



Bulletin
de la

SOCIÉTÉ LINNÉENNE DE LYON



Extrapolation des inventaires de biodiversité incomplets : comment estimer au mieux le nombre d'espèces manquantes et prévoir l'effort additionnel d'échantillonnage requis pour réduire ce nombre

Jean Béguinot

Biogéosciences, Université de Bourgogne, F-21000 Dijon - jean-beguिनot@orange.fr

Résumé. – Approcher l'exhaustivité reste l'objectif idéal de tout inventaire ; cependant la difficulté de réaliser des échantillonnages d'espèces sub-complets est appelée à s'accroître à mesure que se présentent des groupes taxonomiques plus difficiles à traiter (par exemple la plupart des groupes de micro-invertébrés) et que, par ailleurs, de multiples priorités d'investigations viennent entrer en concurrence.

Pour autant, exploités de manière appropriée, les échantillonnages incomplets peuvent en dire plus qu'il ne semblerait au premier abord, et il apparaît possible d'en tirer un meilleur parti sous réserve de pouvoir réaliser des extrapolations numériques fiables.

D'ailleurs, les outils adaptés à l'exploitation dans ce sens des échantillonnages incomplets ne manquent pas et pèchent même plutôt par la diversité des solutions proposées et, corrélativement, par la disparité des résultats auxquels elles conduisent. Heureusement, la mise en évidence récente de contraintes mathématiques restreignant les formulations acceptables de la Courbe d'Accumulation des Espèces au fil de l'accroissement de taille de l'échantillon permet désormais de faire une sélection efficace parmi les différentes solutions proposées, en faveur de l'extrapolation la plus fiable.

Il devient ainsi possible (i) d'estimer avec une assez bonne exactitude le nombre d'espèces manquantes ayant échappé à l'échantillonnage et, partant, le niveau de richesse spécifique totale de l'ensemble inventorié ; (ii) d'extrapoler utilement la cinétique (progressivement ralentie) avec laquelle de nouvelles espèces viendraient s'ajouter à la liste déjà acquise par l'échantillonnage en cours, si celui-ci devait être poursuivi. Ceci permettant, du coup, de jauger objectivement le gain à attendre en termes de nouvelles espèces inventoriées, en fonction de l'effort d'échantillonnage supplémentaire à consentir. Toutes informations essentielles pour décider rationnellement de l'opportunité, soit de poursuivre, soit d'interrompre un échantillonnage en cours, compte tenu (i) du ratio gain attendu/effort additionnel requis, (ii) du degré d'incomplétude de l'échantillon en cours (vis-à-vis de la richesse spécifique totale estimée), (iii) de la concomitance possible (et même courante) de priorités d'inventaires concurrentes, soit en d'autres sites, soit en faveur d'autres groupes taxonomiques.

On illustre ici la mise en œuvre de la procédure proposée d'extrapolation de la courbe d'accumulation des espèces permettant de «prolonger» numériquement l'échantillonnage déjà réalisé en considérant un récent inventaire de la faune cécidogène au sein du Parc National des Écrins. Cet inventaire rassemble les échantillonnages, inévitablement plus ou moins partiels, consacrés à chacun des quatre principaux ordres d'arthropodes cécidogènes : Acariens, Trombidiformes, Hémiptères, Diptères, Hyménoptères.

Mots-clés. – Inventaire, échantillonnage incomplet, extrapolation, Courbe d'Accumulation des Espèces, estimation, richesse spécifique, Jackknife, Chao.

Extrapolating the species accumulation curve for incomplete samplings: how to estimate accurately the number of missing species and reliably forecast the extra sampling effort required to reduce this number

Abstract. – While approaching or even reaching total completeness remains the ideal goal of any species inventory, incomplete samplings are deemed to become more often the rule,

as multiple priorities of investigations accumulate and new taxonomic groups are addressed that are significantly more difficult to cope with (such as (micro-) invertebrates, for example).

However, once post-operated appropriately, incomplete samplings may provide far more data than it would seem and it is indeed possible to obtain valuable extra information from a reliable mode of extrapolation of the so-called Species Accumulation Curve.

In fact, numerous tools designed for the exploitation of incomplete samplings have already been proposed in the literature but the concern, with these numerous different tools, is precisely the disparity of the results they provide respectively. Fortunately, the recent derivation of a specific mathematical relationship strongly constrains and restricts the range of relevant formulations of the Species Accumulation Curve. This, in turn, allows for efficient selection among the various tools proposed, thus highlighting the more reliable type of extrapolation of the Species Accumulation Curve beyond the ongoing sampling.

Accordingly, this makes it possible:

(i) to estimate, with fairly good accuracy, the number of missing species that escaped sampling and hence the total species richness of the sampled assemblage;

(ii) to extrapolate the (gradually slowing) pace at which new species are expected to add progressively to the list of already recorded species, according to the level of additional sampling effort and, thus, to objectively gauge the expected gain in terms of newly collected species according to the degree of additional sampling effort required.

All these are essential information to help making a rational decision about the opportunity either to continue or to stop sampling operation, given (i) the ratio between the expected gain in new species and the required additional sampling effort, (ii) the degree of completeness of the current sample (with respect to the estimated total species richness), (iii) the usual co-occurrence of several other competing priorities of investigations, involving either other sites and/or other taxonomic groups to be further surveyed.

Hereafter, I illustrate the implementation of the suggested procedure, aiming at reliably extrapolating the Species Accumulation Curve so as to “extend” numerically already achieved samplings. For this purpose, I consider a recent survey of the gall-forming fauna of the “Ecrins National Park”, located in south-west French Alps, including partial samplings of each of the four main orders of gall-forming arthropods: Acari-Trombidiformes, Hemiptera, Diptera and Hymenoptera.

Keywords. – Survey, incomplete sampling, completeness, extrapolation, Species Accumulation Curve, estimator, species richness, Jackknife, Chao.

1 - INTRODUCTION

1.1 Enjeux et objectifs

Réaliser des échantillonnages (quasi-) exhaustifs demeure bien entendu le but idéal à atteindre pour les inventaires de biodiversité, que ce soit à l'échelon local ou régional. Cependant, à mesure que l'on en vient à considérer des groupes taxonomiques moins classiquement abordés – tels que, notamment, les (micro-) invertébrés – le souhait d'exhaustivité des échantillonnages devient souvent un objectif quasi inaccessible en pratique. Trois raisons au moins contribuent en effet à limiter la complétude des inventaires : (i) les groupes taxonomiques concernés réunissent souvent localement de très nombreuses espèces, lesquelles (ii) sont fréquemment de tailles individuelles minimales, conséquemment plus ou moins malaisément détectables, enfin (iii) les naturalistes à même de déterminer jusqu'au niveau spécifique au sein de ces groupes «peu attractifs» sont comparativement peu nombreux. Toutes raisons qui font que les répartitions locales et régionales des espèces ressortissant à la plupart de ces groupes taxonomiques demeurent souvent fort méconnues, en sorte qu'il est d'autant plus urgent de s'y intéresser davantage, ce que souligne d'ailleurs opportunément Bruno David, actuel Président du Muséum National d'Histoire Naturelle : *«pour de*

nombreux groupes d'invertébrés, on en est encore au stade des spéculations pour l'évaluation de la biodiversité» (DE WEVER & DAVID, 2014).

Or, la façon de conduire puis d'exploiter de tels échantillonnages incomplets amène à se poser, en particulier, les deux importantes questions suivantes :

(i) *comment choisir rationnellement le moment de mettre un terme à un échantillonnage en cours*, sachant que l'objectif souhaitable d'exhaustivité s'avère pratiquement inaccessible ? La question se pose d'autant plus que, d'une part, la détection de nouvelles espèces va naturellement se ralentissant régulièrement, et très sensiblement, à mesure que l'effort d'échantillonnage progresse et que, d'autre part, diverses autres priorités d'inventaires peuvent venir en concurrence. Quand, et sur quelle base rationnelle, décider dès lors d'arrêter (ou poursuivre) un échantillonnage en cours devient donc une question quasi inévitable dans le cadre d'échantillonnages incomplets, par ailleurs amenés à devenir de plus en plus fréquents.

(ii) après avoir plus 'rationnellement' décidé de mettre un terme à l'opération d'échantillonnage en cours, et faute de pouvoir en savoir davantage sur l'identité des espèces ayant ainsi échappé à l'investigation, *comment estimer au moins leur nombre* et, conséquemment, *le degré d'incomplétude* de l'inventaire réalisé ?

À leur tour, ces deux questions pratiques renvoient à une même troisième : la possibilité d'*extrapoler, de manière aussi fiable que possible*, ce que serait la cinétique de collecte de nouvelles espèces en fonction d'efforts supplémentaires croissant d'échantillonnage, autrement dit la possibilité d'*extrapoler* ce qu'on qualifie de «Courbe d'Accumulation des Espèces» et ce, le plus loin possible au-delà de l'effort d'échantillonnage réalisé.

1.2 Etat de l'art en la matière

S'agissant de l'extrapolation partielle ou complète de la Courbe d'Accumulation des Espèces (C.A.E.), une dizaine de modèles empiriques distincts ont été proposés, visant à rendre compte de ce que pourrait être le prolongement de la C.A.E. pour une taille d'échantillonnage au-delà de l'échantillonnage en cours (voir par exemple THOMPSON *et al.*, 2003). Le calage de ces divers modèles est réalisé, en fonction notamment de la richesse spécifique totale estimée de l'assemblage investigué. Les extrapolations ainsi obtenues sont évidemment toutes différentes.

S'agissant maintenant de l'estimation du nombre d'espèces manquantes, et donc de la richesse spécifique totale, différents estimateurs ont également été proposés, chacun avec une formulation bien distincte ; chacun d'eux délivrant donc une estimation différente (voir revue notamment dans COLWELL & CODDINGTON (1994), GOTELLI & CHAO (2013)).

Les solutions ne manquent donc pas, ni pour l'estimation de la richesse spécifique ni pour l'extrapolation de la C.A.E., mais c'est justement la diversité des différentes méthodes proposées et, corrélativement, le manque de cohérence entre les résultats délivrés par ces différentes méthodes qui fait problème. Il est donc fort souhaitable de voir s'il est possible de sélectionner la meilleure d'entre ces solutions, même s'il est entendu qu'en matière d'extrapolation, aucune de ces méthodes n'est idéalement exacte. Ainsi, pour les estimateurs du nombre d'espèces manquantes, de très nombreuses études comparatives ont été conduites, essentiellement sur une base

empirique (POULIN, 1998 ; HERZOG *et al.*, 2002 ; CHIARUCCI *et al.*, 2003 ; FOGO *et al.*, 2003 ; HORTAL *et al.*, 2006 ; UNTERSEHER *et al.*, 2008 ; BASUALDO, 2011). Aussi intéressantes qu'elles soient, ces études n'ont cependant abouti à aucun consensus, comme on pouvait du reste s'y attendre compte tenu de ce que le degré de fiabilité de chaque type d'estimateur dépend fortement de l'allure de la distribution d'abondance des espèces (qui conditionne l'allure de la C.A.E.), laquelle est spécifique à chaque cas d'inventaire (SOBERON & LLORENTE, 1993 ; COLWELL & CODDINGTON, 1994 ; WALTER & MORAND, 1998 ; HELLMANN & FOWLER, 1999 ; BASUALDO, 2011). L'embarras n'est donc pas moindre ici que pour bien choisir un modèle d'extrapolation de la C.A.E. (évoqué plus haut).

Afin de tenter de lever l'ambiguïté et de mieux étayer le choix optimal parmi les (trop) nombreuses et diverses propositions, il apparaît donc approprié d'envisager de substituer aux démarches empiriques précédentes une approche plus théorique, susceptible par là même, d'aboutir à une méthode de choix *cohérente et généralisable*.

La possibilité d'envisager une telle approche théorique est offerte par la mise en évidence récente d'une relation contraignant, en toute généralité, l'expression mathématique de la Courbe d'Accumulation des Espèces $R(N)$ (où R représente le nombre d'espèces observées en fonction de la taille N de l'échantillonnage) : BÉGUINOT (2014, 2015a). Plus précisément, cette relation contraignante met en correspondance la série des dérivées successives de la C.A.E. avec la série des nombres $f_1, f_2, \dots, f_x$ d'espèces observées 1, 2, ..., x fois au cours de l'échantillonnage (singletons f_1 , doubletons f_2 , etc.) (voir Annexe pour la définition de cette relation).

On présente, ci-après, les indications et conclusions pratiques tirées de cette nouvelle approche, en développant plus particulièrement une série d'exemples d'extrapolation et en soulignant les éléments d'appréciation et de décision qui en résultent.

2 - SÉLECTIONNER L'EXTRAPOLATION LA PLUS FIABLE DE LA C.A.E. ET L'ESTIMATION LA MIEUX CENTRÉE DE LA RICHESSE SPÉCIFIQUE TOTALE DE L'ENSEMBLE ÉCHANTILLONNÉ

Rappelons, en préambule, qu'un échantillonnage de biodiversité peut s'envisager et se réaliser de deux manières bien distinctes :

- soit à l'échelon «local» (station), auquel cas l'échantillonnage consiste à noter l'*abondance* de chaque espèce rencontrée (nombre d'individus échantillonnés relevant de chaque espèce), la taille de l'échantillon étant le nombre total N d'individus récoltés, toutes espèces confondues ;

- soit à échelle plus large (par exemple «régionale»), auquel cas l'échantillonnage consiste à noter l'*incidence* de chaque espèce rencontrée (c'est-à-dire, cette fois, le nombre de stations où l'espèce est présente, quelque soit son abondance dans la station), la taille de l'échantillon étant alors le nombre total N de stations investiguées.

Une espèce est dite observée x fois dans l'échantillon : (i) si elle est représentée par x individus parmi les N dans un échantillon local (échantillonnage d'abondance), ou bien (ii) si elle est présente dans x stations parmi les N d'un échantillon dit régional (échantillonnage d'incidence).

On désigne, dans la suite, par $f_1, f_2, \dots, f_x$ les nombres d'espèces ainsi observées 1, 2, ..., x fois au cours de l'échantillonnage (singletons f_1 , doubletons f_2 , etc.).

La Courbe d'Accumulation des Espèces (C.A.E.) au fil de l'échantillonnage rend compte de l'accroissement du nombre $R(N)$ d'espèces observées en fonction de la taille N de l'échantillon.

L'extrapolation de la C.A.E. d'une part et l'estimation du nombre Δ d'espèces manquantes (et donc l'estimation de la richesse spécifique totale S) d'autre part ont, en fait, *partie liée* : la richesse spécifique totale détermine en effet l'asymptote de la C.A.E., puisque, à mesure que la taille N de l'échantillon grandit, le nombre Δ d'espèces manquantes se réduit et le nombre R d'espèces récoltées s'accroît d'autant, tendant finalement vers la richesse spécifique totale $S (= R + \Delta)$. Choisir de façon optimale l'extrapolation de la C.A.E. et l'estimateur du nombre d'espèces manquantes est donc un choix *lié*. Autrement dit, le processus de choix implique *solidairement* celui de l'estimateur du nombre d'espèces manquantes et celui de l'extrapolation associée de la C.A.E.

Parmi les divers estimateurs du nombre Δ d'espèces manquantes et donc de la richesse spécifique totale (voir revue dans COLWELL & CODDINGTON, 1994 ; GOTELLI & CHAO, 2013), trois estimateurs sont plus particulièrement utilisés : «Chao» ; «Jackknife-1» et «Jackknife-2». Tous trois s'expriment en fonction des nombres f_1 et f_2 d'espèces observées respectivement 1 fois et 2 fois dans l'échantillonnage réalisé :

$$\text{«Chao»} : \Delta = f_1^2 / (2f_2) ;$$

$$\text{«Jackknife-1»} : \Delta = f_1 ;$$

$$\text{«Jackknife-2»} : \Delta = 2f_1 - f_2.$$

Ces formulations, bien distinctes, conduisent à des estimations évidemment différentes, parmi lesquelles l'une est forcément mieux centrée, plus exacte, que les autres. Toute la difficulté réside dans l'établissement de la grille de choix pertinente à cet égard. On a déjà souligné que les démarches empiriques dans ce sens ont échoué, aboutissant à des choix souvent contradictoires selon les inventaires. Difficulté que lève opportunément l'approche théorique. Il a en effet été montré (détails in BÉGUINOT, 2015b) que, pour satisfaire la relation contraignante évoquée en Introduction, l'expression mathématique de l'extrapolation de la C.A.E. doit être choisie en fonction de la valeur du ratio f_1/f_2 . Deux domaines de valeurs sont à distinguer à cet égard :

* le domaine $f_1/f_2 \leq 0,6$, qui correspond essentiellement aux *cas limites* où l'échantillonnage réalisé se trouve être déjà fort proche de l'exhaustivité (ou bien encore le cas très inhabituel où les abondances/incidences des différentes espèces sont égales ou tout au moins voisines)

* le domaine $f_1/f_2 > 0,6$ qui correspond à la *généralité des cas* (c'est-à-dire en dehors notamment des échantillonnages approchant l'exhaustivité).

Pour un échantillon de taille N_0 comportant $R_0 (= R(N_0))$ espèces observées, le nombre estimé Δ d'espèces manquantes, la richesse spécifique totale estimée $S (= R_0 + \Delta)$ et l'expression $R(N)$ de l'extrapolation de la Courbe d'Accumulation des Espèces sont les suivants (BÉGUINOT, 2015b):

$$\rightarrow \text{si } f_1/f_2 \leq 0,6 : \Delta = f_1^2 / (2f_2) \text{ (estimateur 'Chao')} \text{ et donc } S = R_0 + f_1^2 / (2f_2) \\ R(N) = (R_0 + f_1^2 / (2f_2)) \cdot (1 - \exp([\ln((f_1^2 / (2f_2)) / (R_0 + f_1^2 / (2f_2)))] (N/N_0))) \quad [1]$$

$$\rightarrow \text{si } f_1/f_2 > 0,6 : \Delta = 2f_1 - f_2 \text{ (estim. 'Jackknife-2')} \text{ et donc } S = R_0 + 2f_1 - f_2 \\ R(N) = (R_0 + 2f_1 - f_2) - (3f_1 - 2f_2) N_0/N - (f_2 - f_1) N_0^2/N^2 \quad [2]$$

On notera que quand $f_1/f_2 \leq 0,6$, l'extrapolation $R(N)$ (équation [1]) n'a (*ouf!*) plus guère d'intérêt pratique puisqu'on est alors avec un échantillonnage déjà proche de l'exhaustivité.

Il convient de souligner que ces extrapolations s'entendent pour un éventail écologique (c'est-à-dire un éventail de types d'habitats) astreint à demeurer essentiellement inchangé au fil de l'extrapolation d'échantillonnage. Il va en effet de soi que si l'enveloppe écologique investiguée devait être significativement modifiée dans le projet de poursuite de l'inventaire, plus aucune prévision ne serait alors possible, par quelque moyen que ce soit.

Il convient également de ne pas oublier que l'extrapolation complète de la C.A.E. et l'estimation de la richesse spécifique totale ne sont pas dépourvues de biais ; la procédure de choix précédente a justement pour objet de *minimiser ce biais*, sans pour autant pouvoir garantir de l'annuler. À cet égard, il est d'ailleurs recommandé de n'envisager d'extrapolation de la C.A.E. que lorsque le degré de complétude estimé R_0/S est au moins égal à 40-50 %.

Signalons toutefois qu'une procédure d'extrapolation, cette fois effectivement *sans biais*, est cependant possible, mais son ampleur est plus restreinte : elle n'est en effet à considérer que pour un accroissement de N ne dépassant pas 50 % (à la rigueur 75 %) de la taille N_0 de l'échantillonnage réalisé. Cette procédure (dont l'établissement et l'utilisation est détaillée dans BÉGUINOT, 2015a) conduit à l'expression suivante pour l'extrapolation sans biais d'un échantillon de taille originelle N_0 comportant R_0 ($=R(N_0)$) espèces observées :

$$R(N) = R_0 + \sum_x \{(-1)^{(x-1)} f_x [(N-N_0)/N_0]^x\} \quad [3]$$

la sommation $\sum_x$ étant étendue de $x = 1$ à x_m (en pratique, $x_m = 8$ est amplement suffisant). Il s'agit en fait d'un développement en série de Taylor, les dérivées successives de $R(N)$ étant estimées au moyen de la série des f_x grâce à la relation générale susmentionnée (*cf.* Annexe), mettant en correspondance la série des dérivées successives de la C.A.E. avec la série des nombres $f_1, f_2, \dots, f_x$ d'espèces observées 1, 2, ..., x fois au cours de l'échantillonnage.

Enfin, il faut bien garder à l'esprit que la minimisation du biais (extrapolations complètes [1] et [2]) ou l'absence de biais (extrapolation partielle [3]) ne signifie pas pour autant que la prévision fournie ne soit pas sujette à dispersion statistique. En effet la réalisation de tout échantillonnage représente, en fait, un tirage aléatoire au sein de l'ensemble investigué, en sorte que les valeurs de R_0, f_1, f_2 , sont, elles-mêmes, des variables aléatoires, comme telles sujettes à dispersion statistique. Il en est donc de même de l'extrapolation $R(N)$ qui est fonction de $R_0, f_1, f_2, \dots$. Il n'en demeure pas moins qu'il vaut évidemment mieux que cette inévitable dispersion affecte une estimation par ailleurs *non (ou faiblement) biaisée*, plutôt qu'elle ne vienne s'ajouter à une estimation déjà nettement décentrée. Là réside tout l'intérêt de la présente démarche.

On soulignera, enfin, que la grille de sélection proposée ici en faveur du type d'estimateur de richesse spécifique le moins biaisé est assez voisine de celle antérieurement établie, sur une base semi-empirique, par BROSE *et al.* (2003). La présente démarche apporte, en fait, un support supplémentaire, de nature, cette fois, théorique et donc généralisable à la grille de BROSE *et al.* et elle offre, en outre, l'avantage de supprimer les discontinuités entre choix, regrettablement imposées par la grille de BROSE *et al.* [*cf.* argumentation comparative dans BÉGUINOT (2015b)].

3 - EXEMPLE ILLUSTRATIF : ESTIMATIONS (1) DE LA RICHESSE SPÉCIFIQUE TOTALE ET (2) DE L'EFFORT D'ÉCHANTILLONNAGE ADDITIONNEL REQUIS POUR ASSURER UN GAIN DONNÉ EN NOMBRE D'ESPÈCES OBSERVÉES

3.1 Application à l'inventaire partiel de la faune cécidogène du Parc des Écrins

Un inventaire partiel de la faune cécidogène (inductrice de galles végétales) a été réalisé récemment sur une partie du Parc National des Écrins (BÉGUINOT, 2012). Cet inventaire portait sur $N_0 = 147$ sites d'investigations répartis sur le territoire du Parc et rassemblant 337 espèces dont 128 espèces d'Acariens Trombidiformes, 47 espèces d'Hémiptères, 118 espèces de Diptères et 44 espèces d'Hyménoptères. Le tableau I rassemble les principaux éléments numériques relatifs à chacun des inventaires dévolus respectivement à ces quatre ordres d'arthropodes cécidogènes : nombre R_0 d'espèces échantillonnées et nombres f_x d'espèces observées 1, 2, ..., x fois au cours de l'échantillonnage.

	R_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
Acariens	128	45	17	11	9	2	5	5	2
Hémiptères	47	9	9	4	1	4	2	2	0
Diptères	118	38	14	10	11	8	3	2	3
Hyménoptères	44	6	8	9	5	3	1	0	1

Tableau I. Valeurs de R_0 et des huit premiers f_x pour chacun des quatre échantillonnages.

Pour chacun des quatre échantillonnages, le ratio f_1/f_2 est supérieur à 0.6 ; conséquemment l'estimateur du nombre d'espèces manquantes à sélectionner est à chaque fois Jackknife-2 (JK-2 en abrégé) :

$\Delta_{JK-2} = 2f_1 - f_2$ soit $\Delta = 73$, $\Delta = 9$, $\Delta = 62$ et $\Delta = 4$ pour Acariens, Hémiptères, Diptères et Hyménoptères respectivement.

Les estimations de la richesse spécifique totale S et du degré de complétude correspondant de l'échantillonnage, R_0/S (%), sont donc les suivants :

$S = 201$ et $R_0/S = 64$ % pour les Acariens ;

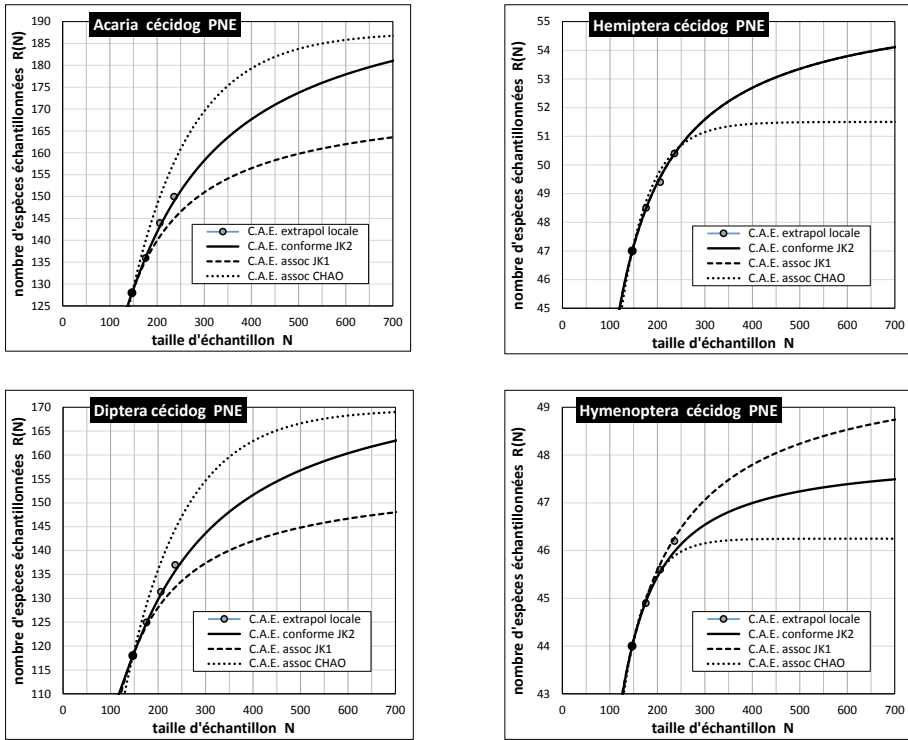
$S = 56$ et $R_0/S = 84$ % pour les Hémiptères ;

$S = 180$ et $R_0/S = 66$ % pour les Diptères ;

$S = 48$ et $R_0/S = 92$ % pour les Hyménoptères.

Pour chacun des quatre échantillonnages l'estimateur du nombre d'espèces manquantes à sélectionner étant à chaque fois Jackknife-2, la formulation à adopter

pour l'extrapolation des Courbes d'Accumulation des Espèces est donc, elle aussi, celle associée à Jackknife-2, c'est-à-dire l'équation [2].



Figures 1 à 4. Extrapolations de la Courbe d'Accumulation des Espèces R(N) en fonction de la taille N de l'inventaire (nombre de sites investigués) : extrapolation limitée et sans biais (selon équation [3]), extrapolation complète à biais minimal (associée à JK-2) et, pour comparaison, extrapolations complètes associées à JK-1 et Chao. L'extrapolation s'applique naturellement à des tailles d'échantillon N au-delà de la taille $N_0 = 147$ de l'inventaire effectivement réalisé.

Les extrapolations des Courbes d'Accumulation des Espèces $R(N > N_0)$ répondent donc aux expressions suivantes :

$$\begin{aligned} R(N) &= 201 - 101.(147/N) + 28.(147/N)^2 && \text{pour l'inventaire des Acariens} \\ R(N) &= 56 - 9.(147/N) - 0.(147/N)^2 && \text{pour l'inventaire des Hémiptères} \\ R(N) &= 180 - 86.(147/N) + 24.(147/N)^2 && \text{pour l'inventaire des Diptères} \\ R(N) &= 48 - 2.(147/N) - 2.(147/N)^2 && \text{pour l'inventaire des Hyménoptères} \end{aligned}$$

Ces extrapolations sont reproduites graphiquement aux figures 1 à 4 (courbes en trait plein).

Aux fins de comparaison avec ces extrapolations, sélectionnées comme assurant un niveau de biais minimal, on a également fait figurer :

- l'extrapolation limitée mais sans biais (série de trois points), fournie par l'équation [3], les valeurs de N_0 , de R_0 , et des huit f_x étant celles indiquées au tableau I.

- les extrapolations associées aux deux autres estimateurs non sélectionnés, Jackknife-1 et Chao (courbes en tirets et pointillée respectivement).

La figure 5 rassemble, sur un même graphe, les extrapolations de la Courbe d'Accumulation des Espèces pour chacun des quatre inventaires, en terme non plus de nombre d'espèces échantillonnées $R(N)$ mais de complétude d'échantillonnage $R(N)/S$ (en %), ce qui permet de comparer directement les «profondeurs d'investigation» de chacun des quatre inventaires. Cette figure est, en même temps un outil de prévision de l'effort additionnel d'échantillonnage requis pour atteindre un niveau accru donné de complétude de l'inventaire.

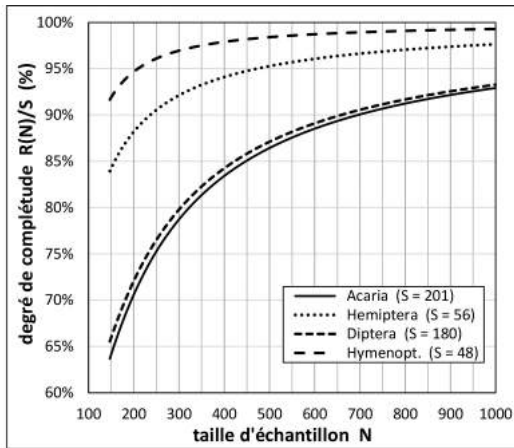
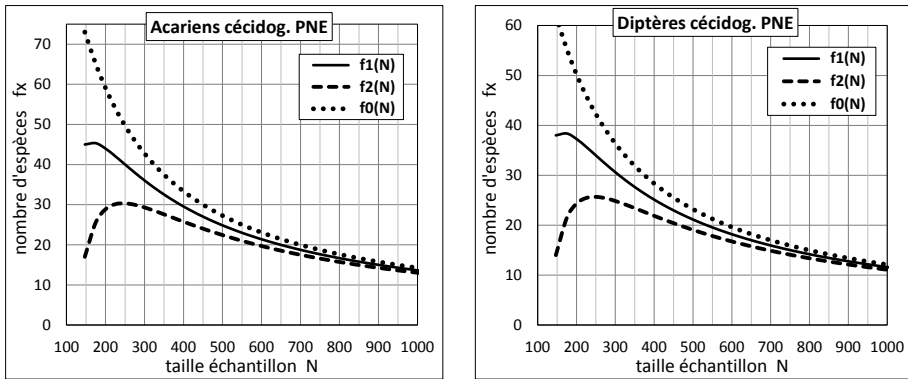


Figure 5. Efforts d'échantillonnage à consentir pour atteindre un degré donné de complétude de l'inventaire : ici, par exemple, pour atteindre un degré de complétude de 90 %, d'ores et déjà obtenu dès $N \approx 130$ pour les Hyménoptères, il faudrait, en revanche, atteindre $N \approx 240$ pour les Hémiptères et des efforts d'échantillonnage bien plus considérables encore, $N_0 \approx 700$, pour les Diptères et les Acariens.

Enfin, en complément de l'extrapolation de la Courbe d'Accumulation des Espèces, il est également intéressant de rentrer un peu davantage dans les détails en considérant comment est appelé à évoluer, à efforts d'échantillonnage croissants, le nombre $f_x(N)$ d'espèces observées x fois dans un échantillon de taille N . L'intérêt porte plus particulièrement sur les espèces « rebelles », faiblement représentées malgré l'accroissement de la taille d'échantillon : disons $x \leq 2$. L'expression de $f_x(N)$ pour $N > N_0$ se déduit immédiatement de la relation générale contraignant l'expression des C.A.E évoquée au § 1, qui met en correspondance la série des dérivées successives de $R(N)$ avec la série des nombres $f_1, f_2, \dots, f_x$ d'espèces observées 1, 2, $\dots$, x fois au cours de l'échantillonnage (voir Annexe pour l'expression de $f_x(N)$ qui s'en déduit). Les figures 6 et 7 rendent compte des évolutions de f_1, f_2 et également de f_0 pour les deux inventaires les moins complets et conséquemment les plus intéressants au point de vue des évolutions extrapolées de f_1, f_2 et f_0 avec la taille d'échantillonnage.



Figures 6 et 7. Extrapolations des nombres f_1 , f_2 , f_0 d'espèces observées 1 fois, 2 fois et non encore collectées, au cours d'un échantillonnage de taille croissante $N > N_0$ ($= 147$).

3.2 Informations pratiques déduites des extrapolations des Courbes d'Accumulation des Espèces pour l'inventaire de la faune cécidogène du Parc des Écrins

Une excellente correspondance est tout d'abord constatée entre d'une part l'extrapolation limitée mais sans biais et d'autre part l'extrapolation complète sélectionnée (ici associée à l'estimateur Jackknife-2), et ce pour chacun des quatre inventaires considérés (figures 1 à 4). Ceci confirme que l'extrapolation sélectionnée est non seulement celle qui conduit au moindre biais d'estimation mais que le biais résiduel doit être relativement restreint.

Au contraire, l'amplitude des écarts entre l'extrapolation associée à Jackknife-2 et chacune des extrapolations respectivement associées à Jackknife-1 et à Chao se révèle considérable (figures 1 à 4), en particulier lorsqu'on considère ceux des inventaires restant les plus éloignés de l'exhaustivité, et pour lesquels l'extrapolation prend justement tout son intérêt : cas, ici, des Acariens et des Diptères. Ceci permet de mesurer concrètement combien il est important d'adopter la formulation d'extrapolation la mieux appropriée (ici donc celle associée à Jackknife-2) pour tenter d'assurer des pronostics suffisamment fiables.

Ceci permet d'accorder crédit aux prévisions résultant des extrapolations des C.A.E pour les quatre groupes taxonomiques considérés.

Ainsi, notamment, il ressort que les degrés d'avancement des quatre inventaires considérés diffèrent très fortement entre eux (figure 5) et ce, *malgré* une pression d'échantillonnage identique. Les niveaux de complétude obtenus pour les inventaires des Acariens et des Diptères se situent très en deçà de ceux des Hémiptères et des Hyménoptères. Et ce, alors même que les espèces récoltées sont bien plus nombreuses pour les premiers que pour les seconds. En sorte que si l'on voulait rétablir des niveaux de complétude d'échantillonnage moins hétérogènes entre les quatre groupes taxonomiques, ce sont, paradoxalement, sur les plus riches en espèces, Acariens et Diptères, qu'il faudrait faire porter prioritairement les efforts d'échantillonnage à venir. Et les écarts en termes d'efforts additionnels à prévoir ne sont pas minces :

ainsi pour rejoindre le niveau de complétude obtenu pour les Hyménoptères (92 % estimé) à l'issue de l'échantillonnage réalisé (147 sites), il faudrait doubler la mise ($N \approx 300$ sites) pour les Hémiptères et plus que quintupler l'effort ($N \approx 850$ sites) pour les Acariens et les Diptères.

Enfin, on a vu qu'il est même possible d'aller encore plus loin en matière d'extrapolation, avec une analyse prévisionnelle de l'évolution des nombres f_0 , f_1 , f_2 des espèces les plus rebelles à la collecte, en réponse à la poursuite de l'échantillonnage. La figure 6 indique ainsi, pour l'échantillonnage des Acariens, par exemple :

- que le nombre f_0 des espèces non encore récoltées se réduirait «hyperboliquement» de $f_0 = 73$ (pour $N_0 = 147$) à $f_0 \approx 16$ pour $N = 850$;

- que le nombre f_1 des espèces rencontrées une seule fois, d'abord croissant, passerait par un maximum (égal à 46) vers $N \approx 160$ pour décroître ensuite régulièrement ($f_1 \approx 15$ pour $N = 850$) ;

- que le nombre f_2 des espèces rencontrées deux fois, d'abord croissant, passerait par un maximum (égal à 30) vers $N \approx 250$ pour décroître ensuite régulièrement ($f_2 \approx 15$ pour $N = 850$) ;

- enfin que le nombre $f_{>2}$ des espèces rencontrées plus de deux fois croîtrait évidemment continument, passant de $f_{>2} = 66$ pour $N_0 = 147$ à $f_{>2} \approx 155$ pour $N = 850$.

L'évolution est qualitativement voisine pour le groupe des Diptères.

4 - CONCLUSION

Une procédure d'extrapolation numérique au service d'une série de questions pratiques

L'extrapolation de la Courbe d'Accumulation des Espèces, qui ne saurait être que purement numérique, ne permet évidemment pas d'accéder, en outre, à *l'identité* des espèces non encore récoltées, dont on a seulement extrapolé le nombre vraisemblable. Mais cette extrapolation numérique permet d'aborder quelques questions pratiques importantes avant de décider, le cas échéant, de se lancer dans une prolongation significative de l'échantillonnage effectif, visant à identifier les espèces non récoltées jusque-là. Questions intéressantes :

- *l'enjeu prévisionnel* d'une prolongation d'échantillonnage : estimation du degré d'incomplétude de l'inventaire réalisé, c'est-à-dire estimation de la richesse spécifique totale de l'assemblage inventorié (valeur asymptotique de l'extrapolation) ;

- *le coût prévisionnel* de la prolongation : estimation de l'effort additionnel d'échantillonnage requis pour accroître le degré de complétude de l'inventaire et atteindre un seuil de complétude donné ;

- *la balance prévisionnelle enjeu / coût* : jauger le «ratio» gain (en espèces nouvelles) sur mise (en terme d'effort additionnel requis) et pouvoir ainsi *juger* sur une base mieux informée de l'opportunité soit de poursuivre soit d'interrompre un échantillonnage en cours. Question qui se pose tout particulièrement lorsque (cas fréquent) coexistent *des priorités d'inventaires concurrentes, soit sur d'autres sites, soit intéressant d'autres groupes taxonomiques*. Un exemple typique entrant dans cette catégorie étant la conduite simultanée de plusieurs inventaires qui, bien que menés conjointement, sont susceptibles néanmoins de progresser à des allures nettement distinctes, posant la question d'une éventuelle conduite future différenciée

selon les inventaires. Exemple que vient illustrer l'inventaire général des arthropodes cécidogènes du Parc des Écrins, évoqué ci-dessus.

La procédure décrite au §2 permet d'aborder rationnellement ces questions en permettant de sélectionner le moins biaisé parmi les trois estimateurs du nombre d'espèces manquantes les plus couramment utilisés : Chao, Jackknife-1 et Jackknife-2. Et, ce faisant, d'en déduire l'estimation la moins biaisée (i) de la richesse spécifique totale et (ii) de l'extrapolation complète de la Courbe d'Accumulation des Espèces, au-delà de l'échantillonnage déjà réalisé : §2, équations [1], [2]. En outre, une extrapolation cette fois strictement sans biais mais d'amplitude limitée (< 75 % d'effort d'échantillonnage supplémentaire) peut être obtenue si souhaitée : §2, équation [3].

ANNEXE

Une équation de validité très générale (démonstration détaillée in BÉGUINOT, 2014 et 2015a) relie :

- d'une part la série des dérivées successives $\partial^x R_{(N)} / \partial N^x$ de la Courbe d'Accumulation des Espèces $R_{(N)}$ (dérivées qui, ensemble, en régissent la forme) et
- d'autre part la série des nombres $f_{x(N)}$ des espèces représentées x fois dans l'échantillon de taille N :

$$[\partial^x R_{(N)} / \partial N^x] = (-1)^{(x-1)} f_{x(N)} / C_{N,x} \approx (-1)^{(x-1)} (x! / N^x) f_{x(N)} \quad [A.1]$$

Appliquant l'équation [A.1] à l'équation [2], on obtient immédiatement les expressions de $f_{0(N)}$, $f_{1(N)}$, $f_{2(N)}$ pour l'extrapolation associée à Jackknife-2 (BÉGUINOT, 2015b) :

$$f_{0(N)} = S - R(N) = (3f_1 - 2f_2) N_0 / N + (f_2 - f_1) N_0^2 / N^2 \quad [A.2]$$

$$f_{1(N)} = N \cdot \partial R_{(N)} / \partial N = (3f_1 - 2f_2) N_0 / N + 2 \cdot (f_2 - f_1) N_0^2 / N^2 \quad [A.3]$$

$$f_{2(N)} = -\frac{1}{2} N^2 \cdot \partial^2 R_{(N)} / \partial N^2 = (3f_1 - 2f_2) N_0 / N + 3 \cdot (f_2 - f_1) N_0^2 / N^2 \quad [A.4]$$

RÉFÉRENCES BIBLIOGRAPHIQUES

- BASUALDO C.V., 2011. Choosing the best non-parametric richness estimator for benthic macroinvertebrates databases. *Revista Sociedad Entomologica Argentina*, 70(1-2): 27-38.
- BÉGUINOT J., 2012. Investigations sur la faune cécidogène d'altitude dans le Parc National des Écrins. *Bulletin mensuel Société Linnéenne de Lyon*, 81 (5-6) : 87-116.
- BÉGUINOT J., 2014. An algebraic derivation of Chao's estimator of the number of species in a community highlights the condition allowing Chao to deliver centered estimates. *International Scholarly Research Notices - Ecology*, vol. 2014: article ID 847328, 6 pages; doi:10.1155/2014/847328.
- BÉGUINOT J., 2015a. When reasonably stop sampling? How to estimate the gain in newly recorded species according to the degree of supplementary sampling effort. *Annual Research & Review in Biology*, 7(5): 300-308; doi: 10.9734/ARRB/2015/18809.
- BÉGUINOT J., 2015b. Extrapolation of the Species Accumulation Curve for incomplete species samplings: a new nonparametric approach to estimate the degree of sample completeness and decide when to stop sampling. *Annual Research & Review in Biology*, 8 (5): 1-9; doi: 10.9734/ARRB/2015/22351
- BROSE U., MARTINEZ N.D. & WILLIAMS R.J., 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*, 84(9): 2364-2377.
- CHIARUCCI A., ENRIGHT N.J., PERRY G.L.W., MILLER B.P. & LAMONT B.B., 2003. Performance of nonparametric species richness estimators in a high diversity plant community. *Diversity and Distributions*, 9: 283-295.
- CHIU C.-H., WANG Y.-T., WALTHER B.A. & CHAO A., 2014. An improved nonparametric lower bound of species richness via a modified Good-Turing frequency formula. *Biometrics*, 70(3): 671-682 ; doi : 10.1111/biom.12200.

- COLWELL R.K. & CODDINGTON J.A., 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society London B*, 345: 101-118.
- DE WEVER P. & DAVID B., 2015. *La Biodiversité de crise en crise*. Albin Michel éd., Paris, 301 p.
- FOGO A., ATTRILL M.J., FROST M.T. & ROWDEN A.A., 2003. Estimating marine species richness: an evaluation of six extrapolative techniques. *Marine Ecology Progress Series*, 248: 15-26.
- GOTELLI N.J. & CHAO A., 2013. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) *Encyclopedia of Biodiversity*, 2nd ed., Vol. 5: 195-211. Waltham, MA: Acad. Press.
- HELLMANN J.J. & FOWLER G.W., 1999. Bias, precision and accuracy of four measures of species richness. *Ecological Applications*, 9(3): 824-834.
- HERZOG S.K., KESSLER M. & CAHILL T.M., 2002. Estimating species richness of tropical bird communities from rapid assessment data. *The Auk*, 119(3): 749-769.
- HORTAL J., BORGES P.A.V & GASPAR C., 2006. Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*, 75: 274-287.
- POULIN R., 1998. Comparison of three estimators of species richness in parasite component communities. *Journal of Parasitology*, 84(3): 485-490.
- SOBERON M.J. & LLORENTE B.J., 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, 7: 480-488.
- THOMPSON G.G., WITHERS P.C., PIANKA E.R. & THOMPSON S.A., 2003. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology*, 28: 361-383.
- UNTERSEHER M., SCHNITTLER M., DORMANN C. & SICKERT A., 2008. Application of species richness estimators for the assessment of fungal diversity. *FEMS Microbiology Letters*, 282: 205-213.
- WALTER B.A. & MORAND S., 1998. Comparative performance of species richness estimation methods. *Parasitology*, 116: 395-405.



SOCIÉTÉ LINNÉENNE DE LYON

Siège social : 33, rue Bossuet, F-69006 LYON

Tél. et fax : +33 (0)4 78 52 14 33

<http://www.linneenne-lyon.org> — email : secretariat@linneenne-lyon.org

Groupe de Roanne : Maison des anciens combattants, 18, rue de Cadore, F-42300 ROANNE

Rédaction : Marie-Claire PIGNAL - Directeur de publication : Bernard GUÉRIN

Conception graphique de couverture : Nicolas VAN VOOREN



Tome 85 Fascicule 7-8 Septembre - Octobre 2016

SOMMAIRE

- Prudhomme J.C. - Une étude locale de la biodiversité: inventaire des coléoptères du domaine de la fondation Pierre Vérots à Saint-Jean-de-Thurigneux (Ain, France). 4. Bruches, charançons, chrysomèles et autres phytophages..... 210-240
- Gruhn G. & Rivoire B. - *Tubulicrinis incrassatus* (Basidiomycota), première récolte en France métropolitaine..... 241-245
- Béguinot J. - Extrapolation des inventaires de biodiversité incomplets : comment estimer au mieux le nombre d'espèces manquantes et prévoir l'effort additionnel d'échantillonnage requis pour réduire ce nombre..... 246-258
- Krieg-Jacquier R., Deliry C. & Roncin P. - Redécouverte de *Ranunculus hederaceus* L. 1753 (*Ranunculales*, *Ranunculaceae*) dans le département de l'Ain (France)..... 261-271

Couverture : L'étang Praillebard et la prairie de Viaire-du-Loup à la Fondation Pierre Vérots à Saint-Jean-de-Thurigneux (Ain), un milieu humide de la Dombes très favorable à la biodiversité entomologique, mai 2016. Crédit : J-P Rabatel

CONTENTS

- Prudhomme J.C. - A local study of biodiversity : inventory of the beetles of the property of Pierre Vérots Foundation in Saint-Jean-de-Thurigneux (Ain, France). 4 : bruchids, weevils, leaf beetles and other phytophagous beetles 210-240
- Gruhn G. & Rivoire B. - *Tubulicrinis incrassatus* (Basidiomycota), a first collection from continental France..... 241-245
- Béguinot J. - Extrapolating the species accumulation curve for incomplete samplings: how to estimate accurately the number of missing species and reliably forecast the extra sampling effort required to reduce this number..... 246-258
- Krieg-Jacquier R., Deliry C. & Roncin P. - Rediscovery of *Ranunculus hederaceus* L. in Ain department, Eastern France (*Ranunculales*, *Ranunculaceae*) 261-271

Prix 10 euros

ISSN 0366-1326 - N° d'inscription à la CPPAP : 0418G85671

Imprimé par Imprimerie Brailly, 69564 Saint-Genis-Laval Cedex

N° d'imprimeur : V0001XX/00 Imprimé en France Dépôt légal : août 2016

Copyright © 2016 SLL. Tous droits réservés pour tous pays sauf accord préalable.