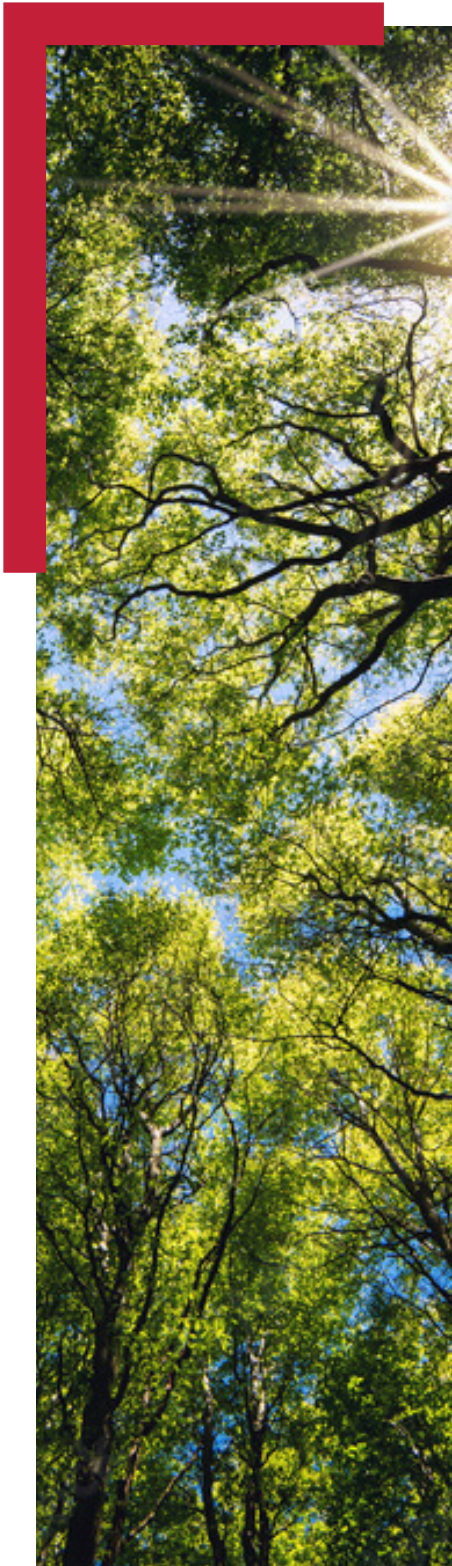




CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE ET LA GESTION DE LA BIODIVERSITÉ



La formation de l'École Pratique des Hautes Études, en analyse de données pour l'écologie et la gestion de la biodiversité ambitionne de vous donner des bases solides en statistiques généralistes afin de vous permettre de construire et d'analyser vos propres jeux de données. Elle vous fournira les compétences nécessaires afin de vous approprier des méthodes plus spécialisées, lorsque vous en aurez besoin : suivis de populations, modèles de distributions d'espèces, analyses génétiques...

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

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

Matériel à prévoir :

- Un ordinateur portable (de préférence sous Windows, mais les Mac sont acceptés) 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 modules complémentaires vous seront communiquées par le formateur une semaine à dix jours avant le début de la formation.



NIVEAU 3 : APPRONFONDISSEMENT

Modélisation statistique de données écologiques : une approche pratique

Ce troisième niveau vous apportera des compétences avancées en analyses de données écologiques. Il vise à vous rendre le plus autonome possible dans la compréhension d'une question statistique, sa traduction en un modèle adapté à des échantillonnages contraints ou structurés dans l'espace ou le temps, à des données non protocolées et/ou multivariées. A l'issue, vous devriez être capable d'analyser en autonomie la plupart des situations courantes que vous rencontrerez en analysant des données classiques en écologie et gestion de la biodiversité. Ce module ne vise pas l'exhaustivité : nous n'aborderons que quelques méthodes, mais le but est que vous vous les appropriiez le plus possible, en sachant identifier les situations auxquelles elles répondent, mais aussi en les exploitant avec un regard critique et constructif. Vous saurez aussi découvrir et mettre en œuvre des méthodes complètement nouvelles à partir d'une exploitation pertinente de la littérature scientifique. Enfin, vous vous habituerez à présenter et justifier une méthodologie d'analyse. Comme dans les deux précédents modules, vous exploiterez des jeux de données issus de cas réels, en augmentant progressivement votre niveau d'autonomie.

Si, dans les deux précédents modules, nous avons exploité des jeux de données volontairement simples ou simplifiés, vous serez cette fois face à la réalité de l'échantillonnage écologique : jeux de données stratifiés, fortes hétérogénéités de pression d'échantillonnage, données opportunistes, détection imparfaite, problèmes d'indépendance, de colinéarité, de distributions asymétriques – bref, tout ce qui peut vous mettre en difficulté dans votre travail d'analyste. Nous accorderons les après-midi à ces cas concrets, implémentés sous le logiciel R. Ce module est orienté vers la pratique telle que vous la connaîtrez en contexte réel. Nous irons peu vers la théorie statistique : en tant qu'utilisateur, votre but premier doit être avant tout de savoir manipuler correctement les concepts et les outils, plus que de les décortiquer dans toute leur complexité. Mettez à profit l'expérience que vous avez désormais acquise pour anticiper vos besoins futurs en analyse de données, les problèmes qui pourraient survenir et la manière de les résoudre.

Et ensuite ?

En fin de niveau 3, vous disposerez d'une compétence généraliste solide en analyse statistique, qui couvrira l'essentiel de vos besoins. D'autres formations, plus spécialisées (modèles de distributions ou de dynamique de population par exemple) existent et pourront ultérieurement compléter vos connaissances en fonction de vos besoins spécifiques. Néanmoins, seule la pratique sur vos propres cas d'étude vous permettra d'acquérir l'expérience nécessaire pour devenir véritablement autonome sur les méthodes et leur implémentation. Sachez exploiter l'ensemble des outils disponibles, mais restez critique : les statistiques, en particulier celles adaptées à l'écologie, restent un champ mouvant dans lequel même des méthodes bien rodées sont régulièrement remises en question. Pratiquez, documentez-vous, mettez-vous à l'épreuve, et formez-vous régulièrement à de nouvelles méthodes afin de compléter votre compétence. Sachez aussi partager vos connaissances : c'est un excellent moyen d'évoluer soi-même.

PLANNING

	Heures	Thèmes	Méthodes abordées	Compétences
Lundi	9h30 12h30	Modéliser des plans d'échantillonnage stratifiés	Modèles hiérarchiques (mixtes)	<ul style="list-style-type: none"> • Adapter la structure d'un modèle de régression aux hétérogénéités d'un jeu de données structuré • Interpréter les résultats d'un modèle mixte • Identifier les compromis relatifs à la construction d'un modèle statistique sur des données complexes • Construire, implémenter et interpréter un modèle hiérarchique sous R • Justifier ses choix de modélisation et les communiquer
	13h30 17h	Mise en pratique : les modèles mixtes sous R		
Mardi	9h30 12h30	Exploiter les méthodes bayésiennes pour analyser des données de comptage complexes	Modèles hiérarchiques bayésiens	<ul style="list-style-type: none"> • Comprendre le principe de l'inférence bayésienne • Élaborer un modèle adapté à l'analyse d'un suivi d'espèce rare ou difficilement observable • Implémenter un modèle hiérarchique pour analyser des données de comptage • Explorer la sensibilité d'un modèle aux conditions de paramétrisation
	13h30 17h	Pratique des modèles hiérarchiques bayésiens sous R		
Mercredi	9h30 17h	Analyser les distributions d'espèces dans un paysage à partir de données d'observations opportunistes	Modèles de processus ponctuels	<ul style="list-style-type: none"> • Élaborer une démarche d'analyse à partir de données inconnues • Proposer des réponses à des problèmes d'analyses complexes • Analyser des données opportunistes sous R
		Modélisation de données opportunistes sous R		
Jeudi	9h30 12h30	Explorer des données de composition de communautés écologiques	Indices de dissimilarité pour l'écologie, dendrogrammes, analyse en coordonnées principales, analyse de coinertie et méthodes inférentielles associées	<ul style="list-style-type: none"> • S'approprier une méthode inconnue à partir de la littérature • Maîtriser l'exploration de données par analyses multivariées • Acquérir une capacité à exploiter des fonctions R complexes en autonomie • Argumenter des choix méthodologiques
	13h30 17h30	Pratique des analyses multivariées sous R		
Vendredi	9h30 12h30	Questions personnalisées	Tous les thèmes abordés dans la semaine	
	13h30 15h30	Évaluation du niveau 3		