



FORMATION

Certificat en analyse de données pour l'écologie

le suivi de la biodiversité
et la gestion des
espaces naturels

f @ in v

www.ephe.psl.eu

Qualiopi
processus certifié
RÉPUBLIQUE FRANÇAISE



École Pratique
des Hautes Études

PSL 



Une formation adaptée aux professionnels de l'étude des milieux naturels, de la gestion des espaces protégés et du soutien aux politiques publiques environnementales.

La demande croissante d'expertise sur les enjeux écologiques requiert une montée en compétence des professionnels chargés de l'acquisition et de l'analyse des données de biodiversité.

Une simple synthèse de données brutes récoltées sur le terrain ne répond plus aux besoins d'anticipation et de suivi des impacts exprimés par les porteurs de projets ou de politiques publiques.

Quantifier des tendances d'indicateurs écologiques, estimer des tailles de populations ou prédire l'aire de distribution d'une espèce font désormais partie des compétences-clés d'un ingénieur écologue. Pour les cadres de structures associatives ou d'aires protégées, la capacité à mettre en place des projets collaboratifs avec des scientifiques passe par une compréhension d'ensemble

des méthodes qui garantissent la fiabilité, la reproductibilité et la transparence de conclusions basées sur les données du terrain. Pour le technicien, il s'agit avant tout d'appréhender son rôle dans la mise en place et l'élaboration des protocoles d'acquisition des données.

Afin de répondre à ces besoins, l'EPHE propose une formation pratique aux méthodes généralistes d'analyse statistique de données, spécifiquement conçue pour répondre aux besoins des écologues et des professionnels de la biodiversité.

Le déroulement du parcours, en trois niveaux, permet à chaque participant d'évoluer selon ses besoins : connaissance des notions essentielles (niveau découverte), mise en pratique des méthodes d'analyse (niveau renforcement), capacité à devenir force de proposition (niveau avancé).

Le principe de la formation repose sur la pratique : des cas concrets associés à des jeux de données réels, un formalisme limité au minimum nécessaire, des séquences d'application en groupes de travail.

L'ensemble de la formation repose sur l'**outil logiciel gratuit R**, l'environnement de travail le plus complet et le plus flexible utilisé par la grande majorité des écologues.

Un certificat, délivré sur validation d'un examen au niveau avancé, atteste de l'acquisition de l'ensemble des compétences développées pendant le parcours de formation. **Il n'y a aucun prérequis pour entreprendre le premier niveau.**



Objectifs de la formation

Le certificat valide une aptitude à :

Formuler dans le **langage des données** des questions portant sur l'évaluation écologique des impacts humains sur la biodiversité, le suivi d'indicateurs écologiques ou la connaissance des espèces et des communautés.

Mettre en place des **protocoles d'échantillonnage** adaptés aux objectifs et aux réalités du terrain.

Synthétiser rigoureusement les données, les explorer et les modéliser afin de répondre aux problèmes posés.

Interpréter avec un **regard critique et constructif** ses propres analyses de données ou celles de la littérature.

Transmettre et communiquer les résultats d'un projet d'analyse en tenant compte du public visé.

Devenir force de proposition dans la mise en place d'une stratégie d'analyse face à un problème ou des données inconnus.

Savoir évoluer vers des méthodes spécialisées (modèles de dynamiques de populations, de séries temporelles ou de distributions d'espèces par exemple).

Modalités de la formation

Inscription : de juin à fin janvier



Les inscriptions

Elles sont ouvertes dès juin et se clôturent à la fin du mois de janvier (fin octobre pour l'année 2025).

Pour une inscription directe aux niveaux *Renforcement* et *Avancé*, un test de positionnement écrit est proposé afin d'évaluer l'aptitude du candidat à suivre la formation ou le rediriger vers le niveau approprié.



Tarification

Plein tarif : 900 € TTC le module*.
À partir du deuxième module suivi à l'EPHE-PSL dans un délai de 2 ans : 650 € TTC.

Autofinancement : 350 € TTC le module.
À partir du deuxième module suivi à l'EPHE-PSL dans un délai de 2 ans : 250 € TTC.



Validation du certificat

Le certificat s'obtient par validation de l'examen du niveau avancé.

- ◆ **Une formation généraliste** en analyse de données, spécifiquement conçue pour les écologues
- ◆ **Une pédagogie tournée vers la pratique**, à partir de cas d'étude réels
- ◆ **Un parcours modulaire**, adapté aux objectifs de chaque participant
- ◆ **Un accès au niveau de découverte sans aucun prérequis**
- ◆ **Un accompagnement personnalisé** tout au long de la formation.

* La formation n'est pas éligible au Compte Personnel de Formation

Public concernés

Ce certificat s'adresse à toute personne souhaitant acquérir une maîtrise pratique des outils statistiques fondamentaux pour l'analyse de données écologiques dans un cadre appliqué.

Professionnels de l'environnement, de l'écologie et du suivi de la biodiversité : chefs de projets, chargés de mission, ingénieurs d'étude ou techniciens de terrain confrontés à la planification, l'exécution et la valorisation des protocoles d'acquisition de données ;

Chercheurs : écologues, biologistes de la conservation, géographes, doctorants et étudiants de niveau master.

Savoir faire pédagogique

La formation est assurée par un enseignant-chercheur de l'EPHE-PSL spécialisé dans l'analyse de données de biodiversité, travaillant régulièrement avec des porteurs d'enjeux publics ou privés, et naturaliste de terrain.

Au cours de votre parcours, le formateur se tient à l'écoute de vos objectifs personnels et vous propose des interactions personnalisées adaptées à vos besoins.

La pédagogie par la pratique limite le formalisme pour se concentrer sur un usage pertinent et raisonné des méthodes abordées à partir de cas réels sur des thématiques et types de données variés : suivis de populations et de distributions

géographiques, indicateurs de biodiversité, composition de communautés, écologie comportementale, traits écologiques, impacts des changements climatiques ou d'utilisation des sols...

La formation alterne séquences de cours basées sur des exemples concrets et travaux en groupe supervisés permettant de mettre en application les connaissances acquises sur des cas d'étude réels. Au fil du parcours, l'accent est mis sur la découverte de l'outil logiciel R (niveau découverte), la mise en application autonome des méthodes (niveau renforcement), puis l'élaboration de réponses appropriées à des situations inconnues (niveau avancé).

Une part conséquente de chaque séquence de travaux en groupe est accordée à l'interprétation et à la communication des résultats, avec un esprit critique et constructif.

L'évaluation, en fin de chaque semaine, vous permet de faire le point sur vos acquis et sur votre marge de progression. En lien avec le formateur, vous identifiez des objectifs pour la suite de votre parcours adaptés à votre contexte professionnel.





Méthodologie

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.



Les trois modules de 30 heures, se déroulent chacun sur 5 jours consécutifs, du lundi midi au vendredi après-midi



FORMATEUR

Jean-Yves Barnagaud
jean-yves.barnagaud@ephe.psl.eu



LIEU DE FORMATION

Format présentiel uniquement

Fondation Maison des Sciences de l'Homme
54 boulevard Raspail - 75006 Paris
(Métro Ligne 12 Sèvres Babylone ou Rennes)



MATÉRIEL À PRÉVOIR

Un ordinateur portable
avec MS Excel ou tout autre logiciel de tableur.

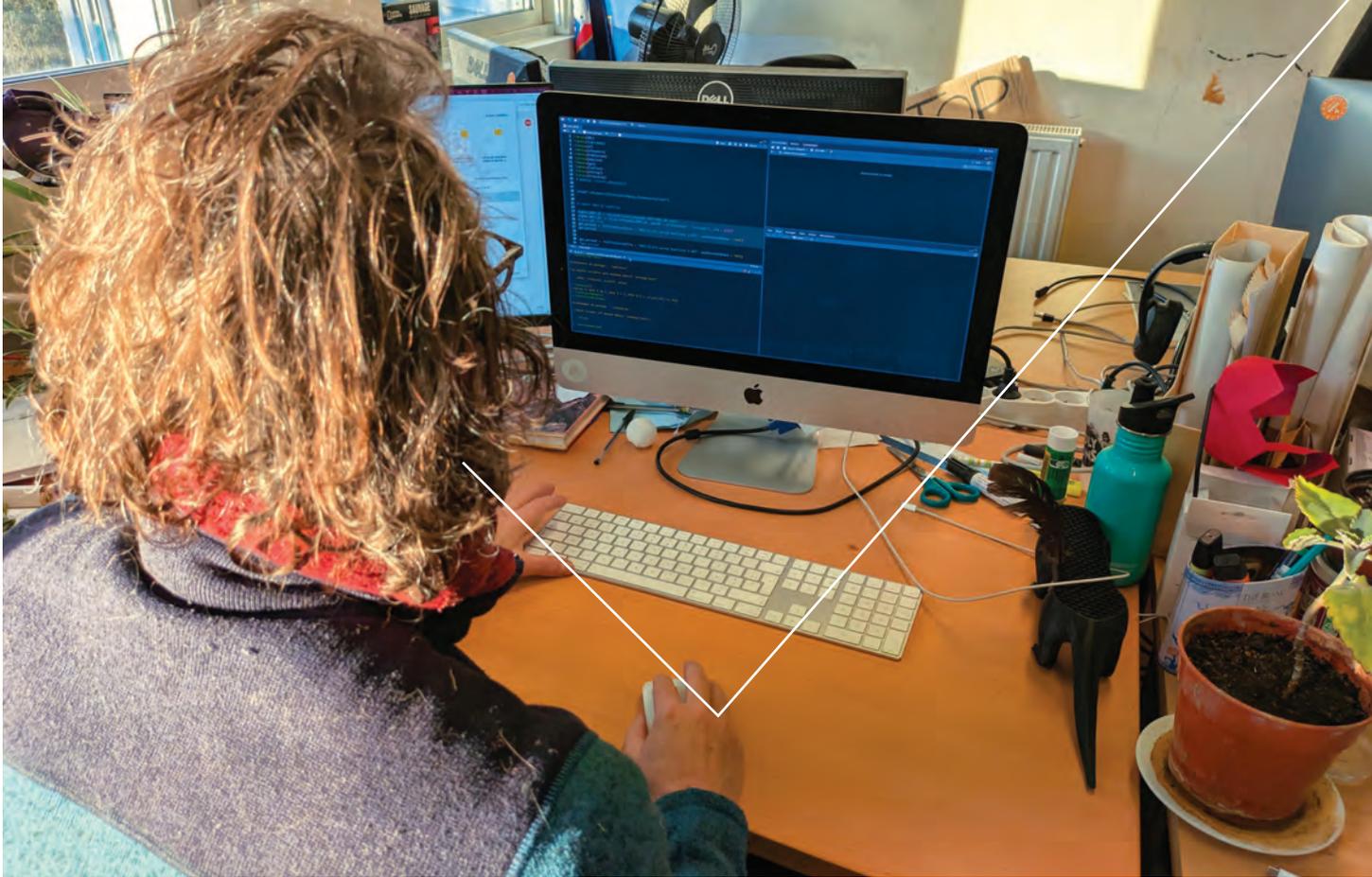
Installer au préalable les logiciels gratuits R et R-studio, téléchargeables à ces adresses :
cran.r-project.org
rstudio.com/products/rstudio/download
(contacter le formateur en cas de difficultés)

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).



INFORMATIONS

+33 (0)1 53 63 61 20
formation.continue@ephe.psl.eu



Niveau 1

Module de découverte

Une introduction à l'analyse statistique et à l'outil logiciel R pour des participants sans aucune expérience préalable de l'analyse de données.

Objectif

Comprendre les grandes étapes du flux d'analyse de données et savoir interpréter des résultats statistiques simples.

Thèmes

Construire un questionnement statistique, élaborer un protocole d'échantillonnage, mettre en forme son jeu de données et le manipuler dans l'environnement R, décrire et synthétiser des données quantitativement et graphiquement, comprendre les notions d'erreur et d'incertitude statistique, découvrir des méthodes d'inférence simples (comparaison d'échantillons, corrélations, introduction à la régression linéaire).

Ce niveau vous est utile si...

vous voulez pouvoir interagir avec des praticiens de l'analyse de données écologiques en maîtrisant leur langage et les principaux enjeux de leurs méthodes, synthétiser vous-mêmes des jeux de données simples, et contribuer à l'élaboration de protocoles de terrain.



30 heures, pendant 1 semaine



PRÉREQUIS
Aucun prérequis

Et ensuite ?

Les statistiques appliquées aux données écologiques sont un champ vaste et changeant, d'autant plus que les questions relatives à l'étude et la gestion de la biodiversité sont elles-mêmes complexes.

Ne vous attendez donc pas à être parfaitement autonome en une semaine de découverte ! À l'issue de ce premier niveau, vous disposerez d'une base solide pour comprendre à quoi sert l'analyse statistique, comment monter un protocole d'échantillonnage et comprendre un résultat statistique. Vous saurez synthétiser des jeux de données simples.

Si vous recherchez une véritable autonomie, il vous faudra maîtriser des méthodes plus avancées permettant d'analyser simultanément plusieurs variables : c'est l'objectif du niveau de renforcement. Pour aller encore plus loin, le niveau avancé vous procurera une véritable compétence pratique afin d'analyser en autonomie des jeux de données à long terme ou large échelle spatiale : suivis de populations, dynamiques de distributions ou communautés écologiques.

	HEURES	THÈMES	MÉTHODES ABORDÉES	COMPÉTENCES
LUNDI	10h 12h	Que sont les statistiques et à quoi servent-elles?	Intérêt des statistiques, définitions, questions et hypothèses, démarches scientifiques, notion d'erreur, description de données, représentations graphiques	<ul style="list-style-type: none"> ■ Mettre en place une démarche d'analyse ■ Connaître le vocabulaire ■ Interpréter des statistiques descriptives ■ Choisir une représentation graphique en fonction des données et de la question posée ■ Structurer un espace de travail ; anticiper le flux de travail
	13h30 17h	Importer et manipuler des données dans l'environnement R	Préparation et vérification de données en vue de leur analyse, mise en place de l'environnement informatique, découverte de R	
MARDI	9h30 12h30	Quantifier la variabilité d'un jeu de données	Descripteurs quantitatifs, impact de la taille d'échantillon, lois de distributions, intervalles de confiance	<ul style="list-style-type: none"> ■ Synthétiser et représenter des données au regard d'une question ■ Interpréter la variabilité d'un jeu de données ; confronter des données à des attendus théoriques ■ Mesurer l'incertitude sur la moyenne d'une série quantitative ■ Communiquer des données par voie graphique ■ Développer un regard critique sur des représentations de données
	13h30 17h	Représentations graphiques de données sous R	Fonctions graphiques essentielles avec la librairie ggplot2, réalisation, interprétation et présentation de graphiques	
MERCREDI	9h30 12h30	Pourquoi et comment échantillonner ?	Motivations de l'échantillonnage protocolé, échantillonnages probabilistes et non probabilistes, cas concrets	<ul style="list-style-type: none"> ■ Construire un plan d'échantillonnage en réponse à une question, adapter un plan d'échantillonnage aux compromis du terrain ■ Maîtriser la notion d'aléa d'échantillonnage et sa traduction sous forme d'intervalle de confiance ■ Comprendre l'idée de test de puissance et la relation incertitude, taille d'échantillon
	13h30 17h	La notion d'intervalle de confiance en pratique – travaux pratiques sur les stratégies d'échantillonnage	Construire et comprendre l'intervalle de confiance, multiples formes d'intervalles de confiance, compromis entre incertitude et faisabilité dans l'élaboration d'un échantillonnage	
JEUDI	9h30 12h30	Relations entre variables : une introduction à la modélisation statistique	Modèle de régression linéaire à une variable, estimation et interprétation des paramètres	<ul style="list-style-type: none"> ■ Répondre à une question à partir de données ■ Construire le modèle adéquat en fonction d'une hypothèse ■ Évaluer l'incertitude sur un résultat ■ Représenter un résultat d'analyse ■ Implémenter son analyse sous R ■ Visualiser et comprendre les résultats ; communiquer le résultat d'une analyse
	13h30 17h	Implémenter un modèle statistique sous R – tests statistiques pour petits échantillons	Modèles linéaires sous R, vérification et inférence sur les résultats, interprétation et critique des résultats, restitution d'analyse	
VENDREDI	9h30 12h30	Questions personnalisées	Tous les thèmes abordés dans la semaine	
	13h30 15h30	Évaluation du niveau 1		



Niveau 2

Module de renforcement

Une approche pratique des modèles linéaires pour des applications en écologie, à destination de participants qui cherchent une capacité à analyser des données en autonomie.

Objectif

savoir construire, vérifier et interpréter des modèles statistiques conçus pour les principaux types de données et de questionnements habituellement rencontrés en écologie.

Thèmes

Explorer des données par des méthodes multivariées, comprendre la notion de modèle statistique, construire des modèles linéaires simples adaptés à une diversité de données (mesures, présences/absences, comptages), savoir représenter une hypothèse biologique sous forme d'équation, maîtriser le flux de construction d'un modèle, interpréter de manière critique les sorties de modèles et les communiquer à une diversité de publics.

Ce niveau vous est utile si...

vous cherchez à devenir autonome sur l'analyse de vos propres jeux de données jusqu'à produire et communiquer vos résultats, élaborer vos modèles au regard de vos questions d'étude ou de recherche, ou acquérir les bases nécessaires à aborder des modèles complexes dans votre champ thématique (biologie des populations, distributions d'espèces, analyses spatio-temporelles de communautés...)



30 heures, pendant 1 semaine



PRÉREQUIS

Validation du niveau 1 ou test de positionnement à l'inscription

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.

	HEURES	THÈMES	MÉTHODES ABORDÉES	COMPÉTENCES
LUNDI	10h 12h	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	Élaborer 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		



Niveau 3

Module avancé

Une ouverture vers la créativité méthodologique par des mises en situation sur données écologiques réelles, pour des participants cherchant à dépasser la simple maîtrise des outils pour développer un regard critique et constructif sur des questions de modélisation statistique complexes.

Objectif

Flexibiliser sa pratique des analyses de données face à des échantillonnages complexes ou des situations inconnues.

Thèmes

Modéliser des jeux de données stratifiés (modèles hiérarchiques et mixtes), construire une stratégie d'analyse sur des données opportunistes, explorer des données de communautés d'espèces par des méthodes d'ordination multivariées, découvrir l'inférence bayésienne, aborder une nouvelle méthode statistique par la littérature, élaborer et diriger un flux de travail sur des jeux de données complexes.

Ce niveau vous est utile si...

vous ambitionnez de devenir force de proposition sur les analyses de données dans votre structure, cherchez à développer une aptitude à élaborer vos propres flux d'analyse sur des jeux de données inconnus, vous préparer à travailler sur des données complexes en autonomie ou en collaboration avec des biostatisticiens.



30 heures, pendant 1 semaine



PRÉREQUIS

Validation du niveau 2 ou test de positionnement à l'inscription

Et ensuite ?

En fin de niveau 3, vous disposerez d'une compétence généraliste solide en analyse statistique, qui couvrira l'essentiel de vos besoins quotidiens comme écologue chargé d'analyser des données dans une structure d'expertise ou de gestion-protection des espaces naturels.

Vous serez aussi prêt à évoluer vers d'autres formations spécialisées, couvrant des thèmes plus spécifiques (modèles de distributions d'espèces, dynamique de population, estimations d'abondances) ou des outils avancés (inférence bayésienne, science reproductible).

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.

	HEURES	THÈMES	MÉTHODES ABORDÉES	COMPÉTENCES
LUNDI	10h 12h30	Modéliser des plans d'échantillonnage stratifiés	Modèles 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	Introduction à l'inférence bayésienne : principe et mise en pratique pour l'analyse de données de comptages	Modèles linéaires en cadre bayésien	<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 ■ Explorer la sensibilité d'un modèle aux conditions de paramétrage
	13h30 17h	Pratique des modèles hiérarchiques bayésiens sous R		
MERCREDI	9 h 30 17h	Analyser la répartition d'un processus écologique dans l'espace à partir de données non protocolées	Analyse de distributions d'espèces, modèles de processus ponctuels, Modèles linéaires généralisés en contexte spatial	<ul style="list-style-type: none"> ■ Élaborer une démarche d'analyse à partir de données inconnues ■ Comprendre les problèmes posés par les données non protocolées issues d'échantillonnages opportunistes ■ Développer une stratégie d'analyse par une démarche collaborative et itérative
JEUDI	9h30 12h30	Analyse d'indicateurs de biodiversité et de composition de communautés écologiques	Indices de dissimilarité pour l'écologie, indicateurs de diversité et de composition, analyses multivariées multi-tables	<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	Analyse de données de communautés sous R		
VENDREDI	9h30 12h30	Questions personnalisées	Tous les thèmes abordés dans la semaine	
	13h30 15h30	Évaluation du niveau 3		

Enseigner la recherche par la pratique de la recherche

EPHE

Les Patios Saint-Jacques
4-14, rue Ferrus 75014 Paris

+33 (0)1 53 63 61 20
formation.continue@ephe.psl.eu



www.ephe.psl.eu



École Pratique
des Hautes Études

