

Comment évaluer la qualité de prospection d'un atlas de biodiversité ?

La méthode des Probabilités d'Absences Cumulées

Émilien Barussaud, Louis Robert du Boislouveau

mai 2023

1. Problématique
2. Présentation de la méthode des Probabilités d'Absences Cumulées (P.A.C)
3. Application de la méthode P.A.C à deux atlas : les papillons diurnes de Bretagne et les amphibiens et reptiles de Loire-Atlantique
4. Résultats et interprétation
5. Perspectives : l'utilisation de la méthode P.A.C pour évaluer et améliorer les atlas de biodiversité

1. Problématique

Les atlas de répartition des espèces animales, à l'échelle régionale comme à l'échelle nationale, sont basés sur le découpage du territoire en mailles carrées. Ces mailles ont généralement dix kilomètres de côté, soit une superficie de cent kilomètres-carrés.

Au cours de la période de prospection d'un atlas (qui varie généralement entre 5 et 15 années), on cherche à inventorier toutes les espèces présentes sur chacune des mailles du territoire. Lors de cet exercice, toutes les mailles ne bénéficient pas d'une pression de prospection identique, loin s'en faut. Cette « *pression de prospection* » peut être définie comme le niveau d'effort déployé pour s'assurer de la présence ou de l'absence d'une espèce sur un territoire donné. Cet effort peut être d'ordre quantitatif (nombre de journées de prospection, nombre d'observateurs, surface prospectée, etc.) ou d'ordre qualitatif (compétence des observateurs, diversité des milieux prospectés, techniques employées, etc.). Plus la pression de prospection est élevée sur une maille, plus les résultats obtenus en termes de présence / absence d'espèces sont significatifs. Quand la pression de prospection est trop faible sur une maille, il est impossible de savoir si l'absence de données pour telle ou telle espèce traduit une absence réelle.

Il paraît donc indispensable de disposer d'un outil pour mesurer ce niveau d'effort. Or, il n'existe actuellement pas de méthode objective pour répondre à la question : telle maille a-t-elle bénéficié d'une pression de prospection suffisante pour que les données de présence / absence y soient interprétables ?

L'indicateur de niveau de prospection le plus couramment utilisé est le nombre de données par maille. Or, cette valeur ne reflète que très imparfaitement l'effort de prospection : on peut par exemple accumuler une grande quantité de données concernant une même espèce sur un même site à différentes dates sans pour autant améliorer la prospection de l'ensemble de la maille pour l'ensemble des espèces.

Dans certains atlas, on présente un graphique du nombre d'espèces par maille en fonction du nombre de données par maille. Ce type de représentation permet de se rendre compte qu'il n'y a pas de relation mathématique évidente entre les deux variables. Toutefois, les auteurs lui superposent souvent une courbe de tendance – généralement de type logarithmique – et concluent qu'au-delà du point d'inflexion de cette courbe « *le nombre de données nécessaire pour détecter une nouvelle espèce augmente rapidement* ». Ainsi, l'abscisse du point en question correspondrait à un niveau de prospection « *suffisant* ». Cette méthode, utilisée faute de mieux, comporte deux inconvénients majeurs : d'une part la série de points est ramenée à une formule mathématique qui est loin de le décrire fidèlement (le r^2 est souvent éloigné de 1) ; d'autre part, le « point d'inflexion » est établi de manière arbitraire.

Appréhender le niveau de prospection par le biais d'un seul paramètre quantitatif (par exemple, le nombre d'observation) est donc insuffisant, les autres paramètres quantitatifs (temps passé, nombre d'observateurs...) étant ignorés, de même que les paramètres qualitatifs (compétences des observateurs, méthodes utilisées...). En outre, la variable « nombre d'observations » donne un poids égal à toutes les espèces. Par exemple, trois observations de la Mésange bleue sur une maille comptent autant qu'une observation de Mésange bleue, une observation de Mésange huppée et une observation de Mésange noire. Or, cette seconde combinaison contient davantage d'information.

La méthode des Probabilités d'Absences Cumulées, présentée dans cet article, s'affranchit de ces biais en définissant le niveau de prospection d'après les résultats de l'atlas, ces derniers étant la résultante de tous les paramètres de prospection, qu'ils soient quantitatifs ou qualitatifs.

2. Présentation de la méthode des Probabilités d'Absence Cumulées (P.A.C)

La méthode des Probabilités d'Absences Cumulées (P.A.C) a été mise au point par Émilien Barussaud à partir de 2020. Elle est inspirée de la théorie de l'information de Shannon, à savoir la mesure de la quantité d'information contenue dans un ensemble de données. Dans notre cas, cet ensemble de données est le tableau de présence / absence croisant les mailles et les espèces. Pour Shannon, l'information est la mesure de l'incertitude calculée à partir de la probabilité d'un évènement. Plus la probabilité qu'un évènement se produise est faible, plus sa réalisation nous offre d'information.

Dans le cas d'un atlas de biodiversité, c'est l'absence d'une espèce très commune sur une maille donnée qui délivre une grande quantité d'information à propos d'une probable sous-prospection sur cette maille. A fortiori, l'absence cumulée de plusieurs espèces très communes sur une même maille.

La méthode P.A.C consiste à calculer la probabilité que plusieurs espèces soient simultanément absentes d'une même maille. P (absences cumulées) = P (absence espèce 1) x P (absence espèce 2) x ... x P (absence espèce n). Toutes les espèces absentes sont prises en compte mais chacune apporte une contribution inversement proportionnelle à sa rareté.

La probabilité d'absence cumulée est comparée à la probabilité unitaire (P unitaire = $1 /$ nombre de mailles), ce qui nous permet de savoir si la combinaison des espèces absentes est statistiquement possible (= si elle peut se réaliser sur au moins une maille) ou statistiquement impossible.

Le résultat s'interprète de la manière suivante :

- si P (absences cumulées) / P unitaire > 1 cela signifie que la **situation est statistiquement possible** et donc que **la prospection est satisfaisante**
- si P (absences cumulées) / P unitaire < 1 cela signifie que la **situation est statistiquement impossible** et donc qu'il y a un **défait de prospection**

Le plus l'indice P.A.C est faible, le plus la sous-prospection est forte :

- indice PAC compris entre 0,1 et 1 = sous-prospection de niveau 1 (= légère)
- indice PAC compris entre 0,01 et 0,1 = sous-prospection de niveau 2 (= moyenne)
- indice PAC compris entre 0,001 et 0,01 = sous-prospection de niveau 3 (= forte)
- etc.

3. Application de la méthode P.A.C à deux atlas : les papillons diurnes de Bretagne et les amphibiens et reptiles de Loire-Atlantique

En 2022, la méthode des Probabilités d'Absences Cumulées (P.A.C) a été testée par Émilien Barussaud et Louis Robert du Boislouveau sur deux atlas :

- Buord Mikaël, David Jean, Garrin Mael, Iliou Bernard, Jouannic Jacques, Pasco Pierre-Yves & Wiza Stéphane (coord.), 2017. Atlas des papillons diurnes de Bretagne. Locus Solus, Lopérec, 324 p.
- Grosselet, O., Gouret, L. & Dusoulier, F. (coord.), 2011. Les Amphibiens et les Reptiles de la Loire-Atlantique à l'aube du XIXème siècle : identification, distribution, conservation. Éditions De mare en mare, Saint-Sébastien-sur-Loire. 207 p.

Les caractéristiques de ces deux atlas sont les suivantes :

- Pour les papillons diurnes de Bretagne : Période de prospection : 2000 à 2015. Au total, 109.537 données produites par 1.133 observateurs sur 354 mailles de 10x10 km couvrant les 4 départements de la Bretagne administrative, pour un total de 86 espèces.
- Pour les Amphibiens et Reptiles de Loire-Atlantique : Période de prospection : 1990 à 2007 . 6.221 données produites par 272 observateurs sur 98 mailles de 10x10 km couvrant le département de la Loire-Atlantique, pour un total de 35 espèces.

La durée de la période de prospection est quasiment identique. Le nombre de données divisé par le nombre de maille donne une moyenne de 309 données par maille pour les papillons de Bretagne et 63 données par maille pour les amphibiens et reptiles de Loire-Atlantique. Il y a donc environ 5 fois plus de données par maille pour les papillons. Ce chiffre s'explique en grande partie par l'émergence des portails participatifs au début des années 2010. Par ailleurs, l'atlas des papillons concerne environ deux fois plus d'espèces que celui des amphibiens et reptiles.

La méthode des Probabilités d'Absences Cumulées a été testée :

- Sur 100 des 354 mailles de l'atlas des papillons de Bretagne. Pour des raisons de temps, nous n'avons pas réalisé l'analyse sur l'ensemble des mailles. Cet échantillonnage n'entraîne aucune conséquence sur les résultats*. Nous avons volontairement laissé de côté les mailles côtières ou frontalières dans lesquels la surface prospectable est trop faible (<75 % de la maille).
- Sur 64 des 98 mailles de l'atlas des amphibiens et reptiles de Loire-Atlantique. Nous avons volontairement laissé de côté les mailles côtières ou frontalières dans lesquels la surface prospectable est trop faible (<75 % de la maille).

Pour chaque maille, nous avons :

- Recensé les espèces absentes : espèce 1, espèce 2... espèce n. Cette étape a demandé beaucoup de temps de travail car nous n'avions pas à notre disposition les données brutes sous forme de tableaux présence / absence.
- Réalisé le calcul : $P(\text{absences cumulées}) = P(\text{absence espèce 1}) \times P(\text{absence espèce 2}) \times \dots \times P(\text{absence espèce n})$.
- Réalisé le calcul : $P(\text{absences cumulées}) / P \text{ unitaire}$. Pour rappel $P \text{ unitaire} = 1 / \text{nb maille total}$

** La Probabilité d'Absence Cumulée peut en effet être calculée pour chaque maille de manière indépendante. La probabilité d'absence d'une espèce est définie par le nombre de mailles qu'elle occupe par rapport au nombre de mailles total de l'atlas. Cette valeur est la même que l'on travaille sur une seule maille ou l'ensemble des mailles.*

4. Résultats et interprétation

Pour l'atlas de papillons diurnes de Bretagne, sur les 100 mailles traitées :

- 19 présentent une prospection satisfaisante (19 %)
- 14 présentent une sous-prospection légère (14 %)
- 15 présentent une sous-prospection moyenne (15 %)
- 52 présentent une sous-prospection forte (52 %)

Pour l'atlas des amphibiens et reptiles de Loire-Atlantique de Loire-Atlantique, sur les 64 mailles traitées :

- 30 présentent une prospection satisfaisante (47 %)
- 21 présentent une sous-prospection légère (33 %)
- 8 présentent une sous-prospection moyenne (12 %)
- 5 présentent une sous-prospection forte (8 %)

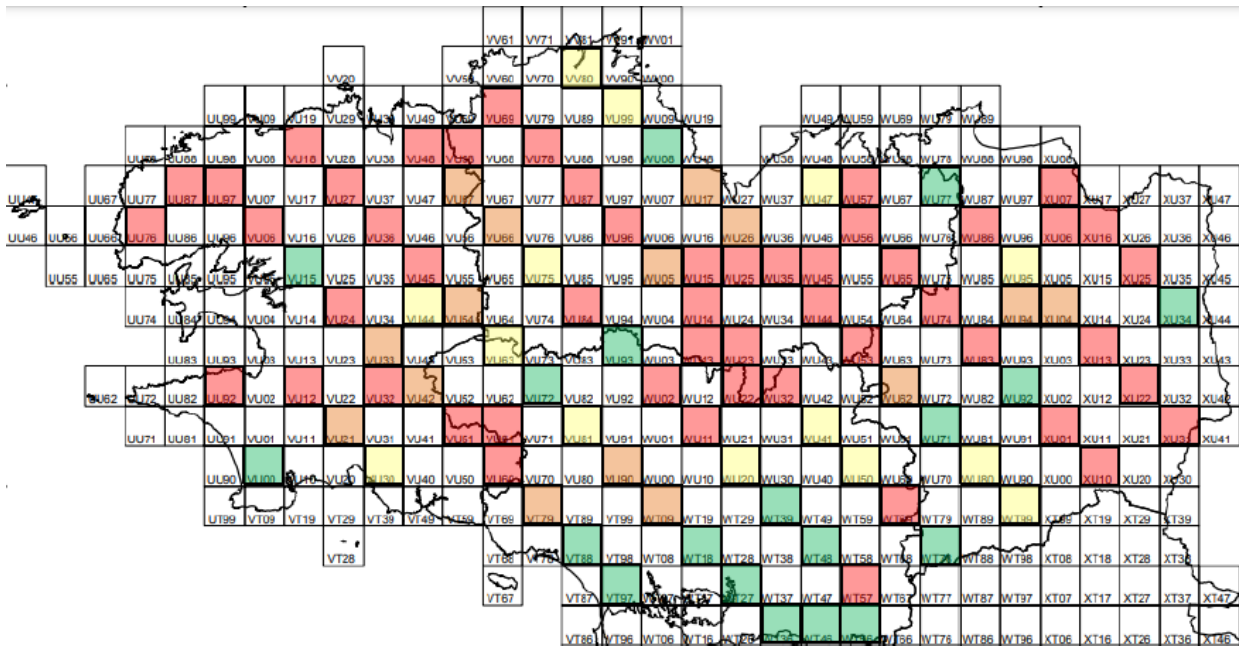
Ces premiers résultats montrent un étonnant contraste entre les deux atlas. L'atlas des papillons bénéficie, en moyenne, de cinq fois plus de données par maille pour seulement deux fois plus d'espèces à rechercher : on pouvait donc s'attendre à ce que la prospection soit meilleure. Or, bien au contraire, l'atlas des amphibiens et reptiles bénéficie d'une prospection satisfaisante sur quasiment la moitié de son territoire (47 % des mailles) contre seulement 19 % pour les papillons.

Une première explication peut être avancée : l'atlas des papillons de Bretagne, plus tardif, a bénéficié des contributions d'un large public via les portails participatifs. Le nombre d'observateurs pour l'atlas des papillons est ainsi quatre fois plus élevé que pour l'atlas des amphibiens et reptiles. Ce dernier, plus ancien, a visiblement été l'œuvre d'une équipe plus restreinte mais aussi plus expérimentée en moyenne.

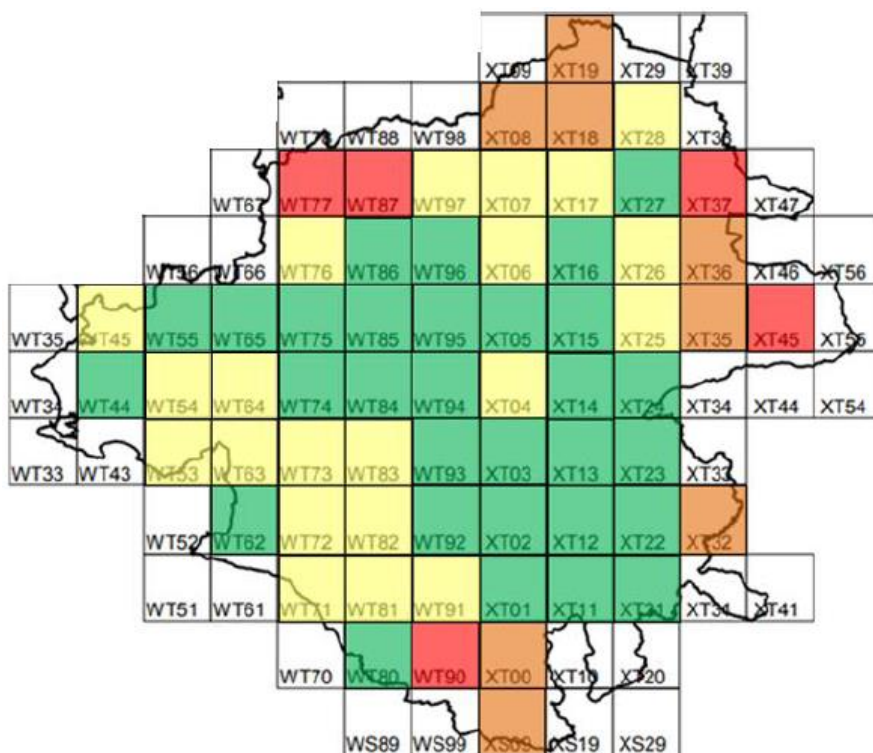
Au niveau de la répartition géographique des mailles sous-prospectées (voir cartes page suivante) :

- Pour les papillons de Bretagne, la prospection est de bonne qualité dans le sud du Morbihan et notamment aux environs de Vannes. D'autres mailles, disséminées dans la région, bénéficient d'une bonne prospection, probablement grâce à la présence d'entomologistes locaux particulièrement impliqués dans le projet. De vastes ensembles de mailles fortement sous-prospectées se dessinent, notamment dans l'intérieur des terres entre Loudéac, Montauban-de-Bretagne et Broons ou encore dans le Léon.
- Pour les amphibiens et reptiles de Loire-Atlantique, un large centre (Nantes et ses environs) ainsi que la côte ouest bénéficient d'une bonne prospection. Les lacunes, peu nombreuses, concernent les marges nord, est et sud du département.

Ces premiers résultats obtenus avec la méthode P.A.C sont en cohérence avec l'expérience vécue par les coordinateurs des atlas de biodiversité. Ces derniers reconnaissent volontiers « *un effort de prospection hétérogène* » (atlas des amphibiens et reptiles de Loire-Atlantique) et des « *observations inégalement réparties dans l'espace* » (atlas des papillons diurnes de Bretagne).



Echantillonnage de 100 mailles sur les 354 que compose le territoire de la Bretagne.



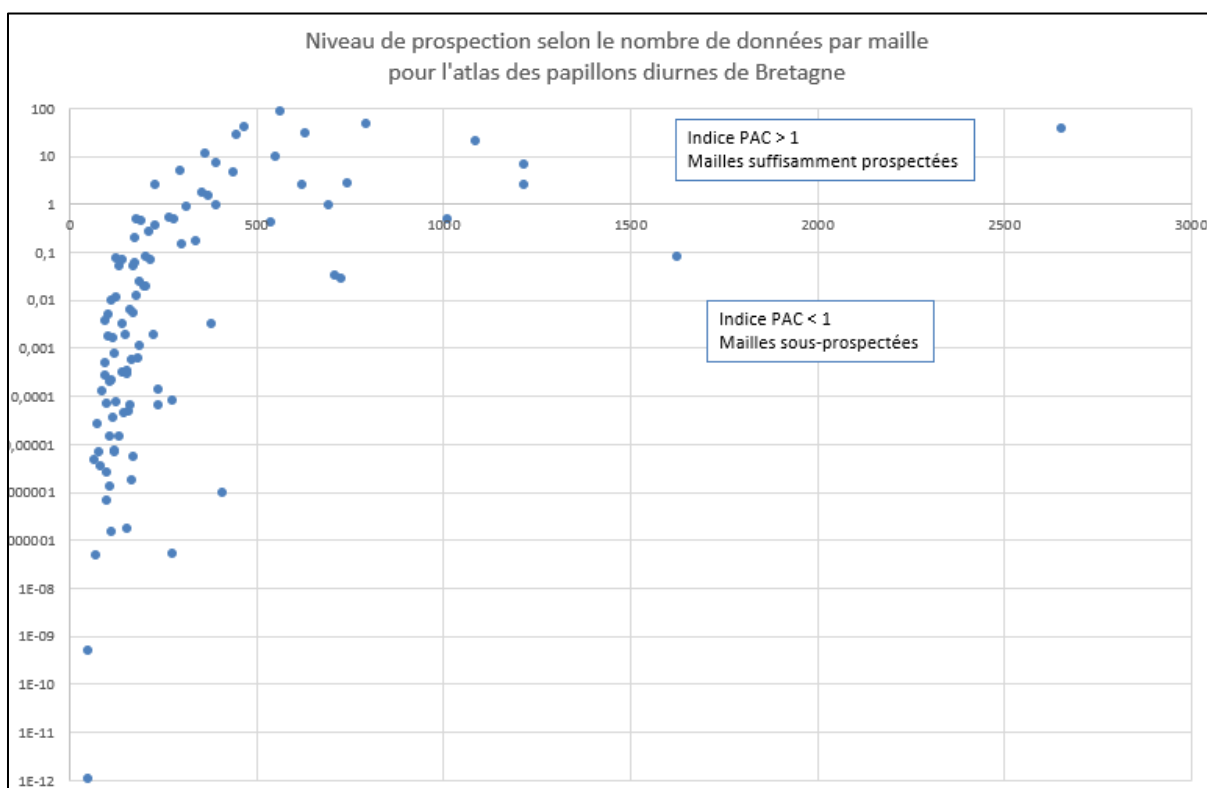
- Prospection suffisante (≥ 1)
- Sous prospection niveau 1 (entre 0,1 et 1)
- Sous prospection niveau 2 (entre 0,01 et 0,1)
- Sous prospection niveau 3 ($< 0,001$)

Cartographie des résultats de la méthode P.A.C pour l'atlas des papillons de Bretagne (en haut) et l'atlas des amphibiens et reptiles de Loire-Atlantique (en bas), réalisée par Louis Robert du Boislouveau (2022)

Observons maintenant la relation entre la valeur de l'indice P.A.C et le nombre de données par maille dans le cas de l'atlas des papillons diurnes de Bretagne. D'une part, et sans grande surprise, les deux mailles ayant le plus faible nombre de données (49 et 51 données) sont aussi les deux mailles ayant l'indice PAC le plus faible, ce qui signifie une très forte sous-prospection. De manière générale, aucune maille comportant moins de 230 données n'est considérée comme suffisamment prospectée. À l'opposé, la maille ayant le nombre de données le plus élevé (2.654 données, du côté de Séné) présente un indice PAC de 38, très largement supérieur à 1 qui représente le seuil de prospection suffisante. Mais d'autres éléments sont plus surprenants. Par exemple, la seconde maille la plus prospectée en termes de nombre de données (1.627 données) est considérée comme insuffisamment prospectée d'après la méthode PAC (indice PAC = 0,082). À l'inverse, parmi les 19 mailles suffisamment prospectées, 9 comportent moins de 500 données. Enfin, le meilleur niveau de prospection (indice PAC = 87) est atteint par une maille qui comporte 565 données.

Si les auteurs de l'atlas jugent que les mailles « *totalisant plus de 250 observations sont bien prospectées* », notre analyse montre que des mailles qui en comptent le double peuvent être insuffisamment prospectées... tandis qu'une maille à 230 données dépasse largement le seuil de prospection suffisante.

Le nombre de données par maille est donc un critère très insuffisant pour juger de la qualité de prospection. L'indice P.A.C permet de prendre en compte un aspect qualitatif (les espèces découvertes ou non) mais aussi d'utiliser un seuil objectif au lieu d'un seuil arbitraire.



5. Perspectives : l'utilisation de la méthode P.A.C pour évaluer et améliorer les atlas de biodiversité

La méthode des Probabilités d'Absences Cumulées (P.A.C) permet d'évaluer le niveau de prospection sur une, plusieurs ou toutes les mailles d'un atlas de biodiversité. Elle pourra être utilisée :

- Dans tous les atlas déjà réalisés pour évaluer leur qualité *a posteriori*
- Dans les futurs atlas de biodiversité pour orienter efficacement les prospections : il suffira de réaliser le calcul après quelques années de prospection pour mettre en évidence les mailles les plus fortement sous-prospectées et ainsi homogénéiser la qualité des atlas régionaux ou départementaux

Par ailleurs, sur les cartes de répartition des espèces, il serait possible de faire figurer le niveau de prospection des mailles afin d'améliorer leur interprétation. On se garderait ainsi d'interpréter de manière erronée les absences sur les mailles fortement sous-prospectées.

La méthode PAC peut donc permettre :

- **Une amélioration de la qualité des futurs atlas de biodiversité en homogénéisant le niveau de prospection sur le territoire**
- **Une meilleure lecture / interprétation des cartes de répartition des espèces**

D'un point de vue technique, le calcul des probabilités d'absences cumulées et même la cartographie pourraient être facilement automatisés à partir de tableaux contenant les données de présence / absence des différentes espèces pour chaque maille.

Appel à contribution : nous invitons toute personne ou structure ayant réalisé un atlas départemental ou régional d'un groupe zoologique (oiseaux, mammifères, insectes...) et qui serait intéressée par notre méthode d'évaluation à nous contacter à l'adresse suivante : info@bet-barussaud.fr