

Cursus Ingénieur Agronome M1

Module d'analyse de données

Rapport de projet

Baptiste BONGIBAUT

Maxence ROCOPLO

Mattéo LE GOUES

Inventaires "Papillons" du campus, peut-on tirer des enseignements de la première année ?

Enjeux et problématiques du projet

Créé en janvier 2022 par les étudiants, la responsable DDRS et les enseignants-chercheurs, le “Projet Taxon” a pour but de réaliser différents inventaires sur le campus rennais de l'Institut Agro Rennes-Angers. Ce projet s'inscrit dans la continuité des travaux d'anciens étudiants PAM-EQ, et celui-ci a abouti à la mise en place d'un plan de mode de gestion différencié des espaces verts du campus.

Toute l'année, les inventaires sont réalisés avec plusieurs objectifs : obtenir des données scientifiques sur le long terme pour mesurer l'évolution de la biodiversité dans un contexte de changement climatique, orienter les choix pour la gestion environnementale du campus, former des volontaires et un réseau d'acteurs en faveur de la biodiversité, ancrer le campus sur son territoire dans une démarche participative.

Aujourd'hui, le projet Taxon regroupe une dizaine de responsables étudiants en plus des volontaires. Nous travaillons avec les acteurs du monde de l'environnement, les chercheurs dans une démarche science-société et continuons à récolter de précieuses données. Depuis le début, 45 inventaires ont été réalisés permettant la récolte de 2500 observations comptant un total de 150 espèces identifiées.

Objectifs et enjeux :

- déterminer la phénologie des espèces de lépidoptères au cours de l'année (présence, absence et évolution sur l'année),
- lien présence d'une espèce en fonction de la zone du campus (prairie, arboretum...)

Problématique : la phénologie des papillons est-elle variable au cours de l'année et dépend-t-elle aussi de la zone du campus (i.e. le plan de réensauvagement de l'arboretum a t il un impact significatif) et de la météo?

Justification et choix des analyses

Concernant la justification des variables actives, nous utilisons les variables “espèce” et “date”. L'objectif est d'étudier la phénologie de ces espèces. Étant donné qu'ils ont une durée de vie très courte de l'ordre de quelques jours à quelques semaines, nous souhaitons étudier la répartition des émergences en fonction des mois. Nous souhaitons aussi ajouter une variable active habitat afin d'étudier la répartition des papillons en fonction des zones du campus.

Nous réalisons des analyses factorielles des correspondances (AFC). En effet, comme expliqué ci-dessus, nous avons deux variables qualitatives “espèce” et “date” puis “espèce” et “habitat”. La première AFC étudie les papillons, la variable “espèce” à 23 modalités et “date” à 11 modalités également. Le nombre total d'individus est de 466 papillons. Cela nous permet ensuite de réaliser une classification. La seconde AFC avec les habitats est réalisée avec les mêmes modalités “espèce” et respectivement 5 modalités “habitat”.

Détails de l'analyse relative aux papillons

A. La phénologie des espèces

Pour l'AFC s'intéressant aux espèces de papillons et aux dates d'inventaires, il est à noter que le χ^2 donne un résultat de 1015 ($p\text{-value} = 2.8E-102$), indiquant que nos variables ne sont pas indépendantes et qu'il existe une distribution particulière des individus entre les modalités. 4 dimensions suffisent pour expliquer 75%. En additionnant les valeurs propres de chaque axe, nous trouvons 2.179. Ce chiffre correspond au Φ^2 et peut se retrouver également en divisant la valeur de χ^2 par le nombre total d'individus ($n=466$).

Le Φ^2 permet de mesurer l'intensité de liaison entre les deux variables et peut prendre une valeur maximale de $\min(I-1, J-1)$ ou $I = \text{nombre de modalité de la variable "espèce"}$ et $J = \text{nombre de modalité de la variable "date"}$). Ici, le Φ^2 pourrait donc atteindre 10. Cela permet de savoir que l'intensité de liaison entre les deux variables est relativement faible. En règle générale, dans notre analyse, les espèces ne s'associent pas préférentiellement avec des dates. Cependant, cela concerne un comportement moyen et il est possible qu'une liaison entre deux modalités soit importante. Cela semble en effet le cas puisque la première dimension comporte une valeur propre de 0.617. La valeur n'est pas égale à 1 donc il n'y a pas d'association exclusive mais il semble y avoir une association préférentielle entre une/des espèces et une/des dates.

En effet, si l'on observe le graphe projetant les espèces et les dates (fig 1A), un groupe semble se détacher sur l'axe 1 avec des coordonnées négatives. Ce groupe rassemble les deux inventaires de juillet et 5 espèces. A première vue, cela indique que ces 5 espèces ont été proportionnellement plus présentes lors des inventaires de juillet par rapport aux autres mois. En observant la qualité de projection ($\cos^2$) des dates sur cet axe, le premier relevé de juillet (Juillet A) est correctement projeté ($\cos^2 = 0.229$), tandis que le second relevé (Juillet B) est très bien projeté ($\cos^2 = 0.918$). Ce dernier est d'ailleurs le meilleur contributeur à la première

dimension (ctr = 60.343). Au niveau des espèces, 4 sur les 5 ont une très bonne projection (supérieure à 0.750). Après étude des projections et du graphe des représentations, il apparaît que l'Amaryllis, le Machaon, le Brun des Pélargoniums et le Moro Sphinx sont des espèces qui émergent particulièrement au mois de juillet.

Deux autres dates semblent correctement projetées sur l'axe 1 (fig 1A). En effet, Octobre A et Octobre B sont deux modalités avec des qualités de projection respectives de 0.321 et 0.262 et environ 10% de contribution chacun. Plusieurs espèces sont projetées dans la même direction que ces dates : le Vulcain, la Piéride du Navet et le Souci. Elles ont une coordonnée supérieure à 1 sur la première dimension. Le Vulcain et le Souci ont une qualité de représentation supérieure à 0.2 alors que la Piéride du Navet a un $\cos^2$ de 0.138 ce qui est faible mais non négligeable. D'autres espèces sont projetées dans la même direction, avec des coordonnées élevées sur l'axe 1 et des qualité de représentation correcte. C'est, par exemple, le cas du Tircis qui a une coordonnée de 0.665 sur la première dimension et un $\cos^2$ de 0.453. Donc, si ce dernier est proportionnellement moins observé sur le mois d'octobre que les espèces citées précédemment, il reste néanmoins fortement observé par rapport au profil moyen.

Cependant, avant de conclure sur les espèces fortement présentes en fin d'année, il faut regarder la seconde dimension. En effet, les deux dates d'octobre ont également des coordonnées importantes sur cet axe (0.8 et 1.0) et des $\cos^2$ élevés (0.245 et 0.241). Il est possible de retrouver ici les trois mêmes espèces (Vulcain, Souci et Piéride du Navet) avec des coordonnées et $\cos^2$ très proches de celles de l'axe 1. C'est également le cas pour le Tircis, cela confirme donc que ces 4 espèces sont proportionnellement plus présentes en fin d'année. L'axe 2 apporte, en plus de cela, des informations sur la Piéride du Chou qui a une forte coordonnée (1.144) et une qualité de représentation de 0.142. Sa coordonnée sur l'axe 1 était correcte (0.450) mais sa qualité de contribution était très faible (<0.025).

Pour finir sur ce premier plan, il faut observer un dernier groupe qui semble se détacher dans les coordonnées négatives de l'axe 2 (fig 1A). Ce groupe est caractérisé par les inventaires réalisés en juin avec Juin A (coord = -0.873, $\cos^2$ = 0.769) et Juin B (coord = -0.967, $\cos^2$ = 0.762). Cela correspond au deux modalités de "date" qui contribuent le plus à l'axe 2 avec respectivement une contribution de 21.622 et 35.909. Quatre espèces sont caractérisées par des coordonnées fortement négatives sur l'axe 2 et une contribution élevée : le Myrtil (coord = -1.170 et $\cos^2$ = 0.896), l'Azuré des Nerpruns (coord = -1.283 et $\cos^2$ = 0.832), le Citron (coord = -1.349 et $\cos^2$ = 0.447) et le Paon du jour (coord = -1.039 et $\cos^2$ = 0.354). Elles sont

donc préférentiellement visibles lors du mois de juin. Comme pour le Tircis précédemment, une espèce a une coordonnée moins extrême que les espèces citées ci-dessus, mais semble avoir une proportion en juin plus importante que le profil moyen. En effet, le Robert-le-Diable a une coordonnée de -0.443 et une contribution de 0.197.

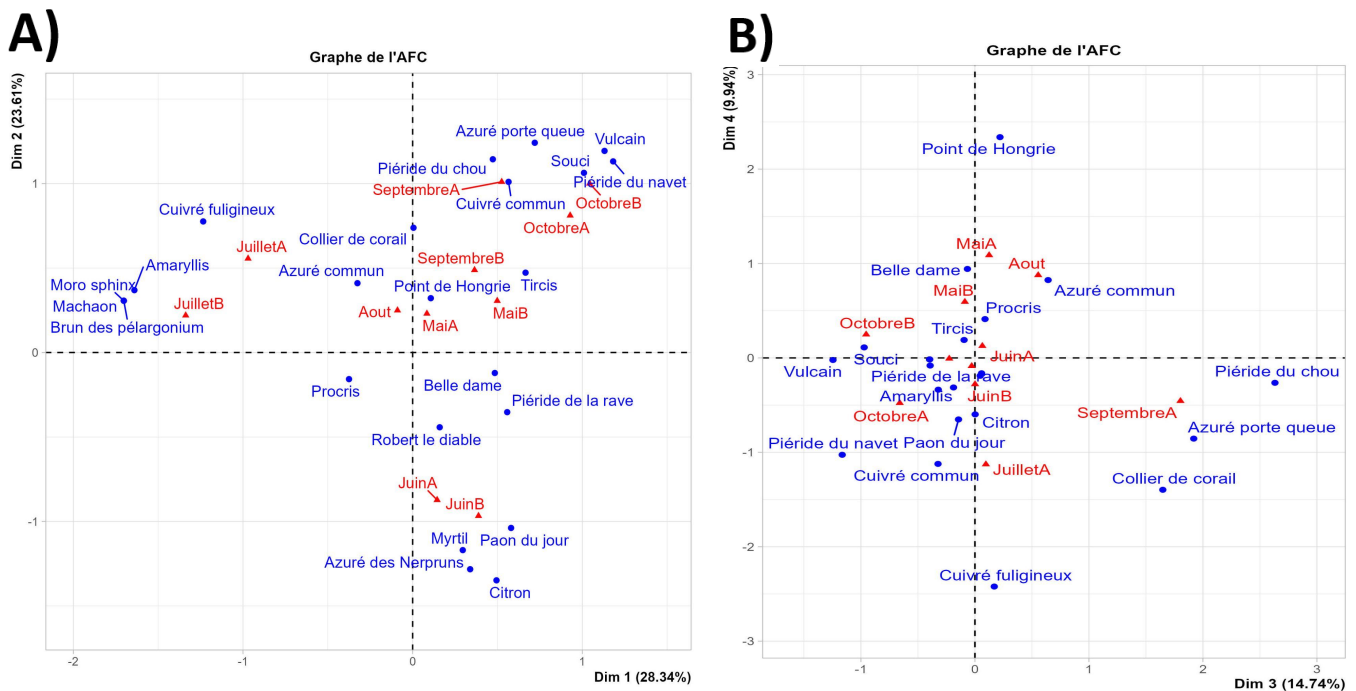


Figure 1 : premier plan de l'AFC portant sur la date d'observation des espèces de papillon
(A) et second plan de la même analyse (B)

Il est maintenant possible de s'intéresser aux dimensions 3 et 4 puisque celles-ci représentent 25 % de la variance de nos données. Le graphe des représentations est plus compact que sur le premier plan mais quelques modalités se détachent du centre de gravité (fig 1B). Sur l'axe 3, la modalité Septembre A est nettement éloignée du centre de gravité (coord = 1.803) et est très bien projetée sur cet axe ($\cos^2 = 0.830$). Trois espèces sont proportionnellement plus visibles que la moyenne lors de cet inventaire : la Piéride du Chou (coord = 2.632 et $\cos^2 = 0.753$), l'Azuré porte queue (coord = 1.919 et $\cos^2 = 0.487$) et le Collier de Corail (coord = 1.648 et $\cos^2 = 0.428$). C'est le premier groupe depuis le début de l'analyse où les deux inventaires correspondant au même mois ne sont pas corrélés positivement. Cela pourrait être un signe de variation importante des espèces sur le mois de septembre.

Enfin, observons la dimension 4 (fig 1B). Mai A et Août ont des coordonnées élevées sur la dimension 4 avec respectivement 1.09 et 0.88. Leur qualité de représentation est également importante ($\cos^2[\text{Mai A}] = 0.33$ et $\cos^2[\text{Août}] = 0.37$). En observant les espèces qui sont projetées dans la même direction, c'est-à-dire avec une forte coordonnée sur l'axe 4, seule

l'Azuré commun a une qualité de représentation supérieure à 0.2 ($\cos^2 = 0.27$). En retournant aux données brutes, nous pouvons constater que 11 sur 24 Azurés commun ont été observés en août soit 46 % des individus de cette espèce. Il y a donc une forte association entre la modalité Août et la modalité Azuré commun. Le point de Hongrie a une très forte coordonnée (2.34) et une qualité de contribution correcte (0.16). En effet, un seul individu a été observé et cela s'est produit lors du premier inventaire de mai (modalité Mai A).

B. L'habitat est-il vraiment important ?

Ci-dessus a été détaillé la répartition des différentes espèces durant la saison. Cela permet de comprendre la phénologie et donc les préférences de chaque espèce d'un point de vue abiotique. Nous nous intéressons maintenant à la préférence d'habitat afin d'observer si, des espèces sont plus fortement réparties dans des zones comprenant une végétation particulière.

Comme expliqué dans l'introduction de ce rapport, nous avons utilisé une couche d'informations d'occupation au sol par type de végétation sur la métropole rennaise. Les données ne couvrent pas l'ensemble du campus et donc le nombre d'individus est ici de 408 au lieu de 466. La couche compte 5 modalités : Bois, Haie, Arbre isolé, Bosquet et Herbacé.

L'AFC réalisée sur les données donne un χ^2 de 138.7 ($p\text{-value} = 4.6E-4$). Φ^2 vaut 0.34 ($138.7/408$). Cela est très faible par rapport au Φ^2 maximum qui pourrait être de 4. Il n'y donc que très peu de liaisons entre la variable "habitat" et la variable "espèce". Le premier plan représente 75 % de la variabilité et la première dimensions a une valeur propre de 0.165. Il semble donc y avoir une faible liaison entre au moins une modalité de chaque variable.

Sur la première dimension (fig 2), 2 modalités d'habitat ont une bonne qualité de projection : Bois ($\cos^2 = 0.954$) et Herbacé ($\cos^2 = 0.533$). Cependant, seul Bois est éloigné du centre de gravité (coord = 2.054). En observant le graphe des représentations, deux espèces ont de forte valeur pour la première dimension et forment un groupe avec la modalité Bois. On retrouve ici le Collier de Corail ($\cos^2 = 0.820$) et l'Azuré des Nerpruns ($\cos^2 = 0.917$). Ce dernier a une coordonnées très extrême sur l'axe avec 5.079. Ces deux espèces semblent donc préférer les zones fortement arborées (n'étant pas considérées comme de réels bois sur le campus). Sur la seconde dimension, Haie et Arbre isolé ont des coordonnées de 0.667 et 0.679 respectivement et des $\cos^2$ de 0.715 et 0.385 (fig 2). Le Robert-le-Diable (coord = 0.912 et $\cos^2 = 0.508$) et le

Cuivré fuligineux (coord = 2.228 et $\cos^2 = 0.600$) vont dans la même direction que ces modalités d'habitat. Il est donc possible de conclure que ces deux espèces sont préférentiellement présentes proches des haies et dans des zones où la strate arborée est faible.

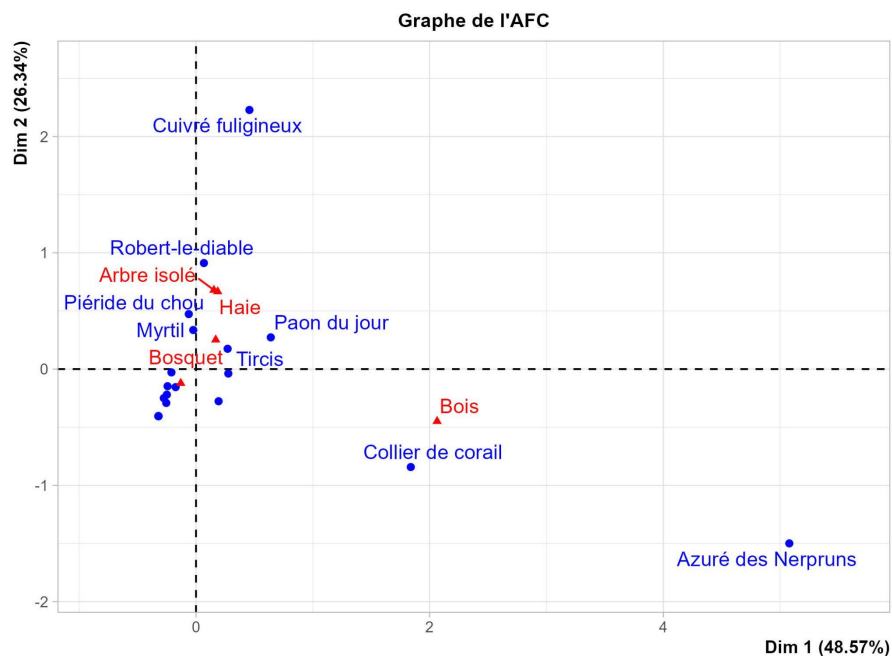


Figure 2 : premier plan de l'AFC portant sur les choix d'habitat (modalité rouge) des espèces de papillon (modalité bleue) du campus

C. Analyse par classification et AFC correspondante

Nous avons réalisé une classification pour compléter l'analyse de la phénologie des espèces de papillons. Cette classification a permis de déterminer 3 classes distinctes. On remarque que la majorité de l'inertie est contenue dans les deux premières dimensions. Les trois classes obtenues montrent des ensembles d'espèces ayant des périodes de vol (et donc de présence) différentes sur le campus à l'état d'imagó (ici, nous ne parlons pas des chenilles).

La première classe de couleur bleue comprend des espèces telles que le Machaon, les Azurés, le Procris, le Cuivre fuligineux et le Brun des pélargoniums (fig 3). Ils se trouvent que ces espèces ont des périodes de vol assez longues pouvant atteindre 1 mois et s'étendent de la fin de l'hiver jusqu'à octobre pour les plus tardifs. Sur le campus, on les trouve majoritairement entre les mois de juin et août.

Quant à la seconde classe (rouge), on observe que les trois espèces qui sont incluses malgré une période de vol aussi étendue que la première, se retrouvent sur le campus seulement après le mois de septembre. En observant le graphique de l'AFC, on observe que

dans le plan factoriel les points composant cette classe sont tous extrêmes contrairement aux deux autres classes (fig 3).

Pour la dernière classe (verte), qui est très distincte de la première et un peu moins de la seconde, on trouve le maximum d'espèces dont les points possèdent des valeurs positives et négatives en dimension 2 et seulement positives en dimension 1 (fig 3). Ces espèces-là sont plutôt fortement abondantes sur le campus tout au long de l'année avec des effectifs en Octobre qui sont les plus élevés toutes espèces confondues. Ces espèces semblent donc être plus généralistes et capables de s'adapter à des conditions météorologiques variables.

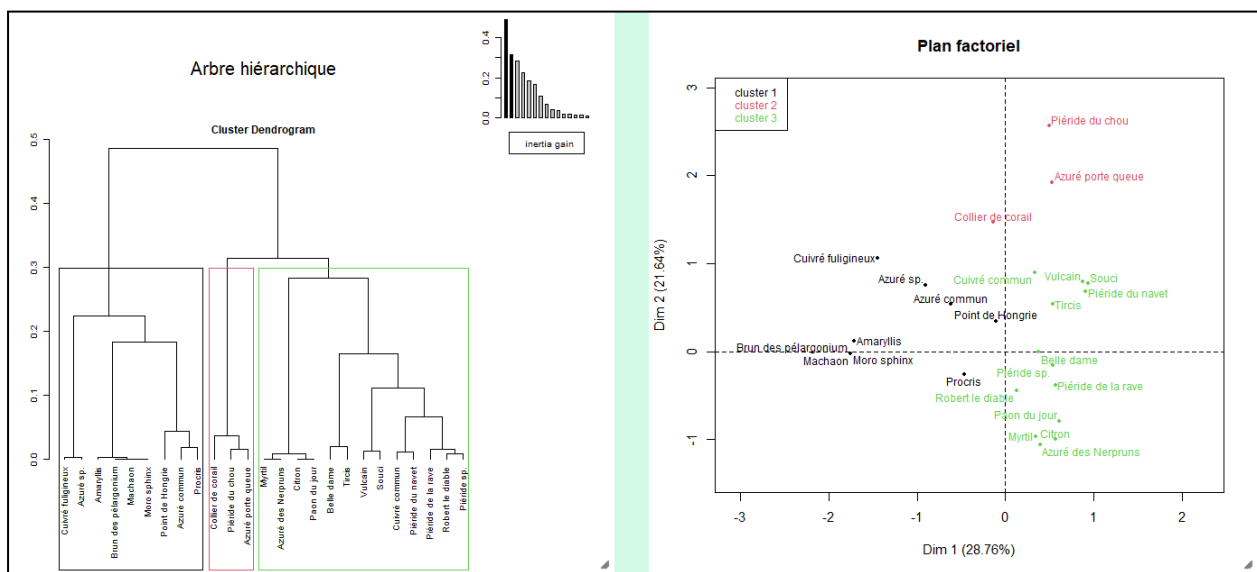


Figure 3 : Arbre hiérarchique et AFC correspondante mettant en avant l'évolution de la présence des espèces de papillons en fonction des mois de l'année.

Conclusion

Ces résultats nous permettent de mettre en évidence les particularités géographiques et saisonnières que le campus offre pour les espèces de lépidoptères. Et cela semble encourageant vis-à-vis de l'installation d'une biodiversité riche sur le campus. Bien sûr ces données vont s'enrichir avec les années et cela nous permettra d'étudier l'impact qu'a le plan de gestion différencié du campus sur l'évolution de sa biodiversité. Lorsque les données seront plus solides il pourrait être fort intéressant de mettre en relation les différents groupes taxonomiques comme par exemple oiseaux et lépidoptères pour établir de potentielles interactions trophiques sur le campus et ainsi construire les réseaux d'interactions trophiques de chaque partie ou micro-habitat qu'offre le campus à la faune et à la flore.