



Projet de Probabilités et Statistiques

Cédric Travelletti et Florian Desmons

2025-2026

Titre du sujet : *Analyse de la faune et de la flore en Suisse*

Groupe : Axel Hall, Jonathan Müller, Fatima Anougmar, Sebastian Morsch

Définition de la problématique

La **faune** et la **flore** sont étroitement liées et entretiennent des **relations de coexistence complexes**. Leur évolution mutuelle influence directement la **stabilité** et la **diversité** des écosystèmes. Cette étude vise à **identifier des tendances** dans l'évolution des données relatives aux espèces, afin d'**envisager la possibilité de prédire** le développement des écosystèmes et d'**évaluer leur stabilité**.

Pour des raisons d'interprétation, la **dynamique d'une espèce** est ici définie comme **l'évaluation de sa présence au sein d'un écosystème et des relations qu'elle entretient avec les autres espèces**.

Elle **n'inclut pas** l'étude du déplacement d'écosystèmes en interaction avec d'autres ensembles écologiques.

Chaque espèce est ainsi évaluée **en fonction de l'écosystème auquel elle appartient**, défini préalablement selon des critères écologiques et géographiques.

L'étude s'axe autour de la problématique suivante :

« La modélisation de la dynamique d'une espèce permet-elle de prédire son évolution au sein de son écosystème ? »

Explication des données obtenues ou à collecter

Les données utilisées proviennent de la plateforme **iNaturalist.org**, l'une des plus grandes bases de référencement d'espèces au monde. Elle recense plus de **500 000 espèces**, soit environ **23 % des espèces connues**. Grâce à sa large communauté d'utilisateurs et à son usage fréquent dans la recherche scientifique, iNaturalist constitue une **source de données crédible et reconnue**.

Les enregistrements s'étendent sur la période **de 1938 à 2024**.

- Les données acquises **entre 1938 et 2011** proviennent de **sources externes** telles que des articles scientifiques ou des études documentant la présence d'espèces.
- À partir de **2011**, les **utilisateurs de la plateforme** — qu'ils soient amateurs ou chercheurs — peuvent **soumettre directement leurs observations**, enrichissant ainsi continuellement la base de données.

Dans le cadre de cette étude, **seules les données concernant la Suisse** ont été retenues, représentant **environ 600 000 observations**, fournies **au format CSV**.

Données à disposition

- Identifiant de l'observation
- Identifiant de l'observateur anonymisé
- Latitude
- Longitude
- Précision positionnelle
- Identifiant de l'espèce (taxon)
- Niveau de l'observation (amateur, recherche)
- Date de l'observation

Pré-traitement

- Filtrage des observations avec
- Normalisation temporelle (éviter le biais lié au nombre d'utilisateurs)
- Informations supplémentaires de l'observation à partir de l'identifiant
 - Call API -> www.inaturalist.org/pages/api+reference
 - Autre (à définir au cours du projet)

Objectifs

- Visualisation en arbre des espèces
 - But :
 - Mettre en évidence des paires d'espèces fréquemment co-observées.
 - Exemple de visualisation [link]
 - Concepts :
 - Co-occurrence
 - Sources :
 - Longitude, latitude, taxon, date

- Cartographie des écosystèmes en Suisse
 - But :
 - Identifier des groupements spatiaux d'espèces et esquisser des « écosystèmes types ».
 - Concepts :
 - Clustering
 - Typage d'écosystèmes
 - Interpolation en zones
 - Sources :
 - Longitude, latitude, taxon, date
 - Sortie :
 - Groupes représentants des écosystèmes

- Etablir les tendances d'évolution d'une espèce dans son écosystème en fonction des autres espèces présentes
 - But :
 - Estimer la probabilité d'occurrence d'une espèce cible selon la composition locale.
 - Concepts :
 - Tendances
 - Facteurs de corrélation
 - Graphes
 - Sources :
 - Groupes d'écosystèmes
 - Sortie :
 - Tendances
 - Facteurs de corrélations sous forme de graphe pondéré

- Evolution des espèces au fil des années
 - But :
 - Quantifier les tendances annuelles des populations.
 - Concepts :
 - Régression
 - Interpolation
 - Sources :
 - Taxon, date
 - Sortie :
 - Courbe d'évolution annuelle des populations

- Evolution des espèces selon les saisons
 - But :
 - Quantifier les tendances saisonnières des populations.
 - Concepts :
 - Régression
 - Interpolation
 - Sources :
 - Taxon, date
 - Sortie :
 - Courbe d'évolution saisonnière des populations

Répartition du travail

Membre	Rôle principal	Responsabilités
Sebastian Morsch	Visualisation de la donnée	Représentation adaptée à l'étude, création d'outils de visualisation (graphes, heatmaps, ...)
Axel Hall	Cartographie des écosystèmes en Suisse	Distinction des espèces par zones, clustering spatial
Jonhatan Muller	Evolution temporelle des espèces (annuel, saisonnier)	Analyse de l'évolution temporelle, analyse de tendances
Fatima Anougmar	Pré-traitement et modélisation	Filtrage/normalisation des données, interprétation des résultats

Gestion du projet

Le projet s'organise selon une planification collaborative visant à assurer la cohérence entre les différentes étapes d'analyse et à garantir une progression régulière du travail.

Planification et étapes clés :

Le déroulement du projet suit la méthodologie suivante :

1. Prétraitement des données : nettoyage, filtrage et normalisation spatiale et temporelle (Fatima).
2. Analyse exploratoire : visualisation des distributions et création de heatmaps pour détecter les motifs principaux (Sebastian).
3. Clustering spatial : identification de regroupements d'espèces et définition d'écosystèmes types (Axel).
4. Analyse des co-occurrences : étude des relations faune–flore dans chaque cluster (Sebastian & Fatima).
5. Modélisation temporelle : identification des tendances annuelles et saisonnières des espèces (Jonathan).
6. Synthèse finale : mise en relation entre types d'écosystèmes et évolution temporelle des espèces, discussion et validation collective.

Chaque membre est responsable de la production et de la validation de ses résultats, tout en assurant une communication régulière avec le reste du groupe.

Suivi et coordination :

- Une réunion hebdomadaire est prévue pour faire le point sur l'avancement, les difficultés rencontrées et la cohérence des résultats.
- Un dépôt GitHub commun est utilisé pour centraliser les scripts, les données et les figures.
- Chaque livrable intermédiaire (visualisation, tableau, présentation) est validé collectivement avant intégration dans la présentation finale.
- Les décisions techniques (format des fichiers, choix des méthodes statistiques, paramètres de clustering) sont documentées dans un journal de projet partagé.

Réflexion sur les résultats attendues

Anticipations :

- Une variation saisonnière importante, en particulier pour les espèces migratrices et florales
- Des associations récurrentes entre certaines espèces animales et végétales
- Attention particulière au biais de causalité / La présence probable de biais spatiaux liés à la concentration d'observateurs dans certaines régions
- Une diversité plus élevée dans les zones naturelles, mise en évidence par le clustering des écosystèmes