



UFR 6

Université Paul Valéry, Montpellier III

Projet TER

Rendu 6 - Visualisations, Graphiques

Alvin VEDEL, Florian DUBOIS, Matis BREILLAD

10 Mars 2024

Résumé

Ce rapport a pour but de présenter les visualisations effectués dans notre projet et de présenter l'outil final permettant une représentation complète des données. Cela comprend 3 sources de données différentes : les données abiotiques tabulaires, les séries temporelles bioclimatiques et les séries temporelles satellites.

Table des matières

Résumé	i
Liste des figures	ii
Introduction	1
1 L’analyse exploratoire des données	3
1.1 Un premier aperçu cartographique	4
1.1.1 Répartition des 50 espèces les plus présentes	4
1.1.2 Comprendre le Spatial Block Hold-Out	4
1.1.3 Visualiser les PO	4
1.2 Graphiques descriptifs	4
1.2.1 Les variables abiotiques	4
2 Développement d’un outil de visualisations	5
2.1 Fonctionnalités	6
2.1.1 Affichage des données	6
2.1.2 Fonctionnalités complémentaires	6
2.2 Choix des idiomes	6
2.2.1 Carte	6
2.2.2 Les variables qualitatives	6
Conclusion	7

Table des figures

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent a sapien tortor. Morbi vel metus a arcu sodales egestas et ac nisi. Maecenas suscipit rutrum urna vel facilisis. Nam finibus, quam vel interdum suscipit, massa eros vehicula mauris, in accumsan orci dui et sapien. Quisque porttitor commodo ultricies. Morbi efficitur magna nisi, eu porta nulla dapibus at. Aenean ac arcu ut purus laoreet ultricies eget eget diam.

Cras efficitur lorem eu fermentum consequat. Sed convallis sed nibh eget ullamcorper. Nulla at tellus porta, aliquet magna non, consectetur orci. Nunc efficitur nulla vitae pulvinar pharetra. Ut porttitor imperdiet lorem, eget pharetra urna maximus sed. Morbi eu magna eget sapien pharetra commodo. Nullam auctor dui ut maximus viverra. Integer venenatis libero vitae enim dignissim dapibus. Nam suscipit mollis purus. Pellentesque rutrum, velit quis imperdiet dictum, elit nulla efficitur libero, a placerat augue tortor ut tortor. Vestibulum et condimentum ligula. Mauris elementum tellus ut efficitur tempor. Praesent id ipsum eleifend, suscipit lectus pulvinar, dapibus felis. In iaculis lectus lorem, in ultricies quam ultricies facilisis. Vestibulum in ante vel mauris tincidunt ornare.

Sed sed molestie ligula. Vivamus sed risus accumsan, convallis neque id, bibendum mi. Donec tristique risus a lectus auctor, sed pulvinar turpis efficitur. Vestibulum purus erat, mattis a lacus non, pellentesque vulputate lectus. Phasellus ut sapien eget lacus laoreet ultricies. Ut vel sodales sapien. Fusce et tortor elit. Proin elementum ante fringilla imperdiet vestibulum. Mauris eros massa, auctor in blandit at, fermentum a massa. Duis vitae elementum nisi, non vehicula eros. Fusce enim purus, interdum et dui sed, ultricies fermentum ex. In in metus ultrices felis dapibus sodales. Morbi eu malesuada sem. Proin aliquet mi id nisi hendrerit sollicitudin.

Sed eget elit id nibh posuere dictum. Morbi rutrum urna id quam iaculis faucibus. Quisque nulla ante, aliquet suscipit mauris vitae, pretium cursus erat. Donec elementum imperdiet tortor, blandit iaculis velit ullamcorper eu. Nullam eget tortor nulla. Nullam sed lacus commodo, fringilla purus eget, ultrices turpis. Phasellus finibus consectetur leo eget accumsan. Praesent vehicula lorem justo, eget dictum justo porttitor sed. Suspendisse molestie nunc a augue sodales pharetra. In hac habitasse platea dictumst. Integer euismod turpis quis sollicitudin dictum.

Vestibulum ornare mi et luctus mattis. Nunc faucibus lorem eros, et luctus magna elementum eu. Sed in odio nec urna blandit dignissim at sed erat. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Vivamus mattis quam quis

lorem porta euismod. Quisque dapibus enim sed eleifend finibus. Maecenas egestas, mi et convallis mollis, elit leo congue libero, vel porta urna felis eget nulla. Proin felis sem, porttitor et velit et, lacinia placerat lorem. Donec eget nunc faucibus, lacinia metus ac, accumsan lectus. Curabitur hendrerit tincidunt purus, non posuere nunc ultrices ac.

Chapitre 1

L'analyse exploratoire des données

Sommaire

1.1	Un premier aperçu cartographique	4
1.1.1	Répartition des 50 espèces les plus présentes	4
1.1.2	Comprendre le Spatial Block Hold-Out	4
1.1.3	Visualiser les PO	4
1.2	Graphiques descriptifs	4
1.2.1	Les variables abiotiques	4

1.1 Un premier aperçu cartographique

1.1.1 Répartition des 50 espèces les plus présentes

Nos données de Présence/Absence possèdent une composante spatiale qui est très importante à considérer dans la construction de modèles, il a été nécessaire de visualiser la répartition des espèces en France et au Royaume-Uni. Pour cela nous avons trié les espèces selon leur nombre d'apparitions et nous avons affiché les 50 espèces les plus présentes sur des cartes séparés. La visualisation a été réalisé en Python à l'aide d'un fichier geojson en tant que fond de carte, la présence d'une espèce était symbolisée par un point rouge.

1.1.2 Comprendre le Spatial Block Hold-Out

Cela nous a permis de voir où se situent les espèces les plus fréquentes ainsi que de se faire une idée sur la localisation des PatchID, nos observations n'étant pas uniformes sur le territoire considéré mais divisé en blocs (Spatial Block Hold-Out). Pour mieux se représenter cette segmentation de l'espace entre les données d'entraînement et de test/validation, nous avons refait une carte sur laquelle un point représente un patchID d'entraînement en rouge et un patchID de test/validation en bleu, cette division spatiale devient alors évidente.

1.1.3 Visualiser les PO

Une carte a également été réalisé recensant l'ensemble des sites d'observation de Présences seules mais quelques problèmes viennent s'ajouter. Dans cette visualisation, un point représentait un patchID or c'est près de 6 millions de points distincts qui sont présents dans le jeu de données. Bien que ces observations se situent dans toutes l'Europe, il devient difficile de distinguer quelque chose et notre carte ne devient qu'un ensemble de points impossibles à analyser.

1.2 Graphiques descriptifs

1.2.1 Les variables abiotiques

Nos visualisations descriptives ont concerné les données tabulaires des Présences Absences. Nous avons procédé à une analyse simple permettant de comprendre leurs distributions.

Chapitre 2

Développement d'un outil de visualisations

Sommaire

2.1	Fonctionnalités	6
2.1.1	Affichage des données	6
2.1.2	Fonctionnalités complémentaires	6
2.2	Choix des idiomes	6
2.2.1	Carte	6
2.2.2	Les variables qualitatives	6

2.1 Fonctionnalités

2.1.1 Affichage des données

Nos données possédant une forte composante géographique il est important de pouvoir les visualiser spatialement, on possède les coordonnées de chaque *"surveyId"* qui sont caractérisés par une combinaison de variables. La fonctionnalité première sera de fournir un outil interactif permettant de visualiser géographiquement les relevés et comprendre leur distribution.

2.1.2 Fonctionnalités complémentaires

Dans un second temps, il sera pertinent de visualiser les variables caractérisant un relevé.

1. Variables abiotiques : un ensemble de 50 variables tabulaires caractérisent un *"surveyId"*, nous allons opter pour une représentation en spirale dans laquelle chaque pixel correspond à un relevé pour un total de 50 spirales. Il sera possible d'identifier les points qui prennent des valeurs similaires sur des variables grâce à la couleur. En sélectionnant une spirale on trie les relevés par cette variable, l'ordre est conservé entre les spirales.
2. Séries temporelles : Des séries temporelles concernant les données bioclimatiques mensuelles et des séries temporelles trimestrielles satellites sont également à notre disposition. Un affichage plutôt simple sous la forme d'un diagramme en ligne pourra être réalisé.
3. De ces données temporelles nous avons créé des images, desquelles nous avons extrait des features grâce à un réseau de neurones. Après avoir performé une ACP sur ces nouvelles variables, il est intéressant de les représenter à l'aide d'une t-SNE sous la forme d'un nuage de points. Les points présentant une espèce seront d'une couleur différentes des autres.

2.2 Choix des idiomes

2.2.1 Carte

La carte est réalisée en leaflet et utilise un fond de carte natif à la librairie. Un *survey* est représenté par un point de couleur bleu.

2.2.2 Les variables qualitatives

Dans les spirales, on utilise une palette séquentielle pour que les couleurs diffèrent suffisamment entre les valeurs sans pour autant tenter de les échelloner. Il s'agit d'une vue

complémentaire à la première qui peut fonctionner en interaction : en cliquant sur un point de la carte, le pixel correspondant dans chaque spirale est mis en évidence. Les séries temporelles seront au nombre de 4 pour les données bioclimatiques mensuelles, chacune identifiée par une couleur. La visualisation sera initialisée avec les valeurs moyennes pour chaque relevé mais il sera possible d'obtenir le détail d'une observation et éventuellement la comparer avec la moyenne. Idem pour les séries landsat qui ont 6 variables différentes, donc 6 couleurs. Concernant la visualisation t-SNE des features, il s'agira d'un nuage de points également bleus avec la possibilité de sélectionner une espèce. Les "surveyId" présentant l'espèce seront alors affichés en rouge pour les visualiser. Deux nuages seront présents, un pour les données bioclimatiques et un pour les données de landsat.

Conclusion

Dans ce projet de TER nous ne nous sommes pas orienté vers des visualisations complexes en privilégiant l'axe prédictif. Cependant il nous semblait important de fournir un outil de visualisation permettant la représentation et l'interprétation des données.

L'outil proposé permet de tenir compte de la problématique spatiale de notre projet et de fournir des descriptifs comparatifs entre les observations pour toutes les variables .