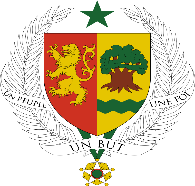
**RÉPUBLIQUE DU SÉNÉGAL**



**Rédigé par**:

FOGWOUNG DJOUFACK Sarah-Laure NIASS Ahmadou NGUEMFOUO NGOUMTSA Célina

SENE Malick

**Statistique exploratoire spatiale Statistiquexploratoire**

**TP 7 : sdmApp (objectifs, fonctions, types de données, résultats, limites…)**

**Sous la supervision de :**

M. Aboubacar HEMA

**Ministère de l’Economie du Plan et de la Coopération**

**Année Académique** :

2023-2024

Ecole nationale de la Statistique et de l’Analyse économique Pierre Ndiaye

**Agence nationale de la Statistique et de la Démographie (ANSDSénégal)**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Un Peuple-Un But-Une Fois**

Table des matières

[I. Objectifs et installation de sdmApp 3](#_Toc184416241)

[1. Fonctionnalités principales du package 3](#_Toc184416242)

[2. Configuration requise 5](#_Toc184416243)

[3. Installation 5](#_Toc184416244)

[II. Objets et fonctions dans sdmApp 6](#_Toc184416245)

[III. Différents onglets de l’applications 8](#_Toc184416246)

[1. Importation des données 8](#_Toc184416247)

[2. Analyse spatiale 9](#_Toc184416248)

[3. Onglet modeling 12](#_Toc184416249)

[4. Onglet R-code 13](#_Toc184416250)

[IV. Limites de l’application sdmApp 13](#_Toc184416251)

[Conclusion 15](#_Toc184416252)

# Objectifs et installation de sdmApp

Le package sdmApp est une application R intégrée à Shiny qui facilite la modélisation de la distribution des espèces, pour des utilisateurs n'ayant pas une expertise avancée en R. Il offre un environnement interactif pour effectuer les étapes nécessaires à la modélisation.

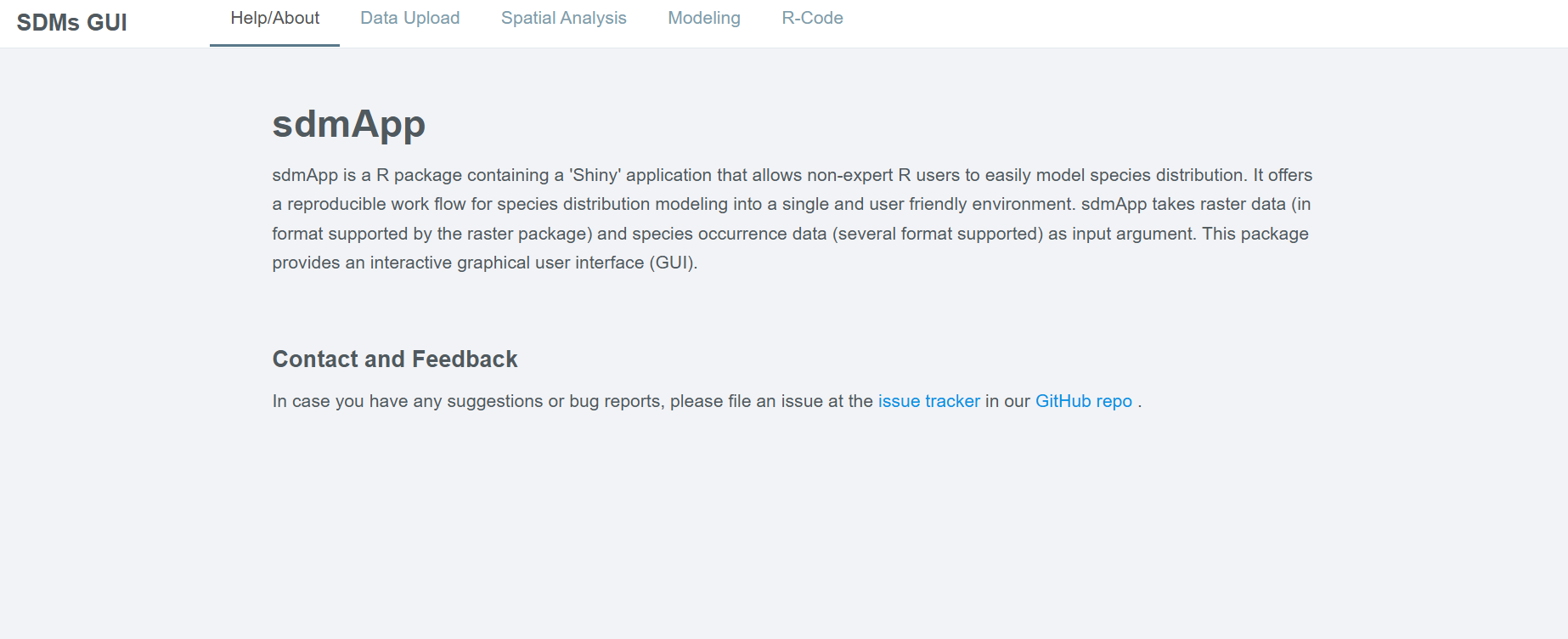
L'un des objectifs principaux de sdmApp est de rendre accessible la modélisation de la distribution des espèces. En regroupant plusieurs outils dans une interface graphique, il permet aux utilisateurs de travailler avec des données environnementales (à partir de fichiers raster) et des données d'occurrence d'espèces (à partir de fichiers CSV ou similaires).

De plus, le package s'inscrit dans une démarche de reproductibilité en permettant aux utilisateurs de télécharger le code R sous-jacent à leurs analyses. Cela garantit que les étapes suivies dans l'interface graphique peuvent être répliquées ou ajustées ultérieurement dans un contexte différent.

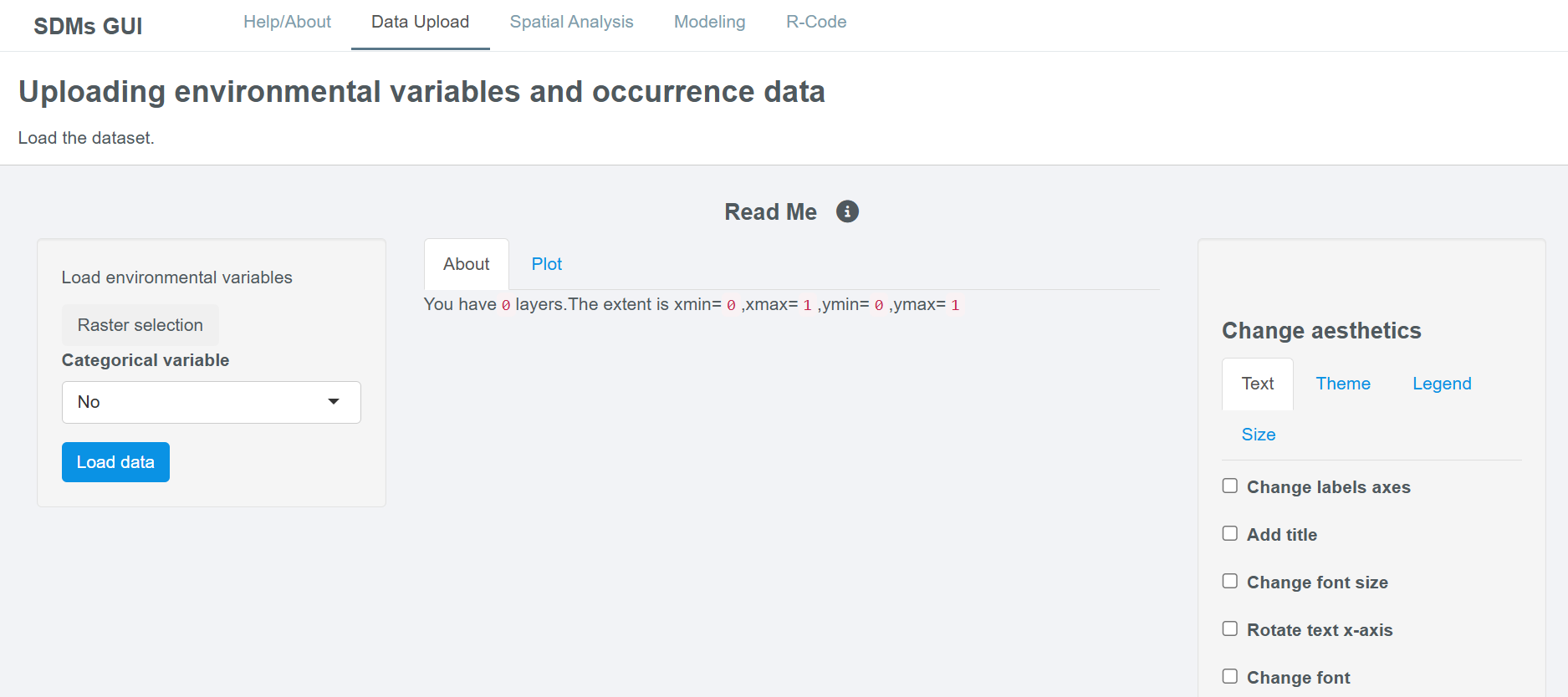
Enfin, sdmApp met à disposition un cadre pour effectuer une modélisation. En plus de permettre l'importation et la visualisation des données, il propose des outils pour la validation croisée spatiale (grâce au package `blockCV`) et la sélection des prédicteurs (avec `CENFA`).

## Fonctionnalités principales du package

Le package sdmApp fournit plusieurs fonctionnalités regroupées autour de cinq fenêtres principales, accessibles depuis une barre de navigation. Voici une description détaillée :



* **Importation des données :**



L'application permet aux utilisateurs de télécharger des fichiers raster contenant des variables environnementales (par exemple, température, précipitations) et des fichiers d'occurrence d'espèces (comme des coordonnées GPS d'observations). Ces données constituent la base de la modélisation de la distribution des espèces.

* **Visualisation des corrélations :**

Une fois les données raster importées, l'application propose une visualisation des corrélations entre les variables environnementales.

* **Sélection des prédicteurs :**

L'application intègre le package `CENFA` pour aider à identifier les variables environnementales les plus influentes sur la distribution des espèces.

* **Validation croisée spatiale :**

L'outil `blockCV` permet de diviser les données spatiales en blocs pour effectuer une validation croisée. Les utilisateurs peuvent choisir d'appliquer ou non cette stratégie selon leurs besoins.

* **Modélisation de la distribution des espèces :**

L'application offre la possibilité de générer des modèles prédictifs avec ou sans validation croisée spatiale. Ces modèles sont basés sur les données d'occurrence et les variables environnementales sélectionnées.

* **Exportation des résultats :**

Les utilisateurs peuvent exporter les résultats obtenus.

* **Reproductibilité :**

Une fonctionnalité clé de l'application est la possibilité de télécharger le code R généré pour chaque étape.

## Configuration requise

Pour utiliser sdmApp, quelques prérequis sont nécessaires :

* **Installation de Java JDK** : Le modèle MaxEnt, utilisé pour la modélisation de la distribution, repose sur une implémentation en Java. Il est donc indispensable que le logiciel Java JDK soit installé.
* **Package rJava :** Ce package R permet d'utiliser Java depuis R. Son installation et sa est également essentielles.
* **Fichier maxent.jar** : Ce fichier doit être placé dans le dossier approprié (dans un emplacement accessible à R) pour que le modèle MaxEnt puisse fonctionner correctement.

## Installation

Le package sdmApp peut être installé à partir de deux sources principales :

* **Depuis GitHub :**

Pour installer la version la plus récente depuis le dépôt GitHub, utilisez les commandes suivantes :

```R

remotes::install\_github("Abson-dev/sdmApp", dependencies = TRUE)

library(sdmApp)

sdmApp()

```

**Depuis CRAN :**

Une version est également disponible sur CRAN et peut être installée comme suit :

```R

install.packages("sdmApp", dependencies = TRUE)

library(sdmApp)

sdmApp()

```

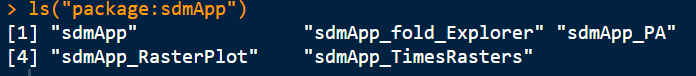
# Objets et fonctions dans sdmApp

Pour explorer les fonctionnalités offertes par le package **sdmApp**, nous avons commencé par examiner les objets et fonctions disponibles à l'aide de la commande R suivante :

```R

ls("package:sdmApp")

```



Ces fonctions représentent les principaux outils du package, permettant de manipuler, visualiser et analyser des données environnementales et des occurrences d'espèces. Dans cette section, nous détaillerons chacune de ces fonctions, leur usage, leurs arguments et les résultats qu'elles produisent. (Ces informations sont obtenus à partir du help de R)

* **sdmApp\_TimesRasters :**

**Description :**

Cette fonction multiplie une carte de probabilité d'occurrence avec une carte de présence/absence pour générer une carte indiquant uniquement les présences.

**Usage:**

sdmApp\_TimesRasters(x, y)

**Valeur retournée :**

Une carte de probabilité d'occurrence modifiée, ne conservant que les zones de présence.

* **sdmApp\_PA**

**Description :**

Cette fonction permet de visualiser une carte de présence/absence en produisant un graphique basé sur **ggplot2**.

**Usage :**

sdmApp\_PA(x)

**Arguments :**

x : Objet Raster représentant une carte de présence/absence.

**Valeur retournée :**

Un objet ggplot affichant la carte de présence/absence.

* **sdmApp\_RasterPlot**

**Description :**

Cette fonction génère un graphique d'une carte raster en utilisant **ggplot2**. Elle est utile pour visualiser les données environnementales ou les résultats d'analyses.

**Usage :**

sdmApp\_RasterPlot(x)

**Arguments :**

x : Objet Raster à visualiser.

**Valeur retournée :**

Un objet ggplot représentant la carte raster.

* **sdmApp**

**Description** : Cette fonction est la principale du package. Elle lance l'application **Shiny** et ouvre l'interface graphique (GUI) pour la modélisation de la distribution des espèces.

* **Usage**

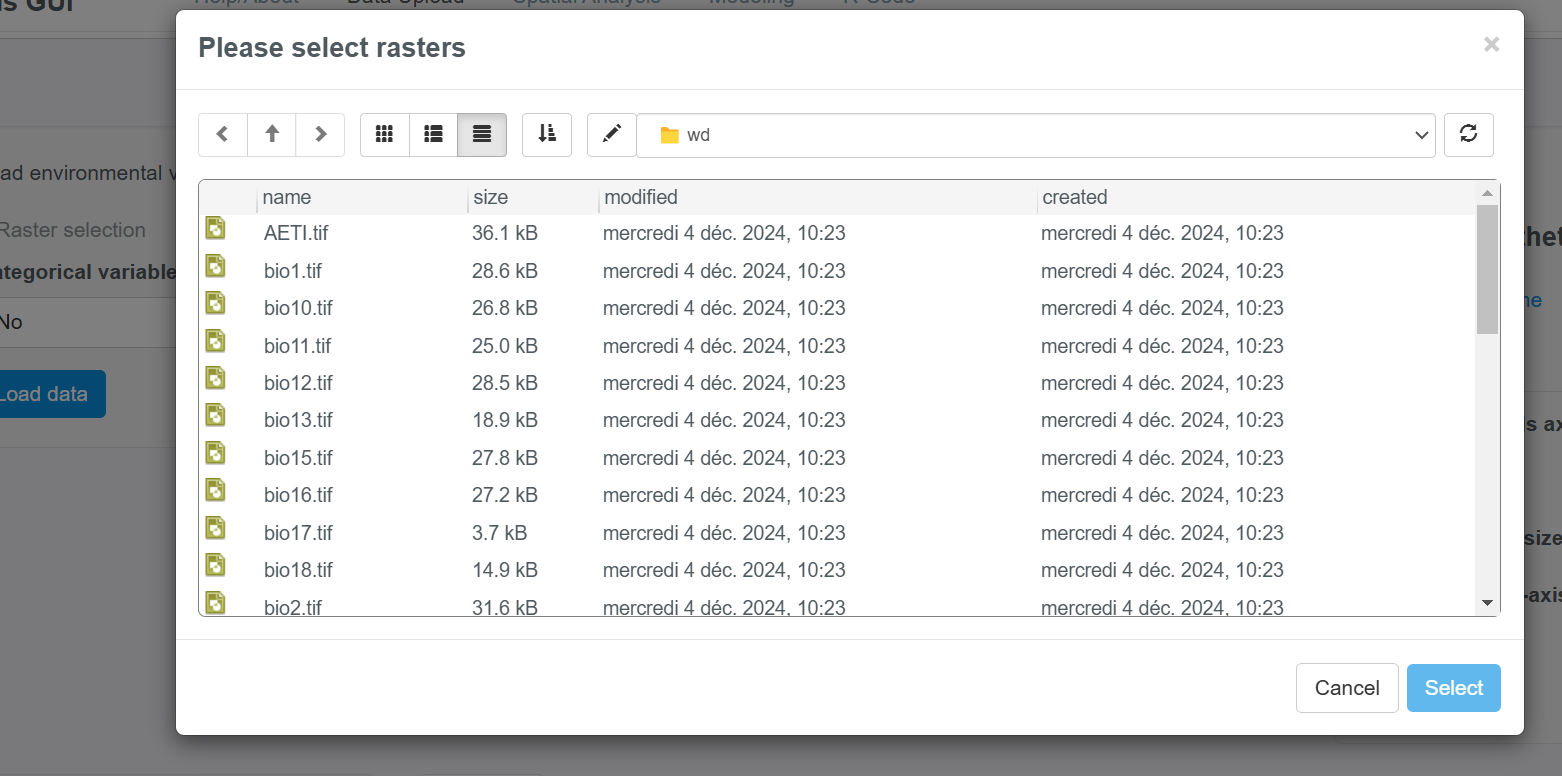
library(sdmApp)

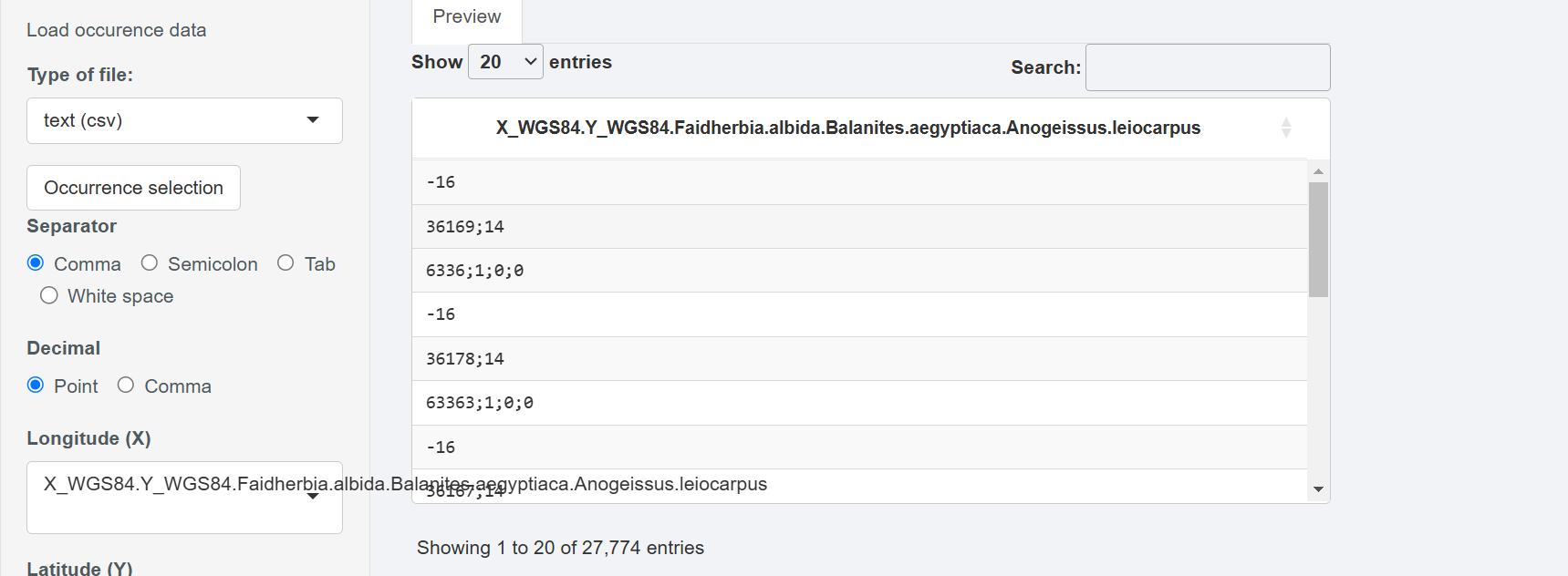
sdmApp()

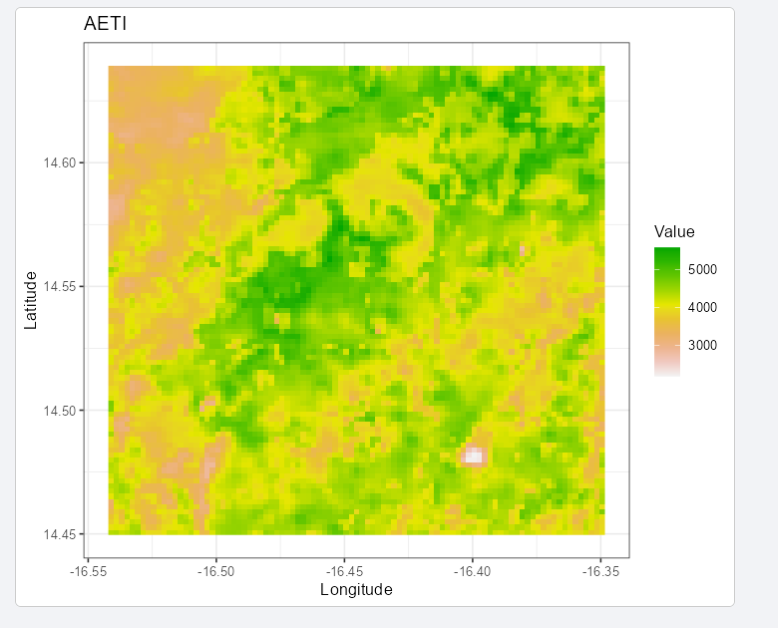
# Différents onglets de l’applications

## Importation des données

Dans cette partie, nous utilisons les données qui sont disponibles dans l’application. Pour cela, nous sélectionnons le raster AETI.tif puis le csv Niakhar.csv. Une fois les donnés sélectionnées, nous pouvons procéder à des visualisations





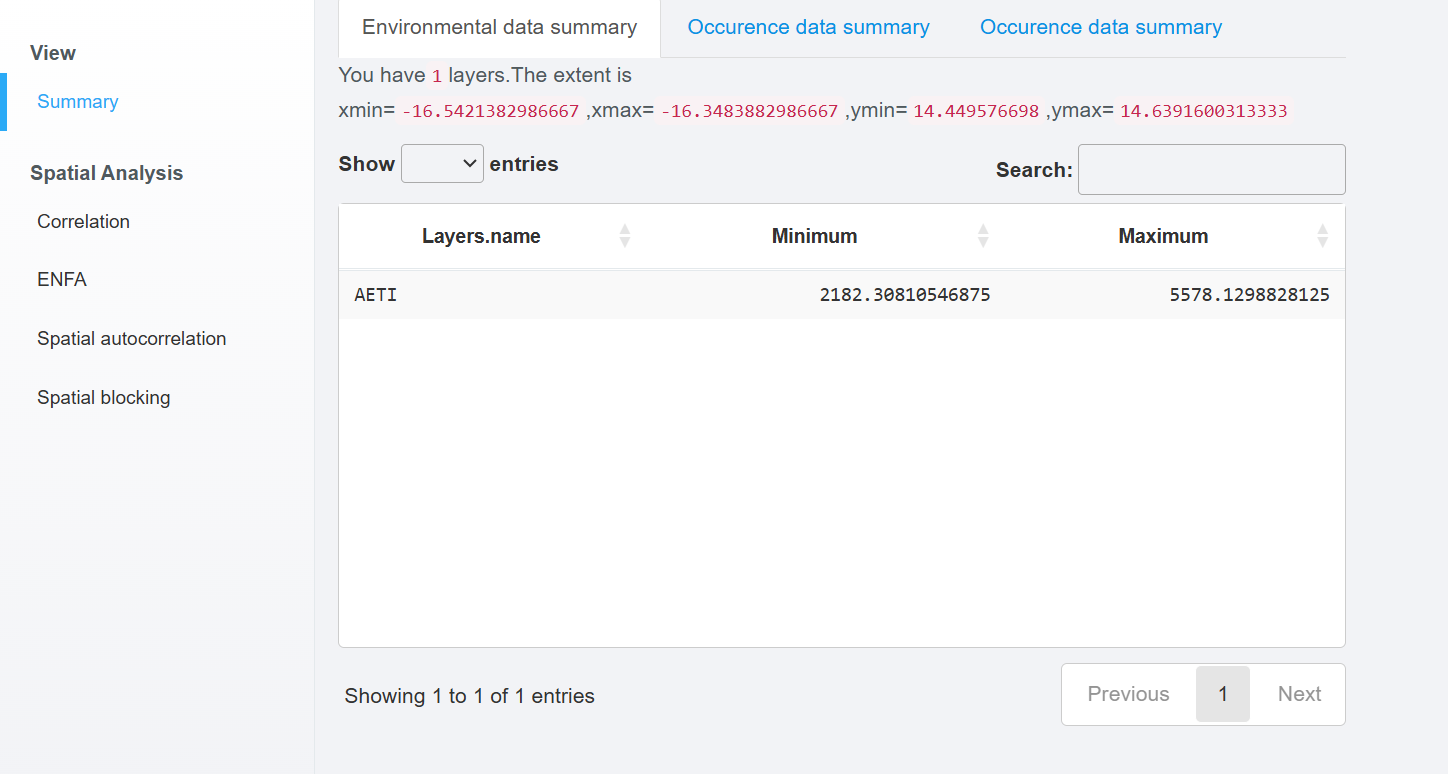


## Analyse spatiale

Cette section du package sdmApp permet d'effectuer une analyse des données environnementales et des occurrences. Les sous-onglets disponibles sont : résumé des données environnementales (summary), Corrélations, ENFA (Ecological Niche Factor Analysis), autocorrélation spatiale, et spatial blocking. Voici un aperçu des fonctionnalités de chacun :

**1. Résumé des Données Environnementales**

Cet onglet affiche un résumé des couches environnementales téléchargées, avec des informations telles que les valeurs minimales, maximales et l'étendue spatiale (xmin, xmax, ymin, ymax).



**2. Corrélations**

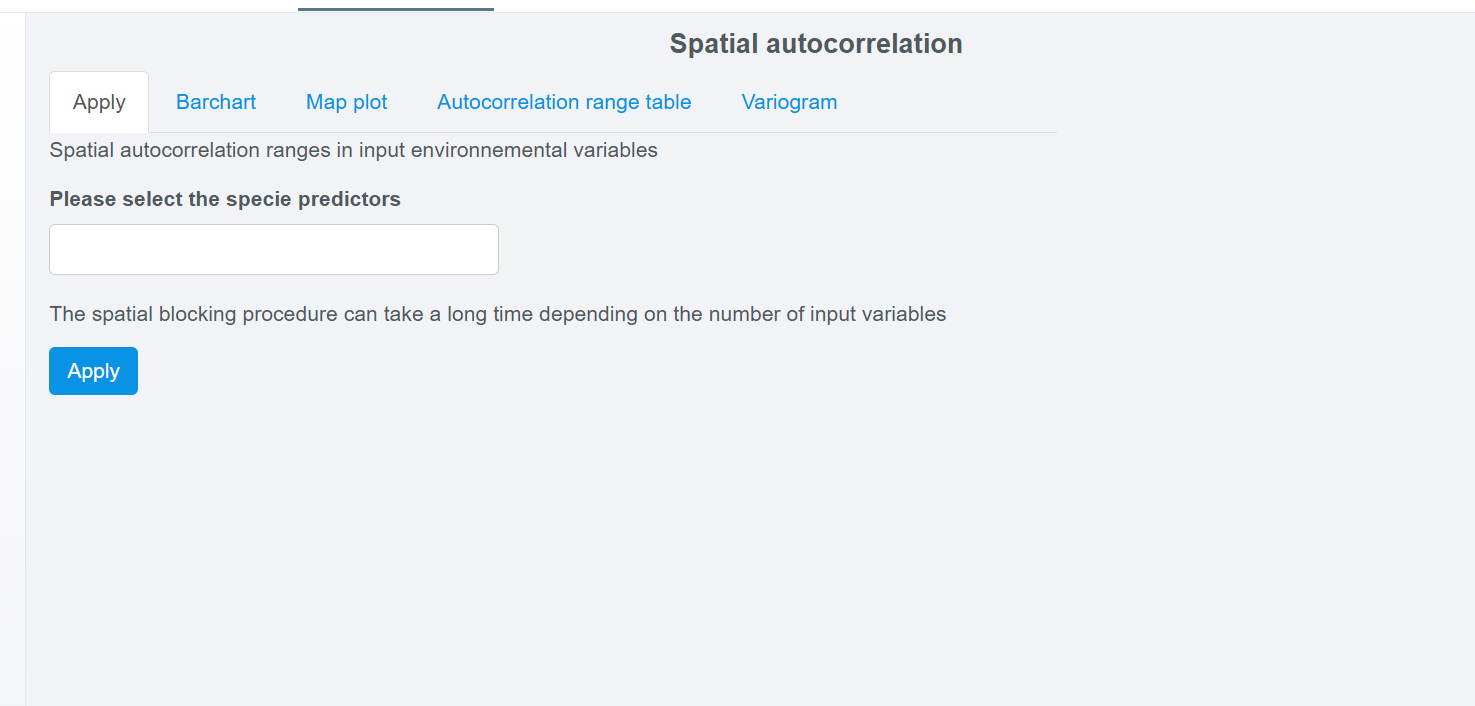
Dans cet onglet, les utilisateurs peuvent analyser les relations entre les variables environnementales via une matrice de corrélation et des visualisations graphiques. Les résultats permettent d'identifier les variables corrélées ou pas.

**3. ENFA (Analyse Factorielle Écologique)**

L'onglet ENFA fournit des outils pour analyser la niche écologique des espèces. Cependant, une erreur peut apparaître si les données ne sont pas correctement préparées ou si les dimensions des matrices sont invalides. Cette fonctionnalité demande des données bien formatées pour produire des résultats exploitables.

**4. Autocorrélation Spatiale**

Cet onglet permet d'analyser la portée de l'autocorrélation spatiale dans les variables environnementales via des diagrammes en barres, des graphiques de variogrammes, et des tables de portée. Ces analyses sont utiles pour comprendre les structures spatiales sous-jacentes des données.

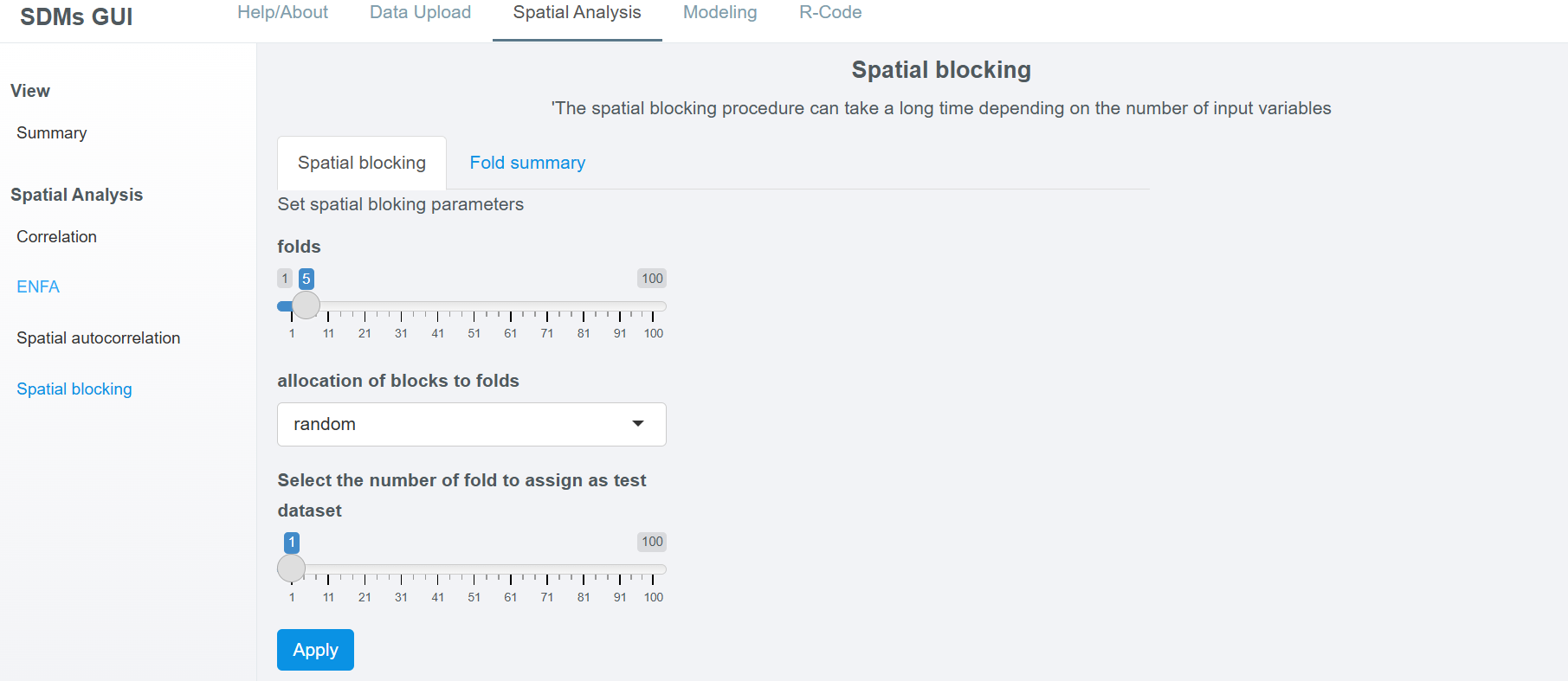


**5. Blocage Spatial**

Le blocage spatial est une méthode essentielle pour préparer des données pour la validation croisée spatiale. Cet onglet permet :

* De définir le nombre de blocs (folds).
* De configurer leur allocation (aléatoire ou selon d'autres stratégies).
* De choisir les blocs destinés au test.

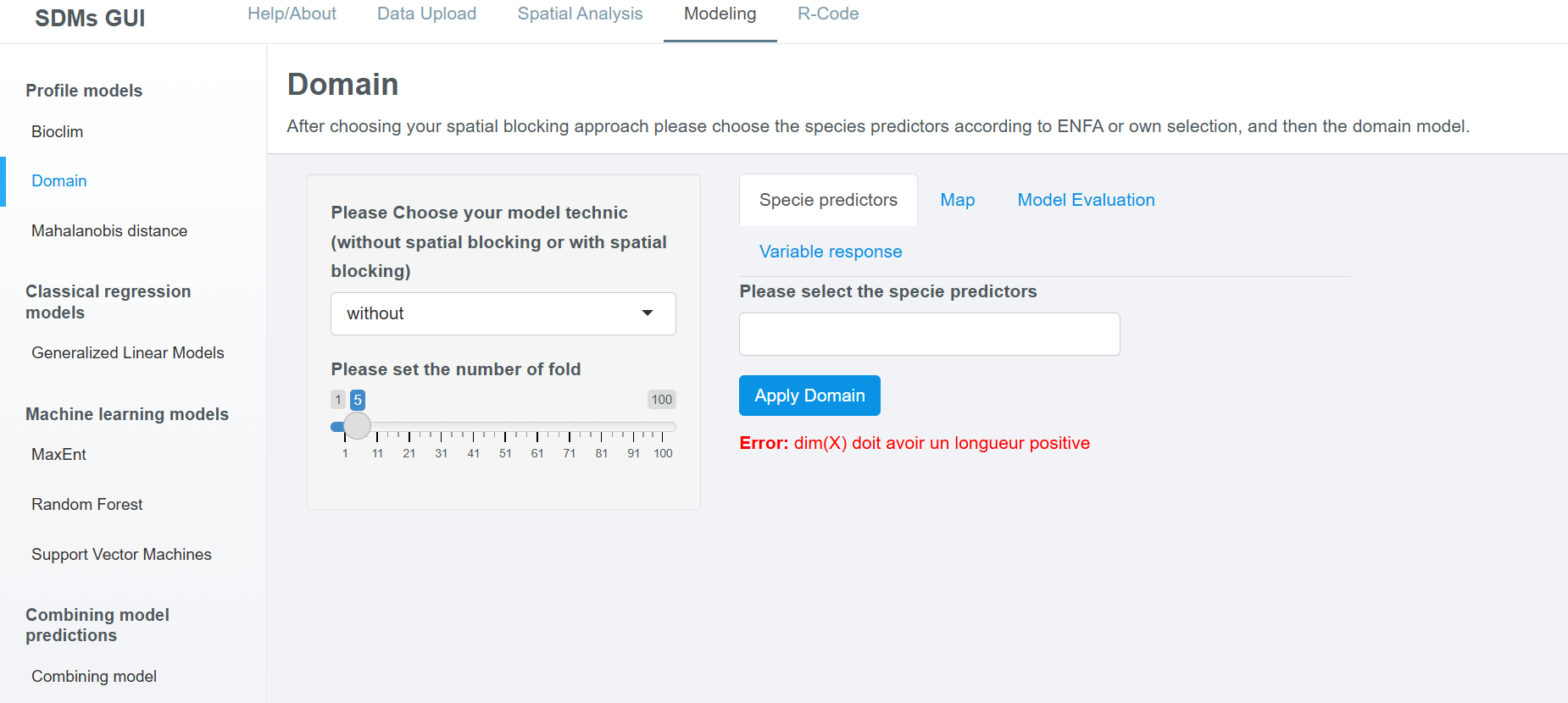
L'objectif principal est d'évaluer les performances des modèles en tenant compte de la structure spatiale, réduisant ainsi les biais liés à l'autocorrélation.



## Onglet modeling

**Description de l'onglet Modeling**

L'onglet "Modeling" permet de sélectionner et d'appliquer différents modèles de prédiction pour l'analyse des distributions d'espèces. Les sous-sections incluent des modèles de profils (comme Bioclim), des modèles de régression classique (GLM), des modèles d'apprentissage automatique (Random Forest, MaxEnt, Support Vector Machines), ainsi que des méthodes de combinaison de modèles.



Les sections principales permettent de :

1. **Choisir une méthode de modélisation :**

* Modèles de profils : Bioclim, Domain, Mahalanobis Distance.
* Modèles de régression classique : GLM.
* Modèles d'apprentissage automatique : Random Forest, SVM, MaxEnt.
* Combinaison de modèles.

1. **Configurer les paramètres spécifiques :**

* Blocage spatial (spatial blocking).
* Nombre de "folds" pour la validation croisée.
* Sélection des prédicteurs environnementaux pour la modélisation.

1. **Évaluer les résultats :**

* Générer des cartes prédictives.
* Analyser la réponse des variables.
* Examiner l'importance des variables.

Chaque méthode propose des ajustements, et des messages d'erreur. Par exemple : dim(X) doit avoir une longueur positive indiquent des paramètres ou données incorrects.

## Onglet R-code

L'onglet "R-Code" est une fonctionnalité pour les utilisateurs qui souhaitent non seulement travailler via l'interface graphique, mais aussi accéder au script R généré automatiquement par l'application. Il affiche le code R correspondant aux étapes effectuées dans l'application, telles que la sélection des modèles, la configuration des paramètres et l'exécution des analyses.

# Limites de l’application sdmApp

Bien que l'application SdmApp propose des outils utiles et accessibles pour la modélisation, des distributions d'espèces, elle présente quelques limites.

1. Limites liées aux fonctionnalités

* **Faible flexibilité dans la gestion des erreurs**

Lorsqu’une erreur se produit, comme l’erreur ’’dim(X) doit avoir une longueur positive’’, ou ’’ pas de méthode pour 'rename' applicable pour un objet de classe "NULL"’’ ou ’’ $ operator is invalid for atomic vectors’’ entre autres, l’application ne fournit pas de contexte suffisant ou de suggestion pour aider l’utilisateur à identifier la source du problème de manière plus aisée.

En outre, des fonctionnalités comme les guides intégrés (comme le help de R par exemple), sont absentes, ce qui peut compliquer la prise en main pour les débutants. A cet effet des informations supplémentaires sur l’application, son fonctionnement, ses fonctions et erreurs potentielles etc peuvent être détaillés davantage dans l’onglet Help/about.

* **Performance et vitesse des calculs**

L'application peut devenir lente lorsque des jeux de données volumineux sont traités (même lors du chargement du jeu de données), en particulier lors de l’utilisation de modèles complexes comme Random Forest ou MaxEnt. Les calculs de modèles ou d'analyses spatiales intensives nécessitent des ressources importantes en termes de mémoire et de processeur. Les utilisateurs travaillant sur des ordinateurs avec des configurations limitées peuvent rencontrer des temps de traitement très longs ou des blocages.

* **Limites liées à l'installation et à la configuration**

Dépendances nombreuses et complexes

L'installation de l'application repose sur de multiples dépendances R, telles que raster, sf, sp, et rgdal, qui doivent toutes être correctement installées et configurées. Si l’un de ces packages n’est pas compatible avec la version de R ou le système d’exploitation, l’application ne fonctionnera pas correctement.

Difficultés liées à Java pour MaxEnt

Le modèle MaxEnt nécessite une installation Java fonctionnelle, ce qui peut être complexe pour certains utilisateurs, surtout si plusieurs versions de Java (ou la mauvaise version) sont installées sur leur système alors qu’il faut une version Java adéquate.

L’intallation semble également necessité un package « CENFA » qui n’est pas disponible directement sur R.

Pas de processus d'installation automatisé

Les utilisateurs doivent manuellement installer l’application et ses dépendances, ce qui augmente les risques d’erreurs et les difficultés. Une solution comme une installation automatique pourrait simplifier ce processus.

**Langue des messages d'erreur**

Les messages d'erreur sont souvent en français, ce qui peut être une barrière pour les utilisateurs non francophones. Par exemple, l’erreur dim(X) doit avoir une longueur positive est difficile à comprendre pour quelqu’un qui ne parle pas français. Il n'y a pas d'option pour basculer vers une langue universelle comme l'anglais.

**Données multi temporelles non prises en charge**

L’application ne semble pas intégrer des analyses dynamiques ou multi temporelles, ce qui limite son utilisation pour des études environnementales évolutives. Par exemple, il serait difficile d'analyser comment les distributions d’espèces changent au fil du temps.

En sus de ces limites, nous avons également pu noter que l’application rencontre des problèmes lors de la sélection des fichiers lorsque l’option ‘Type of file’ est sur l’option ‘text(csv)’. Sur cette option, en appuyant sur l’onglet ‘Load data’, l’application semble arrêter de fonctionner ou rencontrer un bug.

# Conclusion

L'application **SDMs GUI** est une initiative remarquable pour la modélisation des distributions d'espèces, offrant une interface graphique conviviale qui rend accessible un ensemble d'outils couramment utilisés en écologie spatiale. Elle facilite l’accès à des techniques de pointe en proposant des modèles statistiques, des algorithmes de machine learning, et des analyses spatiales intégrées dans une interface simplifiée. Cela en fait un outil potentiellement très utile pour les chercheurs, gestionnaires de la biodiversité et étudiants en écologie. Cependant, comme toute application en développement, elle présente quelques limites qui méritent d’être examinées pour améliorer son potentiel.

Ces limites incluent des performances qui ralentissent pour des jeux de données volumineux, des messages d’erreur peu explicites, et une documentation insuffisante. De plus, les défis liés à l’installation de l’application, en particulier les conflits avec des dépendances comme Java ou des packages R, ajoutent des barrières pour les utilisateurs novices ou non techniques.