



CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE ET LA GESTION DE LA BIODIVERSITÉ



L'École Pratique des Hautes Études vous propose une formation en trois niveaux aux méthodes généralistes d'analyse de données appliquées à l'écologie. Ce parcours vous donnera les compétences fondamentales nécessaires à comprendre des résultats d'analyse statistique, puis à construire, analyser et communiquer vous-mêmes vos propres jeux de données écologiques. Vous aborderez les notions fondamentales et les mettrez en application dans un esprit pratique, à partir de véritables cas d'étude.

Vous deviendrez progressivement autonome, dans la mesure de vos besoins : comprendre l'essentiel, mettre en pratique, puis devenir force de proposition. Vous vous préparerez à aborder des méthodes plus spécialisées : modélisation d'abondances ou de dynamiques de populations, modèles de distributions d'espèces, analyses génétiques ou acoustiques...

La formation est structurée en trois niveaux. Le niveau de découverte est accessible à tous sans prérequis initial. Un test de positionnement vous permet d'évaluer vos compétences préalables si vous souhaitez démarrer aux niveaux suivants.

INFORMATIONS

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

Lieu : Maison des Sciences de l'Homme, 54 boulevard Raspail 75006 Paris.
Métro Sèvres Babylone ou Rennes.

Format présentiel uniquement

Dates en 2024 : 18 - 22 mars

Horaires : 9h30 - 12h30, 13h30 - 17h (sauf lundi : début à 10h, vendredi : fin à 15h30)

Matériel à prévoir :

- Un ordinateur portable avec MS Excel ou tout autre logiciel de tableur.
- Installer au préalable les logiciels gratuits R et R-studio, téléchargeables à ces adresses (contacter le formateur en cas de difficultés) :

<https://cran.r-project.org/>

<https://rstudio.com/products/rstudio/download/>

- Des directives pour l'installation de bibliothèques R complémentaires vous seront communiquées par le formateur une semaine à dix jours avant le début de la formation (droits administrateur nécessaires pour l'installation).



NIVEAU 2 : RENFORCEMENT

Modélisation statistique de données écologiques : une approche pratique

Ce deuxième niveau vous permettra de renforcer vos capacités d'analyse et de synthèse de données, tout en vous apportant la flexibilité nécessaire pour vous adapter à la diversité des questions et des données que vous rencontrerez en écologie. Il vise à vous donner une réelle maîtrise autonome d'une méthode particulière, le modèle linéaire et ses généralisations. S'il est loin d'être le seul chemin possible pour vos analyses et certainement pas le plus à la pointe ou à la mode (comme le sont les méthodes de classification non-supervisées ou l'intelligence artificielle), le modèle linéaire est sans aucun doute la méthode statistique la plus pratiquée par les écologues, qu'ils soient académiques ou issus des structures de gestion. Sa simplicité d'implémentation et son cadre unifié y sont sans doute pour quelque chose. Il est aussi à la base de nombreuses méthodes que vous aurez sans doute à manipuler plus tard : modèles de distributions d'espèces, modèles de dynamique de population, analyse de tendances temporelles... Vous vous entraînerez aussi à communiquer vos résultats d'analyse, en adaptant votre discours à divers publics : chercheurs, chargés d'étude, naturalistes, grand public.

A l'issue de ce niveau, vous serez en mesure de construire des modèles opérationnels répondant à des objectifs d'analyses variés. Vous saurez interpréter les résultats, les présenter, mais vous connaîtrez aussi les limites au-delà desquelles des méthodes plus complexes deviennent indispensables. Comme pour le niveau 1, nous utiliserons essentiellement des jeux de données réels, prêtés pour cette formation par une diversité de chercheurs et d'organismes que vous connaissez. Nous accorderons les après-midi à des cas concrets implémentés sous le logiciel R.

Ce module est orienté vers la pratique des méthodes d'analyses de données auxquelles vous serez réellement confrontés : votre but premier doit être avant tout de savoir les manipuler à bon escient, en ayant conscience de leurs forces et de leurs limites. Profitez de ce temps pour pratiquer, échanger, questionner. Afin de profiter au mieux de cette formation, nous vous incitons à échanger avec le formateur dès le début de semaine sur les objectifs que vous vous êtes fixé.

Et ensuite ?

La modélisation statistique est un vaste domaine qui a pour lui la flexibilité et la robustesse, mais aussi certaines limites. En fin de semaine, vous saurez analyser en autonomie des questions complexes, mais avec des données relativement simples : échantillonnage homogène, erreur non structurée dans l'espace ou le temps, pas d'erreurs d'observation. Cette étape d'appropriation des méthodes de modélisation linéaire est essentielle et vous permettra déjà de répondre à nombre de vos objectifs, mais vous constaterez rapidement que les données écologiques sont complexes : effets régions, tendances interannuelles, erreurs de détection, espèces rares... Le niveau 3 renforcera votre compétence pratique pour analyser en autonomie des jeux de données complexes, structurés dans l'espace ou dans le temps, sujets à des sources d'erreur de mesure ou non protocolés. Il vous donnera aussi la capacité de vous approprier en autonomie des méthodes statistiques nouvelles à partir de la littérature, et de communiquer ces méthodes à d'autres – en somme, devenir une personne-ressources en analyse statistique de données écologiques.

PLANNING

	Heures	Thèmes	Méthodes abordées	Compétences
Lundi	10h 12h	Explorer un jeu de données écologiques : introduction aux analyses multivariées	Analyse factorielle des correspondances, analyse en composantes principales, analyse discriminante	<ul style="list-style-type: none"> • Explorer et synthétiser un jeu de données multivarié • Sélectionner la méthode d'ordination la plus compatible avec un jeu de données • Utiliser les méthodes d'ordination pour l'inférence • Utiliser les méthodes d'ordination pour l'inférence • Communiquer les résultats d'une analyse multivariée
	13h30 17h	Travaux pratiques : les analyses multivariées avec ade4 sous R		
Mardi	9h30 12h30	Elaborer un modèle statistique pour répondre à une question écologique	Régression linéaire simple et multiple	<ul style="list-style-type: none"> • Formaliser une question biologique en un modèle statistique à plusieurs variables • Vérifier et interpréter les résultats d'un modèle • Construire un modèle linéaire au regard d'une question, l'implémenter sous R, représenter graphiquement les résultats, communiquer oralement un travail d'analyse
	13h30 17h	Travaux pratiques : construire un modèle linéaire, le coder sous R et en présenter les résultats		
Mercredi	9h30 12h30	Modéliser des présences/ absences, des fréquences et des comptages	Modèle linéaire généralisé binomial et de Poisson, traitement de comptages surdispensés (modèles quasi-Poisson, GLM négatif-binomial, inflation de zéros)	<ul style="list-style-type: none"> • Adapter le modèle linéaire au traitement de données non gaussiennes fréquentes en écologie Modéliser des données d'espèces ou événements rares • Identifier les contraintes d'un jeu de données écologiques et les transcrire en un modèle approprié • Implémenter, représenter et interpréter un GLM, communiquer des résultats de modèle
	13h30 17h	Travaux pratiques : implémenter un modèle linéaire généralisé sous R		
Jeudi	9h30 12h30	Modéliser des hypothèses biologiques complexes : optimums ou effets variables	Modélisation linéaire de structures quadratiques (polynômes du second degré), termes d'interaction, sélection de modèles, modèles non linéaires	<ul style="list-style-type: none"> • Identifier le compromis entre complexité et parcimonie dans un cadre de modélisation linéaire • Interpréter les résultats d'un modèle linéaire complexe, connaître ses limites et identifier des solutions, communiquer un travail de modélisation au format d'un séminaire
	13h30 17h	Travaux pratiques : modéliser au plus près de l'hypothèse biologique		
Vendredi	9h30 12h30	Questions personnalisées	Tous les thèmes abordés dans la semaine	
	13h30 15h30	Évaluation du niveau 2		