

# Evaluation des biais d'échantillonnage dans les analyses des régressions/extensions dans les Atlas de répartition

**Marc Dufrene**

***Observatoire de la Faune, de la Flore et des Habitats***

Centre de Recherche de la Nature, des Forêts et du Bois (MRW/DGRNE)

Avenue Maréchal Juin, 23

B-5030 Gembloux

Belgique

[M.Dufrene@mrw.wallonie.be](mailto:M.Dufrene@mrw.wallonie.be)

<http://biodiversite.wallonie.be>

## **Introduction**

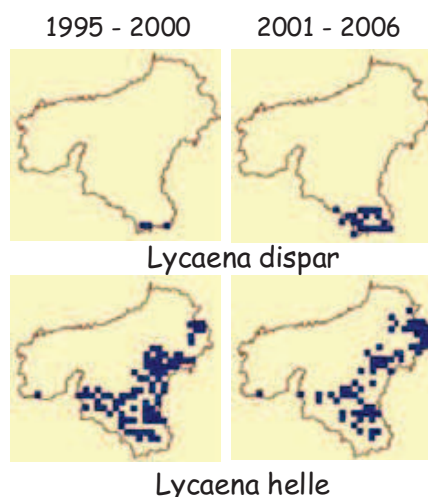
Les Atlas de répartition ont pour premier objectif de faire un bilan de la répartition des espèces d'un groupe biologique à un moment donné en cartographiant la localisation des individus ou des populations dans des mailles géographiques plus ou moins précises. Lorsque de telles synthèses sont réalisées à deux périodes différentes, la tentation est forte de comparer les répartitions des espèces d'un atlas à l'autre pour évaluer leur évolution (régression, stabilité, extension). Les Atlas sont d'ailleurs un des supports essentiels de la réalisation des listes rouges d'espèces menacées (IUCN, 2001) qui sont des outils importants pour identifier les priorités en matière de gestion et de conservation de la biodiversité.

De telles comparaisons dans le temps sont faciles lorsqu'on dispose de jeux de données récoltées avec des protocoles similaires et la même intensité d'échantillonnage locale et régionale. Hors, les Atlas sont, par nature, des ensembles hétérogènes de données récoltées de manière différente à chaque période, avec de nombreux collaborateurs qui inventorient des zones de travail de manière indépendante et avec des techniques plus ou moins spécifiques. Les méthodes d'investigation ont aussi bien évolué d'une période à l'autre. Les Atlas ont souvent été critiqués pour leur caractère hétérogène mais cette hétérogénéité est aussi défendue par d'autres comme un avantage car assimilée à un échantillonnage aléatoire. Il est évident qu'ils recèlent de l'information sur les tendances, mais il faut être prudent pour les mettre en évidence.

Cette contribution a pour but de montrer, avec différents exemples tirés de l'analyse de données de répartition de différents groupes biologiques, les caractéristiques majeures des échantillonnages de données biologiques dans de grandes régions. On résumera les méthodes existantes et on proposera quelques pistes simples à prendre en compte lors de l'analyse des tendances sur de telles données pour comparer des échantillonnages réalisés à des périodes différentes. On montrera comment évaluer de manière simple l'exhaustivité de l'échantillonnage avec la distribution de fréquence du nombre d'espèces par maille et la relation généralement étroite qui existe entre l'intensité de l'échantillonnage et le nombre d'espèces. On vérifiera notamment si des hypothèses comme le fait que toutes les espèces sont échantillonnées de la même manière ou que l'ensemble du groupe peut être considéré comme stable dans le temps sont validées. On montrera aussi comment évaluer la répartition des différences d'échantillonnage dans le temps et l'impact que cela peut avoir sur les hypothèses de base des tests habituels.

## Les méthodes de base

La **Figure 1** montre la répartition de deux espèces de papillons visés par la Directive européenne CE/92/43 (Faune-Flore-Habitats). L'évolution des deux espèces semble à priori être opposée, le Cuivré des marais (*Lycaena dispar*) ayant tendance à augmenter son aire de répartition depuis l'an 2000 alors que le Cuivré de la Bistorte (*Lycaena helle*) semble avoir régressé. La question qui se pose rapidement est celle de la pertinence de cette tendance apparente : s'agit-il d'une tendance réelle ou est-ce dû à un effet d'une variation géographique de l'échantillonnage entre les deux périodes, tout à fait possible dans le cadre de suivi d'espèces emblématiques.



**Figure 1.** Carte de répartition de deux espèces de papillons diurnes dans la région continentale de la Wallonie

De nombreux chercheurs ont essayé de maîtriser ce problème de la variation géographique de l'échantillonnage en développant différentes approches. L'une des premières, souvent attribuée en Belgique à Stroot & Depiereux (1989) mais en fait au moins déjà utilisée par Desender (1986), consiste à comparer le nombre de mailles occupées par une espèce à l'ensemble des mailles occupées par les autres espèces, la mention d'une espèce dans une maille étant un premier indicateur de l'effort d'échantillonnage.

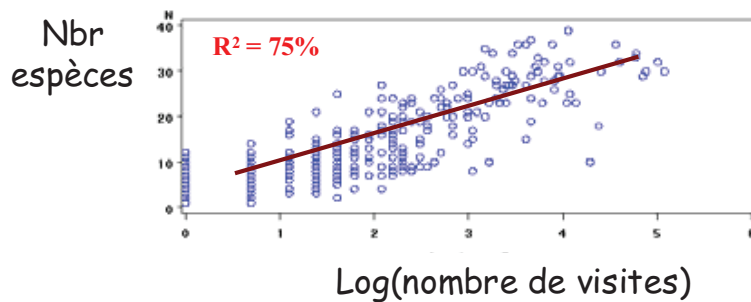
Desender (1986) propose un test de Chicarré (test de G par exemple) pour mesurer l'importance de la différence entre la fréquence observée d'une espèce et sa fréquence attendue calculée en fonction des différences d'échantillonnage. Si une espèce *i* occupe respectivement 58 et 54 carrés pour deux périodes et que l'ensemble des observations des autres espèces est de 4523 et 6345 pour des périodes équivalentes, la valeur de Chicarré est de 4.7, la probabilité associée est de 3% et on peut donc considérer que l'espèce régresse bien. Au lieu d'utiliser toutes les espèces, certains auteurs comme Maes & Van Swaay (1997) utilisent une sélection d'espèces qui sont considérées comme étant de bons indicateurs de l'échantillonnage (espèces répandues qui n'ont à priori pas modifié leur aire de répartition).

Il est évident qu'avec ces approches, le facteur clé est le rapport de la mesure d'échantillonnage. S'il est surévalué, on aura tendance à déclarer de nombreuses espèces comme étant en régression alors que ce n'est pas vrai. S'il est sous-évalué, on sera incapable de détecter des tendances qui nécessiteraient pourtant des actions sur le terrain. Par ailleurs, ces approches relatives supposent que la somme des observations du groupe d'espèces est stable dans le temps. Si, par exemple, toutes les espèces régressent de la même manière,

aucune tendance ne sera détectée. Ces approches supposent aussi que toutes les espèces ont la même probabilité de rencontre, ce qui est loin d'être vrai. Il est donc crucial dans les analyses de tendance d'avoir une maîtrise la plus importante possible de l'échantillonnage.

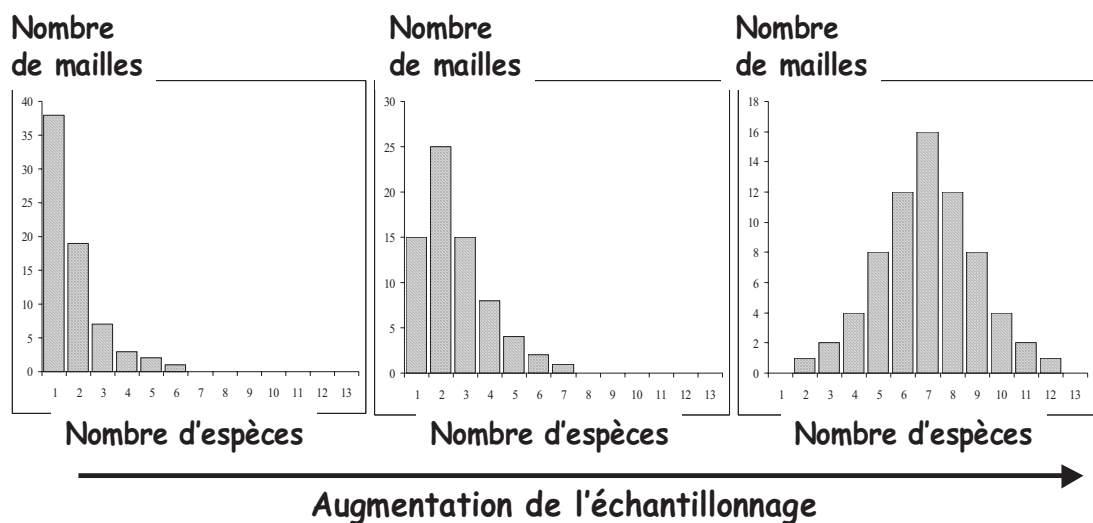
### Comment évaluer l'échantillonnage ?

Une première étape est de vérifier si le nombre de visites réalisées dans des mailles est un bon prédicteur du nombre d'espèces. La **Figure 2** montre une relation générale qui est souvent observée pour de nombreux groupes biologiques, ici les Libellules. Cette relation peut permettre de définir des seuils à partir desquels des mailles peuvent être utilisées pour des comparaisons absolues de tendance.



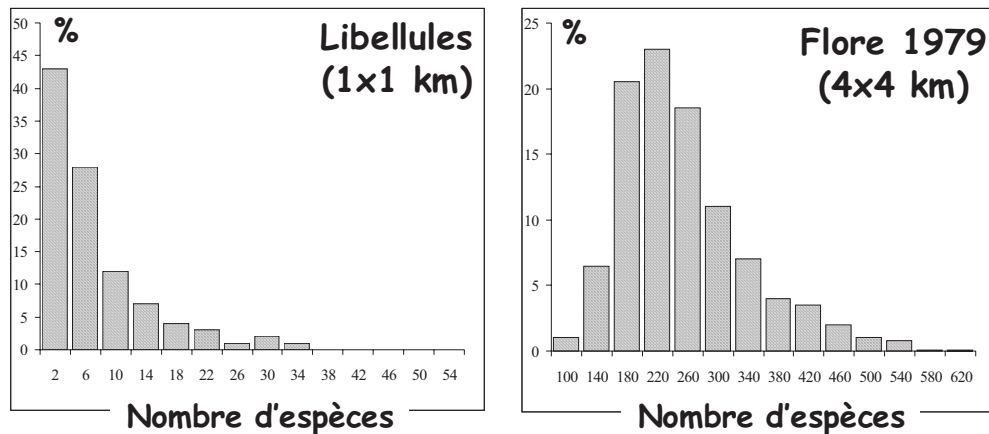
**Figure 2. Relation entre le nombre de visites et le nombre d'espèces pour un jeu de données de Libellules**

Cette relation n'est toutefois mise en évidence que lorsque le nombre moyen de visites par maille est relativement important et c'est loin d'être le cas dans des groupes biologiques où la présence mesurable des espèces sur le terrain n'est pas limitée à quelques semaines (fort renouvellement des espèces au cours du temps). Dans les autres cas, seules des méthodes indirectes peuvent être utilisées.



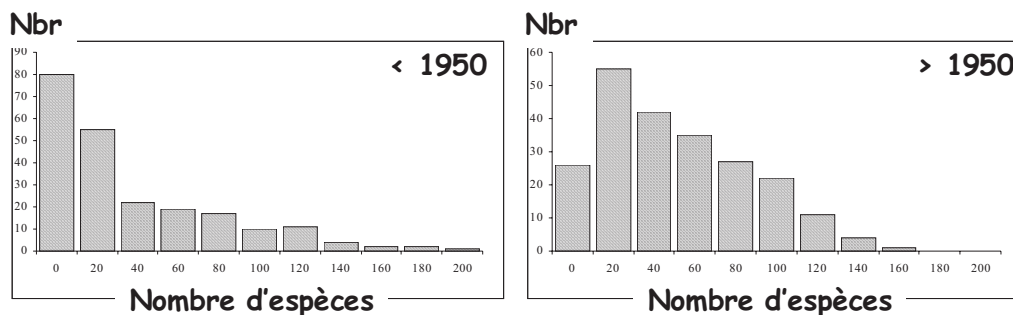
**Figure 3. Evolution de la distribution de fréquence du nombre d'espèces par maille en fonction de l'intensité de l'échantillonnage**

Un élément indirect pour évaluer le caractère exhaustif de l'échantillonnage est la réalisation de courbe de distribution de fréquence du nombre d'espèces par maille. En effet, lorsqu'on commence à faire des inventaires dans une région, on commence par avoir de nombreux carrés avec très peu d'espèces. Plus l'intensité des observations augmente (**Figure 3**), plus le nombre d'espèces par carré augmente et progressivement se découvre une distribution de fréquence du nombre d'espèces en forme de cloche, normale et relativement étroite quand une région est bien échantillonnée et qu'elle est relativement homogène.



**Figure 4. Fréquence relative du nombre d'espèces par maille dans deux groupes inventoriés de manière différente**

Ce type de graphique montre très rapidement (**Figure 4**) que des groupes biologiques inventoriés sans structure d'échantillonnage (Libellules, à l'échelle des carrés 1x1 km) peuvent avoir des profils bien différents de ceux pour lesquels un échantillonnage systématique est en place (Flore, Atlas de 1979). Il est aussi possible de réaliser ce type de graphique par région naturelle pour vérifier si certaines régions sont plus ou moins bien échantillonnées que d'autres et pour tenir compte alors des différences de richesse en espèces qui peuvent les caractériser. On peut aussi tester des effets de la taille des mailles (10, 5 ou 1 km ?) pour identifier l'échelle géographique à laquelle les comparaisons sont les plus pertinentes.



**Figure 5. Différence de fréquence du nombre d'espèces par maille des données de l'atlas des Coléoptères Carabides pour deux périodes**

Appliqué aux deux périodes qu'on souhaite comparer, ce type de représentation montre aussi les différences d'exhaustivité entre les deux périodes. Dans le cas illustré à la **Figure 5**, seule une petite partie des carrés a été bien échantillonnée (seuil fixé habituellement : > 30 espèces)

avant 1950, même si les carrés les plus riches en espèces ont été observés à cette période. Il est alors nécessaire d'être prudent dans les interprétations des tendances des espèces.

### L'effet de la probabilité de rencontre

Lorsqu'on travaille avec un rapport de référence commun à toutes les espèces, on suppose que les espèces ont toutes la même probabilité de rencontre quand on commence à inventorier une maille. C'est évidemment loin d'être le cas sur le terrain. En effet, les espèces rares à l'échelle de la Région wallonne sont aussi très fréquemment plus rares à l'intérieur d'un carré de 10 x 10 km. Même dans les carrés où elles sont présentes, les chances de les observer sont moindres que celles d'espèces largement répandues.

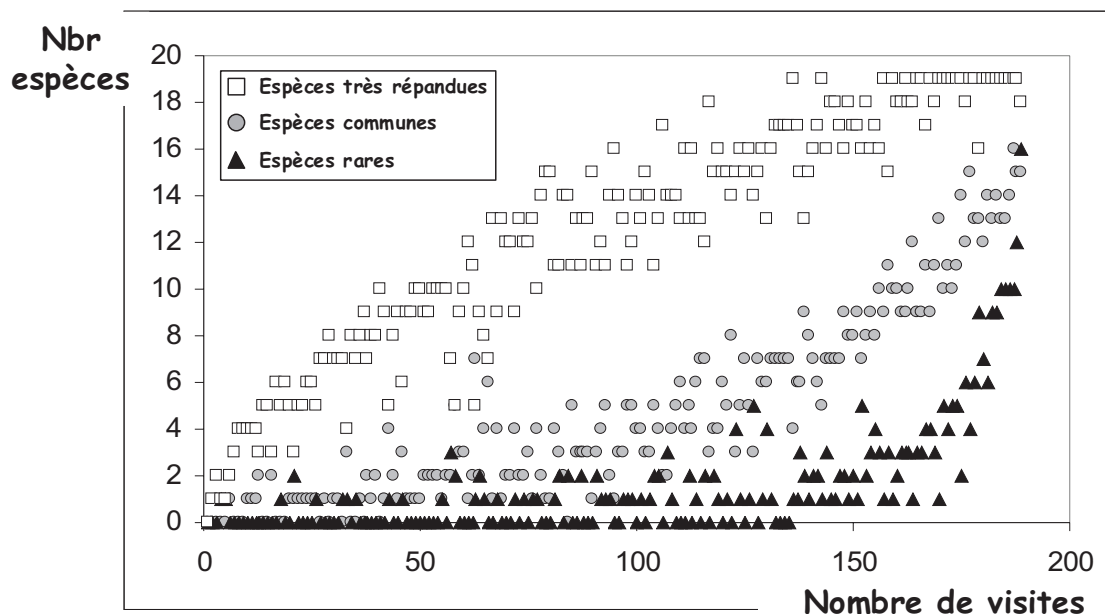
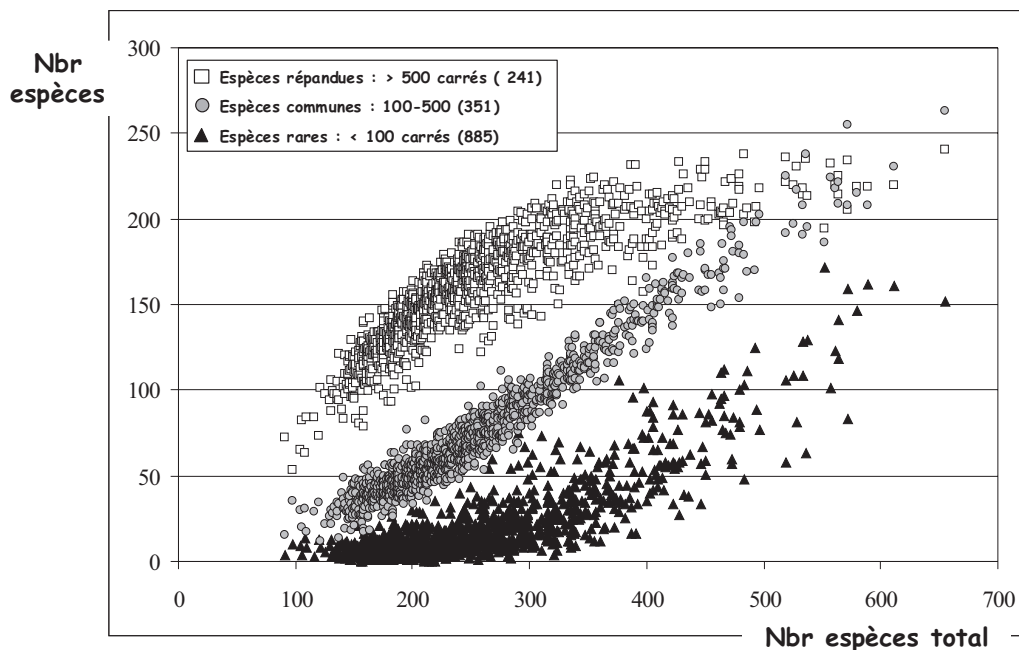


Figure 6. Evolution du nombre d'espèces d'Odonates en fonction du nombre de visites réalisées dans les carrés 10 x 10 km. Les espèces sont subdivisées en 3 groupes en fonction de leur rareté à l'échelle de la région.

La **Figure 6** montre les relations entre le nombre d'espèces d'Odonates observées dans des carrés de 10 x 10 km et la fréquence des visites qui y ont été effectuées depuis 1990. Si on répartit les espèces en trois groupes en fonction de leur rareté régionale (> 80 carrés : 19 espèces très répandues, entre 80 et 25 carrés : 18 espèces communes, < 25 carrés : 29 espèces rares), on observe des évolutions bien différentes qui vont d'une courbe logarithmique avec une asymptote (le nombre total de 19 espèces répandues) vers des courbes apparemment exponentielles.



**Figure 7. Evolution du nombre d'espèces de trois groupes d'espèces de la Flore et du nombre total d'espèces observées dans les carrés IFBL wallons 4 x 4 km.**

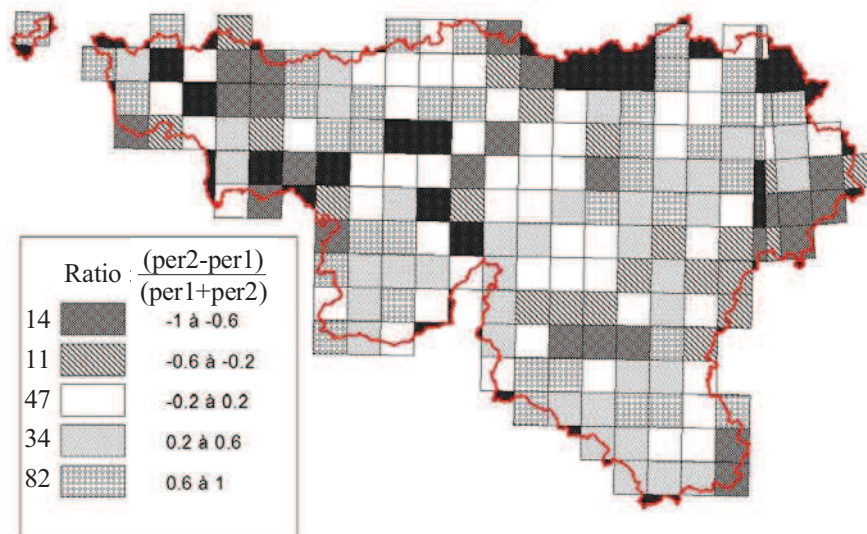
Des profils similaires sont observés dans d'autres groupes biologiques et notamment la Flore (données de l'Atlas de 1979, Van Rompaey & Delvosalle, 1979). Pour ce dernier groupe (**Figure 7**), comme on ne dispose pas d'un indicateur de la fréquence des visites ou de l'exhaustivité de l'échantillonnage, on utilise le nombre total d'espèces comme indicateur indirect (cfr Figure 2). Le graphique montre bien que les carrés caractérisés par un faible nombre d'espèces sont dominés par les espèces répandues. Cela peut être évidemment dû à la diversité des habitats présents dans un carré mais aussi tout simplement à l'intensité des visites dans un carré.

Les deux graphiques confirment bien que plus on échantillonne un carré, plus on observe d'abord des espèces très répandues, puis des espèces communes et enfin des espèces rares. Par définition, les espèces n'ont donc pas la même probabilité d'être observées et il faut en tenir compte dans l'analyse. Il est donc alors difficile d'admettre que la disparition d'une espèce rare compense l'observation d'une espèce répandue dans les analyses de tendance et il est nécessaire d'avoir des indicateurs de l'intensité de l'échantillonnage associés à la récolte de données biologiques.

### **La variabilité géographique de l'échantillonnage**

L'importance de la probabilité de rencontre est essentiellement due au fait qu'entre les larges périodes qui caractérisent généralement les Atlas, la manière dont on échantillonne les carrés ou les mailles a fort évolué. Cela est surtout vrai pour les groupes biologiques qui ne font pas l'objet de prospections systématiques de mailles comme on essaye de le faire avec les Oiseaux ou la Flore. Il est par exemple évident en Wallonie que pour les papillons, les Coléoptères Carabides ou les libellules, le nombre de carrés visités après 1980 ou 1990 est beaucoup plus important qu'il ne l'était avant. La **Figure 8** montre que si on calcule un index symétrique de rapport d'échantillonnage comme  $(per2 - per1) / (per2 + per1)$  avec  $per1$  = nombre d'espèces à la période 1 (ici < 1990) et  $per2$  = nombre d'espèces à la période 2 (ici > 1989), on confirme que de nombreux carrés ont été mieux échantillonnés après 1989.





**Figure 8. Variation géographique de l'échantillonnage dans le temps (avant et après 1990) montrant que le nombre de carrés qui ont augmenté en nombre d'espèces est beaucoup plus important (116 carrés > 0.2) que le nombre de carrés avec moins d'espèces (25 carrés < -0.2).**

Comme ces nouveaux carrés ne sont pas nécessairement échantillonnés de manière très intense et que la probabilité de rencontre favorise largement les espèces très répandues, il est très probable que les espèces répandues soient largement dominantes dans les observations des périodes récentes et que donc, elles ne peuvent servir de référence ou que la somme totale des observations au cours d'une période n'est pas un indicateur correct pour le rapport d'échantillonnage.

Ce problème potentiel est sans doute moins manifeste dans les programmes d'inventaires de mailles qui sont réalisés systématiquement mais à la seule condition que l'intensité de l'échantillonnage soit similaire au cours des deux périodes. C'est en pratique rarement le cas, au moins dans certaines régions.

### **La solution : tenir compte de l'échantillonnage !**

Il est donc pratiquement incontournable de mesurer l'intensité de l'échantillonnage qui a été réalisé dans les différentes mailles au cours des différentes périodes étudiées. Cette évaluation de l'échantillonnage impose que les observations soient stockées de manière assez détaillée afin d'être capable d'identifier la proportion des mailles géographiques qui ont été inventoriées (toponyme, polygone, site, ...) ainsi que le nombre de visites effectuées ou les techniques d'échantillonnage réalisées (inventaire systématique, relevé partiel, observation ponctuelle). Ces informations devraient figurer dans les bases de données biologiques au même titre que les résultats de l'échantillonnage, c'est-à-dire la liste des espèces observées. Il est donc nécessaire que la structure des bases de données puisse gérer cette information. C'est par exemple le cas de la base de données Data Fauna Flora (Barbier et al, 2000) où la structure des trois tables principales pour décrire des observations suppose de définir des localisations géographiques (table des stations) dans lesquelles on effectue des échantillonnages (table des conditions) pour lesquels enfin, on donne les listes d'espèces et informations associées (table des observations).

Dans bien des cas, on ne dispose pas de ces données descriptives de l'échantillonnage dans les anciens relevés qui servent de référence actuellement, mais elles sont pourtant nécessaires pour être capable dans le futur de réaliser des évaluations des tendances qui seront les plus correctes possible.

Une approche espèce par espèce en évaluant l'échantillonnage se dessine aussi comme étant la plus précise et la plus proche de la réalité de terrain. En pratique, l'idée de base est de sélectionner toutes les mailles où une espèce a été observée et de ne garder pour les comparaisons entre périodes que les mailles qui ont été échantillonnées aux deux périodes de manière supposée similaire. Par exemple, Dufrene et al. (2003) ont utilisé cette technique pour les papillons et les libellules en ne prenant en compte que les visites réalisées pendant la période de vol (définie par le jeu de données) de chacune des espèces.

Le tableau suivant montre les différences de résultats obtenus pour une espèce répandue, le Paon du jour (*Inachis io*). Dans le cas de l'approche relative, comme cette espèce pourtant très répandue a été relativement moins observée en deuxième période (après 2000), elle est considérée comme en régression. Or, si on utilise l'approche spécifique, l'espèce apparaît comme stable car elle a été observée dans 167 carrés différents avant le 1<sup>er</sup> janvier 2000 et 159 carrés différents après et le nombre de carrés échantillonnés de la même manière est de 225 carrés.

Ratio relatif	Période 1	Période 2
<i>Inachis io</i>	328	665
Autres espèces	10685	24577

Chi<sup>2</sup> = 4.3    P() = 4%

**Régression**

Ratio spécifique	Période 1	Période 2
<i>Inachis io</i>	167	159
Inventoriés	225	225

Chi<sup>2</sup> = 0.2    P() = 66%

**Stabilité**

Tableau 1. Comparaison de l'approche relative comparant l'évolution des nombres de carrés 1x1 km occupé par *Inachis io* par rapport à ceux occupés par les autres espèces de papillons et de l'approche spécifique en ne sélectionnant que les carrés 1x1 où l'espèce a été mentionnée aux deux périodes et qui ont été échantillonnés correctement (dans le cas présent pendant la période de vol) aux deux périodes.

Il ressort aussi de ces analyses que seulement environ 50% des observations sont effectivement utilisables car, soit très souvent on ne repasse pas aux mêmes endroits qu'auparavant, soit le nombre de nouveaux carrés échantillonnés est très important. Pour maximiser la qualité des futures comparaisons, il sera nécessaire d'organiser les prospections pour maximiser la couverture des populations connues des espèces les plus sensibles. C'est ce qui est en cours dans les programmes d'inventaire et de surveillance des papillons et des libellules en Wallonie (<http://biodiversite.wallonie.be/especes/>).



## Conclusions

Cette contribution n'avait pour but que d'essayer de révéler les biais potentiels des jeux de données utilisés lors de la définition des tendances dans la répartition d'espèces et notamment dans la définition des listes rouges. On ne peut que conseiller de structurer au mieux les programmes d'inventaires de terrain afin qu'ils soient les plus comparables possibles dans la manière dont on a inventorié au cours des périodes étudiées. Les bases de données doivent de leur côté permettre de gérer les informations relatives à l'intensité de l'échantillonnage. On ne peut que se méfier des comparaisons utilisant un rapport relatif commun à toutes les espèces, surtout quand l'intensité des inventaires est loin d'être comparable. Si c'est le cas, il y a un risque non négligeable d'identifier des espèces comme étant menacées alors qu'elles ne le sont pas nécessairement. Toutefois, les populations d'espèces rares sont souvent les mieux surveillées et il est facile de contrôler l'existence ou non de tendances à la régression par un protocole de suivi adapté.

## Remerciements

Les données utilisées pour réaliser les différentes analyses de synthèse résultent, pour les papillons et les libellules, de programmes d'inventaire et de surveillance de l'état de l'environnement wallon financés par la Région wallonne. Nous profitons de cette occasion pour remercier les nombreux bénévoles qui participent aux inventaires et nous ne pouvons que les encourager à continuer à améliorer l'intensité et la qualité de l'échantillonnage.

## Références bibliographiques

Barbier, Y., Rasmont, P., Dufrière, M. & J.-M. Sibert. (2000). Data Fauna-Flora 1.0. Guide d'utilisation. Université de Mons-Hainaut, Mons, Belgique. 106 pp

Desender, K., 1986. Distribution and ecology of Carabid beetles in Belgium (Coleoptera, Carabidae). Parts 1-4. Documents de travail, I.R.Sc.N.B., Bruxelles.

Dufrière, M., Fichet, V. & Goffart, Ph., 2003. Red lists in Wallonia : existing tools and their limits. Report of a communication to EIS - Colloquium organized in Cardiff (UK) September 2003.

IUCN, 2001. Catégories et Critères de l'UICN pour la Liste Rouge : Version 3.1. Commission de la sauvegarde des espèces de l'UICN. UICN, Gland, Suisse et Cambridge, Royaume-Uni. ii + 32 pp.

Stroot, Ph. & Depiereux, E., 1989. Proposition d'une méthodologie pour établir des "Listes Rouges" d'Invertébrés menacés. *Biological Conservation*, 48 : 163-179.

Maes, D. and van Swaay, Ch.A.M., 1997. A new methodology for compiling national Red Lists applied to butterflies (Lepidoptera, Rhopalocera) in Flanders (N-Belgium) and the Netherlands, *Journal of Insect Conservation*, 1 : 113-124.

Van Rompaey, E. & Delvosalle, L. 1979. Atlas de la Flore belge et luxembourgeoise. Jardin Botanique National, Meise.