

Note sur l'approche statistique de la diversité en écologie

Application à l'indice composition pour le
phytoplancton

Note sur l'approche statistique de la diversité en écologie

Application à l'indice composition pour le
phytoplancton

sommaire

1. Contexte	7
2. Eléments bibliographiques	9
2.1. Les principaux outils statistiques en écologie - Les indices de diversité.....	9
2.1.1. Richesse spécifique.....	9
2.1.2. Equitabilité (<i>evenness</i>)	9
2.1.3. Indices hétérogènes	10
2.1.4. Composition spécifique.....	10
2.1.5. Indices de diversité taxonomique	11
2.2. Mesure de la biodiversité dans les évaluations environnementales en milieu marin.....	11
2.3. Synthèse	12
3. Nature des données et de l'indice attendu	12
4. Application	13
4.1. Données.....	13
4.2. Méthodes	13
4.3. Résultats	13
5. Conclusion	14
6. Annexe : graphiques richesse, Berger-Parker, et croisement des deux indices	16

1. Contexte

La liste des éléments de qualité de la directive 2000/60/CE présente la ligne suivante à la fois pour les eaux de transition et les eaux côtières :

Composition, abondance et biomasse du phytoplancton

Les tableaux 1 et 2 sont extraits de la section « Définitions des états écologiques 'très bon', 'bon' et 'moyen' ». Bien que les formulations diffèrent légèrement d'un type de masse d'eau à l'autre, elles sont essentiellement identiques.

L'Ifremer a la responsabilité de la définition de l'indicateur de cet élément de qualité. Il s'appuie sur trois indices identifiés comme « biomasse », « abondance » et « composition ». L'indice biomasse, percentile 90 de la distribution de la chlorophylle *a*, correspond bien à la définition de la directive. L'indice abondance, pourcentage d'échantillons présentant au moins une espèce en état de *bloom*, est en phase avec la fréquence des efflorescences mais pas avec l'intensité des efflorescences. L'indice composition initialement envisagé, pourcentage d'échantillons présentant une espèce toxique pour l'environnement en état de bloom, est apparu comme le moins en phase avec sa définition, de plus il donnait des résultats non-discriminants. Il a donc été abandonné. L'objet de la présente note est de présenter succinctement le traitement statistique de la notion de composition en écologie à travers deux documents :

1. **Les principaux outils statistiques en écologie**, J.C. Gaertner, Master Professionnel Economie Environnement, Ecologie numérique, Année universitaire 2006-2007, Université de la méditerranée Aix-Marseille, Centre d'océanologie de Marseille
2. **Mesure de la biodiversité dans les évaluations environnementales en milieu marin**, Mark J. Costello et al., Centre des sciences de la mer Huntsman, Collection de monographies en recherche et développement, 2001, Numéro de catalogue En105-3/83-2004F ISBN 0-662-76671-7

Tableau 1 : Définitions des états écologiques «très bon», «bon» et «moyen» en ce qui concerne les eaux de transition extraites de la directive 2000/60/CE.

Élément	Très bon état	Bon état	État moyen
Phytoplancton	<p>La composition et l'abondance des taxa phytoplanctoniques correspondent totalement ou presque totalement aux conditions non perturbées.</p> <p>La biomasse moyenne du phytoplancton correspond aux conditions physico-chimiques caractéristiques et n'est pas de nature à détériorer sensiblement les conditions de transparence caractéristiques.</p> <p>L'efflorescence planctonique est d'une fréquence et d'une intensité qui correspondent aux conditions physico-chimiques caractéristiques.</p>	<p>Légères modifications dans la composition et l'abondance des taxa planctoniques.</p> <p>Légères modifications dans la biomasse par rapport aux conditions caractéristiques. Ces modifications n'indiquent pas de croissance accélérée des algues entraînant des perturbations indésirables de l'équilibre des organismes présents dans la masse d'eau ou de la qualité physico-chimique de l'eau.</p> <p>La fréquence et l'intensité de l'efflorescence planctonique peuvent augmenter légèrement.</p>	<p>La composition et l'abondance des taxa planctoniques diffèrent modérément des communautés caractéristiques.</p> <p>La biomasse est modérément perturbée et peut être de nature à produire une forte perturbation indésirable des valeurs des autres éléments de qualité biologique.</p> <p>La fréquence et l'intensité de l'efflorescence planctonique peuvent augmenter modérément. Une efflorescence persistante peut se produire durant les mois d'été.</p>

Tableau 2 : Définitions des états écologiques «très bon», «bon» et «moyen» en ce qui concerne les eaux côtières extraites de la directive 2000/60/CE.

Élément	Très bon état	Bon état	État moyen
Phytoplancton	<p>La composition et l'abondance des taxa phytoplanctoniques correspondent totalement ou presque totalement aux conditions non perturbées.</p> <p>La biomasse moyenne de phytoplancton correspond aux conditions physico-chimiques caractéristiques et n'est pas de nature à détériorer sensiblement les conditions de transparence caractéristiques.</p> <p>L'efflorescence planctonique est d'une fréquence et d'une intensité qui correspondent aux conditions physico-chimiques caractéristiques.</p>	<p>La composition et l'abondance des taxa phytoplanctoniques montrent de légers signes de perturbation.</p> <p>Légères modifications dans la biomasse par rapport aux conditions caractéristiques. Ces changements n'indiquent pas de croissance accélérée des algues entraînant des perturbations indésirables de l'équilibre des organismes présents dans la masse d'eau ou de la qualité de l'eau.</p> <p>La fréquence et l'intensité de l'efflorescence planctonique peuvent augmenter légèrement.</p>	<p>La composition et l'abondance des taxa planctoniques diffèrent modérément de celles des communautés caractéristiques.</p> <p>La biomasse des algues dépasse sensiblement la fourchette associée aux conditions caractéristiques et est de nature à se répercuter sur d'autres éléments de qualité biologique.</p> <p>La fréquence et l'intensité de l'efflorescence planctonique peuvent augmenter modérément. Une efflorescence persistante peut se produire durant les mois d'été.</p>

2. Eléments bibliographiques

2.1. Les principaux outils statistiques en écologie - Les indices de diversité

Ce document est un support de cours d'écologie numérique du Master Professionnel Economie Environnement dispensé à l'université de Marseille en 2006-2007 par J.C. Gaertner. Les indices présentés ci-après n'établissent pas une liste exhaustive.

2.1.1. Richesse spécifique

Elle est définie comme le nombre d'espèces différentes identifiées dans un échantillon. C'est l'indice de diversité le plus utilisé dans la littérature, il est simple à comprendre, à évaluer et à analyser. En revanche, il est influencé par l'effort d'échantillonnage, c'est-à-dire que plus le nombre d'échantillons est important plus les espèces rares seront nombreuses et plus la richesse spécifique sera élevée. Ainsi, dans le contexte de comparaison, spatiale ou temporelle, il est nécessaire de conserver la même stratégie d'échantillonnage. Pour la même raison, les conclusions à une échelle ne sont pas transposables à une autre échelle. Enfin, il ne tient pas compte de l'abondance des espèces. Ainsi, dans un échantillon présentant quatre espèces différentes à raison d'un individu par espèce (e.g. A, B, C, D), la richesse spécifique sera égale à celle d'un autre échantillon présentant également quatre espèces, mais dont l'une comptera plus d'un individu (e.g. A, B, C, D, D, D, D, D, D, D). Les indices de Margaleff et Menhinick tendent à corriger ce point :

- Margalef : $(S - 1) / \ln(N)$;
- Menhinick : S / \sqrt{N} .

avec S , le nombre d'espèces et N , l'abondance totale. Ces deux indices restent influencés par l'effort d'échantillonnage.

2.1.2. Equitabilité (evenness)

L'équitabilité est une mesure de la distribution des individus au sein d'une espèce ; autrement dit, c'est la façon dont l'abondance se répartit au sein des espèces. Le tableau 3 donne certains de ces indices.

Tableau 3 : Indices d'équitabilité.

Nom	Calcul	Avantages	Désavantages
Heip	$e^{H'} / (S - 1)$	sensible aux espèces rares	dépendant de la taille d'échantillonnage quand $S < 10$

Nom	Calcul	Avantages	Désavantages
Pièlou	$H'/\ln(S)$		sensible à S quand > 25
Berger-Parker	N_{\max}/N	simple, facile à calculer, à comprendre et à communiquer	sensible à l'espèce dominante, dépendant de S quand $S > 100$ et quand $S < 100$, augmente avec S
Simpson	$D^{-1}S^{-1}$	compris entre 0 et 1, insensible à la richesse	pas concave, symétrique, sensible aux espèces rares

H , indice hétérogène de Shannon ;

S , nombre d'espèces ;

N_{\max} , nombre d'individus de l'espèce la plus représentée ;

D , indice hétérogène de Simpson ;

Concavité : le résultat sur l'union de tous les échantillons doit être supérieur ou égal à la moyenne des résultats des échantillons ;

Symétrie : même sensibilité aux espèces rares et dominantes.

2.1.3. Indices hétérogènes

Ils intègrent la richesse et l'équitabilité (cf. tableau 4).

Tableau 4 : Indices hétérogènes.

Nom	Calcul	Avantages	Désavantages
Shanon	$H = \sum_i p_i \log(p_i)$	concave, sensible au changement en espèce rare, élément de $[0, \log(S)]$ que l'on peut ramener à $[0, 1]$ en divisant par $\log(S)$: c'est alors l'indice d'équitabilité de Pièlou	biaisé car sensible à la taille de l'échantillon
Simpson	$D = \sum_i p_i^2$	compris dans $[0, 1]$, quand D augmente, la diversité diminue, on peut utiliser $1 - D$ à la place, peu sensible aux variations de taille d'échantillon	D^{-1} n'est pas concave (cf. équitabilité de Simpson et indice de Hill : $D^{-1} \cdot H^{-1}$)

$p_i = N_i/N$, N_i abondance de l'espèce i ;

S , nombre d'espèces ;

Concavité : le résultat sur l'union de tous les échantillons doit être supérieur ou égal à la moyenne des résultats des échantillons ;

Ces indices n'ont pas d'échelle et ne peuvent pas être interprétés sans richesse et dominance.

2.1.4. Composition spécifique

La mesure la plus répandue de diversité concerne un lieu à un moment : c'est l' α -diversité. La β -diversité est la mesure de la diversité dans le temps ou l'espace. Pour un lieu donné, l' α -diversité peut être constante dans le temps. Mais si les espèces changent dans le temps la β -diversité est élevée, sinon elle est faible.

L'indice de Whittaker est défini comme suit :

$$\beta_w = S/\bar{\alpha}$$

avec S le nombre total d'espèces dans le temps ou l'espace et $\bar{\alpha}$ la moyenne du nombre d'espèces. Cette mesure est comprise entre 1 et le nombre total d'échantillons, elle peut être ramenée à l'intervalle [0 ;1]. C'est un indice simple, très utilisé, il donne une très bonne mesure de la β -diversité ;

2.1.5. Indices de diversité taxonomique

Il s'agit d'évaluer la diversité en tenant compte de sa composition. En première approche, la complexité atteinte ici laisse supposer que ce n'est pas utilisable dans notre contexte (e.g. mise en œuvre, communication).

2.2. Mesure de la biodiversité dans les évaluations environnementales en milieu marin

Ce rapport de Costello et *al.* est une recherche bibliographique ayant pour objet l'examen de méthodes d'évaluation et de mesure de la biodiversité dans le contexte des évaluations environnementales. L'inventaire effectué par les auteurs recoupe plus ou moins celui de J.C. Gaertner. Nous reproduisons ci-après des éléments de conclusion en mettant en exergue ceux qui nous semblent les plus importants :

« Les indices à une seule variable peuvent être utiles lorsqu'on cherche à résumer les formes de biodiversité. **Les plus utiles sont ceux qui mesurent la richesse des espèces et la dominance séparément sans les combiner.** Il est impossible d'interpréter les causes du changement dans les indices composites si l'on ne dispose pas de l'information de départ sur la richesse des espèces et la dominance. Selon notre analyse, l'indice de diversité le plus employé est la richesse des espèces. Par conséquent, nous recommandons la richesse des espèces et la dominance comme étant les deux indices les plus appropriés pour évaluer la biodiversité.[...] Nous avons comparé les mesures existantes selon les critères suivants : **clarté de la communication de l'information, facilité d'analyse statistique, rapidité de l'évaluation sur le terrain, au laboratoire et au bureau, et coûts.** Nous concluons que les paramètres minimaux à prendre en compte dans la mesure de la biodiversité marine lors des évaluations environnementales sont la composition spécifique, la richesse des espèces, la rareté, les biotopes présents et des caractéristiques environnementales physiques plus larges (paysages marins) [...] Bien qu'un répertoire des indicateurs écologiques dépasse la portée du présent document, certaines études montrent que **les mesures simples de la biodiversité recommandées ici sont très sensibles à la perturbation et à la pollution** (p. ex. Wildish *et al.*, 2001) **et sont par conséquent utiles dans les cas où l'évaluation environnementale vise à détecter les effets de l'activité humaine.**[...] »

Chez les auteurs, la dominance est définie comme la « proportion d'un échantillon formée par l'espèce la plus abondante » : c'est l'indice de Berger-Parker.

2.3. Synthèse

En première approche, le point de vue de Costello *et al.* pourrait être suivi en choisissant les indices :

- nombre d'espèces pour la richesse;
- indice de Berger-Parker, pour l'abondance.

3. Nature des données et de l'indice attendu

Les données disponibles aujourd'hui et de façon pérenne sont les « flores », comportant une identification et un dénombrement des taxons reconnaissables au microscope optique. Deux types de flores sont déjà utilisés pour le calcul de l'indicateur DCE abondance du phytoplancton :

- les flores totales, dénombrant tous les taxons reconnaissables ;
- les flores indicatrices, dénombrant seulement les taxons présents à une concentration supérieure à 100 000 cellules par litre.

Un indice composition tel qu'envisagé ici ne peut se calculer que sur les flores totales, et donc réduit le nombre d'échantillons potentiellement utilisables.

Dans le plan d'échantillonnage de la DCE la fréquence est d'une donnée par mois. En pratique, on peut s'attendre à ce que le plan présente des données manquantes ou plus d'une mesure par mois. Par ailleurs, les évaluations se font pour un plan de gestion de 6 ans et à l'échelle de la masse d'eau.

La plupart des indices évoqués ci-avant se calculent sur un échantillon. Il faut donc envisager tout à la fois une intégration des données spatialement et temporellement de manière à disposer d'au plus 72 mesures, et la création d'un indice pour l'ensemble du plan de gestion et pour une masse d'eau. Par ailleurs, il faut garder en mémoire les points suivants :

- le nombre de taxons et leurs abondances dépendent de l'expertise de l'observateur ;
- cette expertise varie dans le temps et l'espace ;
- les taxons sont des entités pouvant être décrites à différents niveaux : genres ou espèces le plus souvent, mais parfois familles ou classes ; la diversité des niveaux est liée à la fois à l'expertise de l'observateur, et à la difficulté voire l'impossibilité d'identification inhérente à certains taxons ;
- la richesse et l'abondance des espèces sont saisonnières.

4. Application

4.1. Données

Les données de la période 2004-2009, filtrées en vue d'une application des règles d'évaluation de la DCE, sont utilisées. Seuls les résultats concernant les flores totales ont été conservés. Six points de prélèvement ont été choisis pour illustrer la diversité des cas :

- Point 1 SRN Boulogne, en mer du nord ;
- Men Er Roué, en baie de Quiberon (Bretagne sud) ;
- Le Croisic, entre Vilaine et Loire ;
- L'Eperon Terre, dans les pertuis ;
- Villefranche, en méditerranée est ;
- Calvi, au nord-ouest de la Corse.

4.2. Méthodes

Le nombre d'espèces et l'indice de Berger-Parker sont calculés et représentés sous forme de séries temporelles. De manière à évaluer si le croisement de ces deux variables constitue un espace permettant la création d'un indice unique, la dominance est portée en abscisse et la richesse spécifique en ordonnée. Afin de disposer de points de repères ce graphique est complété, d'une part par la moyenne du nombre d'espèces et les bornes de l'intervalle de confiance à 95% des observations, et d'autre part par les valeurs arbitraires de dominances de 50% et 80%. Tous ces graphiques sont réalisés avec les mêmes échelles afin de faciliter les comparaisons.

4.3. Résultats

Les graphiques sont reportés en annexe.

L'information dispensée par les séries temporelles des richesses spécifiques est facilement compréhensible. Elles présentent une saisonnalité plus ou moins marquée selon les sites. Il est rapidement évident que le nombre de taxons est plus fort en mer du nord et plus faible en méditerranée : le nombre de taxons est élevé, souvent supérieur à 30, en mer du nord, alors que partout ailleurs en Manche-Atlantique il est moins élevé mais cependant souvent supérieur à 20. En méditerranée, il tourne autour de 10 à Villefranche, et est plutôt inférieur à 10 à Calvi. Il s'ensuit que le nombre de taxons est dépendant de la zone géographique considérée (e.g. façade) et ainsi la valeur de la richesse spécifique est une indication relative plutôt qu'absolue. Enfin, la série temporelle du Croisic montre une rupture probablement liée à un changement d'observateur.

Les graphiques de Berger-Parker sont plus difficiles à interpréter. La série de Villefranche montre un indice de Berger-Parker très bas, signe que l'abondance du taxon dominant est généralement très faible, alors que la série de Calvi montre des valeurs plus élevées, et ceci pour des richesses spécifiques relativement similaires. Les valeurs de l'indice de Berger-Parker sont incluses dans l'intervalle $[0 ; 1]$ et ainsi sont comparable d'un site à l'autre. Toutefois, il n'est pas exclu que des zones géographiques présentent des particularités à prendre en compte. Enfin, le décrochement visible dans la série temporelle de la richesse spécifique au Croisic n'apparaît pas sur celle de l'indice de Berger-Parker.

Dans les graphiques croisant richesse spécifique et Berger-Parker, deux positions sont remarquables :

- en haut à gauche, la richesse est importante et le taxon dominant constitue une faible part de l'échantillon ;
- en bas à droite, la richesse est faible et le taxon dominant constitue la plus grande part de l'échantillon.

Pour certains sites (e.g. mer du nord, Pertuis, Le Croisic), la répartition dans les différentes zones du graphique est assez homogène. Les sites de la baie de Quiberon et de Corse ont la particularité d'avoir la zone en bas à droite quasiment vide : quand le nombre de taxons est faible le taxon dominant ne constitue jamais plus de 80% de l'échantillon. Le site de Villefranche montre une répartition très hétérogène, complètement à gauche : le taxon dominant ne constitue jamais plus de 50% de l'échantillon, quel que soit le nombre de taxons.

A noter que l'utilisation de ces deux seuls indices ne permet pas de mettre en évidence un changement des taxa dans le temps.

5. Conclusion

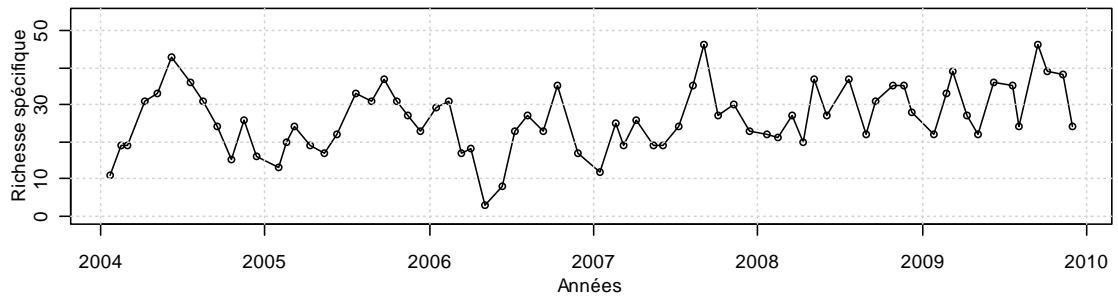
Les éléments récoltés nous incitent à désigner la richesse spécifique et l'indice de Berger-Parker comme pertinents afin d'élaborer un indice composition pour l'élément de qualité phytoplancton, compte tenu de la nature des données disponibles et des qualités attendues de l'indice.

Le traitement réalisé est une approche préliminaire. En dépit de son caractère peu raffiné, elle montre qu'il est possible d'établir un indice à partir de ces métriques. Cet objectif relève d'un travail de plus grande ampleur. Ainsi, il serait sans doute nécessaire de faire une classification ascendante hiérarchique sur la richesse pour au moins différencier la Méditerranée et le reste du littoral métropolitain. Il faut évaluer l'impact de la saisonnalité et la prendre en compte. En ce qui concerne les données, les agrégations temporelles et spatiales doivent être anticipées ainsi que les éléments de confiance et précision. L'ensemble de ce processus devrait aboutir à un EQR qui se combinera facilement aux autres indices.

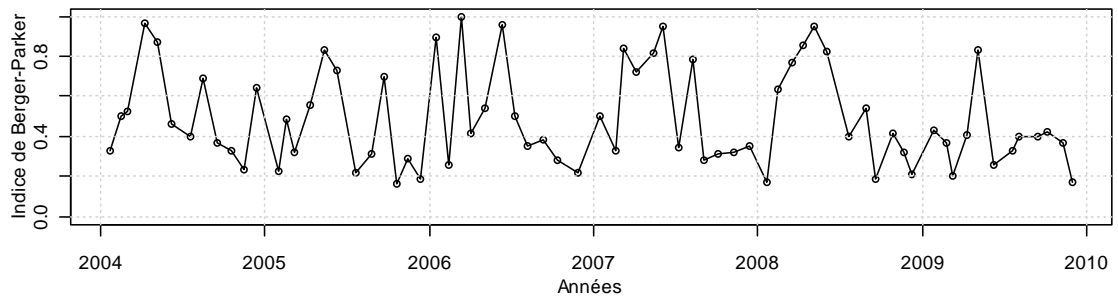
Comme il a été rappelé ci-dessus, ce genre d'indices ne peut être calculé qu'à partir de flores totales. Le choix d'une métrique basée sur ces flores totales -il est difficile d'imaginer définir un indice composition à partir de flores partielles- impliquerait donc une multiplication de ces flores totales. Cette dernière ne peut être envisagée qu'au travers de l'utilisation d'outils d'observation du phytoplancton plus automatiques, tels que la numérisation par FlowCAM associée à la reconnaissance des taxons par Phytolmage. Cette approche aurait pour autre vertu d'uniformiser l'expertise à travers le territoire et assurer sa stabilité dans le temps sans rompre le lien avec les séries historiques.

6. Annexe : graphiques richesse, Berger-Parker, et croisement des deux indices

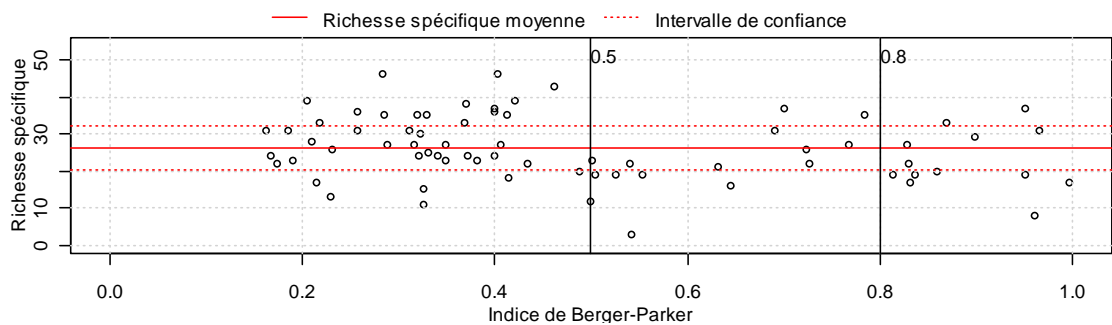
Série temporelle de la richesse spécifique
Point 1 SRN Boulogne



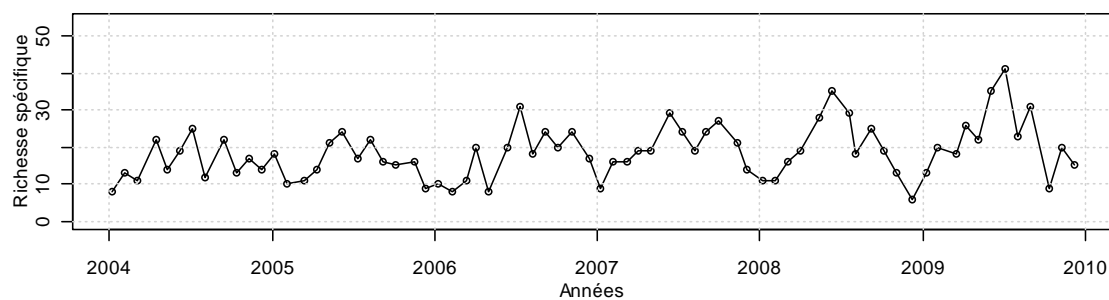
Série temporelle de l'indice de Berger-Parker
Point 1 SRN Boulogne



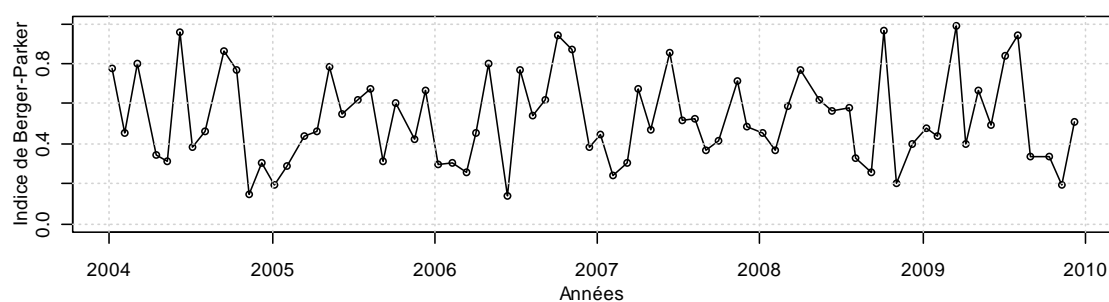
Indice de Berger-Parker et richesse spécifique
Point 1 SRN Boulogne



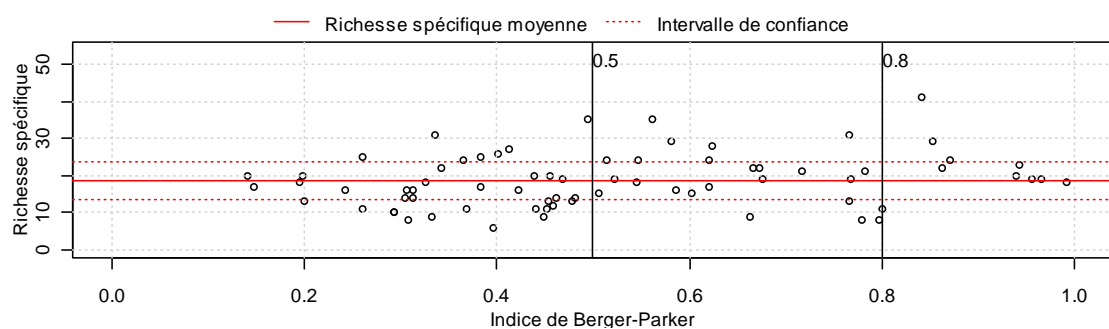
Série temporelle de la richesse spécifique Men er Roue



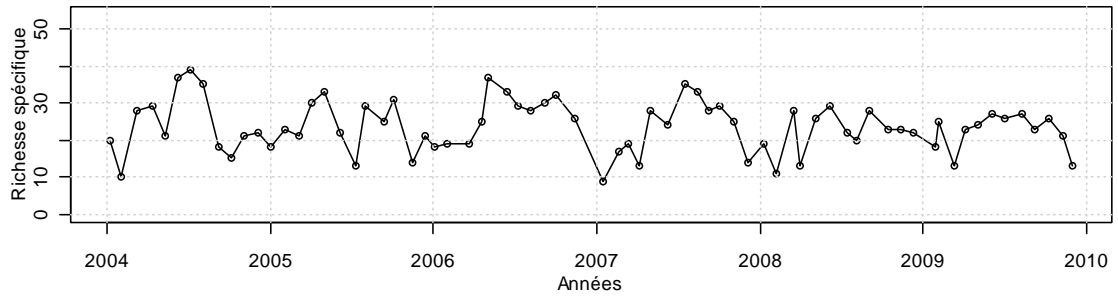
Série temporelle de l'indice de Berger-Parker Men er Roue



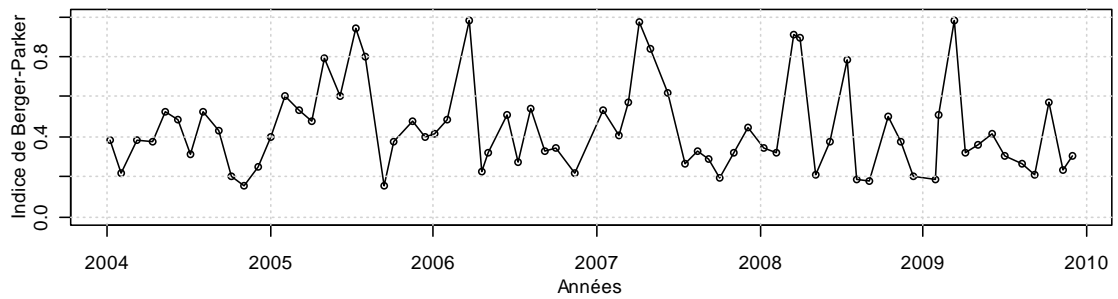
Indice de Berger-Parker et richesse spécifique Men er Roue



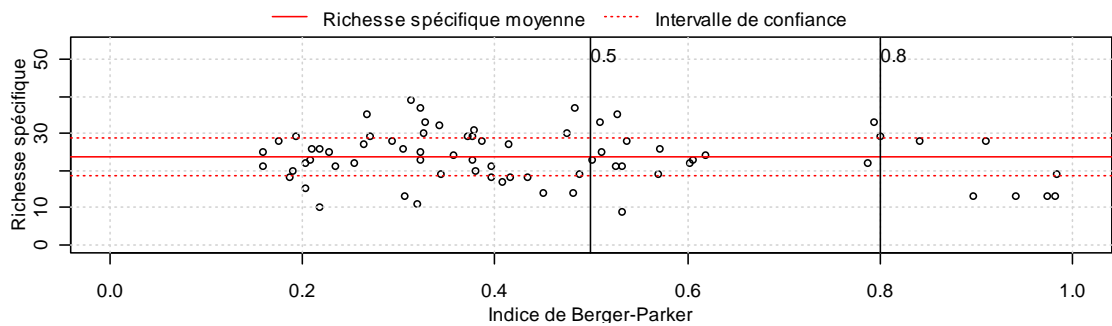
Série temporelle de la richesse spécifique L'Eperon (terre)



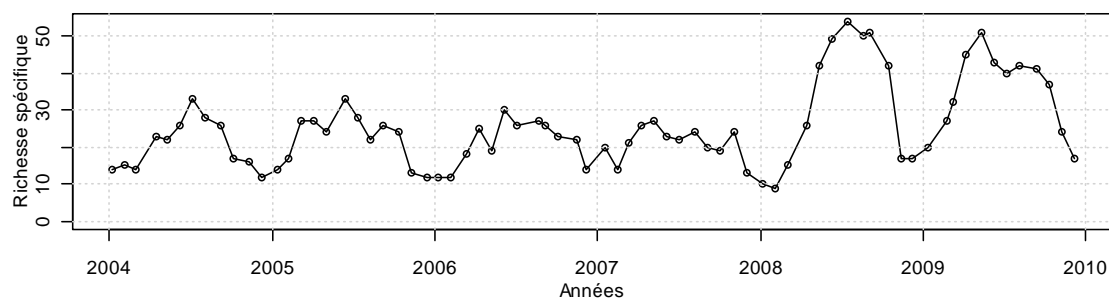
Série temporelle de l'indice de Berger-Parker L'Eperon (terre)



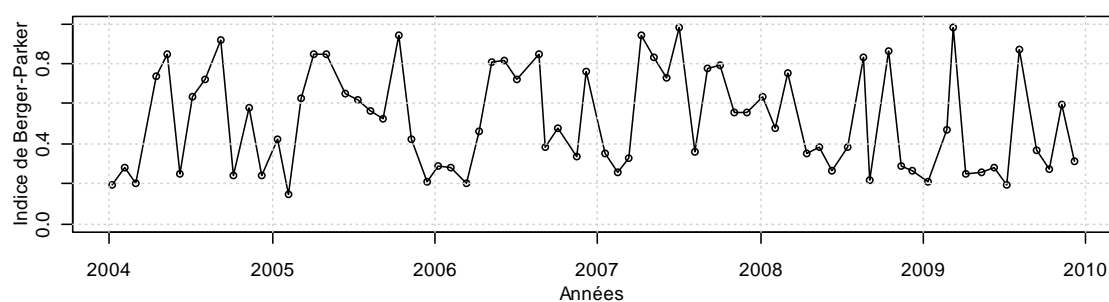
Indice de Berger-Parker et richesse spécifique L'Eperon (terre)



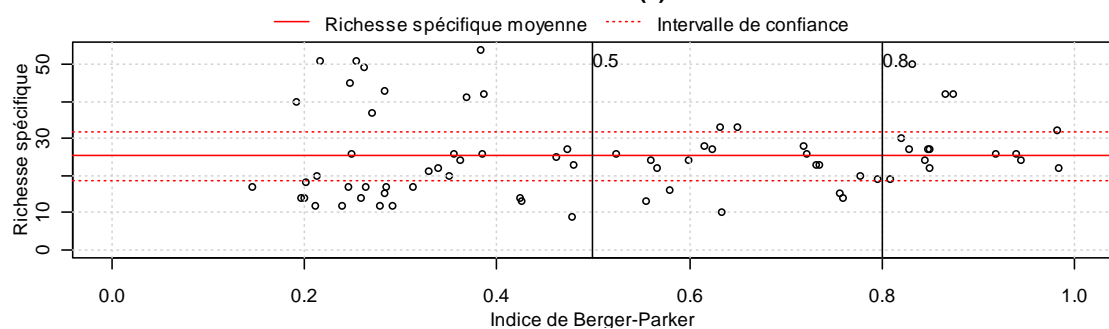
Série temporelle de la richesse spécifique Le Croisic (a)



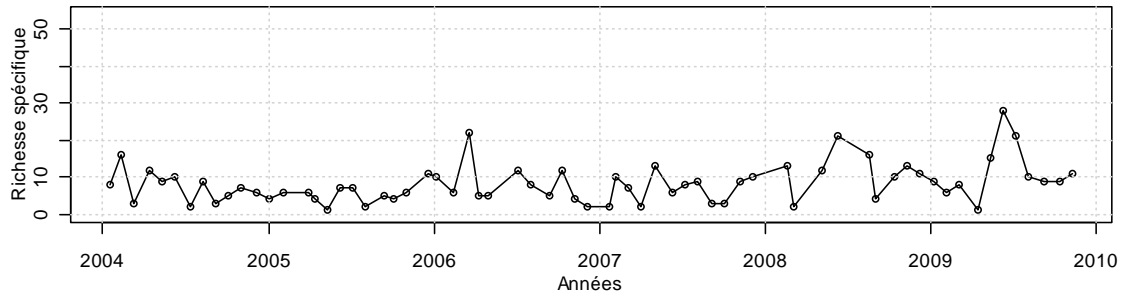
Série temporelle de l'indice de Berger-Parker Le Croisic (a)



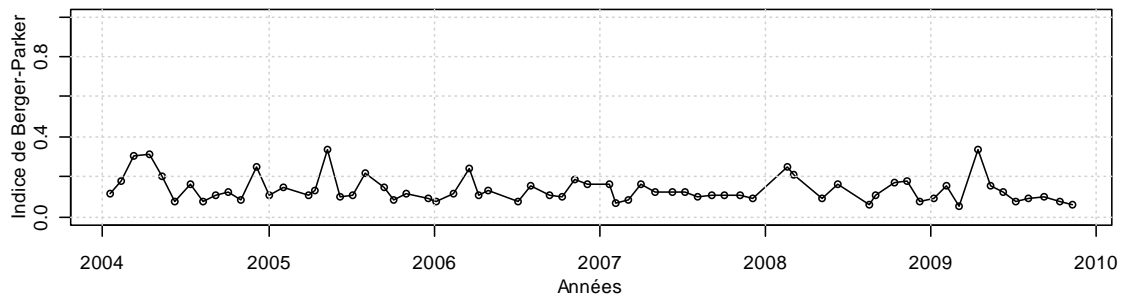
Indice de Berger-Parker et richesse spécifique Le Croisic (a)



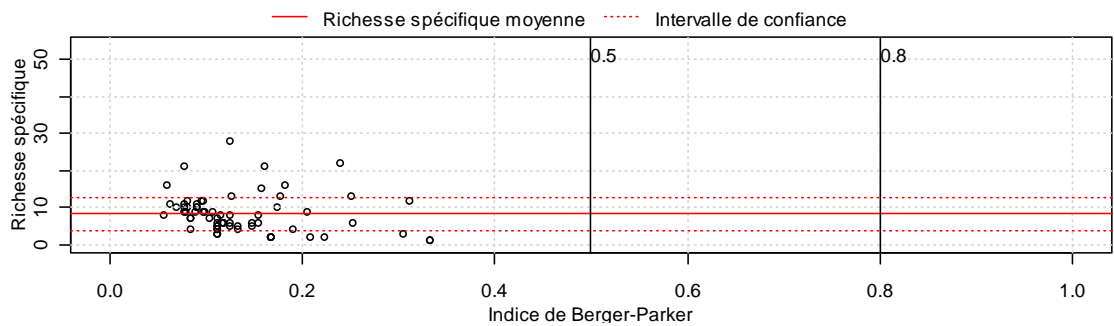
Série temporelle de la richesse spécifique Villefranche



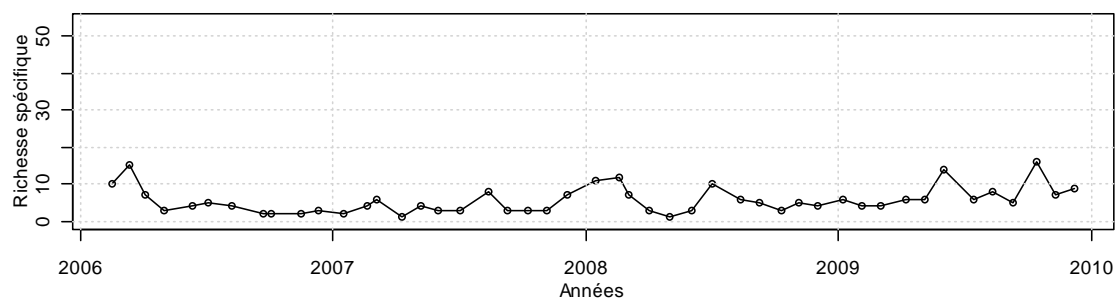
Série temporelle de l'indice de Berger-Parker Villefranche



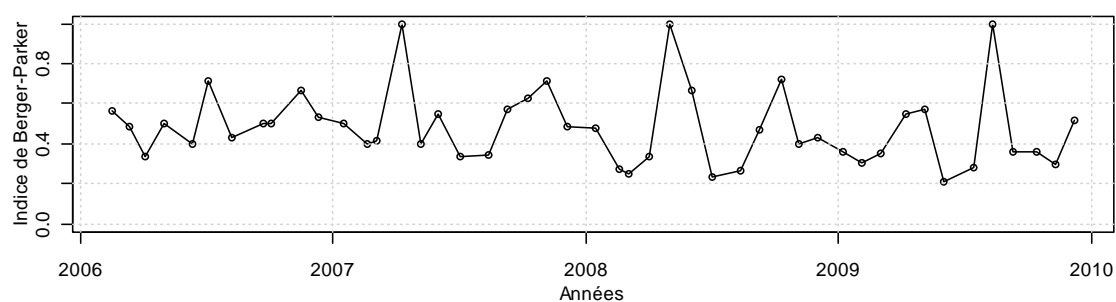
Indice de Berger-Parker et richesse spécifique Villefranche



Série temporelle de la richesse spécifique Calvi



Série temporelle de l'indice de Berger-Parker Calvi



Indice de Berger-Parker et richesse spécifique Calvi

