

Cas pratiques : Analyse multivariée

Jean-Yves Barnagaud (jean-yves.barnagaud@ephe.psl.eu (mailto:jean-yves.barnagaud@ephe.psl.eu))

2023-02-28

Objectifs

- Illustrer les analyses multivariées par des cas concrets
- Pratiquer la manipulation de la librairie `ade4`, dédiée aux ordinations
- Interpréter des graphiques d'ordinations
- Communiquer les résultats

Pour cette séquence, la restitution doit prendre la forme d'un bref exposé de 5 minutes que l'on situera dans le contexte d'une réunion interne (d'équipe, de service) dont le public est constitué de vos collègues : des écologues, mais pas nécessairement connaisseurs de votre sujet ou groupe taxonomique précis.

Séquençage

Vous commencez par prendre connaissance du cas d'étude. Comme souvent, l'analyste de données n'est pas familier du sujet, du site ou du groupe taxonomique étudié : vous devrez faire avec les informations dont vous disposez, qui ne sont que partielles.

Vous explorez ensuite les données avec les outils graphiques et numériques qui vous semblent pertinents : graphiques, tableaux de contingence, résumés divers, matrices de corrélation... Tout ce qui peut vous permettre de vous approprier les données.

Vous réfléchissez au cadre d'analyse le plus pertinent à mettre en place pour répondre à l'objectif. Pour cette séquence, il prend la forme d'une analyse multivariée.

Il faut alors implémenter cette analyse avec la librairie `ade4`. Vous disposez du support de cours et du script associé pour retrouver les fonctions R nécessaires. Dans cette formation, nous avons choisi la librairie `ade4` pour sa flexibilité et parce qu'elle a été spécifiquement conçue pour les besoins des écologues. Vous verrez que cela se paie en termes d'esthétique des graphiques...

S'il vous reste du temps, vous pourrez réfléchir aux limites de votre analyse et essayer d'aller un peu plus loin, sur les indications du formateur.

Cas d'étude n°1 : communautés d'oiseaux en Nouvelle Zélande

Objectif

Déterminer les associations entre espèces et habitat dans des communautés d'oiseaux de l'île du Sud (province de Christchurch), Nouvelle Zélande.

Jeu de données

Vous disposez:

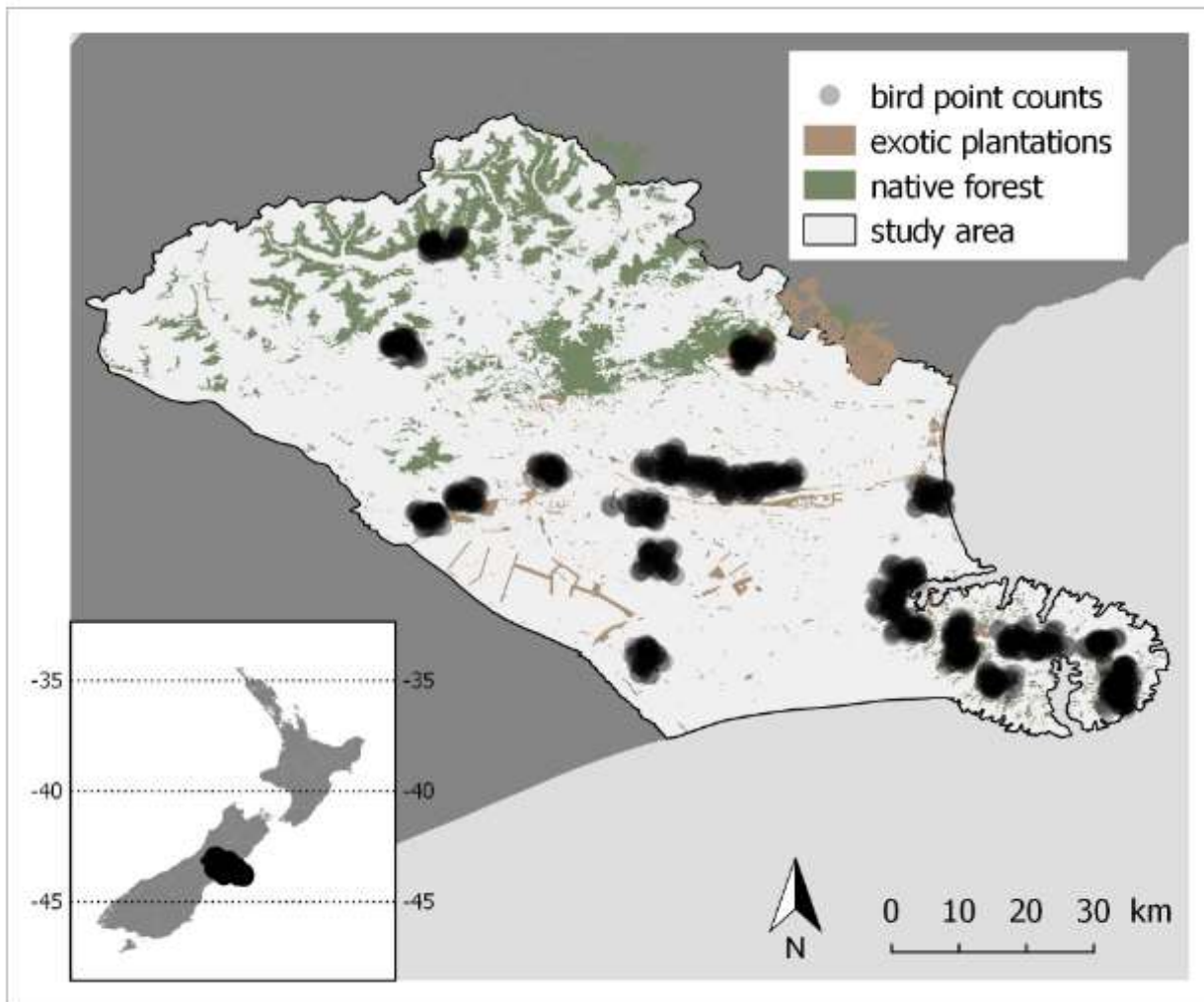
- d’une matrice de points d’écoute (917 sites x 49 espèces : *communautes_oiseaux_NZ.csv*), exprimée en nombre d’individus par point. Tous les points ont été échantillonnés une fois durant une période d’écoute de 15 minutes en 2007-2008, durant lesquelles toutes les espèces d’oiseaux vues ou entendues ont été notées.
- d’une matrice de descripteurs de l’habitat dans un rayon de 500 m autour de chaque point.

acronyme	description
EXOGRAS	végétation herbacée exotique
EXOSCRU	végétation ligneuse exotique basse
MATPLA	plantations matures d’arbres exotiques
NATFOR	forêt native mature
NATGRAS	végétation herbacée native
NATSCRU	végétation ligneuse native basse
YOUPLA	jeunes plantations d’arbres exotiques
SHA500	hétérogénéité des surfaces d’habitat (indice de Shannon)
ALTI	altitude du point d’écoute (m)
VH	hauteur de la végétation autour du point

*Ces variables (sauf les 4 dernières) sont exprimées en % de recouvrement.

La table *especes_NZ.csv* vous donne les noms des espèces et leurs status (introduites, natives, endémiques). Les espèces natives sont présentes en Nouvelle Zélande et ailleurs, les espèces endémiques ne sont présentes qu’en Nouvelle Zélande.

Vous ne disposez pas des coordonnées géographiques des points d’écoute, qui ne sont pas diffusés sur demande du propriétaire des données. Voici la carte de ces points :



Propriété des données : SCION Research / Eckehard Brockerhoff / Luc Barbaro / Marc Deconchat

Attendus

Une exploration des liens espèces - habitats. si vous avez le temps, essayez de discriminer les habitats des espèces natives, exotiques et endémiques.

Cas d'étude n°2 : traits écologiques des oiseaux de Nouvelle Zélande

Objectif

Déterminer les différences de traits écologiques entre espèces endémiques, natives et exotiques de Nouvelle Zélande.

Jeu de données

Vous disposez d'une table de données unique : la table *traits_NZ.csv*. Elle recense 10 traits écologiques pour chaque espèce, qui sont décrits dans le tableur *traits_NZ_reference.xlsx* (en anglais : appelez le formateur si vous avez besoin de traductions). En plus de ces 10 traits, la table contient les noms des espèces, leur statut (native / endémique / exotique, voir cas pratique 1 pour les définitions), et des éléments sur le contexte d'introduction des espèces exotiques.

propriété des données : JY Barnagaud (compilées et publiées dans Barnagaud et al, Diversity & Distributions 2022)

Attendus

Les traits qui discriminent (ou non) les espèces natives et exotiques. Si vous avez le temps, vous pouvez explorer l'existence éventuelle de liens entre les traits des espèces introduites et leur contexte d'introduction.

Cas d'étude n°3 : paysages sonores de Polynésie française

Objectif

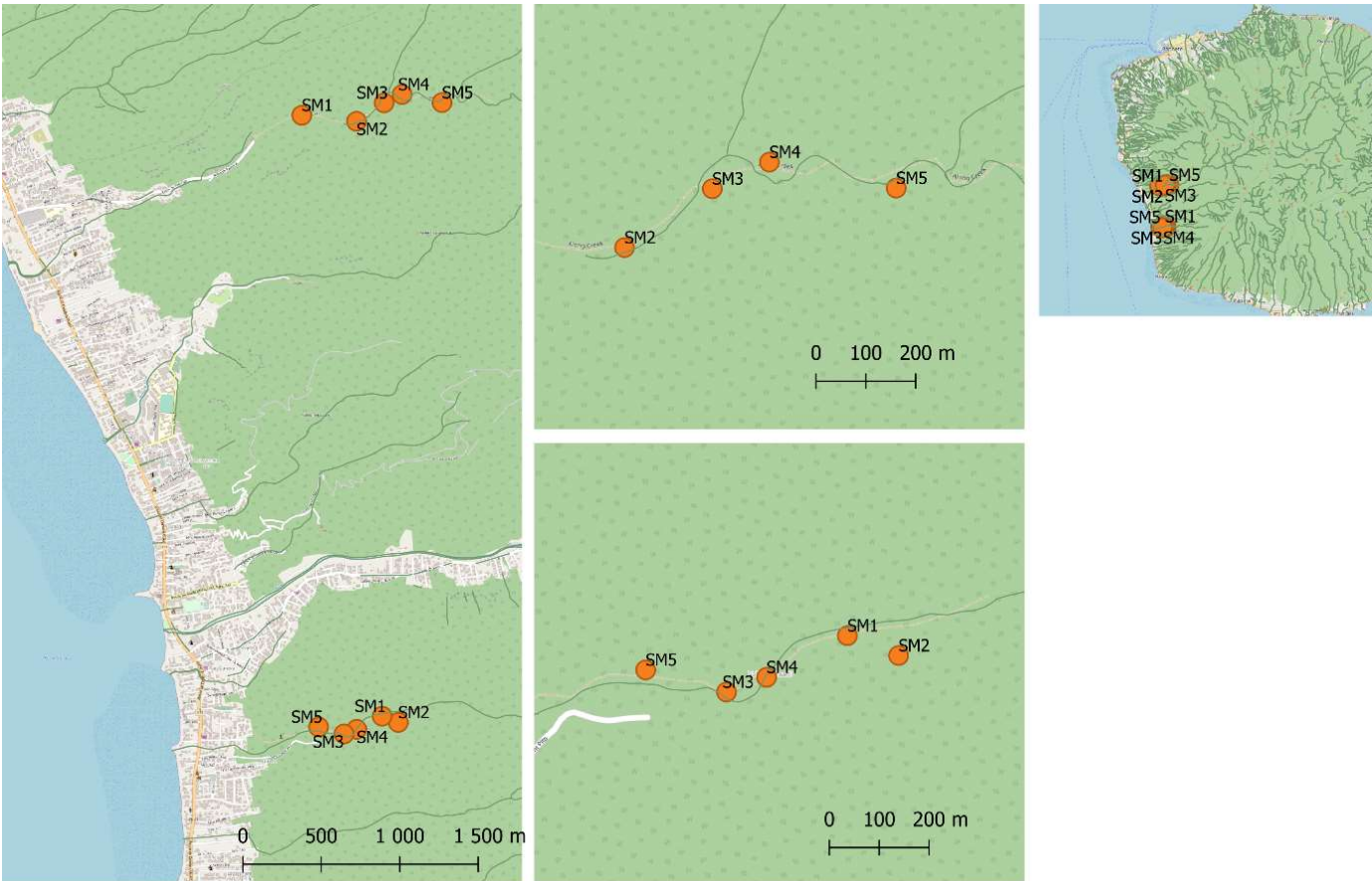
Cette étude vise à comparer les paysages acoustiques dans deux vallées de Tahiti :

- une vallée dite *test* soumise à des mesures de conservation, consistant essentiellement à l'élimination de la végétation exotique et du Bulbul orphée, une espèce invasive prédatrice s'attaquant en particulier aux oeufs du très rare et endémique Monarque de Tahiti
- une vallée dite *control*, proche de la précédente et similaire en topographie et structure de végétation, mais non soumise à des mesures de conservation.

On cherche à savoir si la composition et la diversité du paysage sonore diffère entre les deux vallées.

Jeu de données

Les deux vallées ont été échantillonnées par 5 enregistreurs automatiques Wildlife Acoustics SM Mini disposés d'amont en aval durant 48h (carte ci-dessous). Il y a une session d'enregistrement de 48h par vallée, réalisée en avril 2022. Les deux vallées ont été échantillonnées à 2 jours d'intervalle, mais sans changement de contexte ou de météo susceptible d'avoir affecté le résultat.



Les enregistreurs génèrent un fichier son toutes les 30 minutes. Ces bandes ont été analysées afin de produire des indices de diversité acoustique, qui décrivent quantitativement le paysage acoustique environnant. Le jeu de données est contenu dans *polynesie_acoustique.csv* , dont les colonnes sont :

Champ	Description
FILE	nom du fichier audio
MODALITE	vallée (test ou control)
START	heure de début de l'enregistrement
sites	combinaison vallée - emplacement de l'enregistreur (voir carte ci dessus)
ACI	Acoustic Complexity Index (décrit la variabilité temporelle de l'énergie dans un fichier son)
BIOAC	Bioacoustic index (combinaison du volume sonore et du nombre de bandes de fréquences dans un fichier son)
H	Equitabilité acoustique (régularité de couverture du spectre acoustique)
ADI	Acoustic Diversity Index (diversité des signaux entre bandes de fréquences)
NP	Nombre de pics de hautes fréquences

- N'hésitez pas à discuter l'interprétation de ces indices avec le formateur.

Dans cet exemple, la difficulté réside dans le fait qu'il y a de multiples enregistrements pour chaque site dans chaque vallée.

propriété des données : JY Barnagaud / Ecole Pratique des Hautes Etudes

Attendus

Une description de l'éventuelle différence d'indices acoustiques entre les deux vallées. Si vous avez le temps, essayez de tenir compte des enregistrements répétés par site (discutez-en avec le formateur).

Cas d'étude n°4 : communautés de rhopalocères landaises

Objectif

Cette étude est assez classique en écologie du paysage. Il s'agit d'étudier la répartition de communautés de papillons sur des gradients paysagers, en l'occurrence dans les Landes d'Aquitaine.

Jeu de données

Des transects d'échantillonnage (codés "A" + un nombre) ont été réalisés à la fin des années 2000 dans la forêt landaise. A la demande du propriétaire, les coordonnées des transects ne sont pas divulguées.

Vous disposez de deux jeux de données :

– *aquitaine_rhopaloceres.csv* : une matrice sites x espèces (exprimée en nombre d'observations) – *
read.csv2("aquitaine_rhopaloceres_habitat.csv : une matrice sites x habitats, exprimée en % de couverture de chaque habitat (NB : ces pourcentages ne somment pas à 100 car plusieurs habitats peuvent se recouvrir)

Les noms complets des espèces sont dans *aquitaine_rhopaloceres_index.csv*.

Propriété des données : Luc Barbaro / INRAE, recompilées par JY Barnagaud

Attendus

La relation entre les communautés et les habitats. Cet exemple ne pose pas de difficulté particulière : profitez-en pour aller explorer les résultats un peu plus en profondeur (sont-ils cohérents avec les éléments d'écologie des espèces que vous connaissez ou que vous trouverez sur le net?)

Cas d'étude n°5 : caractéristiques acoustiques et morphologiques des oiseaux de mer

Objectif

On cherche à déterminer s'il existe un lien entre les caractéristiques morphologiques et acoustiques de deux espèces de procellariiformes nichant sur l'Île Verte, archipel de Kerguelen : le prion de la Désolation (PDD) et le pétrel bleu (PB). Le point de départ est l'exemple traité en cours sur les données acoustiques, mais nous allons plus loin en ajoutant une table de mesures morphologiques.

Jeu de données

– *petrel_acoustique.csv* : les mesures acoustiques du cours. Elles ont été acquises par enregistrements manuels sur oiseaux posés chantant depuis leur terrier, dont le numéro de bague a été vérifié juste après l'enregistrement (les oiseaux non bagués ne sont pas enregistrés). Vous pouvez écouter les chants pour vous faire une idée de la différence entre PB et PDD (fichiers sons dans le dossier “donnees” des supports de cours). Une ligne correspond à une phrase de chant, il y en a donc plusieurs par oiseau.

CHAMP	DESCRIPTION
Indiv	identité de l'individu (numéro de bague)
PhCode	identifiant de la phrase de chant
Species	espèce (PDD ou PB)
Ph.NbSy	nombre de syllabes dans la phrase
Ph.Rythme	indicateur du rythme de la phrase
Ph.Du	durée de la phrase
Ph.FO	fréquence fondamentale de la phrase
Ph.Q50	quantile d'énergie médian de la phrase
NbPh	nombre de phrases produites dans la séquence de chant enregistrée

– *petrel_biometrie.csv* : mesures biométriques sur les oiseaux enregistrés. Il n'y a qu'une ligne par oiseau.

CHAMP	DESCRIPTION
Indiv	identité de l'individu (numéro de bague)
Mass	poids (g)
Tarsus	longueur du tarse (mm)
BillLength	longueur du bec (mm)
BillDepth	profondeur du bec (mm)
WingLength	longueur de l'aile (mm)
WingSurface	surface de l'aile (mm ²)

La difficulté de ce jeu de données réside dans le fait que l'individu statistique n'est pas le même dans les deux tables à confronter : il faudra donc trouver une manière d'en homogénéiser les dimensions.

Propriété : Charlène Gémard / Francesco Bonadonna - CNRS

Attendus

La relation morphologie - chant et l'éventuelle différence de cette relation entre espèces. Si vous avez le temps, explorez plusieurs stratégies pour prendre en compte les données acoustiques répétées.

Cas d'étude 6 : peuplements malacologiques de la vallée de Chaudefour, Auvergne

Objectif

Il s'agit d'explorer la distribution de peuplements malacologiques dans la RNN de la vallée de Chaudefour, au pied du Puy de Sancy. La commande initiale était de faire l'inventaire des mollusques de la réserve, donc une étude purement descriptive. L'opérateur a choisi de noter en plus des variables d'habitats afin de se faire une idée de ce qui conditionnait le peuplement malacologique, dont l'écologie est très mal connue. On est donc dans un contexte très exploratoire sur un groupe taxonomique pour lequel il est difficile de faire des hypothèses. Il y en a néanmoins quelques-unes :

- on attend que les peuplements soient structurés par l'altitude, la pente et l'orientation
- on attend une influence de la nature du couvert végétal (herbe, sol nu, roche, litière) – le bois mort a un intérêt particulier car c'est la seule variable que le gestionnaire peut maîtriser (via le mode d'exploitation forestière)

On pense que toutes les espèces sont favorablement influencées par la présence de litière et de bois mort. Certaines, comme *Lehmannia marginata*, peuvent éventuellement aussi bénéficier de couvert rocheux. Mais la malacologie est une discipline où les connaissances sont très lacunaire et beaucoup reste à découvrir... On est donc typiquement dans une étude qui bénéficierait d'une exploration poussée par analyses multivariées.

La méthode d'échantillonnage est la même sur tous les points de relevé : tous les points sont en milieu boisé et consistent en un carré d'1m² de surface. Il est échantillonné par recherche à vue pendant 5 minutes, puis les variables d'habitat sont relevées à vue.

Cet exemple ressemble beaucoup aux cas d'études 1 et 4. La grosse différence est qu'on ne connaît rien à l'écologie des espèces, il s'agit donc d'une étude réellement exploratoire.

Jeu de données

malaco_chaudefour_index.csv contient les noms complets des espèces et des variables environnementales

malaco_chaudefour_coordonnees.csv contient les coordonnées des sites

malaco_chaudefour.csv : peuplements malacologiques (matrice sites x espèces exprimée en nombre d'individus comptés)

environnement_chaudefour.csv contient les variables environnementales.

Propriété des données : Sylvain Vignaud (qui nous a très aimablement proposé ce cas d'étude à l'issue de la formation de 2022)

Commencez par caractériser les variations de composition des communautés, puis ajoutez les données environnementales. S'il vous reste du temps, vous pourrez faire une analyse de niche (voir avec le formateur pour l'aspect technique). Nous vous conseillons, pour faciliter vos interprétations, de cartographier les coordonnées de vos analyses multivariées dans l'espace (vous disposez pour ça des coordonnées géographiques des points en Lambert 93).