



## CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE, LE SUIVI DE LA BIODIVERSITÉ ET LA GESTION DES ESPACES NATURELS

*Une formation à l'analyse de données écologiques 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 :

### 1. Inscription : d'octobre à fin janvier

- Les inscriptions sont ouvertes dès le 1<sup>er</sup> octobre 2024 et se clôturent le 31 janvier 2025.
- Pour une inscription directe aux niveaux 2 et 3, 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.

### 2. Tarification

- Plein tarif : 900 € TTC le module\*. À partir du deuxième module suivi à l'EPHE - PSL dans un délai de 2 ans : 550 € TTC.
- Autofinancement : 350 € TTC le module. À partir du deuxième module suivi à l'EPHE - PSL dans un délai de 2 ans : 200 € TTC.

### 3. Validation du certificat

#### Déroulé :

- Le certificat s'obtient en validant l'examen du niveau avancé.

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

- **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.**

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



## PUBLICS 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.



## PROGRAMME ET CALENDRIER

La formation consiste en **trois modules de 30 heures**, se déroulant **chacun** sur 5 jours consécutifs, du lundi matin au vendredi après-midi.

- **Niveau 1 : module de découverte (du 17 au 21 mars 2025)**

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

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

Thèmes : construire un questionnaire 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.

- **Niveau 2 : module de renforcement (du 24 au 28 mars 2025)**

***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. Prérequis : validation du niveau 1 ou test de positionnement à l'inscription [T3SATIS2].***

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...)



## PROGRAMME ET CALENDRIER

- **Niveau 3 : module avancé (du 31 au 04 avril 2025)**

**Prérequis : validation du niveau 2 ou test de positionnement à l'inscription [T3STAT13].**

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.

### CONTACT

Responsable pédagogique :

**Jean-Yves BARNAGAUD**

[jean-yves.barnagaud@ephe.psl.eu](mailto:jean-yves.barnagaud@ephe.psl.eu)

Maître de conférences

École Pratique des Hautes Études - PSL

**Direction de la Formation continue :**

Inscription administrative, conditions et modalités  
de financement :

[formation.continue@ephe.psl.eu](mailto:formation.continue@ephe.psl.eu)

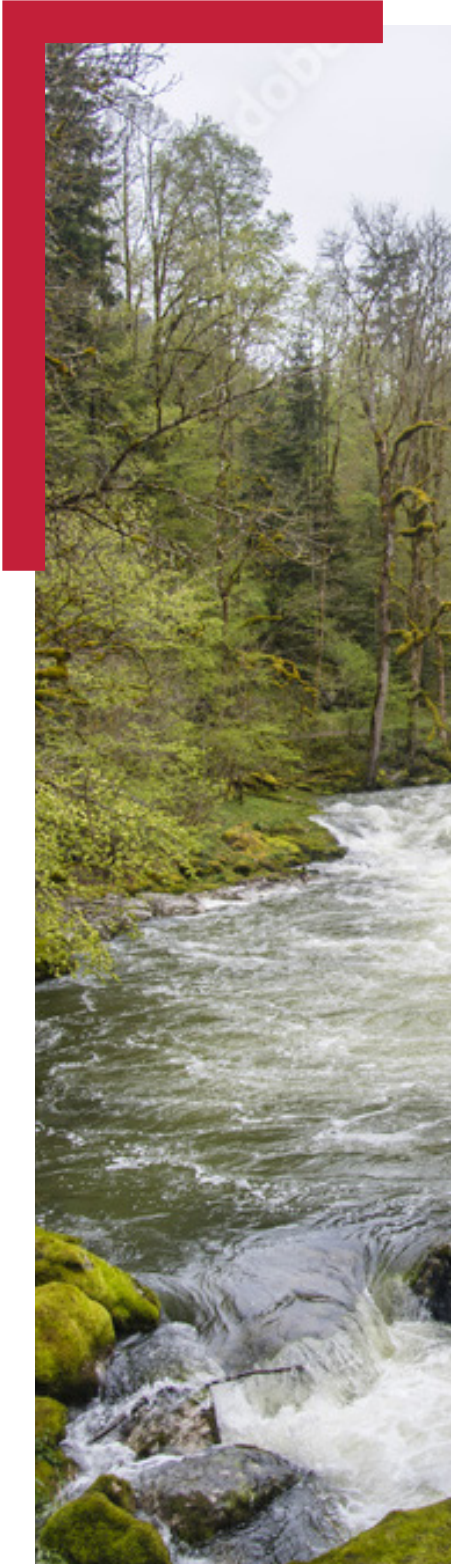


École Pratique  
des Hautes Études

PSL 



## CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE, LE SUIVI DE LA BIODIVERSITÉ ET LA GESTION DES ESPACES NATURELS



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 et Campus Condorcet, Maison des Recherches Nord, salle 0.010.

### Format présentiel uniquement

**Dates en 2025** : Niveau 1: 17 - 21 mars, Niveau 2: 24 - 28 mars,  
Niveau 3: 31 mars - 04 avril

**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 1 : DÉCOUVERTE

### Une première approche pratique de l'analyse de jeux de données écologiques

Ce premier niveau est une **découverte pratique de l'univers des analyses statistiques de données** dans le contexte de l'écologie. Nous aborderons leur raison d'être, leurs enjeux et limites, la manière de construire un échantillonnage, d'explorer et de communiquer des données. En fin de formation, nous commencerons à construire des modèles statistiques en réponse à des questions simples. Nous utiliserons essentiellement des jeux de données réels, prêtés pour cette formation par une diversité de chercheurs et d'organismes que vous connaissez : CNRS, OFB, INRAE, ONF, associations, réserves naturelles...

La formation est résolument axée vers **une approche pragmatique** : nous irons peu vers la théorie statistique, car nous pensons qu'en tant qu'utilisateur, votre but premier doit être avant tout de savoir manipuler correctement les méthodes et les outils, plutôt que de les décortiquer dans toute leur complexité. Nous accorderons les après-midi à une initiation de l'exploration de données sous le logiciel R.

Cette formation est principalement destinée à **des participants qui n'ont aucune expérience préalable en analyse de données** ou qui souhaitent recycler des connaissances et compétences fondamentales. Il n'y a donc aucun prérequis. L'objectif pour vous cette semaine n'est pas de devenir autonome sur vos analyses, mais d'appréhender les principales notions qui vous permettront de comprendre les statistiques présentées dans des rapports ou des articles, ou d'interagir plus facilement avec des chercheurs ou chargés d'études scientifiques qui travaillent au quotidien sur des données.

Nous vous incitons à communiquer dès le début de la formation sur les objectifs que vous vous êtes fixé sur cette semaine, afin de personnaliser au mieux les interactions avec le formateur. Parce que cette formation est résolument orientée vers la pratique, soyez-en acteurs : questionnez, échangez, critiquez, afin que cette semaine soit la plus riche possible.

### 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 ! A 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.

# PLANNING

	Heures	Thèmes	Méthodes abordées	Compétences
Lundi	10h 12h	Pourquoi des analyses statistiques en écologie ?	La démarche statistique, son champ d'application, ce qu'on peut en attendre ou non, grandes notions centrales	<ul style="list-style-type: none"> <li>Comprendre l'intérêt d'une démarche fondée sur l'analyse de données pour l'écologie</li> <li>Connaître les grandes étapes d'une analyse statistique et les principaux résultats à en attendre</li> <li>Maîtriser les éléments clé du vocabulaire</li> <li>Structurer un espace de travail en vue d'une analyse de données</li> </ul>
	13h30 17h	Structurer un questionnaire et un jeu de données en vue d'une analyse	Préparation et vérification de données en vue de leur analyse, mise en place de l'environnement informatique, découverte de R	
Mardi	9h30 12h30	Qu'est-ce que la variabilité de données écologique et pourquoi s'en préoccupe-t-on ?	Indicateurs quantitatifs de séries de données, idée de descripteurs et d'estimateurs, notion d'échantillonnage aléatoire et de taille d'échantillon, utilité des lois de distributions, estimation d'un intervalles de confiance	<ul style="list-style-type: none"> <li>Synthétiser et représenter des données au regard d'une question</li> <li>Interpréter la variabilité d'un jeu de données ; confronter des données à des attendus théoriques</li> <li>Mesurer l'incertitude sur la moyenne d'une série de données et comprendre son interprétation</li> <li>Communiquer des données par voie graphique</li> <li>Développer un regard critique sur des représentations de données</li> </ul>
	13h30 17h	Représentations graphiques de données	Fonctions graphiques essentielles avec la librairie ggplot2 sous R, 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"> <li>Construire un plan d'échantillonnage en réponse à une question, adapter un plan d'échantillonnage aux compromis du terrain</li> <li>Maîtriser la notion d'aléa d'échantillonnage et sa traduction sous forme d'intervalle de confiance</li> <li>Comprendre et exploiter la relation incertitude - taille d'échantillon</li> </ul>
	13h30 17h	L'échantillonnage de données écologiques : planifier et dimensionner	Construction d'un plan d'échantillonnage en réponse à une question, dimensionnement de l'effort en fonction du résultat attendu, idée de test de puissance	
Jeudi	9h30 12h30	Qu'est-ce qu'un modèle statistique ?	Modèle de régression linéaire à une variable, estimation et interprétation des paramètres	<ul style="list-style-type: none"> <li>Répondre à une question à partir de données</li> <li>Comprendre la démarche de modélisation et son intérêt en écologie</li> <li>Construire le modèle adéquat en fonction d'une hypothèse</li> <li>Évaluer l'incertitude sur un résultat</li> <li>Représenter un résultat d'analyse</li> <li>Implémenter son analyse sous R</li> <li>Visualiser et comprendre les résultats ; communiquer le résultat d'une analyse</li> </ul>
	13h30 17h	Implémenter, interpréter et représenter un modèle linéaire sous R	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 : 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"> <li>• Explorer et synthétiser un jeu de données multivarié</li> <li>• Sélectionner la méthode d'ordination la plus compatible avec un jeu de données</li> <li>• Utiliser les méthodes d'ordination pour l'inférence</li> <li>• Utiliser les méthodes d'ordination pour l'inférence</li> <li>• Communiquer les résultats d'une analyse multivariée</li> </ul>
	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"> <li>• Formaliser une question biologique en un modèle statistique à plusieurs variables</li> <li>• Vérifier et interpréter les résultats d'un modèle</li> <li>• Construire un modèle linéaire au regard d'une question, l'implémenter sous R, représenter graphiquement les résultats, communiquer oralement un travail d'analyse</li> </ul>
	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 surdispersés (modèles quasi-Poisson, GLM négatif-binomial, inflation de zéros)	<ul style="list-style-type: none"> <li>• Adapter le modèle linéaire au traitement de données non gaussiennes fréquentes en écologie Modéliser des données d'espèces ou événements rares</li> <li>• Identifier les contraintes d'un jeu de données écologiques et les transcrire en un modèle approprié</li> <li>• Implémenter, représenter et interpréter un GLM, communiquer des résultats de modèle</li> </ul>
	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"> <li>• Identifier le compromis entre complexité et parcimonie dans un cadre de modélisation linéaire</li> <li>• 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</li> </ul>
	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 : APPRONFONDISSEMENT

### Modélisation statistique de données écologiques : une approche pratique

Cette semaine vise à vous rendre aussi autonome que possible dans votre pratique des analyses de données au quotidien, en vous faisant passer de l'utilisation éclairée de méthodes connues (niveau 2 – renforcement) à **la mobilisation de nouvelles méthodes** répondant à des enjeux émergeant de jeux de données complexes : échantillonnages stratifiés (multiples régions ou espèces), données opportunistes issues des sciences participatives, grandes tables de communautés. A l'issue de cette formation, vous devriez **être capable d'élaborer vous-même des stratégies d'analyse**, de les soumettre à la discussion collective, et de mobiliser de nouvelles méthodes à partir de la littérature.

Si, dans les deux premiers niveaux de ce certificat, nous avons surtout ciblé la compréhension des méthodes, ce troisième module mobilisera surtout votre créativité, votre capacité à croiser les informations et à vous approprier des situations inconnues : vous deviendrez ainsi une véritable **personne-ressource** dans votre structure. Parce qu'il s'agit moins de maîtriser un grand nombre de techniques que de développer une flexibilité thématique et méthodologique, nous n'aborderons pas d'outils statistiques fondamentalement nouveaux par rapport au niveau de renforcement. Modèles linéaires et analyses multivariées resteront notre base, mais vous les utiliserez différemment, en vous adaptant au contexte. L'essentiel du module prend la forme de **mises en situation** issues de véritables collaborations entre porteurs d'enjeux et analystes : estimation des effets du changement climatique sur une espèce protégée, étude des effets de changements paysagers sur une espèce rare, approche fonctionnelle de l'impact d'une espèce sur son milieu, analyse des relations communautés – environnement dans le but d'interpréter des indicateurs de biodiversité...

L'analyse de données écologiques est un travail collaboratif qui rassemble naturalistes, écologues et biostatisticiens. Vous développerez donc **vos positionnement** dans des groupes de travail, en identifiant quels sont vos atouts et vos limites, et comment les mettre à profit dans un raisonnement collectif sur les données. Afin de profiter au mieux de cette approche, nous vous incitons à échanger avec le formateur dès le début de semaine sur les objectifs que vous vous êtes fixé au regard de vos besoins professionnels.

Enfin, une composante importante de cette formation réside dans la **communication** tant des résultats que des méthodes : vous travaillerez votre capacité à expliquer clairement vos choix méthodologiques, en justifier la pertinence, et exposer résultats, limites et perspectives de manière équilibrée et constructive.

### 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.

# PLANNING

	Heures	Thèmes	Méthodes abordées	Compétences
Lundi	10h 12h	Adapter un modèle aux contraintes de l'échantillonnage : exemple des plans d'échantillonnage	Modèles hiérarchiques et modèles mixtes	<ul style="list-style-type: none"> <li>• Adapter la structure d'un modèle de régression aux hétérogénéités d'un jeu de données structuré</li> <li>• Interpréter les résultats d'un modèle mixte</li> <li>• Identifier les compromis relatifs à la construction d'un modèle statistique sur des données complexes</li> <li>• Construire, implémenter et interpréter un modèle hiérarchique sous R</li> <li>• Justifier ses choix de modélisation et les communiquer</li> </ul>
	13h30 17h	Mise en pratique : élaborer, implémenter, contrôler et interpréter un modèle mixte		
Mardi	9h30 12h30	Vers une flexibilisation du cadre de modélisation : introduction au cadre bayésien	Modèles linéaires en cadre bayésien	<ul style="list-style-type: none"> <li>• Comprendre le principe de l'inférence bayésienne et son utilité en analyse de données écologiques</li> <li>• Élaborer un modèle adapté à l'analyse d'un suivi d'espèce rare ou difficilement observable</li> <li>• Explorer la sensibilité d'un modèle aux conditions de paramétrage</li> </ul>
	13h30 17h	Pratique des modèles hiérarchiques bayésiens sous R		
Mercredi	9h30 17h	Identifier les enjeux d'analyse et proposer une méthodologie face à un questionnement et un jeu de données inconnus : le cas des données opportunistes en contexte spatial	Analyser la répartition d'un processus écologique dans l'espace à partir de données non protocolées	<ul style="list-style-type: none"> <li>• Élaborer une démarche d'analyse à partir de données inconnues</li> <li>• Comprendre les problèmes posés par les données non protocolées issues d'échantillonnages opportunistes</li> <li>• Développer une stratégie d'analyse créative par une démarche collaborative et itérative</li> </ul>
Jeudi	9h30 12h30	Identifier, comprendre et mettre en pratique une méthodologie d'analyse de données en s'appuyant sur la littérature : exemple des données 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"> <li>• Développer une méthodologie d'analyse face à des données inconnues en s'appuyant sur l'état de l'art</li> <li>• Comprendre les enjeux, caractéristiques et contraintes d'une méthode à partir de la littérature</li> <li>• Elaborer un flux d'analyse dimensionné à la question écologique et aux données disponibles</li> <li>• Argumenter ses choix d'analyse</li> </ul>
	13h30 17h30			
Vendredi	9h30 12h30	Questions personnalisées	Tous les thèmes abordés dans la semaine	
	13h30 15h30	Évaluation du niveau 3		