 100 et 150 jours, c'est-à-dire entre fin mars et fin mai. Pour l'Aurore (*A. cardamines*) et le Citron (*G. rhamni*), le paramètre t_0 est estimé dans les 20 à 70 premiers jours de l'année, soit entre mi-janvier et mi-mars. La courbe d'estimation du t_0 pour le Gazé, est probablement celle qui se rapproche le plus des données. À l'inverse de celle de l'Aurore et du Citron dont les estimations du t_0 sont au moins 50 jours plus tôt que les premières observations issues des données. Pour les cinq espèces, la tendance d'estimation du t_0 suit globalement celle des données. Les estimations du t_0 pour le Gazé (*A. crataegi*) et le Demi-deuil (*M. galathea*) en 2009 se distinguent des autres, avec des dates très précoces, éloignées de la date de première observation extraite des données et une plus grande distance interquantile.

3.4 Durée de la période d'éclosion

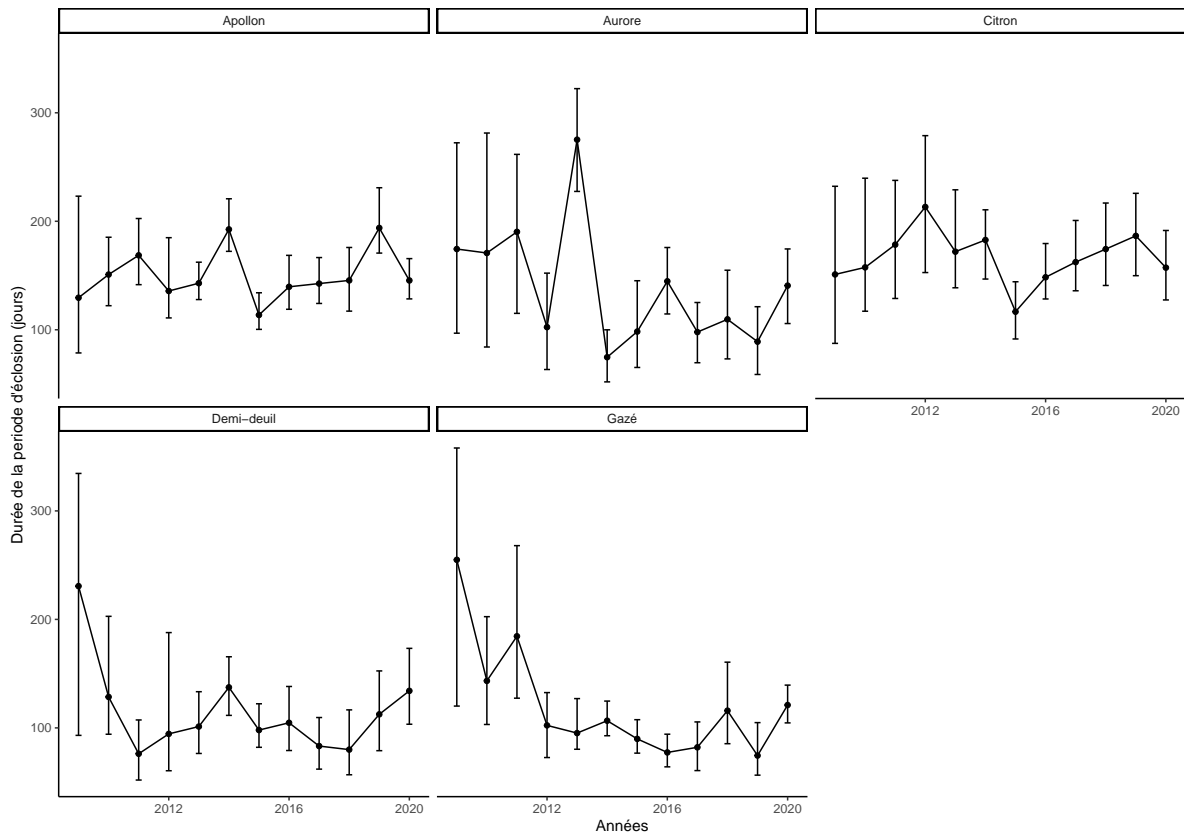


Figure 6 – Estimation de la durée de la période d'éclosion T_E en fonction des années, pour 5 espèces de lépidoptères, dans le parc national des Écrins. Les barres d'erreurs indiquent l'incertitude de l'estimation avec les quantiles a posteriori à 97,5% et 2,5%.

Pour les 5 espèces confondues, le modèle estime les périodes d'éclosions T_E (i.e. la période où de nouveaux individus apparaissent), allant d'une durée de 80 à 200 jours (figure 6). Le Citron (*G. rhamni*) et l'Apollon (*P. apollo*) ont une variation relativement similaire de la durée de la période d'éclosion en fonction des années, avec des T_E estimés entre 100 et 200 jours. L'estimation du T_E pour le Gazé (*A. crataegi*) et le Demi-deuil (*M. galathea*) en 2009, se distingue des autres, avec une plus grande incertitude. L'Aurore (*A. cardamines*) en 2014 présente également un T_E très élevé, supérieur à 250 jours.

3.5 Effet des variables environnementales sur la distribution des espèces

Tableau 1 – Probabilité *a posteriori* que l'effet de l'altitude sur l'espèce en ligne soit supérieure à celui sur l'espèce en colonne.

	Aurore	Citron	Gazé	Apollon	Demi-deuil
Aurore	/	0.99	1	1	0.16
Citron	0.006	/	0.83	1	0
Gazé	0	0.17	/	0.99	0
Apollon	0	0	0.007	/	0
Demi-deuil	0.83	1	1	1	/

Tableau 2 – Probabilité *a posteriori* que l'effet de la forêt sur l'espèce en ligne soit supérieure à celui sur l'espèce en colonne.

	Aurore	Citron	Gazé	Apollon	Demi-deuil
Aurore	/	0.19	0.01	0	0
Citron	0.81	/	0.10	0	0.02
Gazé	0.98	0.89	/	0.04	0.26
Apollon	1	1	0.95	/	0.82
Demi-deuil	1	0.98	0.74	0.18	/

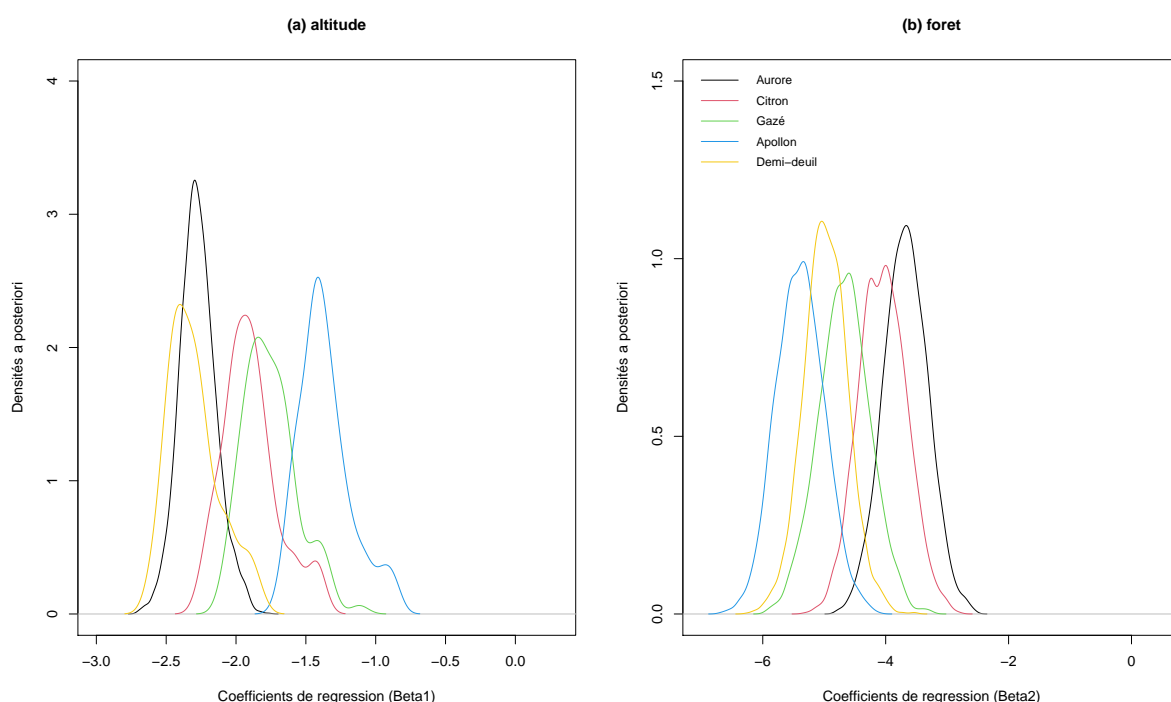


Figure 7 – Densités a posteriori des coefficients de régression β_1 et β_2 , correspondant respectivement à la régression entre la présence des espèces et les variables environnementales : altitude ou forêt. Pour 5 espèces de lépidoptères dans le parc national des Écrins.

Les densités a posteriori des paramètres β_1 et β_2 indiquent dans quelle mesure la présence des espèces varie avec la variation des valeurs des variables environnementales (figure 7). De manière générale, la présence des 5 espèces de papillons diminue lorsque l'altitude et la forêt augmentent (figure 7). Les probabilités a posteriori calculées dans le tableau 1 et le tableau 2 permettent de comparer l'effet des variables environnementales sur l'abondance des espèces. Le Demi-deuil et l'Aurore sont les plus impactés lorsque l'altitude augmente. Elles sont suivies du Citron, et du Gazé, puis de l'Apollon qui semble être l'espèce la moins impactée par l'altitude. En revanche, lorsqu'on regarde l'impact de forêt sur l'Apollon, il présente l'abondance la plus faible. Pour les 5 espèces, une proportion élevée de forêt au sein d'une cellule diminue leur abondance. Les deux espèces qui semblent moins impactées par la présence de forêt sont l'Aurore et le Citron.

4 Discussion

Nous avons développé un modèle spatio-temporel pour analyser les émergences de lépidoptères univoltins à l'échelle du parc national des Écrins à partir de données opportunistes. Le modèle prend en compte à la fois l'hétérogénéité de l'effort d'échantillonnage et la dynamique intra et inter-annuelle des papillons le long de gradients environnementaux. Ci-dessous nous discutons point par point les hypothèses faites et donnons des perspectives d'amélioration du modèle.

4.1 Modélisation de l'effort d'échantillonnage

L'effort d'échantillonnage estimé est spatialement hétérogène et la probabilité d'observer une des espèces de papillon augmente lorsque le nombre d'observations réelles dans une cellule augmente (figure 3). Ces résultats supposent une estimation satisfaisante bien qu'il soit difficile de la valider. La variation spatiale de l'effort dépend directement du nombre, de la répartition des observateurs sur le terrain, et de l'accessibilité des secteurs du parc (Giraud *et al.*, 2016). La condition d'un effort constant dans le temps est cruciale pour démêler l'abondance des espèces de l'effort d'échantillonnage (Botella *et al.*, 2021). Nous avons restreint l'analyse aux années 2009-2020, correspondant à l'arrivée du numérique pour le recueil des données opportunistes. D'après les experts du parc, l'effort d'échantillonnage est constant durant cette plage d'années dans le parc des Écrins. Les expériences de simulation de Botella *et al.* (2021) révèlent un autre biais : lorsqu'une caractéristique environnementale influence la densité des espèces de la même façon que le véritable effort d'échantillonnage. Pour éviter ce biais, nous pourrions prendre un autre support pour estimer l'effort que la grille utilisée pour l'abondance. Nous pourrions remplacer la grille par les délimitations des secteurs du parc par exemple. Ce changement de résolution aiderait à distinguer l'effort de l'abondance.

4.1.1 La détectabilité des espèces

Dans le modèle, l'effort d'échantillonnage influence de manière similaire l'observation de plusieurs espèces, sans prendre en compte une détectabilité hétérogène entre espèces. Cette hypothèse facilite la mise en place d'une première version du modèle, mais n'est pas satisfaisante (Chen *et al.*, 2013). Plusieurs facteurs influencent la détectabilité d'une espèce (Guillera-Arroita, 2017). Dans un premier temps, les caractéristiques de l'espèce, comme son apparence physique ou son comportement. Par exemple, la taille et la couleur des papillons jouent un rôle important dans sa détectabilité. Ensuite, les caractéristiques du site et les types d'occupations du sol. Un papillon sera plus difficilement détectable dans une végétation dense par exemple. La détectabilité varie également en fonction de variations temporelles comme les conditions météorologiques, la saison et l'heure de la journée. Enfin, la détectabilité dépend de la méthode d'échantillonnage. Dans le but d'améliorer la précision de l'effort d'échantillonnage, il faudrait intégrer un terme de détectabilité en considérant qu'un individu à un point z est rapporté avec une probabilité $R_{iS}(z)$ où R_i est la probabilité de détection de l'espèce i (Botella *et al.*, 2021).

4.2 Modèle mécaniste pour la dynamique d'émergence des papillons

4.2.1 La dynamique d'abondance

Nous prenons l'exemple de l'Apollon (*Parnassius apollo*), pour discuter de l'ajustement des données aux courbes d'émergence (figure 4). La forme de la courbe d'abondance est cohérente et concorde plutôt bien avec les données d'observation. L'utilisation de métriques d'ajustement permettrait de quantifier la capacité du modèle à reproduire ces dynamiques (Soulsby & Thomas, 2012).

4.2.2 La date de première émergence et la durée de la période d'éclosion

Les valeurs de t_0 obtenues permettent de séparer les espèces en deux groupes. Nous avons d'un côté le Gazé (*A. crataegi*), l'Apollon (*P. apollo*) et le Demi-deuil (*M. galathea*) dont la majorité des t_0 sont estimés entre fin-mars et fin-mai (figure 5). Les dates de première émergence estimées sont légèrement précoces car ce sont des espèces généralement visibles de mai à août (Lafranchis & Jutzeler, 2014). De l'autre côté, nous avons l'Aurore (*A. cardamines*) et le Citron (*G. rhamni*), qui ont des t_0 estimés entre mi-janvier et mi-mars. Le Citron est parfois observé en janvier ou en février, notamment parce qu'il s'agit d'une espèce très longévive à l'état imago et donc en réalité visible toute l'année (Gutiérrez & Thomas, 2000). L'Aurore est aussi une espèce précoce pouvant généralement être observée tôt dans l'année (Davies, 2019). D'après les experts du parc, les résultats sont relativement précoces dans l'année, mais leurs tendances sont cohérentes. Il est normal que les t_0 soient légèrement inférieurs par rapport aux dates de première observation d'une espèce dans une année. Ce résultat est à l'origine du système d'équations utilisé pour modéliser la dynamique d'abondance $\lambda_{i,t,z,y}$. En effet, nous imposons une courbe d'abondance nulle avant t_0 (Soulsby & Thomas, 2012). Par conséquent, la date réelle d'éclosion du premier adulte est quelques jours après le t_0 estimé. Néanmoins, les dates de premières éclosions sont certainement trop anticipées.

Concernant la durée de la période d'éclosion T_E , nous obtenons des valeurs trop élevées, s'étendant de 60 à 200 jours (figure 6). Le T_E correspond à la période où de nouveaux individus éclosent et non pas à la période de vol (i.e. le nombre total de jours de présence des adultes). La durée de la période d'éclosion se situe généralement de 25 à 45 jours (Komonen *et al.*, 2004). Prenons comme exemple l'Apollon, le modèle estime sa période d'éclosion entre 100 et 200 jours. La période de vol de l'Apollon est d'une trentaine de jours d'après Bubová *et al.* (2016). Les résultats ne sont pas cohérents avec la réalité, la durée d'éclosion estimée est identique, voire plus grande que la période de vol d'un individu. Pour prendre en compte ces connaissances, les lois a priori sur les paramètres T_E doivent être modifiées.

Plusieurs facteurs peuvent expliquer les incohérences du modèle. L'estimation des paramètres est bien plus complexe que dans l'article de Soulsby & Thomas (2012), et dépend notamment de l'effort d'échantillonnage, lui-même estimé. En particulier, il est possible que la date de première émergence et la durée de la période d'éclosion soient parfois confondues. Si le T_E estimé est grand, nous obtenons un t_0 petit et à l'inverse avec un T_E petit, nous obtenons un grand t_0 et une courbe plus piquée. Une étude de simulation permettrait de mieux caractériser les problèmes d'identifiabilité des paramètres. De la même façon que pour T_E , la durée de vie T pourrait être fixée et non estimée en se référant à la littérature (Beck & Fiedler, 2008). Enfin, une autre piste d'amélioration serait d'obtenir des données de présence-absence recueillies dans le cadre de relevés systématiques et planifiés (Phillips *et al.*, 2009). Réussir à créer un

modèle mixte avec l'intégration, en plus de nos données, de vraies présences/absences aiderait à calibrer l'effort et estimer de façon plus précise les paramètres t_0 et T_E (Fithian *et al.*, 2015).

4.2.3 Rôle de la température dans l'émergence des papillons

Pour interpréter les tendances quant à une avancée ou un recul de l'émergence des papillons au cours des années, nous avons besoin de la température (O'Connor *et al.*, 2012). En effet, il existe une forte relation entre les changements dans les premières apparitions des papillons et la température (Bale *et al.*, 2002; Roy & Sparks, 2000). Plus précisément, les degrés-jours de croissance (GDD : growing degree days), sont une mesure d'accumulation thermique qui lie directement les variations climatiques et la phénologie des papillons (Cayton *et al.*, 2015). Pour améliorer la compréhension des résultats, il est nécessaire de regarder les GDD de manière spatio-temporelle, par années et par cellules (Hodgson *et al.*, 2011). Dans cet objectif, nous avons extrait les données climat SAFRAN de 1980 à 2019 sur la zone du parc (56 mailles de 8km par 8 km) avec des températures journalières. Par la suite, nous pensions calculer les GDD à partir du 1^{er} janvier de chaque année, de sorte que chaque date d'émergence soit associée à une valeur GDD spécifique au site et à l'année. Nous voudrions intégrer ces données dans l'interprétation des résultats du modèle. Un autre facteur non-négligeable est l'influence de l'altitude sur les températures en montagne (Illán *et al.*, 2012). Les périodes d'émergence des papillons se produisant plus tard à des altitudes plus élevées, il serait pertinent de faire dépendre les paramètres du modèle des tranches d'altitudes.

4.3 Variables environnementales pour la répartition des espèces

Le résultat global de la régression log-linéaire des variables d'environnement indique une association négative des espèces avec la présence de forêts et hautes altitudes (figure 7). Gallou *et al.* (2017) ont étudié la richesse en espèces de papillons le long du gradient d'altitude (200–2700 m) dans les Alpes françaises. Ils documentent une forte diminution de l'abondance des espèces après 1900 m, ce qui suggère que seules les espèces spécialistes survivent aux fortes contraintes climatiques à haute altitude. L'absence totale des espèces en haute altitude doit constituer un poids fort dans la régression, ce qui explique que toutes les espèces ont une relation négative, même si certaines sont parfois observées en altitude. Le classement allant de la plus à la moins impactée par l'altitude et la forêt (tableau 1 et 2) est cohérent avec l'écologie des espèces considérées. L'Apollon est l'espèce la moins touchée par l'altitude et, parmi les 5 espèces étudiées, celle que l'on retrouve le plus entre 1000 et 2000 mètres d'altitude (Fred *et al.*, 2006). Le Gazé est également l'une des espèces trouvée en altitude (Jugovic *et al.*, 2017). Le Citron est placé à la position intermédiaire, c'est une espèce avec une plasticité importante, mais pas pour autant une espèce d'altitude (Stasek *et al.*, 2008). Enfin, le Demi-deuil est majoritairement retrouvé en plaine et dans les champs, d'où son association plus négative que les autres avec les hautes altitudes (Vandewoestijne *et al.*, 2004). L'Aurore est l'espèce la moins impactée par les sites à dominante boisée (Shreeve *et al.*, 2001). L'Apollon, plutôt en altitude et le Demi-deuil en prairie, sont moins enclins à être vus en forêt. Le Citron quant à lui, a une réponse plastique face aux habitats à dominante boisée. Certains papillons suivent de près l'aire de répartition de leurs plantes hôtes naturelles, c'est le cas de du Citron (*G. rhamnii*), une espèce de papillon limitée en hôtes (Gutiérrez & Thomas, 2000; Quinn *et al.*, 1998). Il aurait été intéressant

d'utiliser les plantes hôtes comme prédicteur environnemental dans le modèle (Phillimore *et al.*, 2012).

4.4 Apport du modèle pour la gestion et la conservation des espèces

Cette première version du modèle fournit une méthode d'estimation de l'effort d'échantillonnage, permettant l'utilisation de données opportunistes de présence seule, sans connaissances préalables sur la variation de l'effort. Le modèle nécessite néanmoins une grande base de données multi-espèces et un effort d'échantillonnage supposé constant dans le temps. De nombreuses pistes d'améliorations sont envisagées pour le modèle, comme l'ajout d'une probabilité de détection et l'intégration de données climatiques. Le modèle mécaniste intégré au modèle probabiliste, permet l'extraction de paramètres tels que la courbe d'abondance des adultes en vol en fonction du temps $\lambda_{i,t,z,y}$ et la date de première émergence t_0 . Le paramètre t_0 pourrait contribuer à l'identification de potentiels décalages phénologiques chez des espèces cibles. Le modèle permet également d'obtenir l'abondance relative des espèces et leur réponse à divers types d'environnement, permettant ainsi de connaître, si et comment (1) les types d'habitat agissent sur la répartition des espèces, et (2) les dynamiques d'émergences évoluent en fonction du temps.

Références

- Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecology and Evolution*, 3(1) :177–187.
- Bale, J. S., Masters, G. J., Hodkinson, I. D., Awmack, C., Bezemer, T. M., Brown, V. K., Butterfield, J., Buse, A., Coulson, J. C., Farrar, J., Good, J. E. G., Harrington, R., Hartley, S., Jones, T. H., Lindroth, R. L., Press, M. C., Symrnioudis, I., Watt, A. D., & Whittaker, J. B. (2002). Herbivory in global climate change research : direct effects of rising temperature on insect herbivores. *Global Change Biology*, 8(1) :1–16.
- Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19 :10–15.
- Beck, J. & Fiedler, K. (2008). Adult life spans of butterflies (Lepidoptera : Papilionoidea + Hesperioidea) : broadscale contingencies with adult and larval traits in multi-species comparisons. *Biological Journal of the Linnean Society*, 96(1) :166–184.
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted Views of Biodiversity : Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, 8(6) :e1000385.
- Botella, C., Joly, A., Bonnet, P., Munoz, F., & Monestiez, P. (2021). Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data. *Methods in Ecology and Evolution*, 12(5) :933–945.
- Brooks, S. P. & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4) :434–455.
- Bubová, T., Kulma, M., Vrabec, V., & Nowicki, P. (2016). Adult longevity and its relationship with conservation status in European butterflies. *Journal of Insect Conservation*, 20(6) :1021–1032.
- Callaghan, C. T. & Gawlik, D. E. (2015). Efficacy of eBird data as an aid in conservation planning and monitoring. *Journal of Field Ornithology*, 86(4) :298–304.
- Cayton, H. L., Haddad, N. M., Gross, K., Diamond, S. E., & Ries, L. (2015). Do growing degree days predict phenology across butterfly species? *Ecology*, 96(6) :1473–1479.
- Chen, G., Kéry, M., Plattner, M., Ma, K., & Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, 101(1) :183–191.
- Cima, V., Fontaine, B., Witté, I., Dupont, P., Jeanmougin, M., & Touroult, J. (2020). A test of six simple indices to display the phenology of butterflies using a large multi-source database. *Ecological Indicators*, 110 :105885.
- Davies, W. J. (2019). Multiple temperature effects on phenology and body size in wild butterflies predict a complex response to climate change. *Ecology*, 100(4) :e02612.
- De Solan, T., Renner, I., Cheylan, M., Geniez, P., & Barnagaud, J. Y. (2019). Opportunistic records reveal Mediterranean reptiles' scale-dependent responses to anthropogenic land use. *Ecography*, 42(3) :608–620.
- de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodríguez, A., Temple Lang, D., & Paganin, S. (2021). *NIMBLE : MCMC, Particle Filtering, and Programmable Hierarchical Modeling*. R package version 0.11.1.

- Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen Science as an Ecological Research Tool : Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41 :149–172.
- Dobson, A., Milner-Gulland, E., Aebischer, N., Beale, C., Brozovic, R., Coals, P., Critchlow, R., Dancer, A., Greve, M., Hinsley, A., Ibbett, H., Johnston, A., Kuiper, T., Comber, S., Mahood, S., Moore, J., Nilsen, E., Pocock, M., Quinn, A., & Keane, A. (2020). Making Messy Data Work for Conservation. *One Earth*, 2(5) :455–465.
- Dorazio, R. M. (2012). Predicting the Geographic Distribution of a Species from Presence-Only Data Subject to Detection Errors. *Biometrics*, 68(4) :1303–1312.
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2) :129–151.
- Elith, J. & Leathwick, J. R. (2009). Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1) :677–697.
- Ferrier, S. (2002). Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning : Where to from Here? *Systematic Biology*, 51(2) :331–363.
- Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models : pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6(4) :424–438.
- Fred, M. S., O'Hara, R. B., & Brommer, J. E. (2006). Consequences of the spatial configuration of resources for the distribution and dynamics of the endangered *Parnassius apollo* butterfly. *Biological Conservation*, 130(2) :183–192.
- Gallou, A., Baillet, Y., Ficetola, G. F., & Després, L. (2017). Elevational gradient and human effects on butterfly species richness in the French Alps. *Ecology and Evolution*, 7(11) :3672–3681.
- Gelman, A. & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4) :457–472.
- Giraud, C., Calenge, C., Coron, C., & Julliard, R. (2016). Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics*, 72(2) :649–658.
- Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection : advances, challenges and opportunities. *Ecography*, 40(2) :281–295.
- Gutiérrez, D. & Thomas, C. D. (2000). Marginal range expansion in a host-limited butterfly species *Gonepteryx rhamni* : Range expansion in a butterfly. *Ecological Entomology*, 25(2) :165–170.
- Henckel, L., Bradter, U., Jönsson, M., Isaac, N. J. B., & Snäll, T. (2020). Assessing the usefulness of citizen science data for habitat suitability modelling : Opportunistic reporting versus sampling based on a systematic protocol. *Diversity and Distributions*, 26(10) :1276–1290.
- Hodgson, J. A., Thomas, C. D., Oliver, T. H., Anderson, B. J., Brereton, T. M., & Crone, E. E. (2011). Predicting insect phenology across space and time. *Global Change Biology*, 17(3) :1289–1300.
- Horns, J. J., Adler, F. R., & Sekercioglu, C. H. (2018). Using opportunistic citizen science data to estimate avian population trends. *Biological Conservation*, 221 :151–159.

- Illán, J. G., Gutiérrez, D., Díez, S. B., & Wilson, R. J. (2012). Elevational trends in butterfly phenology : implications for species responses to climate change. *Ecological Entomology*, 37(2) :134–144.
- Isaac, N. J. B., Van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science : extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10) :1052–1060.
- Jugovic, J., Grando, M., & Genov, T. (2017). Microhabitat selection of *Aporia crataegi* (Lepidoptera : Pieridae) larvae in a traditionally managed landscape. *Journal of Insect Conservation*, 21(2) :307–318.
- Kery, M., Royle, J. A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., & Zbinden, N. (2010). Site-Occupancy Distribution Modeling to Correct Population-Trend Estimates Derived from Opportunistic Observations. *Conservation Biology*, 24(5) :1388–1397.
- Knight, A. T., Cowling, R. M., Rouget, M., Balmford, A., Lombard, A. T., & Campbell, B. M. (2008). Knowing But Not Doing : Selecting Priority Conservation Areas and the Research–Implementation Gap. *Conservation Biology*, 22(3) :610–617.
- Komonen, A., Grapputo, A., Kaitala, V., Kotiaho, J. S., & Päävinen, J. (2004). The role of niche breadth, resource availability and range position on the life history of butterflies. *Oikos*, 105(1) :41–54.
- Lafranchis, T. & Jutzeler, D. (2014). *Papillons de France guide de détermination des papillons diurnes (Rhopalocères, Zygènes et Hétérocères diurnes)*. Diatheo.
- Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, 6(1) :337–348.
- Norris, K. (2004). Managing threatened species : the ecological toolbox, evolutionary theory and declining-population paradigm. *Journal of Applied Ecology*, 41(3) :413–426.
- O’Connor, M. I., Selig, E. R., Pinsky, M. L., & Altermatt, F. (2012). Toward a conceptual synthesis for climate change responses : Theory for climate change ecology. *Global Ecology and Biogeography*, 21(7) :693–703.
- Pearce, J. L. & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43(3) :405–412.
- Pebesma, E. (2018). Simple Features for R : Standardized Support for Spatial Vector Data. *The R Journal*, 10(1) :439–446.
- Phillimore, A. B., Stålhandske, S., Smithers, R. J., & Bernard, R. (2012). Dissecting the Contributions of Plasticity and Local Adaptation to the Phenology of a Butterfly and Its Host Plants. *The American Naturalist*, 180(5) :655–670.
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models : implications for background and pseudo-absence data. *Ecological Applications*, 19(1) :181–197.
- Powney, G. D. & Isaac, N. J. B. (2015). Beyond maps : a review of the applications of biological records. *Biological Journal of the Linnean Society*, 115(3) :532–542.
- Quinn, R. M., Caston, K. J., & Roy, D. B. (1998). Coincidence in the distributions of butterflies and their food-plants. *Ecography*, 21(3) :279–288.

- Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G., & Warton, D. I. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6(4) :366–379.
- Roy, D. B. & Sparks, T. H. (2000). Phenology of British butterflies and climate change : phenology of butterflies and climate change. *Global Change Biology*, 6(4) :407–416.
- Rushton, S. P., Ormerod, S. J., & Kerby, G. (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41(2) :193–200.
- Schmeller, D. S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Déri, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van Swaay, C., Kobler, A., Babij, V., Papastergiadou, E., & Henle, K. (2009). Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conservation Biology*, 23(2) :307–316.
- Shreeve, T., Dennis, R., Roy, D., & Moss, D. (2001). An Ecological Classification of British Butterflies : Ecological Attributes and Biotope Occupancy. *Journal of Insect Conservation*, 5(3) :145–161.
- Soulsby, R. L. & Thomas, J. A. (2012). Insect population curves : modelling and application to butterfly transect data. *Methods in Ecology and Evolution*, 3(5) :832–841.
- Stasek, D. J., Bean, C., & Crist, T. O. (2008). Butterfly Abundance and Movements Among Prairie Patches : The Roles of Habitat Quality, Edge, and Forest Matrix Permeability. *Environmental Entomology*, 37(4) :897–906.
- Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution*, 19(6) :305–308.
- Swaay, C. v., Warren, M., & Loïs, G. (2006). Biotope Use and Trends of European Butterflies. *Journal of Insect Conservation*, 10(2) :189–209.
- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., De Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L., & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, 427(6970) :145–148.
- Van Strien, A. J., Van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6) :1450–1458.
- Vandewoestijne, S., Martin, T., Liégeois, S., & Baguette, M. (2004). Dispersal, landscape occupancy and population structure in the butterfly *Melanargia galathea*. *Basic and Applied Ecology*, 5(6) :581–591.
- Warton, D., Renner, I., & Ramp, D. (2013). Model-Based Control of Observer Bias for the Analysis of Presence-Only Data in Ecology. *PloS one*, 8(11) :e79168.
- Warton, D. I. & Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *The Annals of Applied Statistics*, 4(3) :1383–1402.