



ELSEVIER

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Comptes Rendus Biologies

www.sciencedirect.com

Biodiversité/Biodiversity

Effort d'échantillonnage et atlas floristiques – exhaustivité des mailles et caractérisation des lacunes dans la connaissance

*Sampling effort and floristic atlases: Survey completeness of localities and description of knowledge gaps*Jeanne Vallet ^{a,*}, Maëlle Rambaud ^b, Loïc Coquel ^a, Laurent Poncet ^c, Frédéric Hendoux ^b^a Conservatoire botanique national du Bassin parisien – délégation Sarthe, UMS 2699 CNRS-MNHN inventaire et suivi de la biodiversité, Muséum national d'histoire naturelle, 43, rue de l'Estérel, 72000 Le Mans, France^b Conservatoire botanique national du Bassin parisien, UMS 2699 CNRS-MNHN inventaire et suivi de la biodiversité, Muséum national d'histoire naturelle, 61, rue Buffon, 75005 Paris, France^c Service du patrimoine naturel, Muséum national d'histoire naturelle, 36, rue Geoffroy-St-Hilaire, CP 41, 75231 Paris cedex 05, France

I N F O A R T I C L E

Historique de l'article :

Reçu le 29 mai 2012

Accepté après révision le 22 novembre 2012

Disponible sur internet le 25 décembre 2012

Mots clés :

Exhaustivité

Effort d'échantillonnage

Priorités de prospection

Inventaires floristiques

Estimateurs Jackknife

R É S U M É

Les atlas floristiques sont des outils centraux pour les politiques de conservation de la flore même si l'effort d'échantillonnage est rarement homogène sur les territoires concernés. Ce travail vise à développer une méthode pour estimer le taux d'exhaustivité des unités géographiques d'échantillonnage d'atlas floristiques. Il propose également un outil d'aide à la prospection pour parer à l'hétérogénéité de l'effort d'échantillonnage et optimiser l'efficacité de nouveaux inventaires. Une synthèse bibliographique couplée à des tests sur un jeu de données a abouti au choix d'un estimateur non paramétrique, le Jackknife 1, pour estimer la richesse réelle des unités géographiques d'échantillonnage. Le nombre de données de chaque unité géographique d'échantillonnage est utilisé comme estimateur de l'effort d'échantillonnage. Le rapport entre la richesse observée et cette richesse estimée donne le taux d'exhaustivité de l'inventaire de chaque unité géographique d'échantillonnage. Dix-huit variables ont ensuite été sélectionnées pour décrire les inventaires et orienter les futures prospections ; ces variables renseignent les localisations, les périodes et les espèces à privilégier pour de futurs inventaires.

© 2012 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

A B S T R A C T

Keywords:

Survey completeness

Sampling effort

Survey priorities

Floristic inventories

Jackknife estimators

Floristic atlases have an important input to flora conservation planning even though their data quality varied greatly across countries. This study aimed to assess survey completeness of cells of floristic atlases. Then, a surveying guide is designed to overcome as efficiently as possible sampling biases. A review and analyses on a wide dataset were carried out to select an estimator of the true species richness of surveyed cells. The Jackknife 1, a non-parametric estimator, appeared as the best compromise for regional floristic atlases. The number of records in each cell was used as an estimator of sampling effort. The ratio between the observed species richness and the estimated species richness measures the completeness of inventories in each surveyed cell. Eighteen variables were selected to describe current inventories and design new surveys. These variables highlight locations, periods and species to be given priority in future studies.

© 2012 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

* Auteur correspondant.

Adresse e-mail : jvallet@mnhn.fr (J. Vallet).

Abridged English version

Floristic atlases are of central importance to provide information and conserve biodiversity. Nevertheless, data quantity and quality (sampling effort, sampling methods standardization, data validation...) are often heterogeneous across localities in such projects. Thus, quality protocol assessments must be set up before using this type of information for basic and applied purposes. The discrimination of localities that can be considered relatively well surveyed from those not sufficiently surveyed is a key step in this protocol and can be attained by calculating survey completeness using different estimators. Firstly, this study aimed to select the most suitable method to assess survey completeness of localities mapped in floristic atlases. Secondly, we developed a surveying guide to describe recorded data of each locality in order to identify knowledge gaps and direct future samplings toward periods, places and habitats which may be the most favorable to find unrecorded species.

Our approach consisted in a bibliographic review to identify the main methods suggested to assess survey completeness. Then, we tested the most relevant of them on an atlas of 2459 cells of 5 km × 5 km from the National Botanical Conservatory of the Paris Basin. We compared their results and the correlations between them.

Four strategies were identified to assess the survey completeness: modelization of the species-area relationship; modelization of species richness by using environmental variables; modelization of species accumulation curves and use of non-parametric estimators. Non-parametric estimators are based on the postulate that the number of unrecorded species in a locality depends on the number of species rarely recorded. The literature shows that non-parametric estimators give the best results in the majority of cases and their conditions of application match well with floristic atlases characteristics. For their calculation, we used the number of records of each cell as sampling effort estimator. Four non-parametric estimators were applied to the studied dataset; their results in terms of survey completeness were highly correlated. A non-parametric estimator, the Jackknife 1, appeared to be the most suitable for estimating survey completeness of each cell of the studied atlas. We then introduced a surveying table containing 18 variables to describe current data in each atlas cell. A first serie of variables underlined for each atlas cell the least sampled areas or habitats. A second serie highlighted the least sampled months or a lack of regularity in sampled years. Finally, other variables focused on recorded species by examining which proportion of common species was recorded and the possible presence of a gap between the number of rare species recorded in the past and the number currently recorded.

The selected method to assess survey completeness can be broadly applied to other floristic atlases because the required data for its calculation and the required conditions of application are fulfilled in most floristic atlases. In the same way, the variables selected to describe data in each atlas cell sum up the different characteristics of floristic surveys. We hope these tools will favour a more

efficient use of floristic atlases in both basic and applied biodiversity conservation.

1. Introduction

La connaissance de la distribution des taxons est une information de première importance pour connaître la biodiversité et mettre en place des stratégies de conservation [1,2]. Une approche classique d'acquisition de données sur la distribution d'espèces est d'inventorier des unités géographiques d'échantillonnage (maillages Universal Transverse Mercator [UTM] de diverses résolutions ou limites administratives type communes) et de réaliser des inventaires des espèces au sein des différents milieux présents dans ces unités. Les cartes de distribution en résultant sont classiquement publiées sous forme d'atlas ou diffusées sur Internet [1,3]. Dans ce cadre, les inventaires de type atlas sont souvent un outil central des démarches de conservation afin d'évaluer quels sont les taxons les plus rares ou les plus menacés. Cependant, la qualité des données disponibles pour de telles démarches n'est souvent pas homogène sur l'ensemble des territoires concernés [4]. Il est donc essentiel de connaître les biais existant dans le degré de connaissance des territoires puis de chercher à améliorer les connaissances sur les zones les moins inventoriées.

Le taux d'exhaustivité des inventaires, c'est-à-dire le ratio entre la richesse observée et la richesse maximale attendue, peut être utilisé pour estimer le biais existant dans le niveau de connaissance de différentes unités géographiques d'échantillonnage. Pour estimer la richesse maximale attendue, un grand nombre de méthodes ont été développées mais il n'existe pas de méthode universellement reconnue [5,6]. Le choix de la méthode va dépendre de la biologie du groupe taxonomique étudié, de l'échelle spatiale considérée et de la méthode de recueil des données. Un des objectifs de ce travail est de déterminer quelle est la méthode la mieux adaptée pour estimer l'exhaustivité d'inventaires floristiques réalisés dans le cadre de projets d'atlas départementaux ou régionaux. La méthode choisie se devra d'être applicable quelle que soit l'unité géographique d'échantillonnage de l'atlas (mailles UTM, communes...); elle ne devra s'appuyer que sur des données facilement disponibles pour ce type de projet d'atlas et ne devra être mise en œuvre qu'avec des méthodes et des outils sans contraintes d'accessibilité.

En France, la réalisation d'atlas floristiques est une mission centrale des conservatoires botaniques nationaux (Loi Grenelle II, article 48, chapitre 4, section 3). Le Conservatoire botanique national du Bassin parisien (CBNBP) est une de ces structures dont le territoire d'intervention est représenté sur la Fig. 1. Suite à la publication d'atlas communaux sur la majorité de son territoire [7–14], le CBNBP met actuellement en place une nouvelle génération d'inventaires sur son territoire afin d'assurer le renouvellement des données sur la base de mailles Lambert 93 5 km × 5 km [18]. La base de données du CBNBP et ce projet de nouvelle génération d'inventaires ont servi de support pour développer un outil d'aide à la prospection qui permet :

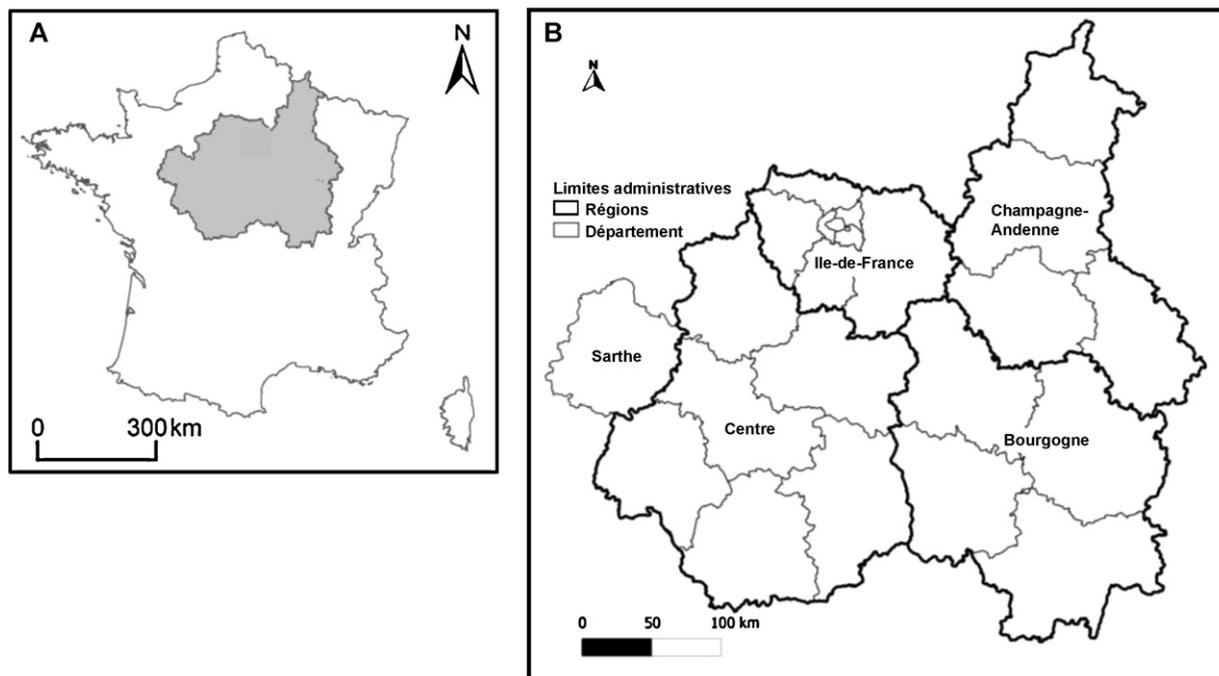


Fig. 1. Cartes représentant en (A) en gris, le territoire d'agrément du Conservatoire botanique national du Bassin parisien (CBNBP) au sein de la France, en (B) la localisation des quatre régions administratives et du département composant le territoire d'agrément. La suite du travail portera sur le département de la Sarthe et les régions de l'Île-de-France et du Centre.

- de hiérarchiser l'ordre de prospection des mailles en fonction du taux d'exhaustivité de leurs inventaires ;
- de décrire les caractéristiques géographiques, temporelles et taxonomiques des inventaires existants afin d'optimiser les futures prospections et acquisitions de données au sein des unités géographiques d'échantillonnage.

2. Matériel et méthodes

2.1. Données

Classiquement, dans le cadre de programmes d'atlas floristiques, plusieurs relevés sont répartis au sein des unités géographiques d'échantillonnage de façon à restituer une image fidèle de la composition floristique des principaux grands milieux présents (ou complexes d'habitats tels que les forêts, prairies, etc.) qu'ils soient semi-naturels ou rudéraux. Les botanistes enregistrent tous les taxons qu'ils observent au sein de relevés qui sont cartographiés puis gérés sous système d'information géographique. Les données recueillies pour chaque relevé sont les données de base : lieu/date/observateur/espèce mais aussi des données sur le ou les milieux dans le(s)quel(s) le relevé est effectué. Le référentiel utilisé pour recueillir cette information est le référentiel Corine Biotope [15,16] ; le degré de précision généralement utilisé dans cette typologie hiérarchique est au minimum de deux chiffres. Les codes à un chiffre présents sur le territoire du CBNBP (CC1 à CC8) sont recensés dans le [Tableau 1](#). Ces inventaires constituent le cœur de la base de données du CBNBP dénommée FLORA.

À ces derniers, s'ajoutent des données acquises dans d'autres cadres. Il y a des données transmises par le réseau de bénévoles du conservatoire et d'autres qui sont acquises via d'autres programmes d'inventaires (exemples : connaissance des espèces rares et des habitats ; inventaire d'espaces remarquables...). Des données bibliographiques, notamment historiques, issues de la saisie de publications naturalistes, d'herbiers ou de rapports d'études sont aussi intégrées à la base. Il en résulte une base de données conséquentes de près de cinq millions de données mais assez hétérogène quant à l'origine des données, aux protocoles de recueil et à l'effort d'échantillonnage [17].

Dans le but d'actualiser les données de cette base, les données considérées contemporaines seront uniquement les données postérieures à l'an 2000. C'est l'exhaustivité des inventaires floristiques de ce jeu de données que l'on cherchera à évaluer. Les données antérieures à l'an 2000 seront considérées « historiques ». Les données de la base de données du CBNBP ont été projetées sur le réseau de mailles 5 km × 5 km basé sur le système de projection Lambert 93 [18]. Les règles de projection pour attribuer les données aux mailles sont les suivantes :

- les données des polygones inclus dans les mailles sont automatiquement attribuées à la maille ;
- les données des polygones à cheval sur plusieurs mailles sont attribuées à la maille qui contient le plus important pourcentage de recouvrement du polygone à condition que le recouvrement du polygone sur la maille soit d'au moins 25 % ;
- pour les données communales non géoréférencées, les données sont rattachées au polygone constitué par la

Tableau 1

Variation de description des données existant au sein de chaque maille. La dernière colonne explicite la façon dont chaque variable doit être interprétée en termes d'orientation des prospections.

	Abréviation	Description de la variable	Objectif pour l'orientation des prospections
Où ?	% mailles de 1 km ² avec relevé	Proportion de mailles 1 km × 1 km (25 par maille de 5 km × 5 km) qui contient un polygone d'inventaire	Description de la répartition des relevés au sein de la maille pour repérer si certains secteurs de la maille n'ont pas été prospectés
	Nombre data CC2 : milieux aquatiques Nombre data CC3 : landes, fruticées, pelouses, prairies Nombre data CC4 : forêts Nombre data CC5 : tourbières et marais Nombre data CC6 : rochers, éboulis, sables Nombre data CC8 : agricole et artificiel	Nombre de données pour chacun des habitats décrits par des codes Corine Biotope à la précision d'un chiffre	Identification des types d'habitat sous-inventoriés
Quand ?	Nombre data printemps 1 : 1 mars–15 mai	Nombre de données recueillies entre le 1 ^{er} mars et le 15 mai	Période permettant la détection des espèces vernales
	Nombre data printemps 2 – 16 mai–30 juin	Nombre de données recueillies entre le 16 mai et le 30 juin	Période favorable à la détection d'un maximum d'espèces dont certaines ne sont identifiables qu'à cette saison (orchidées notamment)
	Nombre data été – 1 ^{er} juillet–31 août	Nombre de données recueillies entre le 1 ^{er} juillet et le 31 août	Période favorable à la détection des espèces à floraison estivale
	Nombre data automne – 1 ^{er} septembre–31 octobre	Nombre de données recueillies entre le 1 ^{er} septembre et le 31 octobre	Période permettant la détection des espèces les plus tardives notamment les espèces liées aux milieux qui s'exondent ainsi que les espèces non indigènes originaires de contrées plus chaudes
	Nombre data Biblio (< 2000)	Nombre de données issues d'observations antérieures à l'an 2000	Description de l'âge des données et de la fréquence des inventaires. Ces variables doivent permettre de prioriser les mailles afin d'assurer un renouvellement des inventaires. Pour un même taux d'exhaustivité, les mailles n'ayant pas de données récentes sont prioritaires sur des mailles dont les données sont plus récentes. Ces indices peuvent aussi encourager la saisie de données bibliographiques lorsqu'elles existent afin d'améliorer la connaissance historique de la maille
	Nombre data 2000–2005	Nombre de données récoltées entre 2000 et 2005	
	Nombre data > 2005	Nombre de données postérieures à 2005	
	% années visitées post 2000	Proportion d'années après 2000 pour lesquelles il existe au moins une donnée	
Quelles espèces ?	Nombre espèces rares tot	Nombre d'espèces rares observées toutes périodes confondues dans la maille. La liste des espèces considérées comme rares correspond aux 795 espèces les plus rares à l'échelle du territoire d'agrément du CBNBP [41]	Les variables sur les espèces rares permettent d'envisager des prospections spécifiques sur certaines espèces rares lorsque la différence entre le nombre d'espèces rares connues toutes périodes confondues est largement supérieure au nombre d'espèces rares observées postérieurement à 2000
	Nombre espèces rares post 2000	Idem mais seules les observations postérieures à l'an 2000 sont comptabilisées	
	% espèces ubiquistes post 2000	Proportion des 85 espèces les plus fréquentes du territoire d'agrément du CBNBP [41] présentes dans la maille	Une faible proportion d'espèces ubiquistes peut indiquer que certains habitats et notamment les habitats les plus communs sont sous-prospectés

commune et les règles précédemment énoncées sont appliquées à celui-ci pour le rattachement de ses données à une maille.

Ensuite, afin de limiter certains biais et avoir des données les plus homogènes possibles, nous avons réalisé au préalable une sélection des données

(1 donnée = 1 espèce × 1 lieu × 1 date × 1 observateur) par rapport :

- au niveau taxonomique : dans FLORA les taxons peuvent être renseignés à différents niveaux (genre, groupe, espèce, sous-espèce, variété, forme). Les données sont actuellement gérées sous le référentiel national français

Taxref v.3 [19]. Nous avons choisi le rang de l'espèce en éliminant les données renseignées au niveau supraspécifique et en compilant les données infraspécifiques au rang spécifique. En effet, le mélange de taxons de rang spécifique et infraspécifique pourrait conduire à compter en double un même taxon. Les espèces hybrides ont également été écartées. Certaines espèces posant des difficultés d'identification ont été regroupées en sections plus largement reconnues (*Rubus* gr. *fruticosus*, *Taraxacum* gr. *ruderalia*...). Ces regroupements permettent de lisser d'éventuels biais liés à la variabilité des compétences botaniques des différents prospecteurs ;

- au statut d'indigénat : seules les espèces indigènes et naturalisées ont été conservées car les espèces subspontanées, accidentelles et cultivées, restent ponctuellement relevées. Les définitions des statuts d'indigénat sont données dans Filoche et al. [20].

Les données qui seront présentées dans cette étude couvrent trois délégations du territoire du CBNBP : deux délégations régionales, l'Île-de-France et le Centre, et une délégation départementale, la Sarthe (Fig. 1), soit un territoire de 57 775 km².

De ces sélections résulte un jeu de données de 2 145 362 données pour le territoire considéré (Sarthe, Île-de-France, Centre) réparti sur 2459 mailles 5 km × 5 km.

2.2. Revue des grands types de méthodes pour estimer l'exhaustivité d'inventaires

Le taux d'exhaustivité d'une unité géographique d'échantillonnage est le rapport entre la richesse observée et la richesse maximale attendue de l'unité géographique d'échantillonnage multiplié par cent pour obtenir un pourcentage. L'enjeu est d'estimer cette richesse maximale attendue qui sera nommée dans la suite de l'article R_{est} . Les différentes méthodes proposées pour calculer la richesse maximale attendue d'un inventaire ont été recensées et leur cadre d'application cerné afin d'identifier lesquelles pourraient être pertinentes pour des atlas floristiques. Ainsi, une recherche bibliographique des mots clés *completeness*, *biodiversity inventories* a été effectuée dans la base de données Web of knowledge (<http://wokinfo.com/>). En premier lieu ont été examinés les articles méthodologiques (en particulier ceux comparant différentes méthodes d'estimation de l'exhaustivité des inventaires) et les articles portant sur des jeux de données de type atlas. Cette recherche a été complétée par des références citées au sein de ce premier lot d'articles.

2.3. Orientation des prospections au sein des unités géographiques d'échantillonnage

Une liste de variables permettant de répondre aux questions suivantes a été dressée :

- où mener les nouveaux inventaires au sein de chaque unité géographique d'échantillonnage ?
- à quelle période les mener ?
- quelles espèces cibler (communes, rares) ?

Lorsque plusieurs variables pouvaient être utilisées pour traduire une même information, la variable la plus aisée à interpréter pour guider les botanistes sur le terrain a été retenue.

2.4. Analyse des données

Suite à la revue bibliographique, les méthodes semblant pertinentes pour calculer l'exhaustivité des inventaires de mailles d'atlas floristiques ont été appliquées au jeu de données du CBNBP. En effet, le choix d'une méthode d'estimation de l'exhaustivité des inventaires est extrêmement dépendant des données et de leurs modes d'acquisition. Colwell et Coddington [45] avancent que la meilleure approche consiste à en tester un certain nombre et à comparer leurs résultats.

Ainsi, quatre estimateurs non paramétriques ont été testés : le Jackknife 1, le Jackknife 2, le Chao et le Bootstrap. Leurs formules sont présentées dans le Tableau 2. Les résultats issus des différentes méthodes ont été comparés à l'aide de statistiques descriptives (moyenne, étendue) puis en examinant leurs corrélations. La corrélation entre les différents estimateurs retenus a été testée par un test de corrélation de rang de Kendall qui ne nécessite pas une distribution normale des données. De plus, ce test non paramétrique permet de mettre en évidence une relation monotone entre deux variables y compris dans les cas où celle-ci n'est pas linéaire [23,43]. Afin d'aider à la visualisation des relations entre les variables, des courbes de lissage ont été ajoutées aux graphiques. Ces courbes de lissage sont issues de la méthode de régression locale LOWESS (paramètre de lissage appliqué de 2/3) ; cette méthode ne pose aucune hypothèse quant à la forme de la relation entre les variables et permet simplement dans le cas présent d'explorer la forme de la relation entre deux variables [44].

Un autre paramètre influençant le choix des estimateurs a aussi été calculé : l'équitabilité de Shannon. Sa formule est la suivante :

$$-\frac{\sum_{i=1}^{R_{obs}} p_i \cdot \ln(p_i)}{\ln(R_{obs})}$$

Tableau 2

Formules des quatre estimateurs non paramétriques appliqués au jeu de données de l'étude.

Nom de l'estimateur	Formule
Chao	$R_{est} = R_{obs} + \frac{n_1^2}{2 \cdot n_2}$
Jackknife d'ordre 1	$R_{est} = R_{obs} + n_1 \frac{N-1}{N}$
Jackknife d'ordre 2	$R_{est} = R_{obs} + n_1 \frac{2 \cdot N - 3}{N} - n_2 \frac{(N-2)^2}{N(N-1)}$
Bootstrap	$R_{est} = R_{obs} + \sum_{i=1}^{R_{obs}} (1 - p_i)^N$

R_{est} : nombre d'espèces de l'unité géographique d'échantillonnage estimé par l'estimateur non paramétrique ; R_{obs} : nombre d'espèces observées dans l'unité géographique d'échantillonnage ; N : nombre de données de l'unité géographique d'échantillonnage ; n_1 : nombre d'espèces observées une seule fois dans l'unité géographique d'échantillonnage ; n_2 : nombre d'espèces observées seulement deux fois dans l'unité géographique d'échantillonnage ; p_i : fréquence d'observation de l'espèce i par rapport au nombre de données N de l'unité géographique d'échantillonnage.

où R_{obs} est la richesse observée dans l'unité géographique d'échantillonnage ; p_i est la fréquence d'observation de l'espèce i par rapport au nombre de données dans l'unité géographique d'échantillonnage.

Les analyses ont été réalisées avec le logiciel libre R 2.11.1 [21]. Plus précisément, le calcul des estimateurs non paramétriques a été réalisé à l'aide du package *vegan* [22]. La fonction *specpool* de ce package a dû être modifiée afin de pouvoir intégrer de gros jeux de données qui ne peuvent être mis sous la forme d'un tableau croisé espèces \times relevés. La version modifiée du script permet l'utilisation d'un jeu de données se présentant sous la forme d'une liste des données (une ligne = une donnée = 1 espèce \times 1 lieu \times 1 date \times 1 observateur) se trouve en [Appendice électronique 1](#). Quant aux tests de corrélation de Kendall, ils ont été réalisés à l'aide de la fonction *corr.test* du package *psych* [42].

3. Résultats

3.1. Revue des grands types de méthodes d'estimation de la richesse spécifique

Dix-neuf articles ont été examinés et sont cités ci-après. Quatre types de méthodes existent pour estimer la richesse d'une unité géographique d'échantillonnage à partir de données de présence-absence :

- extrapolation de la courbe aire-espèces (*species-area relationship* ou SAR). La relation entre la surface échantillonnée et sa richesse spécifique est connue de longue date et se modélise avec une fonction logarithmique [24] ou puissance [25]. Elle permet à partir de données récoltées sur de petites surfaces (classiquement des quadrats) d'estimer la richesse de l'ensemble de la zone d'étude ;
- modélisation de la richesse spécifique en fonction des variables environnementales de la zone d'étude. Le principe consiste dans un premier temps à sélectionner des unités géographiques d'échantillonnage que l'on estime bien connues. À partir de ces unités, un modèle liant des variables environnementales importantes pour la distribution du taxon étudié (pluviométrie, occupation des sols...) et leur richesse spécifique est construit. Ce modèle permet ensuite d'estimer la richesse spécifique des autres unités géographiques d'échantillonnage à partir de leurs variables environnementales.

Les deux autres méthodes utilisent la relation existant entre l'effort d'échantillonnage et la richesse spécifique. L'effort d'échantillonnage est une mesure rarement collectée lors d'inventaire de type atlas de répartition et de toutes les façons difficile à évaluer [1,4]. Ainsi, on peut utiliser le nombre de relevés effectués [28], le nombre de jours ou d'années de prospection [27,29,30], ou encore le nombre de données [2,31]. Quand l'information biologique provient de sources de données issues de différentes méthodologies, la meilleure solution est clairement d'utiliser le nombre de données comme indicateur de l'effort d'échantillonnage [31,37] :

- estimateurs paramétriques (ou *species accumulation curves*: SAC). Le principe de cette méthode est de modéliser l'asymptote de la courbe liant richesse spécifique et effort d'échantillonnage. Le postulat est que plus l'effort d'échantillonnage est important, plus le nombre d'espèces trouvées est élevé, la relation entre les deux étant asymptotique. Différentes méthodes pour calculer cette asymptote existent parmi lesquelles les plus utilisées sont : la fonction de Clench (appelé aussi modèle de Michaelis-Menten) et le modèle exponentiel [26,27] ;
- estimateurs non paramétriques. Ce type d'estimateur est basé sur la forme de la courbe de fréquence des espèces. Le postulat est que le nombre d'espèces détectées un faible nombre de fois diminue avec l'augmentation de l'effort d'échantillonnage. Les estimateurs les plus utilisés sont : Chao 1 et 2, Bootstrap, ICE et Jackknife 1 et 2 [5,26].

L'efficacité des estimateurs paramétriques et non paramétriques a fait l'objet de nombreuses comparaisons dans la bibliographie [5,6,26,28,32–34]. Il apparaît assez clairement dans ces études [5,26,38] que les estimateurs paramétriques donnent des résultats moins précis que les estimateurs non paramétriques car ils sont particulièrement sensibles à l'effort d'échantillonnage.

Ces comparaisons montrent aussi que les différents estimateurs non paramétriques sont plus ou moins performants en fonction de la structure du jeu de données au sein de chaque unité géographique d'échantillonnage, à savoir : la richesse observée ; le nombre de données ; la forme de la courbe de distribution des espèces en fonction de leur nombre d'observation [26,32]. Selon ces études, parmi les estimateurs non paramétriques les plus fréquemment utilisés (Jackknife, Chao, Bootstrap, ICE), les estimateurs Jackknife apparaissent généralement comme le meilleur compromis, à savoir que ce sont eux qui présentent les résultats les moins biaisés et les plus précis dans le plus grand nombre de situations (hétérogénéité des protocoles de recueil de données ; variation de l'effort d'échantillonnage, du taux d'exhaustivité des inventaires ; valeur de la richesse observée). Parmi les estimateurs Jackknife, le choix de l'ordre de l'estimateur dépend de l'ordre de grandeur du taux d'exhaustivité des inventaires ainsi que de l'équitabilité de l'inventaire [26]. Ainsi, pour les faibles taux d'exhaustivité seront privilégiés les estimateurs d'ordre important (3 et plus) tandis que pour des taux d'exhaustivité supérieurs à 50 %, les estimateurs Jackknife d'ordre 1 ou 2 sont plus adaptés [26]. Cependant, Lam et Kleinn [40] ont montré que les estimateurs Jackknife d'ordre supérieur à 3, recommandés par Brose et al. [26] quand le taux d'échantillonnage est faible, pouvaient être fortement imprécis et estimer des richesses inférieures à la richesse observée. Concernant l'équitabilité, Brose et al. [26] indiquent que pour une faible équitabilité (de l'ordre de 0,5 à 0,6), les estimateurs Jackknife d'ordre 2 à 4 sont à privilégier ; pour une forte équitabilité (supérieure à 0,8), on préférera les indices Jackknife d'ordre 1 ou 2.

3.2. Choix d'un estimateur de la richesse spécifique pour des atlas floristiques

3.2.1. Premier tri des différentes méthodes

La première méthode basée sur une extrapolation de la courbe aire–espèces est généralement utilisée à partir de données issues de relevés de même surface. Les relevés classiquement effectués dans les inventaires de type atlas sont de superficie variable, souvent en lien avec la nature de l'habitat dans lequel le relevé est effectué. De plus, dans les atlas, certaines données notamment les données bibliographiques, ne sont pas associées à des relevés mais simplement à l'unité géographique d'échantillonnage. En plus de ces problèmes méthodologiques, la forme de la courbe aire–espèces peut avoir une forme différente selon l'échelle spatiale considérée ; aussi Palmer [5] recommande de ne pas extrapoler les courbes au-delà de la gamme de surfaces étudiées (i.e. se contenter de faire de l'interpolation). Or la surface inventoriée par commune dans le cadre des atlas est en moyenne de $2,6 \% \pm 0,2$ e.s. Cette méthode n'est donc pas appropriée pour des atlas floristiques.

La seconde méthode basée sur une modélisation de la richesse spécifique en fonction de variables environnementales a déjà été appliquée dans le cadre d'inventaires de type atlas [2,3,35]. La qualité du modèle explicatif dépend fortement de la capacité à réunir un lot de données géographiques pertinent pour expliquer la richesse floristique à l'échelle désirée. Pour la flore, ce sont généralement des variables climatique, géologique, pédologique et d'occupation des sols qui sont prépondérantes [36]. Il est assez compliqué actuellement de trouver des jeux de données à une échelle suffisamment fine pour expliquer pertinemment la richesse floristique d'unités géographiques d'échantillonnage de l'ordre de quelques dizaines de km^2 et à une échelle régionale. De plus, cette méthode présente l'inconvénient que la richesse estimée est souvent inférieure à la richesse observée [2,3,35] alors qu'il est bien reconnu que dans de tels inventaires la richesse observée sous-estime toujours la richesse réelle. Aussi, bien qu'envisageable, cette méthode n'est pas facilement applicable dans le contexte de l'étude.

Les estimateurs paramétriques étant plus sensibles à l'effort d'échantillonnage que les estimateurs non paramétriques [5,26,38], les estimateurs non paramétriques apparaissent mieux adaptés pour calculer l'exhaustivité des inventaires des unités géographiques d'échantillonnage d'atlas. Parmi les estimateurs non paramétriques les plus fréquemment utilisés (Jackknife, Chao, Bootstrap, ICE), les estimateurs Jackknife apparaissent généralement être le meilleur compromis [26,32]. Cependant, comme les propriétés des estimateurs non paramétriques varient en fonction de la structure de chaque jeu de données, nous avons comparé les résultats obtenus pour quatre estimateurs non paramétriques afin d'examiner la cohérence de leurs résultats : le Jackknife 1, le Jackknife 2, le Chao et le Bootstrap.

Dans le cas d'un inventaire en cours de réalisation, une difficulté supplémentaire est que le taux d'exhaustivité des inventaires est très variable entre les unités géographiques d'échantillonnage ; cependant, pour pouvoir comparer les unités géographiques d'échantillonnage les unes avec les

autres, il est nécessaire de n'utiliser qu'un seul estimateur quelque soit le taux d'exhaustivité des inventaires. Il n'est donc pas envisageable de choisir un estimateur (et notamment l'ordre d'un estimateur Jackknife) en fonction du taux d'exhaustivité de chaque unité géographique d'échantillonnage. En outre, avec les estimateurs non paramétriques, la richesse estimée ne peut être supérieure à deux fois la richesse observée [39] ; aussi il faut avoir déjà recensé une part relativement importante des espèces de l'unité géographique d'échantillonnage pour appliquer ces méthodes. Le choix d'un estimateur pour les unités géographiques d'échantillonnage faiblement prospectées est donc difficile [40]. Pour parer à ce problème des unités géographiques d'échantillonnage faiblement prospectées, nous avons établi que les estimateurs non paramétriques ne seraient appliqués que sur les mailles de $5 \text{ km} \times 5 \text{ km}$ présentant plus de 250 données. Ce nombre de données correspond environ à un temps de prospection d'une demi-journée (données internes non publiées) ; ce temps de prospection permet d'inventorier trois à cinq grands types d'habitats (forêt, milieux herbacés, milieux humides, milieux rudéraux...), ce qui permet déjà d'avoir un bon aperçu de la composition floristique d'une maille de 25 km^2 . Ce seuil serait à ajuster pour des unités géographiques d'échantillonnage de surface différente. Toute maille de 25 km^2 à moins de 250 données sera considérée d'emblée comme mal connue et prioritaire pour la prospection.

3.2.2. Résultats des tests réalisés sur les données du Conservatoire botanique national du Bassin parisien

L'équitabilité de Shannon est très forte (moyenne = 0,97) et peu variable entre les différentes mailles ce qui oriente, parmi les estimateurs Jackknife, vers ceux d'ordre 1 ou 2 [26].

Le calcul des quatre estimateurs non paramétriques étudiés a été réalisé sur les 2133 mailles $5 \text{ km} \times 5 \text{ km}$ présentant plus de 250 données. Les taux d'exhaustivité moyens (Fig. 2) obtenus par le Jackknife 1 et le Chao sont proches (respectivement 69 % et 67 %) alors qu'ils sont plus faibles pour le Jackknife 2 (60 %) et au contraire plus élevés pour le Bootstrap (83 %). L'étendue des valeurs est la plus importante pour le Chao (94 %) puis viennent le Jackknife 2 (65 %), le Jackknife 1 (39 %) et enfin le Bootstrap (21 %).

Les relations entre la richesse spécifique et les taux d'exhaustivité issus des différents estimateurs avec le nombre de données sont, comme attendues, asymptotiques (Fig. 3). L'asymptote semble se profiler entre 2000 et 4000 données selon les estimateurs. Les corrélations de Kendall entre les différents estimateurs sont hautement significatives (Fig. 3). Le Chao est l'estimateur donnant les résultats les plus différents des trois autres ; sa corrélation avec les autres indices n'est en effet pas linéaire. Cela rejoint les observations de Brose et al. [26] qui montrent que le Chao est biaisé pour des faibles et des forts efforts d'échantillonnage. La corrélation observée la plus forte est celle entre le Jackknife d'ordre 1 et celui d'ordre 2 ($\tau = 0,91$). Les fortes corrélations entre les différents estimateurs indiquent que bien que les valeurs des taux d'exhaustivité puissent être assez différentes selon l'estimateur choisi, les valeurs relatives des taux d'exhaustivité des mailles sont peu affectées par le choix de l'estimateur.

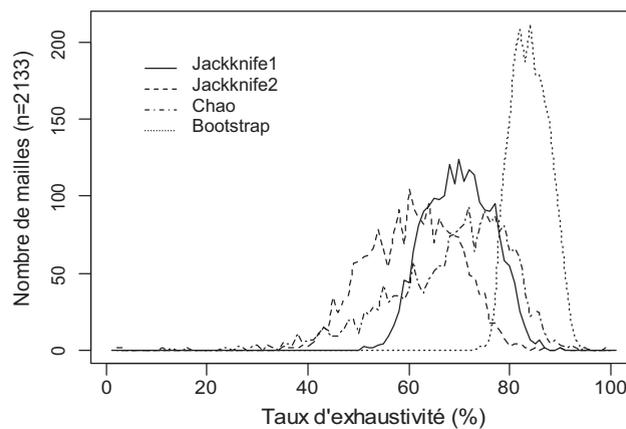


Fig. 2. Courbe de distribution de fréquence des mailles 5 km \times 5 km en fonction de leur taux d'exhaustivité pour chacun des quatre estimateurs non paramétriques étudiés. Sont considérées les mailles ayant au moins 250 données.

Suite à ces résultats, le choix a été fait de retenir l'estimateur qui présente généralement, d'après la bibliographie, la meilleure exactitude : le Jackknife 1 [26,32]. Les valeurs d'exhaustivité obtenues à partir de cet estimateur ont ensuite été classées (méthode des seuils naturels de Jenks) afin d'en faire une représentation spatiale (Fig. 4). Il apparaît une assez forte hétérogénéité dans le niveau de prospection entre les régions, la région Île-de-France apparaissant comme nettement mieux prospectée que les autres.

3.3. Variables sélectionnées pour orienter les prospections au sein des mailles

Dix-huit variables ont été retenues pour décrire les inventaires de façon synthétique (Tableau 1). Dans ce tableau, l'objectif de chaque variable en termes d'orientation des prospections est précisé. Concernant les variables décrivant la saisonnalité des inventaires, les bornes des saisons ont été choisies pour traduire la réalité de la phénologie des espèces mais aussi pour permettre une répartition relativement équilibrée des données entre les quatre classes ce qui permet de mieux discriminer les mailles par la suite. La période du 31 octobre au 1^{er} mars n'a pas été considérée car très peu de prospections se déroulent pendant ces mois d'hiver et il n'y a quasiment pas d'espèces uniquement détectables à cette période.

Un extrait du tableau de prospection pour le jeu de données étudié (Appendice électronique 2) expose concrètement comment le tableau peut être mis en forme en faisant ressortir les terciles des valeurs les plus faibles et les plus fortes afin de guider les futures prospections.

4. Discussion

Ce travail a permis de choisir une méthode pour calculer l'exhaustivité d'une unité d'échantillonnage d'atlas floristique, le Jackknife 1. Ce taux d'exhaustivité permettra d'interpréter les données des atlas au regard du niveau de connaissance de la flore dans l'aire géographique concernée. De plus il permettra, pour la suite du travail de prospection, de déterminer quelles mailles doivent être

prospectées en priorité. Une fois les mailles à prospector choisies, le tableau de variables décrivant l'inventaire existant permet de déterminer la stratégie la plus efficace pour conduire les prospections en précisant les localisations, périodes et espèces à privilégier.

4.1. Intérêts et limites de l'estimation du taux d'exhaustivité des inventaires par le Jackknife 1

Un premier intérêt des estimateurs non paramétriques est que les données nécessaires en entrée pour leur calcul sont extrêmement simples. Pour le Jackknife 1, il suffit de connaître pour chaque unité géographique d'échantillonnage le nombre total d'espèces observées, le nombre d'espèces observées une seule fois et le nombre total de données. Aussi la méthode est largement applicable à d'autres structures menant des projets d'inventaires floristiques. Le second intérêt est que ce type d'estimateur peut aussi être appliqué à d'autres maillages que le maillage 5 km \times 5 km y compris des mailles de taille non identique (exemple : communes). En effet, le postulat sur lequel est basé le Jackknife 1 à savoir que le nombre d'espèces non contactées dans un inventaire dépend du nombre d'espèces contactées une seule fois ne pose pas de problèmes particuliers pour une application sur des unités géographiques d'échantillonnage de superficie variable. Ainsi, en faisant l'hypothèse que la richesse spécifique dépend de la superficie de la commune, pour un même effort d'échantillonnage (en temps passé par exemple) pour deux communes de superficie différentes, l'inventaire de la commune de plus grande superficie devrait posséder plus d'espèces contactées une seule fois d'où un taux d'exhaustivité plus faible que la commune de faible superficie. Un test du Jackknife 1 sur le jeu de données du CBNBP projeté sur le maillage communal a d'ailleurs donné des résultats corrects, c'est-à-dire traduisant les biais d'échantillonnage déjà connus intuitivement (résultats non présentés).

Les limites du Jackknife 1 (ou d'un autre estimateur non paramétrique) sont liées à des biais méthodologiques divers dans le recueil de données résultant en une structure non homogène du jeu de données entre les unités géographiques d'échantillonnage. Le choix de l'estimateur Jackknife 1 est

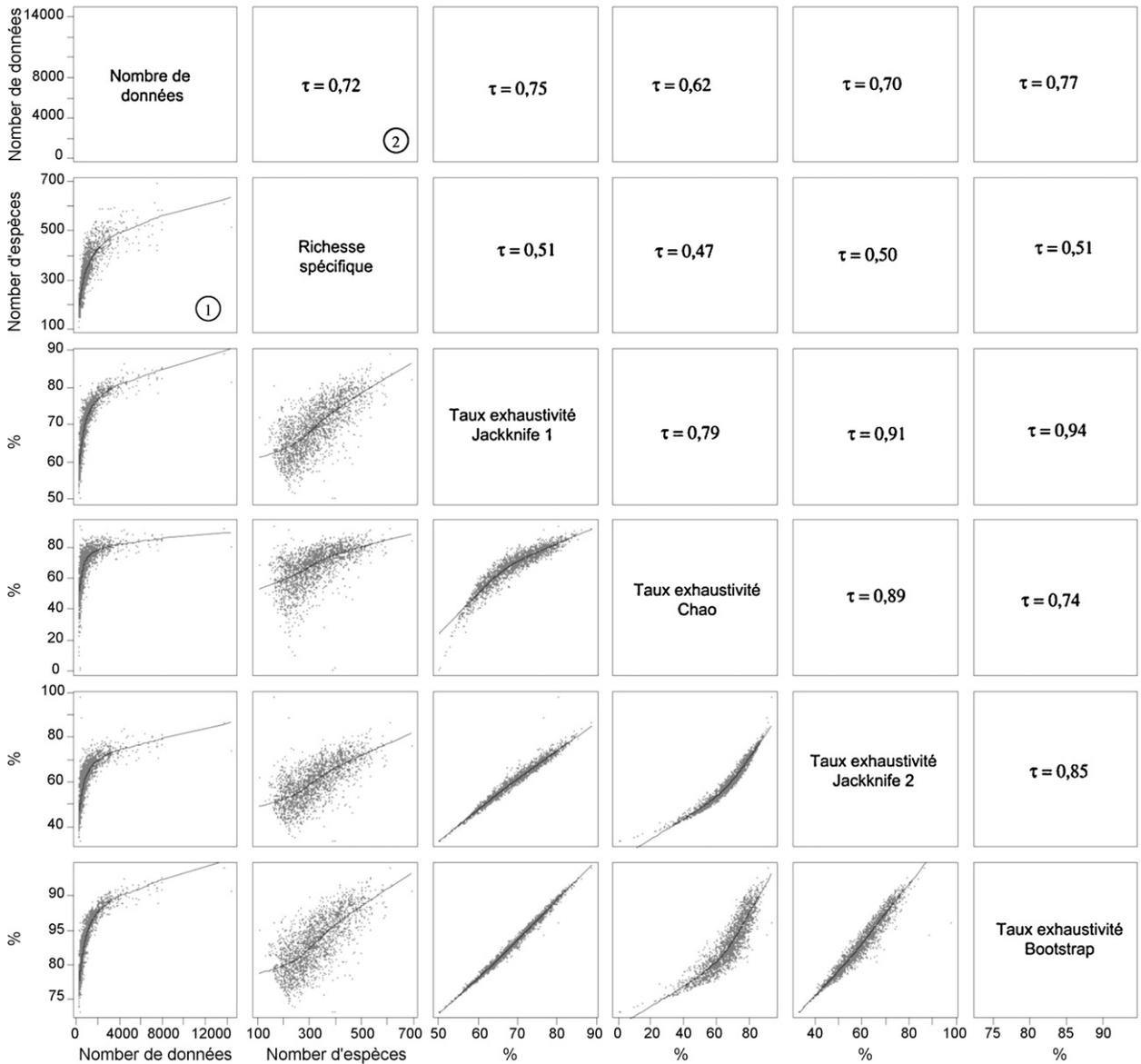


Fig. 3. Relations deux à deux entre le nombre de données des mailles, leur richesse spécifique et les taux d'exhaustivité des inventaires des quatre estimateurs non paramétriques testés. Les coefficients de corrélation de Kendall (τ) sont indiqués sur la partie à droite de la diagonale ; toutes les corrélations sont hautement significatives ($p < 0,001$). Par exemple, le graphique étiqueté 1 représente la relation entre le nombre de données (abscisses) et la richesse spécifique (ordonnées) ; le coefficient de corrélation de Kendall correspondant se trouve dans le cadre numéroté 2. Les courbes noires sont les courbes de lissage (méthode LOWESS, paramètre de lissage de 2/3).

un compromis entre différents paramètres ; aussi les valeurs d'exhaustivité doivent être considérées avec un certain esprit critique. Les taux d'exhaustivité des unités géographiques d'échantillonnage ayant fait l'objet de prospections dans un autre cadre que le protocole atlas standard doivent en particulier être regardés avec précaution. En effet, du fait d'un protocole spécifique, il peut en résulter une structure de jeu de données très particulière pour laquelle le Jackknife 1 est un mauvais estimateur. Par exemple, si les données sont recueillies de façon intensive sur une longue période, l'hypothèse clé sur laquelle se basent les méthodes d'estimation de la richesse, à savoir que chaque unité géographique d'échantillonnage contient un nombre fini

d'espèces, est violée. Un très fort effort d'échantillonnage permet en effet de capturer les espèces qui apparaissent temporairement sur l'unité géographique d'échantillonnage ou celles qui immigrent depuis un habitat adjacent favorable ; la communauté n'est alors plus fermée [4,46]. Quelques unités géographiques d'échantillonnage ayant fait l'objet d'un tel inventaire sont présentes dans le jeu de données du CBNBP ; il apparaît alors effectivement que le Jackknife 1 surestime assez largement la richesse estimée ce qui rejoint la bibliographie [39]. Une autre limite de la méthode soulignée par Aranda et al. [37] est que les naturalistes peuvent avoir tendance à ne pas systématiquement noter les espèces communes pour chaque

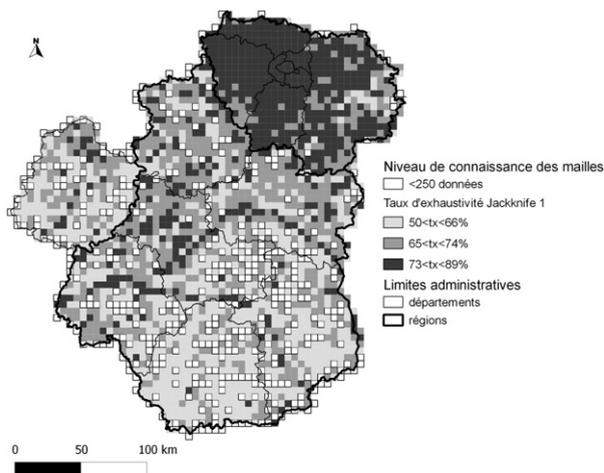


Fig. 4. Représentation spatiale du taux d'exhaustivité des inventaires contemporains (postérieurs à 2000) des mailles 5 km × 5 km sur le département de la Sarthe et les régions Île-de-France et Centre. Les limites des classes du taux d'exhaustivité issu du Jackknife 1 ont été obtenues par la méthode des seuils naturels de Jenks.

relevé, notamment lorsque ces dernières ont déjà été recensées par ailleurs dans l'unité géographique d'échantillonnage. Cette pratique est à éviter au maximum car, pratiquée de façon importante, elle invalide l'utilisation de cette méthode et limite fortement l'exploitation du jeu de données pour d'autres analyses, la forme de la courbe de fréquence des espèces étant une caractéristique de communautés largement reconnue et sur laquelle se basent par conséquent nombre de méthodes statistiques.

L'examen de variables caractérisant la structure du jeu de données en termes spatial, temporel et de nature des espèces, notamment celles présentées dans le tableau d'orientation des prospections (Tableau 1 et Appendice électronique 2), peut aider à repérer d'éventuelles situations pour lesquelles le taux d'exhaustivité issu du Jackknife 1 est mauvais.

4.2. Intérêts et limites des variables d'orientation des prospections au sein des mailles

Cette méthode a l'avantage d'être simple à mettre en œuvre et aboutit à un tableau de variables très synthétique ce qui permet d'en faire un examen rapide avant une prospection. De plus, elle peut être adaptée à la disponibilité des variables recueillies dans le cadre des inventaires. Par exemple, dans le cas où l'information sur les codes Corine Biotope ne serait pas recueillie lors des inventaires, des informations sur les habitats inventoriés pourraient être acquises en utilisant des listes d'espèces indicatrices par grand type d'habitats et, pour chaque liste, de calculer la proportion d'espèces présentes dans chaque unité géographique d'échantillonnage.

La description des données au sein des unités géographiques d'échantillonnage peut être biaisée par des informations incomplètes dans la base de données. Par exemple, les données pour lesquelles la date est incomplète (seulement l'année) ne sont pas prises en compte

pour estimer l'effort d'échantillonnage par saison. Enfin, les informations sur le nombre de données par habitat doivent être réinterprétées par un examen de documents cartographiques. En effet, il peut y avoir deux causes pour expliquer un faible nombre de données par habitat. Par exemple, un faible nombre de données en habitat forestier (code Corine Biotope n° 4) signifie que l'habitat forestier a été sous-inventorié ou que l'habitat forestier est très peu présent voire absent sur l'unité géographique d'échantillonnage. Dans le premier cas, la connaissance de la flore de l'unité géographique d'échantillonnage pourra être améliorée en inventoriant des forêts alors que dans le second cas, cela traduit une réalité de la composition floristique de l'unité géographique d'échantillonnage. Un intérêt d'utiliser la typologie Corine Biotope avec seulement un chiffre de précision est que ces grands types d'habitat (Tableau 1) peuvent être en bonne partie repérés sur des documents cartographiques (cartes topographiques au 25 000^e, orthophotoplans notamment) ; il est donc possible d'interpréter la variable avant d'aller inventorier sur le terrain.

5. Conclusion

Les atlas naturalistes sont des outils essentiels à la mise en place des politiques de biodiversité mais complexes à exploiter car les données sur lesquelles ils sont fondés sont souvent issues de sources diverses et de protocoles assez peu standardisés. Il en résulte souvent un effort d'échantillonnage non homogène. La méthode développée dans ce travail permet d'estimer le taux d'exhaustivité d'unités géographiques d'échantillonnage pour des atlas floristiques. La méthode retenue, l'estimateur non paramétrique Jackknife 1, est fondée sur des données très simples en entrée et pourra donc être appliquée à de nombreux atlas quelque soit l'unité géographique d'échantillonnage sur laquelle ils sont fondés (mailles, communes...).

Cet indicateur permettra en outre d'exploiter les données des atlas à des fins conservatoires (identification de points chauds de diversité...) en ayant conscience des limites des patrons observés en termes de variabilité de l'effort d'échantillonnage. Par ailleurs, cet indicateur permettra aussi d'orienter les futures prospections pour homogénéiser l'effort d'échantillonnage entre les mailles. De plus, les auteurs proposent d'associer à cet indicateur d'autres indicateurs pour décrire le jeu de données et optimiser les futures prospections au sein des unités géographiques d'échantillonnage pour détecter le plus efficacement possible de nouvelles espèces.

Déclaration d'intérêts

Les auteurs déclarent ne pas avoir de conflits d'intérêts en relation avec cet article.

Remerciements

Les auteurs remercient l'ensemble des botanistes et informaticiens du CBNBP qui ont récolté, saisi et aidé à l'exploitation des données. Nous sommes également reconnaissants envers les membres du Groupe de Travail

Flore du CBNBP (Pascal Amblard, Olivier Bardet, Anne Beylot, Rémi Dupré et Sébastien Filoche), ainsi qu'envers des membres du CBN de Brest (Arnaud Cochard, Fabien Dortel, Julien Geslin, Pascal Lacroix et Sylvie Magnanon) pour leurs échanges sur le sujet. Nous sommes également très reconnaissants à Fiona Lehane d'avoir amélioré l'anglais du papier. Les auteurs remercient vivement le lecteur du manuscrit pour ses remarques très pertinentes qui ont permis d'améliorer la qualité de cet article. Ce travail a bénéficié du soutien de la région Pays-de-la-Loire, du FNADT et du FEDER.

Annexe. Matériel complémentaire

Le matériel complémentaire accompagnant la version en ligne de cet article est disponible sur <http://www.sciencedirect.com> et <http://dx.doi.org/10.1016/j.crv.2012.11.005>.

Références

- [1] M.P. Robertson, G.S. Cumming, B.F.N. Erasmus, Getting the most out of atlas data, *Divers. Distrib.* 16 (2010) 363–375.
- [2] D. Sánchez-Fernández, J.M. Lobo, P. Abellán, A. Millán, How to identify future sampling areas when information is biased and scarce: an example using predictive models for species richness of Iberian water beetles? *J. Nat. Conserv.* 19 (2011) 54–59.
- [3] D. Callaghan, P. Ashton, Knowledge gaps in bryophyte distribution and prediction of species-richness, *J. Bryol.* 30 (2008) 147–158.
- [4] M.O. Hill, Local frequency as a key to interpreting species occurrence data when recording effort is not known, *Methods. Ecol. Evol.* 3 (2012) 195–205.
- [5] M.W. Palmer, The estimation of species richness by extrapolation, *Ecology* 71 (1990) 1195–1198.
- [6] E. Cam, J.D. Nichols, J.R. Sauer, J.E. Hines, On the estimation of species richness based on the accumulation of previously unrecorded species, *Ecography* 25 (2002) 102–108.
- [7] G. Arnal, J. Guittet, Atlas de la flore sauvage du département de l'Essonne, Biotope/Muséum national d'histoire naturelle, Mèze (Collection Parthénope)/Paris, 2004.
- [8] S. Filoche, G. Arnal, J. Moret, La biodiversité du département de la Seine-Saint-Denis : atlas de la flore sauvage, Biotope/Muséum national d'histoire naturelle, Mèze (Collection Parthénope)/Paris, 2006.
- [9] D. Pujol, J. Cordier, J. Moret, Atlas de la flore sauvage du département du Loiret, Biotope/Muséum national d'histoire naturelle, (Collection Parthénope)/Paris, 2007.
- [10] O. Bardet, E. Fédoroff, G. Causse, J. Moret, Atlas de la flore sauvage de Bourgogne, Conservatoire botanique national du Bassin parisien, Biotope/Muséum national d'histoire naturelle, Mèze (Collection Parthénope)/Paris, 2008.
- [11] R. Dupré, P. Boudier, P. Delahaye, M. Joly, J. Cordier, J. Moret, Atlas de la flore sauvage du département d'Eure-et-Loir, Biotope/Muséum national d'histoire naturelle, Mèze (Collection Parthénope)/Paris, 2009.
- [12] G. Hunault, J. Moret, Atlas de la flore sauvage du département de la Sarthe, Biotope/Muséum national d'histoire naturelle, Mèze (Collection Parthénope)/Paris, 2009.
- [13] F. Perriat, S. Filoche, J. Moret, Atlas de la flore sauvage du département du Val-de-Marne, Biotope/Muséum national d'histoire naturelle, Mèze (Collection Parthénope)/Paris, 2009.
- [14] S. Filoche, F. Perriat, J. Moret, Atlas de la flore sauvage de Seine-et-Marne, Conseil général de Seine-et-Marne, Deauville, 2010.
- [15] P. Devillers, J. Devillers-Terschuren, J.P. Ledant, CORINE Biotopes Manuel. Habitats of the European Community. Data specifications – Part 2. EUR 12587/3 EN, European Commission, Luxembourg, 1991.
- [16] M. Bissardon, L. Guibal, Corine biotopes. Version originale. Types d'habitats français, École Nationale du Génie Rural des Eaux et Forêts, Nancy, 1997.
- [17] A. Muratet, E. Porcher, J.C. Abadie, L. Poncet, J. Moret, A. Robert, Use of extensive habitat inventories in biodiversity studies, *Biodivers. Conserv.* 18 (2009) 3115–3125.
- [18] Muséum national d'histoire naturelle, Service du patrimoine naturel. INPN, Inventaire National du Patrimoine Naturel, Cartes et information géographique, 2010. Available at: <http://inpn.mnhn.fr/telechargement/cartes-et-information-geographique>. Consulté le 2 mai 2012.
- [19] Muséum national d'histoire naturelle, Service du patrimoine naturel. INPN, Inventaire National du Patrimoine Naturel, référentiels taxonomiques, 2010. Available at: <http://inpn.mnhn.fr/telechargement/referentielEspece>. Consulté le 2 mai 2012.
- [20] S. Filoche, M. Rambaud, S. Auvert, A. Beylot, F. Hendoux, Catalogue de la flore vasculaire d'Île-de-France (rareté, protections, menaces et status). Rapport du Conservatoire botanique national du Bassin parisien, version complète 1a, 2011.
- [21] Development Core Team, R 2.11.1: A language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011.
- [22] J. Oksanen, G. Blanchet, K. Roeland, P. Legendre, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, H.H. Stevens, H. Wagner, Community Ecology Package: « vegan 2.0-2 », R Foundation for Statistical Computing, Vienna, Austria, 2011.
- [23] P. Legendre, L. Legendre, Numerical Ecology, Elsevier, New York, 1998.
- [24] H.A. Gleason, On the relation between species and area, *Ecology* 3 (1922) 158–162.
- [25] O. Arrhenius, Statistical investigations in the constitution of plant associations, *Ecology* 4 (1923) 68–73.
- [26] U. Brose, N.D. Martinez, R.J. Williams, Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns, *Ecology* 84 (2003) 2364–2377.
- [27] S.C. Aranda, R. Gabriel, P.A.V. Borges, J.M. Lobo, Assessing the completeness of bryophytes inventories: an oceanic island as a case study (Terceira, Azorean archipelago), *Biodivers. Conserv.* 19 (2010) 2469–2484.
- [28] J.A. González-Oreja, C. Garbisu, S. Mendarte, A. Ibarra, I. Albizu, Assessing the performance of nonparametric estimators of species richness in meadows, *Biodivers. Conserv.* 19 (2009) 1417–1436.
- [29] C.E. Moreno, G. Halffter, Assessing the completeness of bat biodiversity inventories using species accumulation curves, *J. Appl. Ecol.* 37 (2000) 149–158.
- [30] J. Soberón, R. Jiménez, J. Golubov, P. Koleff, Assessing completeness of biodiversity databases at different spatial scales, *Ecography* 30 (2007) 152–160.
- [31] J.M. Lobo, Database records as a surrogate for sampling effort provide higher species richness estimations, *Biodivers. Conserv.* 17 (2008) 873–881.
- [32] B.A. Walther, J.L. Moore, The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance, *Ecography* 28 (2005) 815–829.
- [33] A. Chiarucci, N.J. Enright, G.L.W. Perry, B.P. Miller, B.B. Lamont, Performance of nonparametric species richness estimators in a high diversity plant community, *Divers. Distrib.* 9 (2003) 283–295.
- [34] J. Hortal, P.A.V. Borges, C. Gaspar, Evaluating the performance of species richness estimators: sensitivity to sample grain size, *J. Anim. Ecol.* 75 (2006) 274–287.
- [35] T. Wohlgemuth, Modelling floristic species richness on a regional scale: a case study in Switzerland, *Biodivers. Conserv.* 7 (1998) 159–177.
- [36] J. Gurevitch, The Ecology of Plants, Sinauer Associates Inc., Sunderland, 2006.
- [37] S.C. Aranda, R. Gabriel, P.A.V. Borges, E.B. de Azevedo, J.M. Lobo, Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores), *Bryologist* 114 (2011) 611–624.
- [38] U. Brose, Estimating species richness of pitfall catches by non-parametric estimators, *Pedobiologia* 46 (2002) 101–107.
- [39] E.P. Smith, G. van Belle, Nonparametric estimation of species richness, *Biometrics* 40 (1984) 119–129.
- [40] T.Y. Lam, C. Kleinn, Estimation of tree species richness from large area forest inventory data: Evaluation and comparison of jackknife estimators, *For. Ecol. Manage.* 255 (2008) 1002–1010.
- [41] M. Rambaud, F. Hendoux, S. Filoche, Vers un indice de rareté robuste hiérarchisant les actions de conservation de la flore, *J. Botanique* 57 (2012) 49–58.
- [42] W. Revelle, Procedures for Psychological, Psychometric and Personality Research: « psych », R Foundation for Statistical Computing, Vienna, Austria, 2011.
- [43] G. Millot, Comprendre et réaliser les tests statistiques avec R : manuel pour les débutants, De Boeck Supérieur, Bruxelles, 2009.
- [44] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.* 74 (1979) 829–836.
- [45] R.K. Colwell, J.A. Coddington, Estimating terrestrial biodiversity through extrapolation, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 345 (1994) 101–118.
- [46] P.B. Adler, W.K. Lauenroth, The power of time: spatiotemporal scaling of species diversity, *Ecol. Lett.* 6 (2003) 749–756.