

ZODIAC PRO :

Une Application pour
l'Analyse Agro-climatique au
Service de l'Agriculture
Iséroise



LEUCHI ILIAS

Tuteur : GARAT Philippe



SOMMAIRE

I. INTRODUCTION	
1. Contexte	4
2. Objectifs	4
II. PRESENTATION DES DONNEES	
1. Données simulées, les modèles DRIAS	5
2. Données réelles, Infoclimat	6
3. Transformation des données.....	8
III. IAC & ANALYSES STATISTIQUES	
1. Catégorie et calcul d'IAC.....	10
2. Comparaison statistique des données	13
3. Zonage géographique.....	16
IV. L'APPLICATION ZODIAC PRO	
1. Découpage de l'application	20
2. Présentation de ses fonctionnalités	21
V. CONCLUSION	
1. Récapitulation des réalisations.....	30
2. Perspectives d'amélioration	30
3. Bilan du stage	31
VI. ANNEXES	
1. Fonction de scraping	32
2. Extrait des données	32
3. Fonction de traitement des données Infoclimat.....	33
4. Fonction de calcul des IAC.....	34
5. Fonction des tests statistiques	35
6. Fonction d'interpolation	36

REMERCIEMENTS

Avant de commencer, je tiens à remercier l'ensemble des personnes qui ont été présentes tout au long de mes trois mois de stage. Tout d'abord, je souhaite exprimer ma gratitude envers mon superviseur, GARAT Philippe, pour son soutien, ses conseils et son encadrement tout au long de cette période. Je tiens également à remercier chaleureusement l'ensemble du personnel du LJK pour leur accueil au sein du laboratoire, en particulier les autres stagiaires avec qui nous avons partagé le même bureau, et qui ont grandement facilité mon intégration dès mon premier jour de stage. Pour finir, je tiens à exprimer ma profonde gratitude envers la chambre d'agriculture de l'Isère pour leur confiance. Ils ont su exprimer leurs besoins et n'ont pas hésité à proposer des pistes d'améliorations afin d'obtenir une application à leur image. Je suis sincèrement reconnaissant envers toutes les personnes qui ont contribuées à faire de ce stage une expérience enrichissante et mémorable, et je conserverai précieusement les souvenirs de cette période de formation.

GLOSSAIRE

Termes	Définitions
Agrégation	Désigne le processus de regroupement de multiples valeurs de données individuelles en une seule entité. Elle peut impliquer diverses opérations telles que la sommation, la moyenne, le comptage, le maximum, le minimum, appliquées aux données regroupées selon des critères spécifiques tels que le temps, la catégorie, la géographie, etc.
Boxplot	Graphique qui résume la distribution d'un ensemble de données en illustrant la médiane, les quartiles et les valeurs aberrantes éventuelles d'une manière visuellement concise.
Clé API	Code unique et sécurisé utilisé pour permettre à des applications, des services ou des développeurs d'accéder à certaines fonctionnalités ou données d'un service en ligne de manière contrôlée et autorisée.
DRIAS	Plateforme qui a pour vocation de mettre à disposition des projections climatiques régionalisées réalisées par Météo France et dans les laboratoires français de modélisation du climat.
ETP	L'évapotranspiration potentielle : l'évapotranspiration est le processus par lequel l'eau liquide terrestre est renvoyée dans l'atmosphère environnante sous forme gazeuse, du fait de la respiratoire des plantes et la transpiration du sol. On parle d'évapotranspiration potentielle pour désigner le calcul théorique de l'évapotranspiration réelle.
IAC	Indicateur Agro-Climatique : mesure utilisée pour évaluer les conditions climatiques spécifiques à l'agriculture.
RCP	Scénarios d'émissions de gaz à effet de serre et de concentrations atmosphériques qui sont utilisés dans les modèles climatiques pour évaluer les impacts du changement climatique futur. Il existe quatre RCP principaux : RCP2.6, RCP4.5, RCP6.0 et RCP8.5, allant du plus optimiste au plus pessimiste.
RData	Format de fichier utilisé par le logiciel R pour compresser et sauvegarder des objets R tels que des données, des variables ou des modèles de manière efficace et comprimée. Cela permet de réutiliser facilement ces objets dans des sessions R ultérieures sans avoir à les recréer à partir de zéro.

I. INTRODUCTION

1. Contexte

Les prévisions climatiques actuelles, établies par divers modèles climatiques, convergent vers un scénario inéluctable : une augmentation moyenne des températures allant de 2 à 5 °C d'ici la fin du siècle, accompagnée d'une redistribution des précipitations et d'une augmentation des événements climatiques extrêmes. Ces changements auront certainement un impact significatif sur les productions agricoles, soulignant ainsi l'importance d'anticiper ces effets à l'échelle locale. Face à cette situation, les experts ont développé un outil de sensibilisation et d'aide à la décision pour les agriculteurs, nommé CANARI. Cette application offre une visualisation rapide et accessible en ligne de plus d'une centaine d'indicateurs agro-climatiques (IAC), couvrant ainsi les besoins spécifiques des différentes filières agricoles (blé, maïs, vergers, prairies, etc.). Chaque IAC peut être calculé spécifiquement pour chaque localité en France métropolitaine, pour les périodes du passé récent (1985-2020), futur proche (2021-2050) et du futur lointain (2051-2100), et selon un RCP choisis entre le 4.5 et 8.5. Ce calcul repose sur une approche multi-modèles, grâce aux différents modèles disponibles sur DRIAS, ce qui permet de mieux appréhender la diversité potentielle des évolutions climatiques à venir. Toutefois, CANARI présente quelques inconvénients, principalement liés à la présence de plusieurs IAC similaires dans la conception, dans différentes parties de l'application. Par exemple, nous pouvons être amenés à calculer un nombre de jours supérieur à un seuil dans des parties différentes, où seule la valeur du seuil diffère. Cette redondance peut générer une certaine confusion et complexité lors de l'utilisation de l'outil.

2. Objectifs

Dans ce contexte, la chambre d'agriculture de l'Isère a décidé de développer sa propre application de calcul et visualisation pour les différents types d'IAC. Cette application sera conçue en premier lieu pour répondre aux besoins de l'Isère et doit offrir une plus grande flexibilité dans le paramétrage des IAC. De plus, la chambre d'agriculture souhaite intégrer un zonage du département, basé sur un ou plusieurs indicateurs pour une meilleure compréhension et adaptation aux particularités locales. L'objectif du stage est de concevoir un prototype de cette application nommé ZODIAC Pro pour : ZOnage Des Indicateurs Agro-Climatiques Prospectif. La programmation R et Rshiny a été retenue. Pour réussir ce projet, plusieurs étapes clé doivent être suivies. Dans un premier temps, il est essentiel de se familiariser avec l'application CANARI et de comprendre en détail le processus de calcul de tous les IAC qui y sont implémentés. Ensuite, il faudra collecter l'ensemble des données nécessaires pour permettre une analyse approfondie et une visualisation pertinente de ces données. Une fois ces étapes préliminaires achevées, le développement de l'application ZODIAC Pro pourra commencer, en se concentrant sur la création des différentes fonctions nécessaires au calcul des IAC. Ici, nous expliquerons d'abord le processus de traitement des données, en commençant par la récupération des données DRIAS, puis nous aborderons la collecte de données réelles ainsi que le traitement effectué sur l'ensemble de ces données. Par la suite, nous évoquerons les analyses statistiques réalisées en utilisant les différents IAC. Nous présenterons les différents types d'indicateurs utilisés, les comparaisons statistiques effectuées, et les deux méthodes de zonage mises en place. Pour finir, nous nous pencherons en détail sur l'application ZODIAC Pro, en expliquant le découpage réalisé et en présentant les différentes fonctionnalités mises en œuvre. Commençons notre présentation par l'exploration des données utilisées.

II. PRÉSENTATION DES DONNEES

L'ensemble des données nécessaires comprennent différentes mesures climatiques telles que : la température moyenne, la température minimale, la température maximale, les précipitations totales, la chute de neige à grande échelle, l'humidité, la vitesse du vent, le rayonnement visible, le rayonnement infrarouge et l'ETP. Toutes ces mesures sont celles proposées par les modèles DRIAS et que nous essayerons de retrouver pour les données réelles. Chacune de ces mesures correspond à une variable dans notre ensemble de données. De plus, nous ajoutons la variable « bilan hydrique », qui représente la différence entre les précipitations totales et l'ETP. Regardons maintenant comment la récupération des données DRIAS s'est réalisée.

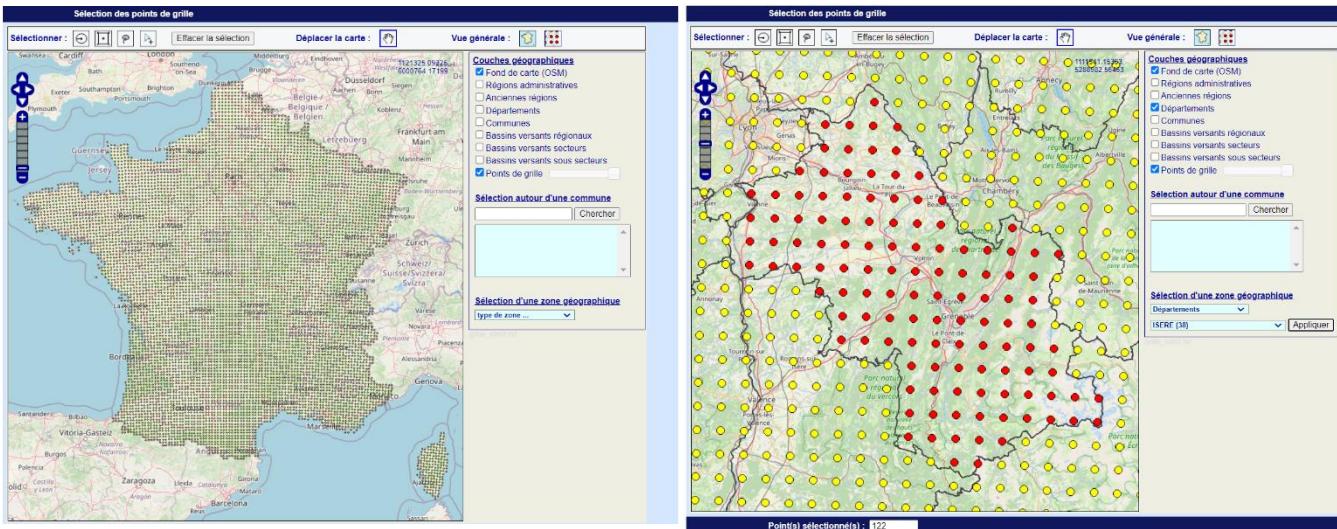
1. Données simulées, les modèles DRIAS

DRIAS propose aux utilisateurs plusieurs modélisations du climat réalisées dans les différents laboratoires français. Parmi tous ces modèles, la chambre d'agriculture décide d'en conserver un unique intitulé modèle ALADIN63_CNRM-CM5 (en abrégé : ALADIN).

DRIAS les futurs du climat

Catalogue des produits	Expérience	Identifiant	Scénario d'émission	Période	Modèles GCM / RCM - correction ADAMONT (France)	Institution RCM
Mon panier	DRIAS-2020	ALADIN63_CNRM-CM5	Référence	1951-2005	CNRM-CERFACS-CNRM-CM5 / CNRM-ALADIN63	Météo-France / Centre National de Recherches Météorologiques
Mes requêtes	DRIAS-2020	ALADIN63_CNRM-CM5	RCP2.6	2006-2100	CNRM-CERFACS-CNRM-CM5 / CNRM-ALADIN63	Météo-France / Centre National de Recherches Météorologiques
Utilisateur connecté	DRIAS-2020	ALADIN63_CNRM-CM5	RCP4.5	2006-2100	CNRM-CERFACS-CNRM-CM5 / CNRM-ALADIN63	Météo-France / Centre National de Recherches Météorologiques
Compte: Ilias	DRIAS-2020	ALADIN63_CNRM-CM5	RCP8.5	2006-2100	CNRM-CERFACS-CNRM-CM5 / CNRM-ALADIN63	Météo-France / Centre National de Recherches Météorologiques
Identifiant: iliasleuchi@gmail.com	DRIAS-2020	RACMO22E_CNRM-CM5	Référence	1950-2005	CNRM-CERFACS-CNRM-CM5 / KNMI-RACMO22E	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
Nom: LEUCHI	DRIAS-2020	RACMO22E_CNRM-CM5	RCP2.6	2006-2100	CNRM-CERFACS-CNRM-CM5 / KNMI-RACMO22E	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
Prénom: Ilias	DRIAS-2020	RACMO22E_CNRM-CM5	RCP4.5	2006-2100	CNRM-CERFACS-CNRM-CM5 / KNMI-RACMO22E	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
Mail: iliasleuchi@gmail.com	DRIAS-2020	RACMO22E_CNRM-CM5	RCP8.5	2006-2100	CNRM-CERFACS-CNRM-CM5 / KNMI-RACMO22E	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
Se déconnecter	DRIAS-2020	WRF381P_IPSL-CM5A	Référence	1951-2005	IPSL-IPSL-CM5A-MR / IPSL-WRF381P	Institut Pierre-Simon Laplace
	DRIAS-2020	WRF381P_IPSL-CM5A	RCP4.5	2006-2100	IPSL-IPSL-CM5A-MR / IPSL-WRF381P	Institut Pierre-Simon Laplace

Ci-dessus, nous pouvons voir l'organisation des différents modèles, nous remarquons qu'ils ne sont pas tous dans la même période et qu'on retrouve plusieurs fois le même modèle, mais pour des RCP différents. Le RCP 2.6 étant jugé comme trop peu réaliste, nous avons décidé de conserver les RCP 4.5 et 8.5, bien que le RCP 4.5 commence lui aussi à perdre en pertinence. Ces simulations sont réalisées entre 2006 et 2100. Nous récupérons également les données du modèle ALADIN appliqué entre 1951 et 2005, qui reflètent un scénario similaire à la réalité observée durant cette période. Ces données seront utilisées pour une comparaison avec des données réelles non simulées.



Après avoir choisi le modèle, la prochaine étape consiste à déterminer la zone géographique à étudier. Dans notre cas, nous sélectionnons les données de l'Isère qui représente un total de 122 points spatiaux. Ils suivent le maillage SAFRAN, qui correspond à des mailles ayant une dimension de 8 km de côté.

La dernière étape implique de choisir les variables pertinentes à conserver pour notre étude et de les exprimer dans les unités appropriées. Une fois cela réalisé, nous procédons à la sélection du format des positions géographiques. Nous décidons donc de choisir l'intégralité des variables et d'extraire la latitude, la longitude et l'altitude de chaque point de données.

Les données sont enregistrées à l'échelle journalière, ce qui signifie qu'il y a une ligne pour chaque point spatiaux et pour chaque jour. Cependant, en raison des limitations de téléchargement imposées par DRIAS, il est nécessaire de procéder à plusieurs extractions pour obtenir l'ensemble des données souhaité. Nous nous retrouvons donc avec plusieurs fichiers texte (.txt), pour chaque RCP, à traiter sous R.

⚠️ Les Paramètres sont fournis par défaut dans l'unité dans laquelle ils sont archivés

<input checked="" type="checkbox"/> Températures ... [* Sélectionnez l'unité]	[tasminAdjust] [tasmaxAdjust] [tasAdjust]
<input checked="" type="checkbox"/> Température minimale journalière à 2 m	K C F ⓘ
<input checked="" type="checkbox"/> Température maximale journalière à 2 m	K C F ⓘ
<input checked="" type="checkbox"/> Température moyenne journalière à 2 m	K C F ⓘ
<input checked="" type="checkbox"/> Précipitations ... [* Sélectionnez l'unité]	[tototAdjust]
<input checked="" type="checkbox"/> Précipitations totales	kg/m ² /s mm ⓘ
<input checked="" type="checkbox"/> Chute de neige à grande échelle	kg/m ² /s mm ⓘ
<input checked="" type="checkbox"/> Humidité ... [* Sélectionnez l'unité]	[rhussAdjust]
<input checked="" type="checkbox"/> Humidité spécifique à 2 m	kg/kg g/kg ⓘ
<input checked="" type="checkbox"/> Rayonnement ... [** uniquement avec les modèles 'ALADIN63_CNRM-CM5, RACMO22E_CNRM-CM5, RACMO22E_EC-EARTH']	[rsdsAdjust] [ridsAdjust]
<input type="checkbox"/> Rayonnement visible incident à la surface	W/m ² ⓘ
<input type="checkbox"/> Rayonnement infra-rouge incident à la surface	W/m ² ⓘ
<input checked="" type="checkbox"/> Vent ... [* Sélectionnez l'unité]	[sfcWindAdjust]
<input type="checkbox"/> Vitesse du vent à 10 m	m/s km/h ⓘ
<input checked="" type="checkbox"/> Evapo-Transpiration Potentielle ... [* Sélectionnez l'unité]	[evpsb/potAdjust]
<input type="checkbox"/> Evapotranspiration potentielle (méthode Hargreaves)	kg/m ² /s mm ⓘ

2. Données réelles, Infoclimat

Une fois les données DRIAS générées, l'objectif est de pouvoir comparer les résultats du modèle avec les données simulées pour la période allant de 1951 à aujourd'hui. Le but de cette approche est de s'assurer que les simulations du modèle ALADIN présentent une bonne cohérence statistique avec les données réellement observées. Pour cela, il faut trouver une source de données capable de nous fournir un maximum d'informations similaire à celle disponible dans nos données simulées.

Infoclimat est un site web et une communauté en ligne qui fournit des informations météorologiques et climatiques en temps réel pour la France et d'autres régions. C'est une plateforme collaborative où les passionnés de météo, les amateurs, les professionnels et les chercheurs partagent leurs observations.

Grâce à cela, nous pouvons avoir accès à des archives climatiques pour différentes villes et stations météorologiques. Dans notre cas, ce sont les données historiques pour les stations iséroises qui répondent au besoin. Nous optons pour cette source de données, car elle permet une extraction gratuite des informations de diverses stations météorologiques, qui peuvent être facilement associées aux points spatiaux des données DRIAS. Toutefois, Infoclimat ne propose qu'un nombre limité de mesures climatiques, avec uniquement 27 stations disponibles en Isère, et les périodes de données accessibles varient d'une station à une autre.

1. Zone géographique

Recherchez des stations...

2. Période

du au

7 jours consécutifs maximum, sauf si vous êtes propriétaire de la station

3. Téléchargement

JSON CSV

Sélectionnez des stations et dates pour obtenir le lien API

Je ne suis pas un robot reCAPTCHA Confidentialité - Conditions

TÉLÉCHARGER

Voici la page qui permet d'extraire les données d'Infoclimat. Cette interface génère un lien de téléchargement pour le format, les stations, et la période sélectionnée. Cependant, cette méthode limite la période maximale à 7 jours, alors que nous souhaitons générer des données sur une période beaucoup plus étendue. Pour résoudre ce problème, nous utilisons le scraping de données, qui consiste à automatiser l'extraction en utilisant un programme informatique. Sur R, nous développons une boucle qui utilise le lien généré sur Infoclimat pour changer automatiquement les dates de début et de fin à chaque itération (annexe 1). Ce qui permet alors de récupérer un maximum de données par station. Voici à quoi ressemble un lien Infoclimat lorsque nous demandons un fichier csv :

[https://www.infoclimat.fr/opendata/?method=get&format=csv&stations\[\]=00014&start=2020-01-01&end=2020-01-07&token=U4StkDUIT94L1JViziKubiCg2MKj4rV8in6TU16o7hk3bBhGGQoDdA](https://www.infoclimat.fr/opendata/?method=get&format=csv&stations[]=00014&start=2020-01-01&end=2020-01-07&token=U4StkDUIT94L1JViziKubiCg2MKj4rV8in6TU16o7hk3bBhGGQoDdA)

En rouge, nous retrouvons les dates de début et de fin qui se changeront automatiquement dans notre programme. En orange, l'identifiant unique associé à chacune des stations, et en vert, une clé API nécessaire pour l'extraction des données. Nous réalisons cette opération pour chaque station en Isère et enregistrons l'ensemble des données de chaque station dans un fichier RData.

Infoclimat ne propose pas de données journalières, mais plutôt des données en temps réel avec des fréquences différentes selon les stations. De plus, certaines de ces stations comportent des données manquantes. Tout cela nécessite donc une transformation des données afin de les rendre utilisables et comparables.

3. Transformation des données

Afin de rendre toutes nos données exploitables dans l'application, il est nécessaire de les standardiser. Pour ce faire, nous avons opté pour la forme des données DRIAS, étant donné qu'elles représentent la majeure partie de notre ensemble de données. Nous prévoyons également d'ajouter des informations complémentaires qui se révéleront précieuses pour la progression du projet. Regardons les modifications apportées aux données simulées.

A. Données DRIAS

Actuellement, nous disposons de plusieurs ensembles de données contenant différentes variables pour chaque RCP. La première étape consiste donc à fusionner ces ensembles de données afin d'obtenir un jeu de données global comprenant toutes les variables. Nous effectuons cette opération pour chaque RCP, aboutissant ainsi à deux jeux de données, un pour chaque scénario. Dans chacun de ces jeux de données, nous incluons les données historiques simulées.



Concernant le nom des variables, nous conservons ceux proposés par défaut dans DRIAS, que nous utiliserons également lors de la modification de nos données Infoclimat. De même, pour nos individus, chaque ligne représente un jour et un point spatial. Les modifications des données simulées sont donc principalement liées à l'ajout d'informations supplémentaires par jointure de tables attributaires.

Tout d'abord, nous devons attribuer un identifiant à chaque point spatial. Pour cela, nous ajoutons une variable appelée "ID" qui contient un nom unique pour chaque latitude, longitude et altitude. La colonne "ID" sera composée de noms allant de "p001" à "p122" pour chaque point spatial des données DRIAS. Nous extrayons également un jeu de données dans un RData qui stocke les coordonnées associées à chaque identifiant. Cela nous permettra de conserver facilement les informations sur la localisation de chaque point spatial pour une utilisation ultérieure dans notre projet.

Le reste des modifications portent sur les dates. Les sorties DRIAS nous fournissent des informations précises sur les jours. À partir de ces jours, nous extrayons deux informations supplémentaires : l'année, et le mois. Par exemple, pour la date du 09/10/2000, nous enregistrons dans différentes variables l'année 2000, et le mois d'octobre. Grâce à ces détails, nous aurons plus de facilité pour effectuer des analyses et générer des visualisations adaptées à différentes échelles temporelles selon les besoins de l'utilisateur.

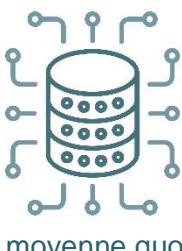
Les données DRIAS sont maintenant complètes, ce qui nous donne deux jeux de données, chacun avec 19 colonnes et 6 684 014 lignes, sans aucune donnée manquante (annexe 2.A). Nous les stockons chacun dans un fichier RData différent. Elles sont prêtes à être utilisées dans l'application ZODIAC Pro. Nous pouvons à présent passer à l'étape de transformation des données provenant d'Infoclimat.

B. Données Infoclimat

Les données d'Infoclimat sont pour le moment stockées dans un fichier RData, avec un jeu de données par station météorologique. Cependant, les informations contenues dans ces données diffèrent d'une

station à une autre. L'objectif va être de les uniformiser en adoptant un format similaire aux données DRIAS.

Pour accomplir cette tâche, nous développons une fonction qui exécute les divers ajustements en plusieurs étapes (annexe 3). La première étape implique la sélection des variables qui peuvent être comparées aux données simulées. Nous retenons ainsi la température, les précipitations et la vitesse du vent. Nous préservons également l'information temporelle à laquelle nous appliquons une transformation afin de conserver uniquement la date du jour. Cette manipulation nous conduit à obtenir plusieurs lignes (données horaires) qui correspondent à la même date, ce qui nous permettra une agrégation des données.



En ce qui concerne la température, nous effectuons plusieurs regroupements en utilisant diverses fonctions. Tout d'abord, nous calculons la moyenne pour obtenir la température moyenne. Ensuite, nous déterminons le maximum pour obtenir la température maximale journalière, et également le minimum pour obtenir la température minimale journalière. En ce qui concerne les précipitations, nous les regroupons en effectuant une somme. Et pour la vitesse du vent, nous calculons la moyenne quotidienne.

En parallèle de l'agrégation, nous effectuons un traitement des données manquantes en vérifiant si une variable possède un nombre suffisant de données. Nous considérons qu'il est nécessaire d'avoir au moins 12 valeurs pour une date, avec au moins 50% de ces valeurs non-manquantes. Si ces critères ne sont pas respectés, nous attribuons la valeur "NA" (*Not available*).

Chaque agrégation est effectuée individuellement, puis les résultats sont combinés pour former un ensemble de données. À ces données, nous ajoutons l'année et le mois, ainsi qu'une colonne "ID" qui représente l'identifiant de la station sur Infoclimat. Nous renommons les variables et obtenons un jeu de données complet pour une station.

Nous appliquons cette fonction pour les autres stations, puis regroupons l'ensemble pour obtenir la totalité des données réelles. Ainsi, nous obtenons un jeu de données comportant 10 colonnes et 54 466 lignes que nous stockons dans un RData (annexe 2.B). Bien que le nombre de données soit inférieur à nos simulations en raison de la quantité réduite de stations sur Infoclimat et de la période plus courte, ce nombre reste suffisant pour réaliser nos comparaisons. Pour chaque station, nous collectons manuellement les coordonnées de latitude et de longitude, en les associant à l'identifiant dans un jeu de données que nous ajoutons au fichier RData contenant les coordonnées. Avec toutes les données en main et soigneusement préparées, nous pouvons maintenant passer à la phase d'analyse.

III. IAC & ANALYSES STATISTIQUES

La partie décisive, afin de satisfaire les demandes de la chambre d'agriculture, réside dans l'analyse statistique des données. La première étape consistera à cibler, définir, et voir comment sont calculés les différents types d'IAC. Une fois ce travail accompli, nous pourrons traiter des analyses statistiques demandées. Dans un premier temps, nous nous focaliserons sur les comparaisons des données réelles et simulées (tests d'homogénéité). Par la suite, nous aborderons l'analyse spatiale et le zonage géographique que nous avons élaboré, qui se divise en deux approches distinctes. Tout d'abord, une cartographie univariée (basée sur un IAC), suivie d'une analyse multivariée (classification des points spatiaux en fonction de plusieurs IAC).

1. Catégorie et calcul d'IAC

Commençons par expliquer les IAC nécessaires à la réalisation du projet. On compte trois catégories d'IAC que nous avons appelé : quantiles, seuil critique et la quantité critique. Explorons de manière plus approfondie les significations cachées derrière chacune de ces appellations.



Les indicateurs de type "quantiles" sont des valeurs qui divisent les observations d'une variable en parties de fréquences égales. Ils nous aident à comprendre comment les valeurs de la variable étudiées sont réparties. Dans notre cas, nous conservons cinq quantiles Q05, Q25, Q50, Q75, et Q95. Par exemple, le Q75 signifie que 75% des valeurs se trouvent en dessous de ce quantile, et donc 25% des valeurs se trouvent au-dessus. Les quantiles nous offrent la possibilité de visualiser l'évolution au fil du temps. Chaque quantile est calculé sur une période annuelle spécifique.

Un indicateur de type "seuil critique" concerne les valeurs qui se situent au-dessus ou en dessous d'un niveau déterminé. L'indicateur est alors défini en termes de nombre de jours ou bien de cumul. Par exemple, nous pouvons être amenés à rechercher les valeurs où la température moyenne dépasse les 15°C au cours d'une période spécifique, puis de calculer soit le nombre de jours, soit le cumul des degrés au-delà de ce seuil critique. Il peut également être requis de déterminer la durée maximale de jours consécutifs au sein de cette période où cette condition est satisfaite.

Pour finir, nous définissons les indicateurs de type "quantité critique". Ici aussi, nous regardons les valeurs d'une variable situées au-dessus ou en dessous d'un seuil, mais contrairement à la précédente approche, nous ne nous plaçons pas dans une période spécifique, mais plutôt nous visons à trouver, à partir d'une date de départ, le moment où le cumul atteindra une valeur cible prédéfinie. Il est possible d'avoir une valeur d'écrêtage, elle consiste à limiter certaines données en les restreignant à la valeur spécifique. Le nombre de jours pour atteindre la cible est la seule information à extraire pour cette catégorie d'indicateur. Par exemple, nous pouvons rechercher, à partir du 09 octobre le nombre de jours où le cumul des températures maximales au-dessus de 20°C, avec une limite à 35°C, atteint les 100°C d'excédent.

Chaque catégorie d'IAC implique la sélection d'une variable. Pour les quantiles, nous ajoutons le choix de la période dans l'année. De même pour le seuil critique. Cette catégorie demande également la sélection d'un seuil, ainsi que le type de comparaison à effectuer (inférieur à ou supérieur à). La quantité critique, quant à elle, nécessite la sélection d'une date de départ, d'un seuil, du type de comparaison, de la cible à atteindre, et si nécessaire, d'une valeur d'écrêtage. Le calcul des IAC repose sur une fonction

spécifique (annexe 4) qui requiert en entrée un jeu de données (RCP 8.5 ou 4.5), l'IAC à calculer et tous les éléments énumérés ci-dessus. Selon l'IAC à évaluer, les valeurs nécessaires sont introduites, et les paramètres non requis pour ce calcul prennent la valeur NULL. La fonction se charge ensuite d'effectuer le calcul de l'indicateur et renvoie un tableau avec une ligne par année, contenant les valeurs de l'IAC pour la période sélectionnée.

A l'aide des valeurs de ce tableau, nous réalisons un lissage des données afin d'extraire une tendance générale. Ce lissage est réalisé grâce à une fonction R appelée "smooth.spline". Un des paramètres de cette fonction est le degré de lissage. Il s'agit de la quantité de réduction des variations des données. C'est une valeur entre 0 et 1. Plus elle est proche de 1 plus les variations seront réduite, permettant ainsi de mieux se rendre compte de la tendance des données. Ce degré de lissage apparaît comme un paramètre qu'un utilisateur pourra faire varier dans l'application. Prenons des exemples de visualisations d'IAC afin d'illustrer ces informations. Commençons avec les graphiques disponibles pour les quantiles.

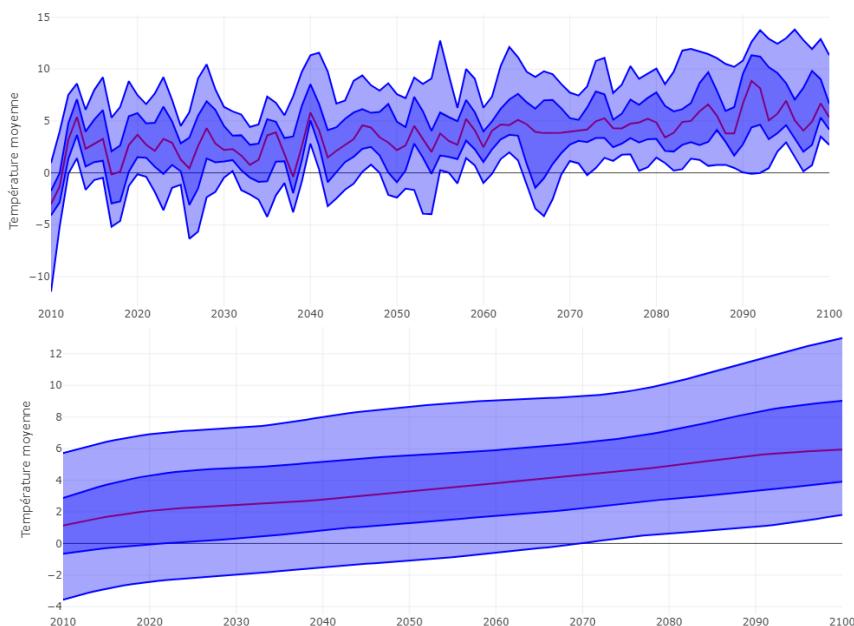


Figure 1 : Evolution des quantiles de la température moyenne de janvier, entre 2010 et 2100, du RCP 8.5 pour la ville de Saint-Martin-d'Hères

Prenons l'exemple de la température moyenne du mois de janvier, en examinant l'évolution entre 2010 et 2100 pour les données du scénario 8.5 de DRIAS, pour la ville de Saint-Martin-d'Hères. Nous produisons deux graphiques en appliquant deux niveaux de lissage différents : le premier à 0.3 et le deuxième à 0.8. En observant la Figure 1, nous constatons une augmentation tendancielle des températures. Cette augmentation est plus notable sur le deuxième graphique en raison du niveau de lissage plus élevé, ce qui améliore la clarté de la tendance.

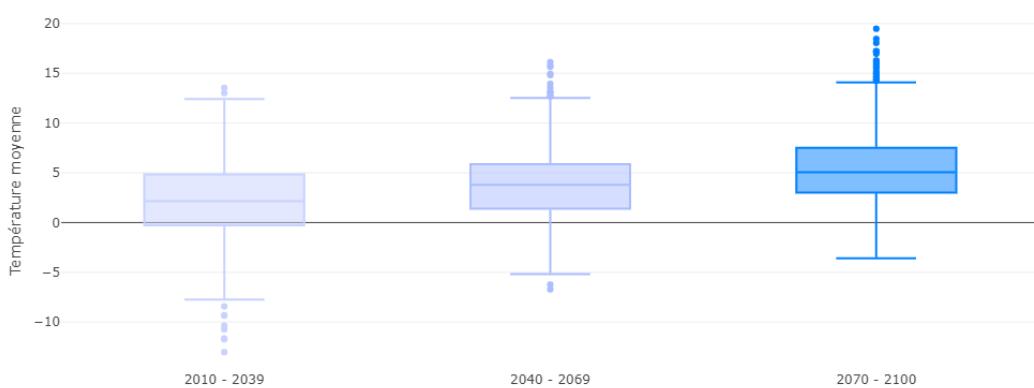


Figure 2 : Boxplot de la température moyenne de janvier, entre 2010 et 2100, du RCP 8.5 pour la ville de Saint-Martin-d'Hères

Une autre représentation graphique est accessible pour visualiser les quantiles. Il s'agit d'un boxplot qui renferme les mêmes données que le graphique précédent, mais organisées en intervalles de 30 ans. Certains utilisateurs se montrent plus compétents dans l'interprétation de ce genre de graphique. Poursuivons avec des exemples de visualisation pour les IAC de type seuil critique. Nous en retrouvons un pour la version *nombre de jours*, et un pour la version *cumul*. Les graphiques sont similaires, explorons-les plus en profondeur afin de saisir leur mode d'interprétation.

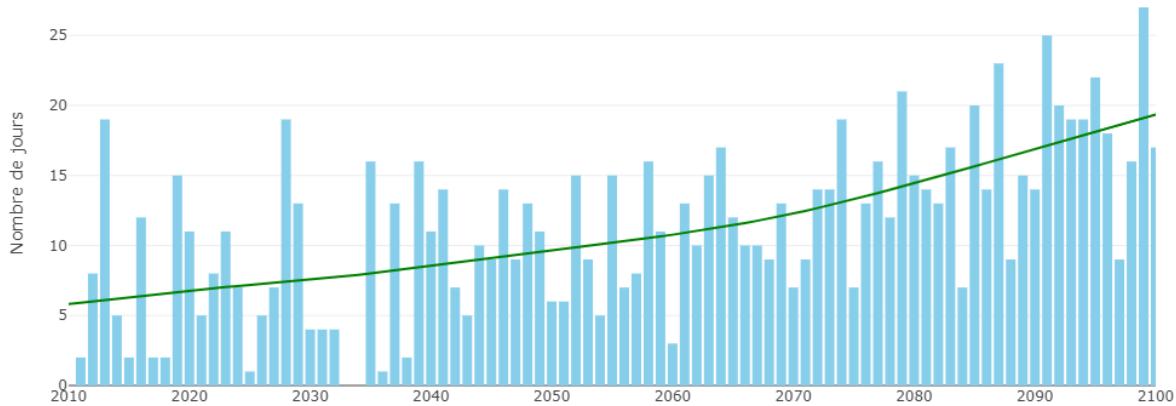


Figure 3 : Evolution du nombre de jours avec une température moyenne de janvier supérieure à 5°C, entre 2010 et 2100, pour le RCP 8.5 de la ville de Saint-Martin-d'Hères

Voici le graphique proposé pour voir l'évolution du nombre de jours par rapport à un seuil critique. Dans la Figure 3, nous observons l'évolution entre 2010 et 2100, du nombre de jours où les températures moyennes sont supérieures à 5°C, pour le mois de janvier. La courbe verte représente la tendance, nous utilisons un degré de lissage de 0.8. Les barres correspondent aux valeurs exactes calculées. Grâce à cette visualisation, nous arrivons à lire la tendance de nos données. Ici, nous remarquons que le nombre de jours avec une température moyenne au-dessus de 5°C en janvier, risque d'augmenter considérablement d'après les simulations, avec le scénario 8.5. Ce type de graphique est aussi utilisé pour le cumul, et la quantité critique.

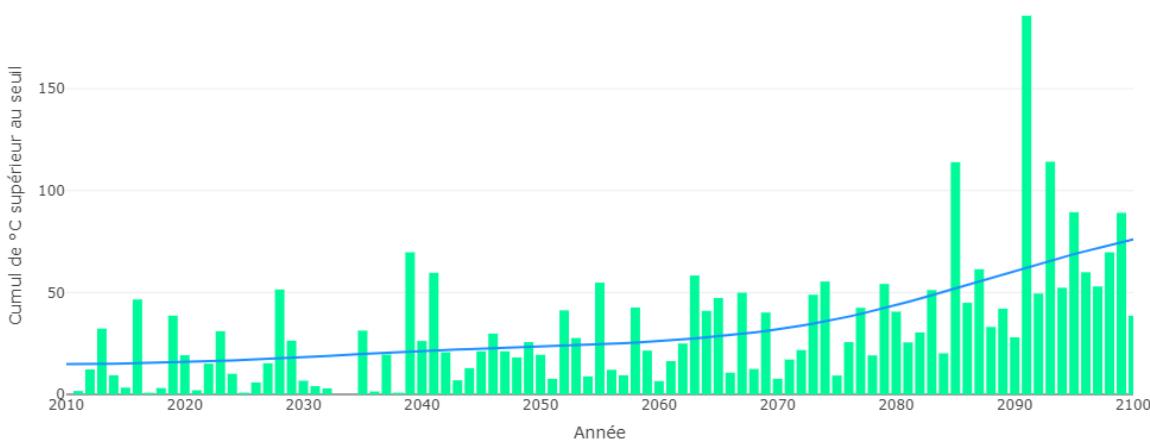


Figure 4 : Evolution du cumul de degrés supérieurs à 5°C pour les températures moyennes de janvier, entre 2010 et 2100, pour le RCP 8.5 de la ville de Saint-Martin-d'Hères

Pour les cumuls, nous prenons le même exemple. Ici, la tendance est représentée par la courbe bleu, tandis que les valeurs calculées sont en vert. Nous observons avec la Figure 4, l'évolution entre 2010 et 2100, des cumuls des degrés au-dessus de 5°C, pour le mois de janvier. Nous remarquons que la tendance est à la hausse. C'est-à-dire que d'après le modèle ALADIN pour le RCP 8.5, il y aura des excès plus importants concernant les températures lorsque nous dépasserons les 5°C, pour les températures moyennes de janvier.

Nous venons de voir les deux graphiques principaux proposés pour le seuil critique. Terminons notre tour des visualisations avec la dernière catégorie d'IAC, la quantité critique.

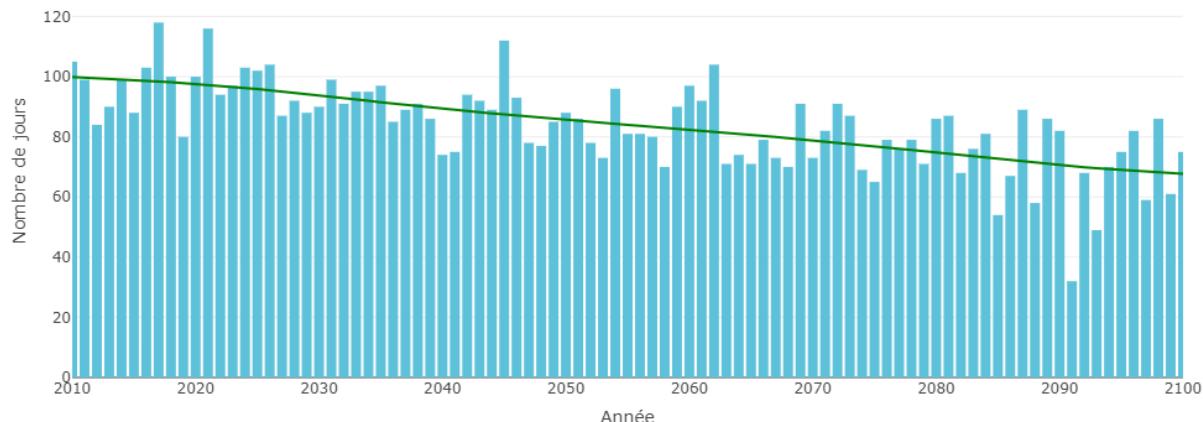


Figure 5 : Evolution du nombre de jours pour atteindre les 200°C, pour les températures moyennes supérieures à 5°C, à partir du 1^{er} janvier, entre 2010 et 2100, avec le RCP 8.5 pour la ville de Saint-Martin-d'Hères

Pour cette catégorie, nous proposons une seule visualisation, similaire aux précédentes. Mais l'information apportée n'est pas la même. Dans la Figure 5, nous observons l'évolution entre 2010 et 2100, du nombre de jours pour atteindre un cumul de 200°C, pour les températures moyennes supérieures à 5°C, à partir de la date départ fixée ici au 1er janvier. Nous utilisons les données du scénario 8.5. Il n'y a pas de valeur d'écrêtage dans cet exemple. Nous remarquons que la tendance, ici en vert avec un degré de lissage à 0.8, est décroissante. Ce qui signifie que de moins en moins de jours seront nécessaires pour atteindre un cumul de 200°C, quand nous observons les températures moyennes supérieures à 5°C.

L'ensemble de ces IAC sont d'une grande importance pour le secteur agricole, car ils offrent la possibilité de s'ajuster aux nouvelles conditions climatiques à venir, facilitant ainsi la prise de décisions pour la planification des cultures (semis, irrigation, récolte, etc.). A présent, nous comprenons comment nos IAC sont définis, calculés, et illustrés. Penchons-nous maintenant sur les analyses statistiques réalisées.

2. Comparaison statistique des données

La première analyse consiste à confronter les données réelles aux données simulées. Pour rappel, notre objectif est de garantir que les données simulées ne s'écartent pas excessivement de la réalité. Dans cette optique, nous avons initialement décidé de confronter les données historiques des simulations aux données réelles provenant d'Infoclimat. Cependant, après réflexion, il apparaît plus pertinent de pouvoir également comparer les données simulées entre elles, plus spécifiquement entre les deux scénarios différents. Les comparaisons s'opèrent entre deux ensembles d'échantillons. Chaque ensemble représente un jeu de données différents, avec les données relatifs à la même variable et à la même période temporelle. Pour réaliser cette tâche, nous mettons en œuvre deux approches de comparaison : une visuelle à travers la création de graphiques, et une statistique via l'application de quatre tests d'homogénéité pour comparer les deux échantillons.

Pour la première approche, nous utilisons deux graphiques distincts : le premier affiche les deux séries temporelles de la variable que nous souhaitons comparer, tandis que le deuxième confronte les distributions (densités de probabilités) des deux échantillons de données, avec deux barres verticales pour visualiser les moyennes de chaque échantillon. Si dans chacun des graphiques les courbes se

rapprochent, ou pourra considérer que les deux échantillons sont issus d'une même distribution, et inversement.

La deuxième approche basé sur des tests d'homogénéité permet de confirmer ou réfuter les observations graphiques. En effet, les tests d'homogénéité sont des méthodes statistiques utilisées pour évaluer la similitude entre deux distributions de données, ils fournissent des mesures quantitatives pour évaluer à quel point les données s'ajustent, notamment la p-valeur. Chaque test se base sur une hypothèse nulle qui correspond à la caractéristique des données attendues. La p-valeur quant à elle est une probabilité qui quantifie la crédibilité de l'hypothèse nulle. Dans notre cas, nous mettons en place quatre tests différents. Le premier est un test de Kolmogorov qui permet de vérifier l'égalité des distributions. L'hypothèse nulle est que nos deux échantillons sont issus d'une même distribution. Le second est un test sur les variances de nos échantillons, avec pour hypothèse nulle l'égalité des variances. Ensuite, nous effectuons un test sur les moyennes. Ici, nous cherchons à déterminer si les moyennes des deux échantillons sont égales. Pour finir, nous réalisons un test de corrélation, l'hypothèse nulle étant que la corrélation entre les deux distributions n'est pas significativement supérieure à zéro. Une p-valeur proche de 1 indiquera que l'hypothèse nulle est validée, ce qui est favorable pour tous nos tests à l'exception du test de corrélation. Dans le contexte actuel, une p-valeur proche de 1 dans le test de corrélation suggérera un lien négatif ou nul. Cependant, nous cherchons à maintenir la cohérence entre tous nos tests en interprétant une p-valeur de 1 comme une confirmation de la proximité entre nos deux échantillons. Pour ce faire, nous optons pour une manipulation consistant à soustraire 1 de la p-valeur du test de corrélation. Cette transformation permettra d'obtenir une valeur proche de 1 lorsque la corrélation entre les deux distributions est significativement positive. En conséquence, l'ensemble de nos tests aboutira à une mesure de probabilité : si cette mesure est proche de 0, nous estimerons que nos échantillons sont différents, alors qu'une mesure proche de 1 nous permet de considérer que nos deux échantillons présentent des similitudes. Pour réaliser tous nos tests, nous développons une fonction (annexe 5) qui retourne un tableau avec les différentes probabilités pour chacun des tests. Les cases du tableau prennent une teinte de plus en plus verte à mesure que les valeurs se rapprochent de 1.

Pour apporter davantage de clarté, prenons un exemple de comparaison entre deux échantillons. Nous sélectionnons alors deux sites. Le premier est le point p048 de nos données DRIAS pour le RCP 8.5, qui correspond à la ville de Saint-Martin-d'Hères. Le deuxième provient des données d'Infoclimat et correspond à la même ville. Nous examinons les données de 2009 à 2020 pour la température moyenne. Commençons avec les graphiques.

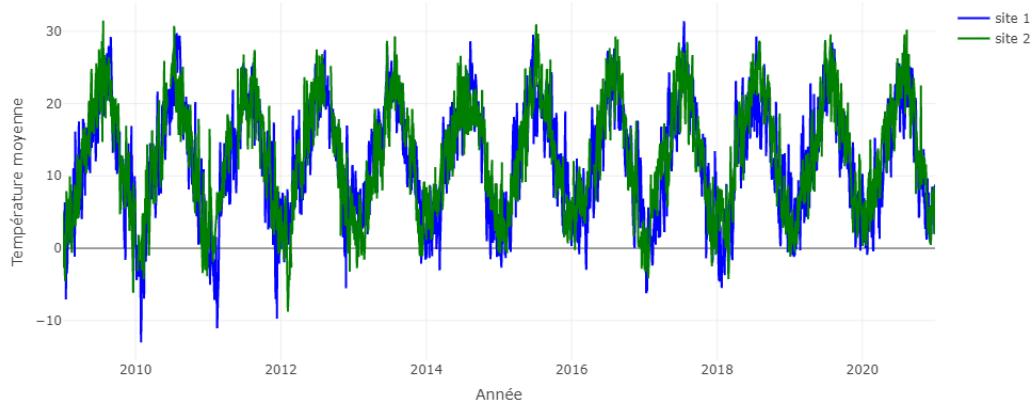


Figure 6 : Série temporelle entre 2009 et 2020, des températures moyennes des données DRIAS RC8.5 et Infoclimat, pour la ville de Saint-Martin-d'Hères

Voici comment se présente une série temporelle. Elle permet de visualiser les deux échantillons sur un même graphique. Cette méthode facilite l'observation pour déterminer si les deux ensembles de données suivent des tendances et des variations saisonnières similaires. Dans le cas illustré ici, les deux courbes se chevauchent, ce qui suggère que les données simulées sont en accord avec celles provenant d'Infoclimat.

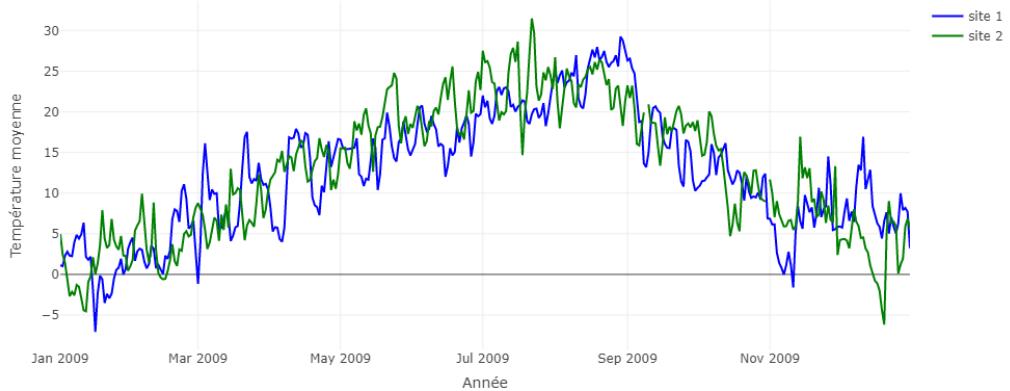


Figure 7 : Série temporelle en 2009, des températures moyennes des données DRIAS RC8.5 et Infoclimat, pour la ville de Saint-Martin-d'Hères

Nous pouvons également voir comment se comportent nos échantillons pour une année précise. Ici, nous nous focalisons sur 2009 avec la Figure 7. À cette échelle, les distinctions entre chaque valeur deviennent plus apparentes. En ce qui concerne les températures moyennes de cette année, nous pouvons constater une similitude dans la tendance et la saisonnalité, car les deux courbes présentent une forme générale semblable, malgré des différences notables à certains points.

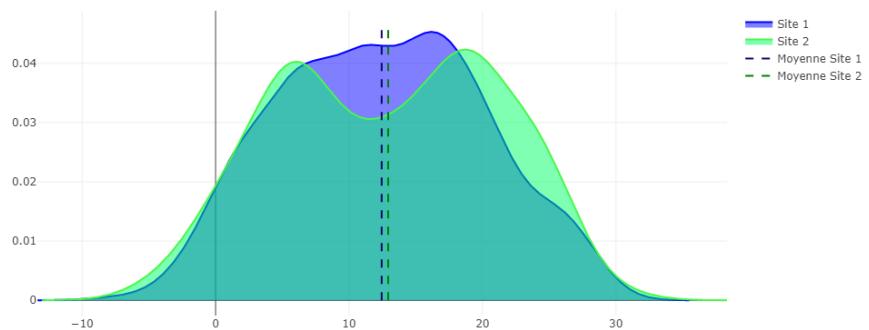


Figure 8 : Densité de distribution en 2009, des températures moyennes des données DRIAS RC8.5 et Infoclimat, pour la ville de Saint-Martin-d'Hères

La Figure 8 nous permet d'observer la densité de distribution pour une année donnée. Dans cet exemple, nous examinons les températures moyennes de l'année 2009 complète. Nous pouvons remarquer que les données réelles présentent une distribution en deux parties (distribution bimodale), discernables grâce à deux pics distincts, alors que les données simulées suivent une distribution plus gaussienne, c'est-à-dire une répartition en cloche autour de la moyenne. Les deux moyennes semblent être relativement proches. Il est important de noter que des échantillons avec des effectifs différents peuvent être présents, en raison de possibles valeurs manquantes dans les données provenant d'Infoclimat. Pour quantifier les hypothèses faites graphiquement, regardons les résultats des tests statistiques.

Tableau 1: Résultat des tests pour les températures moyennes des données DRIAS RC8.5 et Infoclimat, pour la ville de Saint-Martin-d'Hères entre 2009 et 2020

	Egalité_Distributions	Egalité_Variances	Egalité_Moyennes	Test_de_Correlation
2009	0.041	0.099	0.402	1
2010	0.542	0.013	0.507	1
2011	0.003	0.008	0.004	1
2012	0.108	0.003	0.394	1
2013	0.022	0.008	0.297	1
2014	0	0	0.045	1
2015	0.224	0.345	0.644	1
2016	0.457	0.262	0.83	1
2017	0.174	0.311	0.295	1
2018	0.001	0.566	0	1

Le Tableau 1 présente les résultats obtenus à partir des quatre tests statistiques appliquées aux températures moyennes entre 2009 et 2020. Pour rappel, une valeur se rapprochant de 1 indique une similitude entre nos deux échantillons. L'analyse des corrélations sur l'ensemble de la période met en évidence un lien fort et positif entre les données réelles et simulées. Pour le reste, les résultats divergent selon les années. En se concentrant sur l'année 2009, le test d'égalité des distributions reflète les tendances de densité observées dans la figure précédente, avec une valeur qui se rapproche de 0, indiquant que les distributions ne peuvent pas être considérées comme égales. Cette observation se répercute également dans les résultats des tests de variance. En ce qui concerne la moyenne, la valeur se situe près de 0.5, ce qui ne suffit pas à retenir une égalité significative des moyennes, bien que la proximité reste notable, comme observée dans la Figure 8.

L'ensemble de ces résultats n'empêcheront pas la réalisation des simulations. Ils permettent essentiellement de prendre du recul sur les données mises à disposition, afin de ne pas travailler avec des données simulées trop aberrantes.

3. Zonage géographique

L'aspect zonage géographique est une partie capitale pour la Chambre d'Agriculture de l'Isère. Il s'agit d'un des atouts majeurs que l'application ZODIAC Pro doit offrir par rapport à CANARI. Explorons les deux approches de zonage géographique mises en œuvre pour répondre à cet objectif. Tout d'abord, nous aborderons la méthode de cartographie d'IAC, aussi désignée sous le terme d'interpolation spatiale.



A. Interpolation spatiale

L'interpolation spatiale fait référence aux techniques utilisées pour estimer des valeurs inconnues dans l'espace à partir de données disponibles à des emplacements spécifiques. En d'autres termes, l'interpolation spatiale consiste à créer une représentation continue d'une variable dans une région où seules quelques mesures sont disponibles. Dans notre cas, nous voulons créer une carte de l'Isère selon un IAC, à partir des valeurs que nous obtenons pour les différents points spatiaux. Pour réaliser les estimations, nous utilisons la méthode GWR (Geographically Weighted Regression). C'est une technique d'analyse spatiale utilisée pour modéliser les relations entre les points dans l'espace en

tenant en compte l'emplacement géographique. Pour cela, nous partons d'une grille de l'Isère où chaque case (ou pixel) doit obtenir une valeur. Nous ne connaissons que les valeurs des 122 points spatiaux. Pour chaque point spatial, une régression linéaire locale est effectuée en utilisant uniquement les observations situées autour de ce point. Les paramètres de régression sont estimés localement pour chaque point. La GWR utilise des noyaux (kernels) pour pondérer les observations en fonction de leur distance par rapport au point d'intérêt. Pour résumer, nous estimons les points inconnus à l'aide de plusieurs régressions locales, auxquelles un poids est attribué selon la distance du point par rapport aux observations que nous avons déjà.

En R, nous développons une fonction (annexe 6) qui permettra de réaliser l'interpolation. Cette fonction calcule les estimations pour chaque case de la grille de l'Isère. En utilisant ces valeurs estimées, il est possible de générer une carte de l'Isère représentant la distribution spatiale d'un IAC. En plus des quantiles, du seuil critique et de la quantité critique, nous ajoutons la possibilité de créer des cartes basées sur la moyenne et la somme des variables. Deux types de cartes sont proposés : une avec une variation continue de couleur en fonction des valeurs de l'IAC, et une carte à isolignes qui divise la région étudiée (Isère) en différentes zones selon des intervalles de valeurs. Pour mieux comprendre, prenons un exemple d'interpolation.

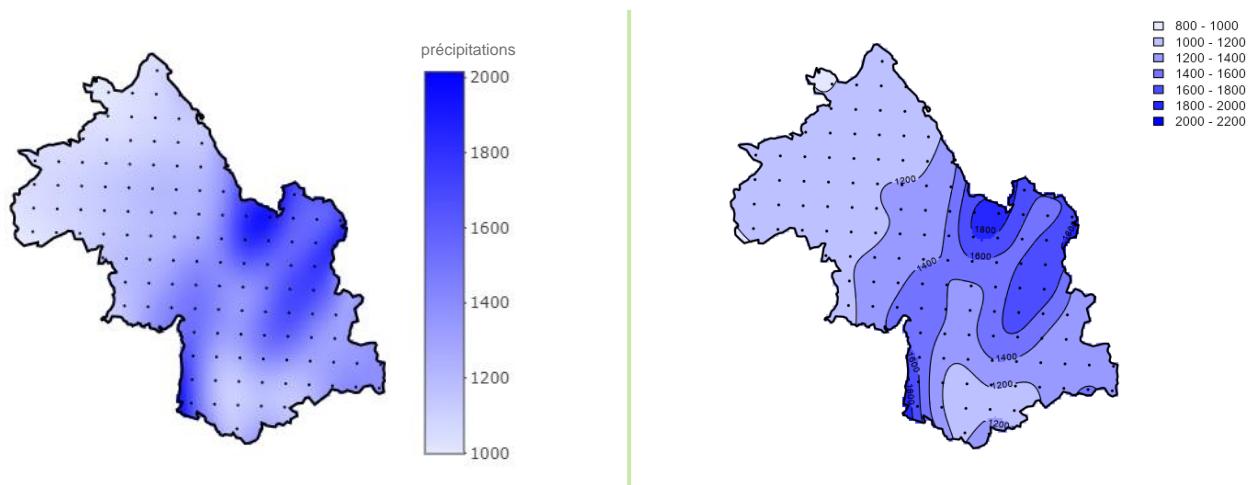


Figure 9 : Cartes d'interpolation de la somme des précipitations annuelles, en Isère, entre 2025 et 2030

La Figure 9 présente deux cartes générées grâce à la fonction d'interpolation. Ici, nous avons réalisé la somme des précipitations annuelles entre 2025 et 2030 en utilisant les données du scénario 8.5. Pour chaque année et chaque point spatial, nous calculons la somme des précipitations. Nous faisons ensuite la moyenne de ces sommes sur l'ensemble des années. Cette moyenne est réalisée peu importe l'indicateur statistique utilisé. Ainsi, nous obtenons une valeur pour chaque point spatial, que nous utilisons ensuite dans la méthode GWR pour l'interpolation. Dans notre exemple, nous utilisons une palette de couleurs allant du blanc au bleu. Cette palette varie en fonction de la variable utilisée. Les deux cartes apportent la même information, mais elles sont présentées différemment. Nous pouvons remarquer que, globalement, le nord de l'Isère connaît des précipitations moins fortes que le sud. En réalisant plusieurs cartes pour différentes années, nous pourrons observer comment le climat évolue selon les différentes zones de l'Isère. Maintenant que nous savons comment cartographier un IAC, essayons de classer nos points spatiaux selon plusieurs IAC.

B. Classification multicritère

La classification vise à regrouper les individus similaires en fonction de leurs caractéristiques communes. Dans notre contexte, les individus sont nos 122 points spatiaux DRIAS. Pour réaliser une classification, nous décidons d'utiliser une CAH (Classification Ascendante Hiérarchique). C'est une méthode d'analyse statistique utilisée pour explorer et organiser des données multivariées en groupes. Ici, les variables sont les IAC choisis par l'utilisateur. Pour réaliser la CAH, il faut donc générer un jeu de données qui contient les calculs de plusieurs IAC pour l'ensemble de nos points spatiaux. Une fois ce jeu de données mis à disposition, il suffit d'appliquer les fonctions R permettant de réaliser une CAH, en choisissant le nombre de classes désiré, et d'extraire les différents groupes créés par la classification. Nous pouvons ensuite ressortir une carte avec une couleur pour chaque point selon son groupe d'appartenance. Les points du même groupe seront donc les points avec des caractéristiques similaires par rapport aux IAC sélectionnés. Prenons un exemple de classification.

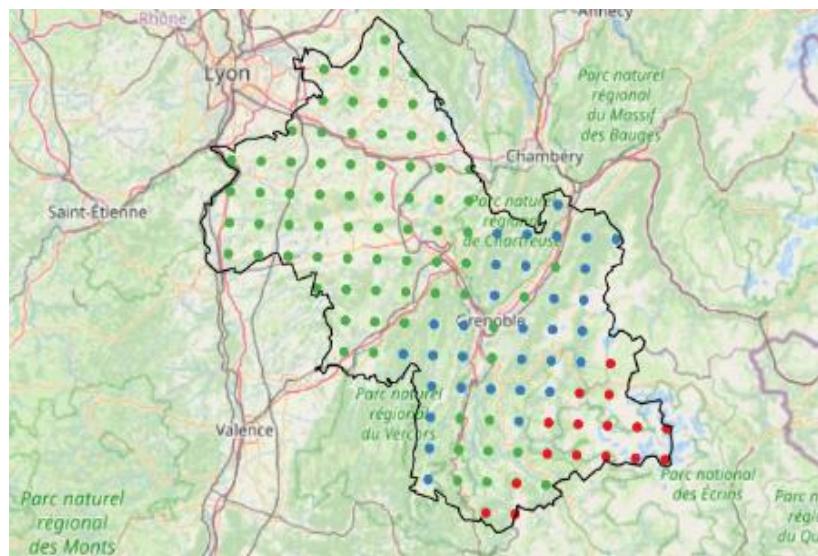


Figure 10 : Classification des points DRIAS selon trois IAC

Dans cet exemple, nous décidons de retenir trois IAC. L'IAC 1 représente la moyenne des températures minimales en janvier, l'IAC 2 correspond au nombre de jours où la température maximale dépasse 7°C en janvier, et l'IAC 3 est la somme totale des précipitations pour le mois de janvier. Nous collectons les données pour la période allant de 2025 à 2030 et choisissons de les classifier en trois classes. Le résultat se traduit par la création d'une carte, comme illustré dans la figure 10, où trois couleurs sont utilisées pour représenter les trois classes. Nous observons que ces groupes affichent une cohérence géographique. Plus spécifiquement, le département de l'Isère est principalement divisé en trois zones distinctes : le nord, le centre et le sud. Cependant, il est à noter que certains points situés dans la région centrale présentent des caractéristiques plus proches de la zone nord de l'Isère.

Avec cette carte, nous arrivons à visualiser les différents groupes, mais nous ne savons pas quelles valeurs des IAC les caractérisent. Nous réalisons donc une ACP (Analyse en Composantes Principales). L'ACP est une technique statistique utilisée pour réduire la dimensionnalité des données tout en conservant les informations essentielles, ce qui facilite leur interprétation, leur visualisation et leur analyse. Dans notre cas, nous utiliserons l'ACP pour produire un graphique spécial nommé biplot. Cette visualisation permet de représenter à la fois les individus et les variables d'un ensemble de données en deux dimensions sur un même graphique (plan factoriel). Grâce à ça, nous pourrons comprendre les relations entre les variables et les observations. Nous laisserons le choix à l'utilisateur de projeter les deux dimensions (axes factoriels) qu'il désire.

Un diagramme en barre des pourcentages de variance expliquée disponible pour faciliter ce choix. Reprenons l'exemple précédent, et essayons de trouver quel sont les variables qui caractérisent les classes.

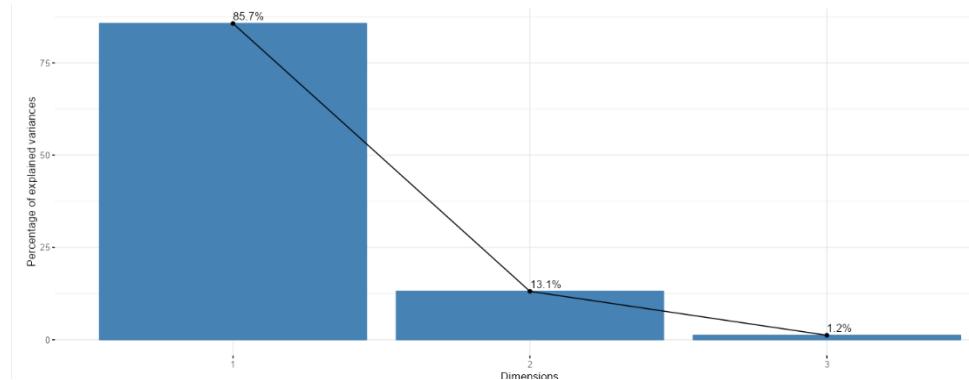


Figure 11 : Diagramme en barre des pourcentages de variance expliquée

On commence par chercher quels axes apportent le plus d'information dans l'ACP. Avec la Figure 11, nous remarquons que les dimensions 1 et 2 sont suffisantes pour conserver plus de 98% de l'information. Nous décidons donc de représenter ces deux axes sur le biplot.

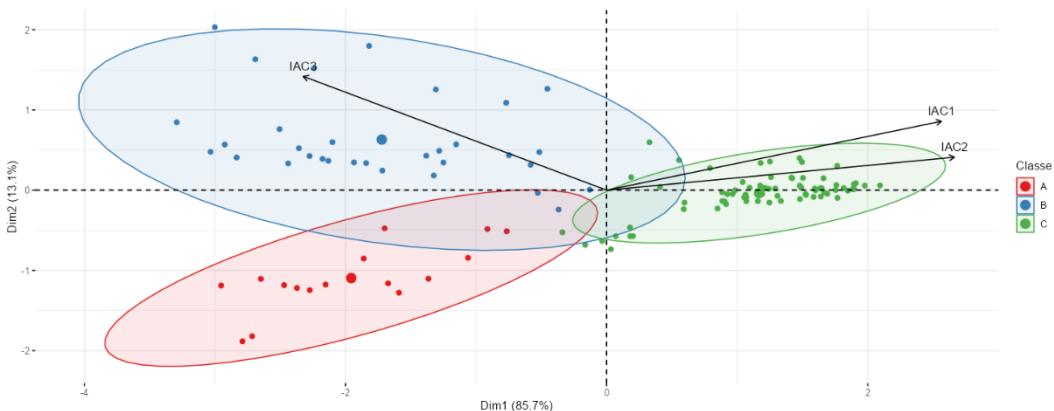


Figure 12 : Biplot de l'ACP

La Figure 12 illustre le biplot obtenu à partir de notre exemple. Chaque point représente l'un des 122 points DRIAS, les flèches correspondent aux IAC. Les ellipses et les couleurs sont utilisées pour représenter les différents groupes. Nous pouvons observer que le groupe vert, qui correspond au nord de l'Isère dans la Figure 10, est associé aux points présentant des valeurs élevées pour l'IAC 1 et l'IAC 2. Ces indicateurs représentent respectivement la moyenne des températures minimales et le nombre de jours où la température est supérieur à 7°C en janvier. Le groupe bleu se distingue par des valeurs élevées pour l'IAC 3, qui représente la somme des précipitations du mois de janvier. Quant au groupe rouge, il se positionne à l'opposé de l'IAC 1 et 2, ce qui signifie qu'il est caractérisé par des valeurs faibles pour ces indicateurs.

À présent, nous possédons les éléments nécessaires pour classifier nos points spatiaux et identifier les IAC qui définissent ces classes. La phase de zonage géographique est à présent opérationnelle. Nous comprenons la signification des IAC, et maîtrisons la réalisation de l'ensemble des analyses statistiques. Abordons désormais la dernière étape, qui concerne la conception de l'application ZODAC Pro.

IV. L'APPLICATION ZODIAC PRO

L'application ZODIAC Pro a pour but d'explorer et de visualiser les données des simulations climatiques de manière conviviale, sans nécessité de compétences avancées en programmation web. Elle permet principalement de faciliter la mise en place des analyses statistiques vues précédemment (calcul d'indicateurs agro-climatiques, tests statistiques, zonage géographique). Voyons comment se présente l'application. Dans un premier temps, nous observons les principales sections de l'application. Par la suite, nous établirons un guide d'utilisation avec une vue d'ensemble de ses fonctionnalités.



1. Découpage de l'application

Pour permettre à l'utilisateur de s'y retrouver dans l'application, nous devons opter pour un découpage en plusieurs parties distinctes, simple à comprendre. Il faut aussi éviter de mélanger des informations différentes dans la même partie. Pour cela, nous devons trouver la façon la plus optimale de présenter l'ensemble des informations.

Nos analyses précédentes étaient scindées en trois principales parties. La première était l'observation d'un IAC sur un site au choix (point DRIAS ou site météo Infoclimat), suivie d'une comparaison entre deux sites, et enfin le traitement de l'ensemble des points spatiaux DRIAS. Ce même schéma est reflété dans notre application, qui se divise donc en trois grandes parties : une dédiée à l'observation d'un point DRIAS, une autre à la comparaison de deux sites (point DRIAS ou site Infoclimat) , et enfin la dernière partie au zonage géographique de l'ensemble des points spatiaux DRIAS. Chacune de ces sections renferme plusieurs sous-parties distinctes.

ZODIAC Pro	
Observation d'un site	<
Comparaison de deux sites	<
Observation de plusieurs sites	<

Observation d'un site	▼
» Carte	
» Visualisation	
» Quantiles	
» Seuil critique	
» Quantité critique	

Pour l'observation d'un site, nous retrouvons tous les éléments relatifs aux IAC. Dans un premier temps, il y a la section *carte* qui offre la possibilité de sélectionner un scénario et un emplacement DRIAS. Vient ensuite la section *visualisation*, qui a pour but d'illustrer les données sélectionnées par un graphique ou un tableau. Viennent ensuite les sections dédiées aux grandes catégories d'IAC. Chacune permet d'avoir accès aux visualisations relatives aux catégories d'IAC vues dans la partie des analyses statistiques. Pour la catégorie "quantiles" par exemple, nous retrouvons le boxplot ainsi que le graphique permettant de voir l'évolution des différents quantiles dans le temps.

Comparaison de deux sites	
» Cartes	
» Visualisation	
» Comparaison statistique	

Ensuite, nous entrons dans la phase de comparaison entre deux sites. Cette étape suit une structure similaire à la précédente. La section *carte* est utilisée pour choisir les deux jeux de données ainsi que les sites à comparer. La section *visualisation* permet d'accéder aux graphiques des séries temporelles et des distributions (densités de probabilité). Nous retrouvons également les deux tableaux de données relatifs aux sélections. Enfin, la section de *comparaison statistique* génère le tableau des résultats des tests d'homogénéité effectués, comme présenté précédemment dans la section des analyses statistiques.

Observation de plusieurs sites ▾
» Cartographie
» Paramètres classification
» Classification

La troisième grande partie, dédiée au zonage géographique, se divise en trois sections distinctes. La première, intitulée *cartographie*, traite de l'interpolation spatiale. Elle comprend les paramètres à saisir et présente les cartes générées. En ce qui concerne la classification, l'entrée des paramètres est plus complexe du fait de la nécessité de sélectionner plusieurs IAC. C'est pourquoi cette partie est subdivisée en deux sous-sections distinctes : l'une pour la saisie des divers paramètres et l'autre pour l'affichage des résultats, notamment la carte et le biplot.

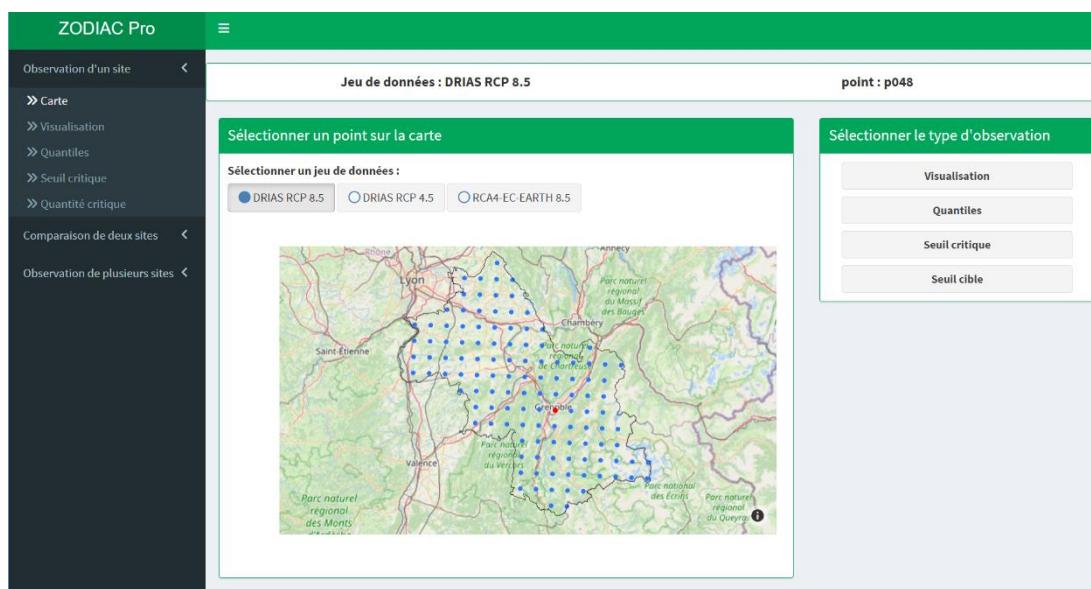
Voici comment l'application ZODIAC Pro divise ces différentes sections. L'élément clé à noter est que chaque partie est liée au nombre de points DRIAS à prendre en compte, que ce soit un seul point, deux points ou la totalité des points. Il ne nous reste plus qu'à explorer les détails cachés derrière chacune de ces parties.

2. Présentation de ses fonctionnalités

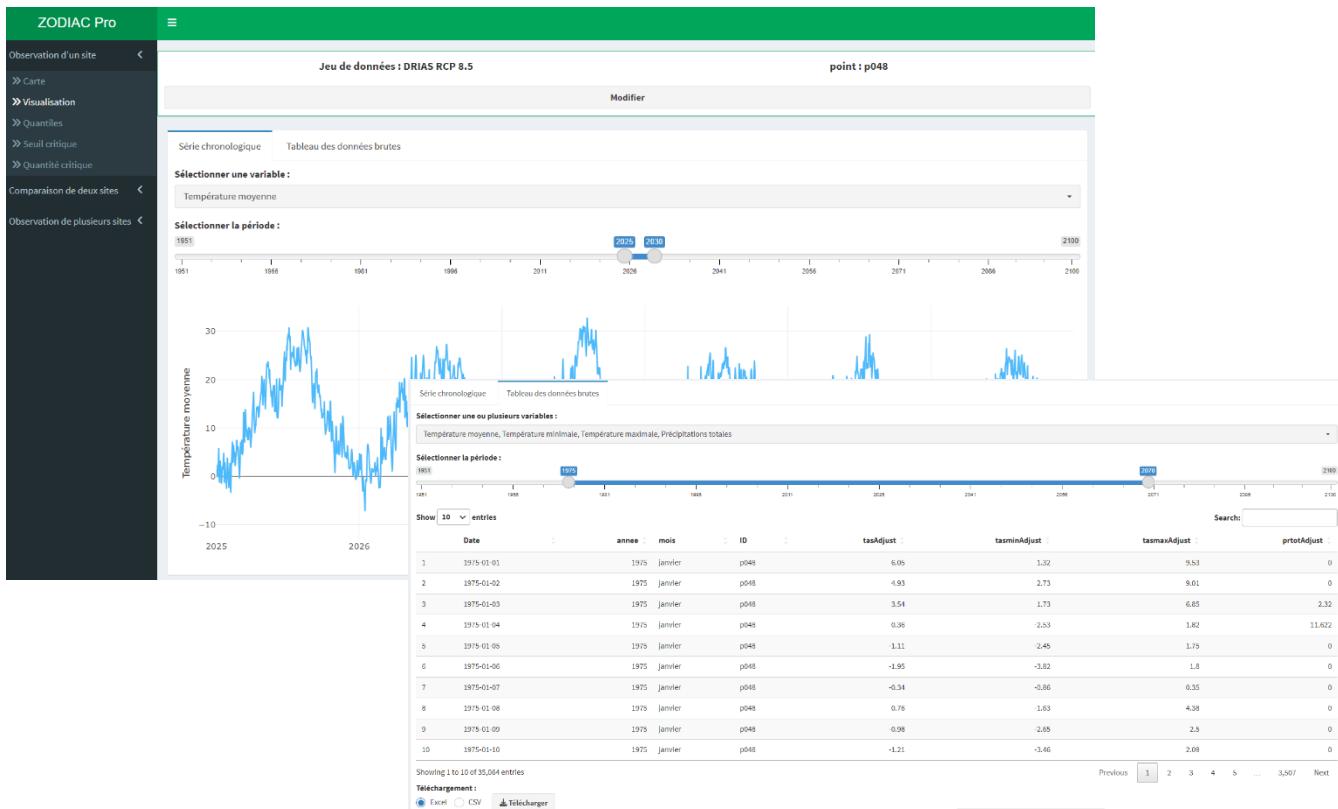
Nous avons maintenant couvert tous les éléments essentiels pour assurer le bon fonctionnement de l'application ZODIAC Pro. Il ne nous reste plus qu'à exposer l'ensemble des fonctionnalités disponibles. Dans cette optique, nous élaborons un guide qui permettra une utilisation simplifiée de l'application. Voyons ensemble le corps de chaque partie présenté précédemment. Procédons dans l'ordre, et débutons avec l'observation d'un site.

A. Observation d'un site

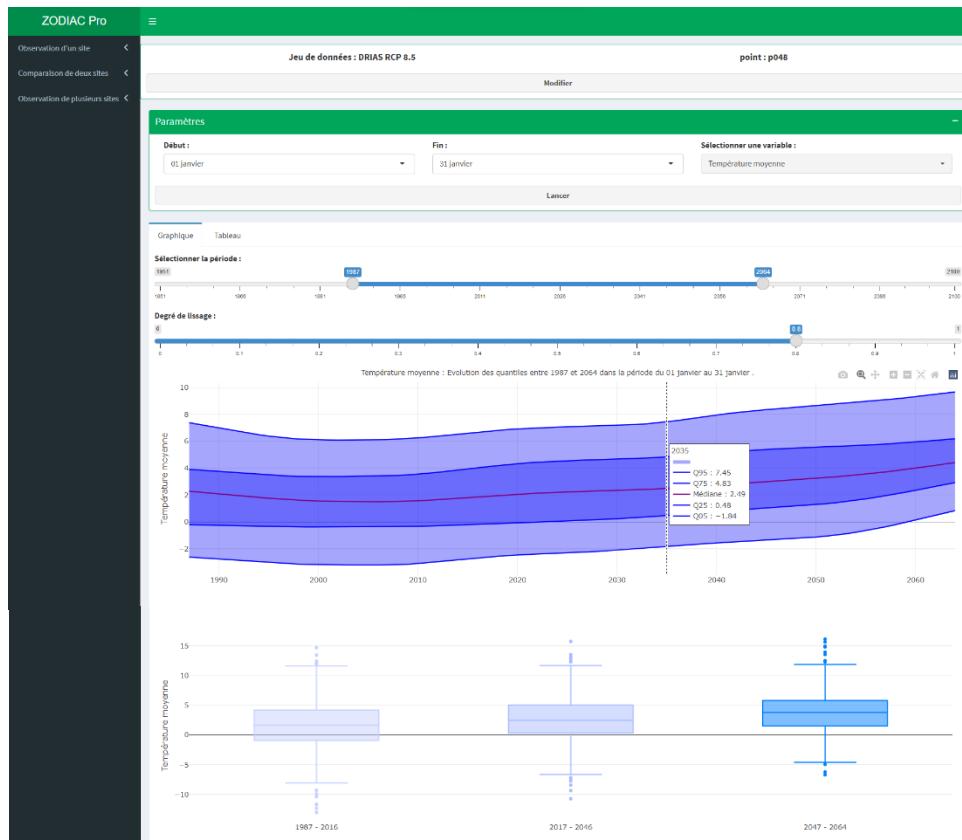
Cinq pages composent cette partie, regardons le contenu de chacune d'entre elles.



La première est la section *carte*. Elle permet la sélection d'un point DRIAS pour la visualisation des IAC. Tout d'abord, il est nécessaire de choisir un jeu de données, puis de cliquer sur le point que l'on souhaite analyser. Ce point deviendra alors rouge. En haut de la page, un résumé du jeu de données et du point sélectionné est affiché. Si aucun point n'est choisi, un message en rouge apparaît, invitant l'utilisateur à faire une sélection. Ce panel en haut de page est présent pour toutes les sections, afin de toujours se souvenir des données utilisées. Ensuite, il suffit de choisir le type d'observation. Nous pouvons soit appuyer sur un des boutons situé à droite, soit changer de page en utilisant le menu à gauche.



Dans la section *visualisation* des données, nous retrouvons deux panels différents. Le premier est un graphique d'une série temporelle. Il faut alors procéder à la sélection d'une variable et une période pour obtenir la visualisation. Le deuxième panel affiche le tableau de données. Là aussi, il faut procéder au choix de la période, et choisir toutes les variables que nous souhaitons afficher. Il est possible de télécharger ces données grâce au bouton "Télécharger", il faut au préalable choisir le format de sortie. Les données sont relatives au choix réalisé dans la page précédente et sont rappelées en haut de page. Le bouton "Modifier" permet de retourner à la page "Carte".

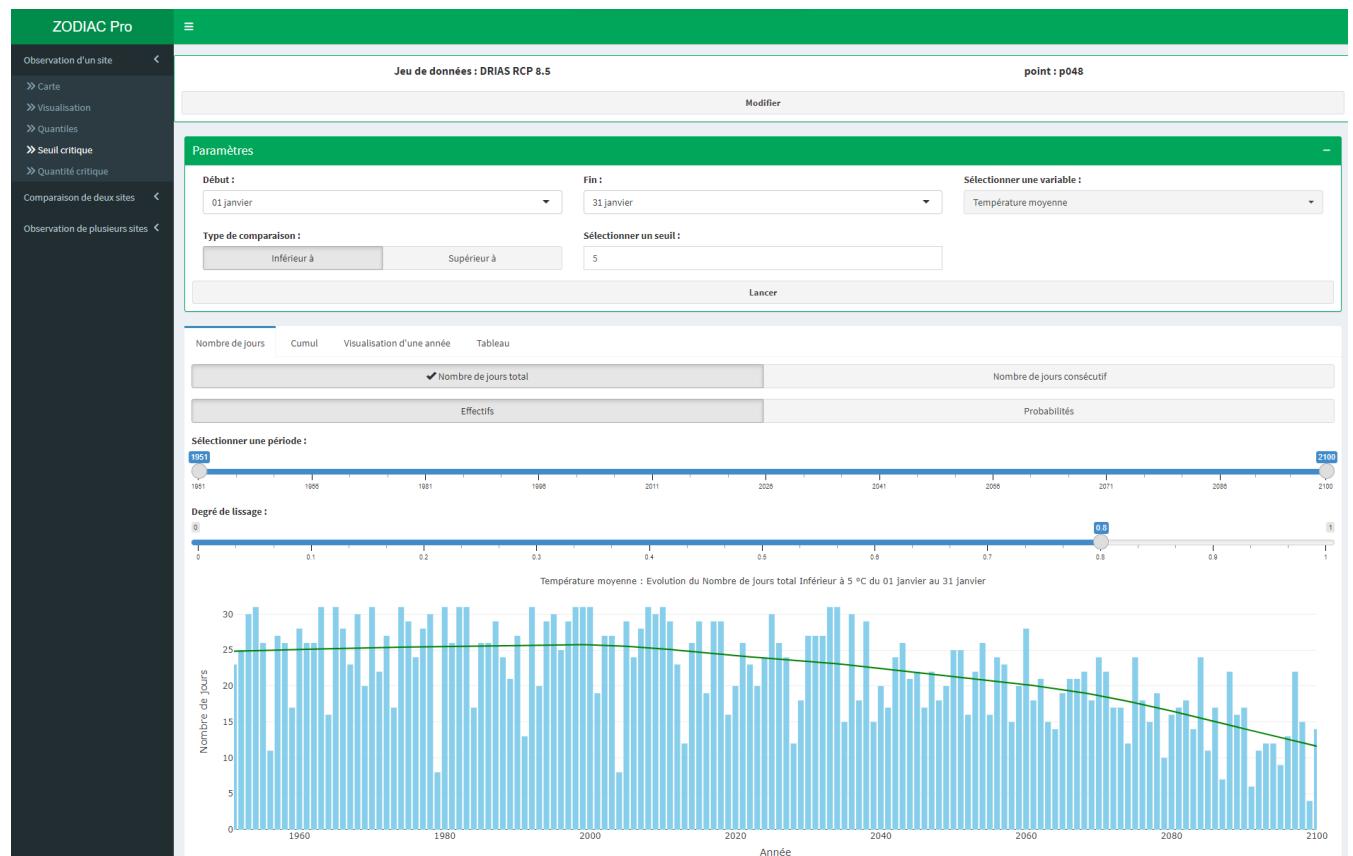


Passons aux calculs et visualisations des IAC, avec la page *quantiles*. Ici aussi, nous retrouvons 2 panels. D'abord, nous avons les deux graphiques pour représenter les quantiles. Il faut au préalable choisir une date de début, une date de fin, et une variable. Une fois ce choix réalisé, il faut cliquer sur le bouton "Lancer" pour obtenir les calculs et visualisations. Nous pouvons faire varier l'intervalle des années. Le premier graphique a une barre de survole afin d'obtenir les quantiles pour une année. Le boxplot adaptera son nombre de boîtes en fonction de la période. Il crée des ensemble regroupant 30 années au maximum. Si la période sélectionnée contient moins de 30 années, l'application réalise une unique boîte. Nous retrouvons également une barre permettant de modifier le degré de lissage pour le premier graphique.

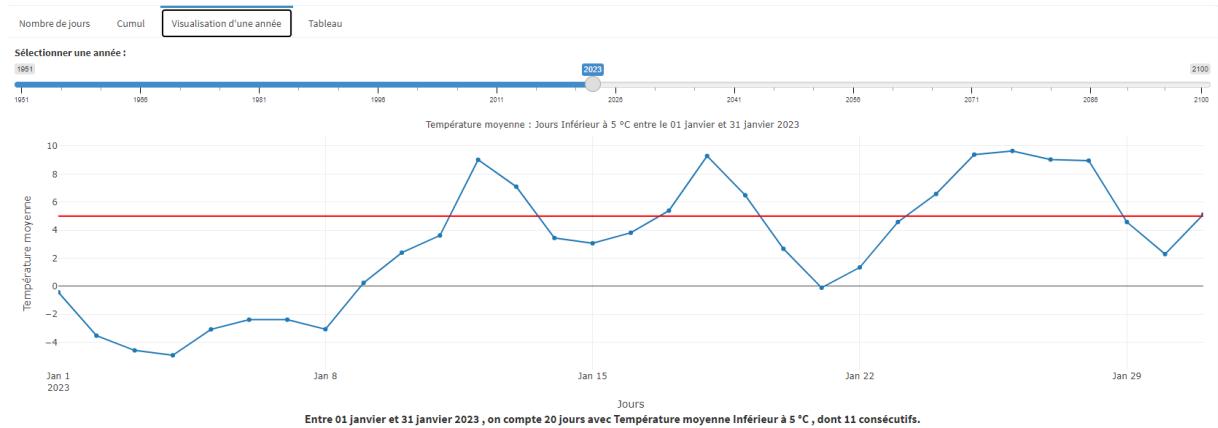
The screenshot shows a two-panel interface. The top panel is a timeline titled 'Sélectionner la période:' with markers for 1951, 1985, 2006, and 2100. Below it is a table titled 'Show 10 entries' with columns for 'année', 'Q05', 'Q25', 'Med.', 'Q75', and 'Q95'. The table lists 10 rows of data from 1951 to 1994. The bottom panel contains download buttons for Excel and CSV, and a 'Télécharger' button.

année	Q05	Q25	Med.	Q75	Q95
1951	-3.13	-0.79	2.66	6.475	8.96
1956	1.125	4.035	5.45	7	9.025
1961	-1.18	1.42	4.06	5.71	11.805
1966	0.655	3.195	5.84	8.99	13.805
1971	0.09	2.335	6.37	8.85	11.805
1976	-0.855	1.615	3.63	4.405	12.015
1981	1.465	6.38	8.61	11.07	13.095
1986	-2.015	1.89	4.26	5.18	8.095
1991	1.135	2.465	5.25	7.78	9.225
1996	0.215	2.11	3.88	5.66	8.07

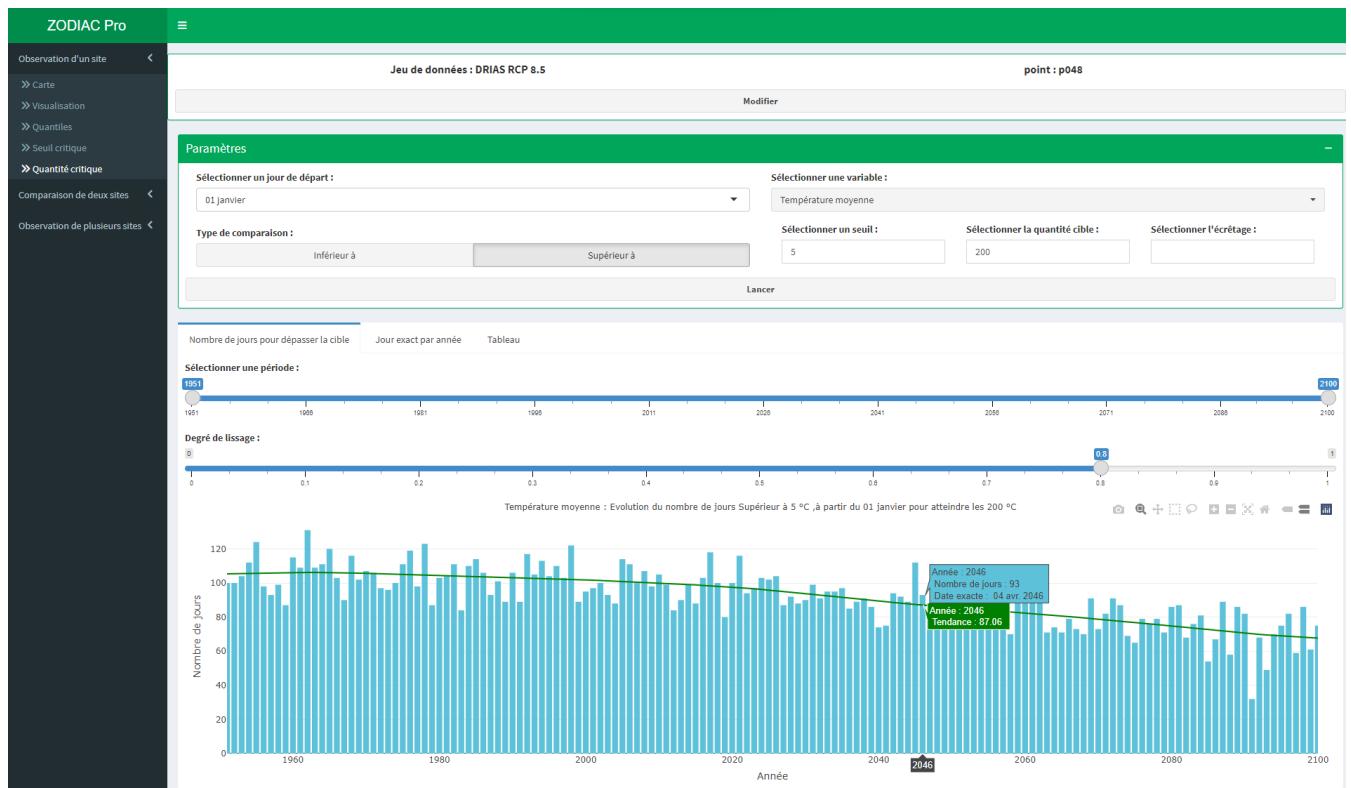
Le deuxième panel est un tableau qui nous donne les valeurs des quantiles selon les paramètres sélectionnées. La période est également modulable. Pour chaque année, nous retrouvons une valeur par quantile. De plus, un bouton de téléchargement est inclus. Il est similaire à celui observé dans la section précédente. Pour chaque catégorie d'IAC, un tableau de ce type est disponible.



Vient, ensuite, la page *seuil critique*. Nous trouvons ici quatre panels. Le nombre de jours et le cumul sont semblables dans la conception. Nous y retrouvons les graphiques des IAC respectifs, mais avant de pouvoir les visualiser, il faut sélectionner les différents paramètres et cliquer sur "Lancer". Nous choisissons alors une date de début et de fin, une variable, le type de comparaison à faire, et bien évidemment le seuil critique. Nous pouvons également faire varier les années ainsi que le degré de lissage. Pour le panel du nombre de jours, nous ajoutons la possibilité de choisir si nous voulons les jours consécutifs ou non, et si le graphique doit être affiché en effectif ou en fréquence.

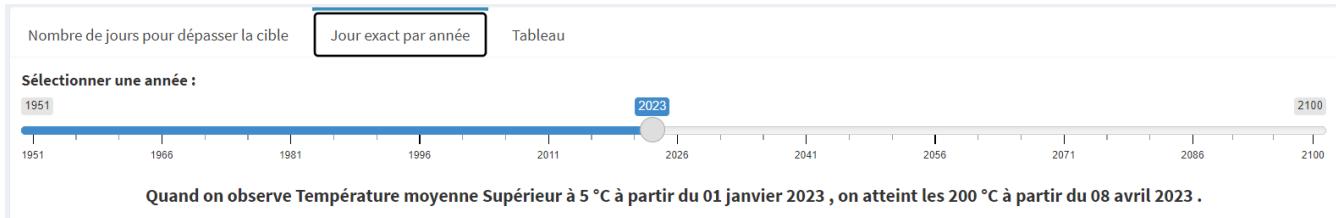


Le panel "Visualisation d'une année", affiche les données de la sélection, avec le seuil représenté par une barre horizontale rouge. Il permet de voir quels sont les jours précis où le seuil est atteint. Un message indique combien de jours se situent en dessous ou au-dessus du seuil, selon le type de comparaison choisi, et donne le nombre de jours consécutifs maximum. Le dernier panel affiche un tableau avec les valeurs des IAC, similaire au tableau des quantiles.



Pour finir avec les observations d'un site, voyons la page sur la quantité critique. Elle ressemble beaucoup à la page précédente. Ici, les paramètres à choisir sont la date de départ, la variable, le type de comparaison, le seuil critique, la quantité cible, et si nécessaire une valeur d'écrêtage. Si l'écrêtage entré n'est pas en adéquation avec le type de comparaison et le seuil, un message s'affiche indiquant

que la valeur est incorrecte. Dans cette partie, nous retrouvons trois panels. D'abord, nous retrouvons le graphique de l'IAC. Nous pouvons modifier l'intervalle d'années ainsi que le degré de lissage pour la tendance. En le survolant, nous obtenons des informations supplémentaires comme la date exacte ou la quantité cible est atteinte, et la valeur de la tendance selon le degré de lissage choisis.



Le panel "Jour exact par année", permet de sélectionner une année afin d'obtenir la date exacte ou la quantité cible est atteinte. Il renvoie un message précis avec toutes les informations pour l'année. Le dernier panel est un tableau qui nous donne l'ensemble de ces informations pour un intervalle d'année choisi.

B. Comparaison de deux sites

Passons maintenant à la deuxième grande partie de l'application, avec la comparaison de deux sites.

ZODIAC Pro

Observation d'un site

Comparaison de deux sites

Cartes

Visualisation

Comparaison statistique

Observation de plusieurs sites

Site 1 : p048 DRIAS RCP 8.5
Site 2 : Saint-Martin-d'Hères Info Climat

Sélectionner le premier site

Sélectionner un jeu de données :

DRIAS RCP 8.5 DRIAS RCP 4.5 RCA4-EC-EARTH 8.5 Info Climat

Sélectionner le deuxième site

Sélectionner un jeu de données :

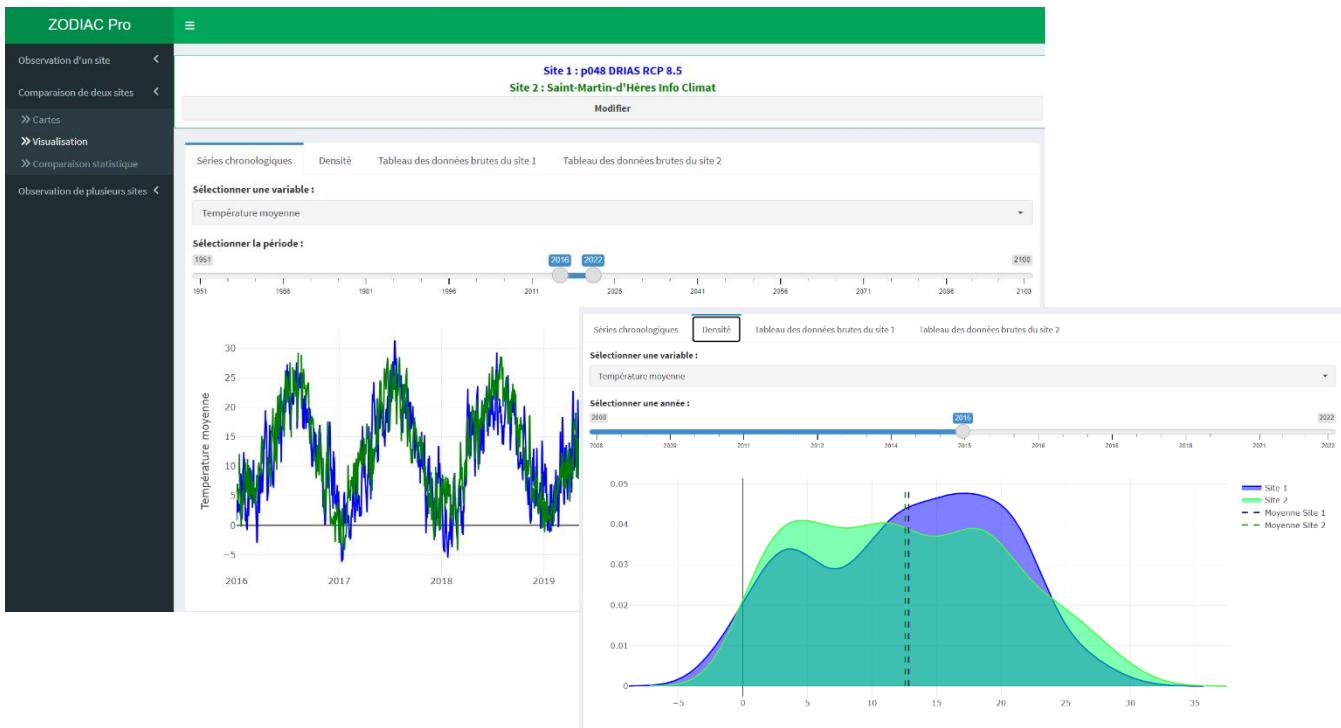
DRIAS RCP 8.5 DRIAS RCP 4.5 RCA4-EC-EARTH 8.5 Info Climat

Sélectionner le type d'observation

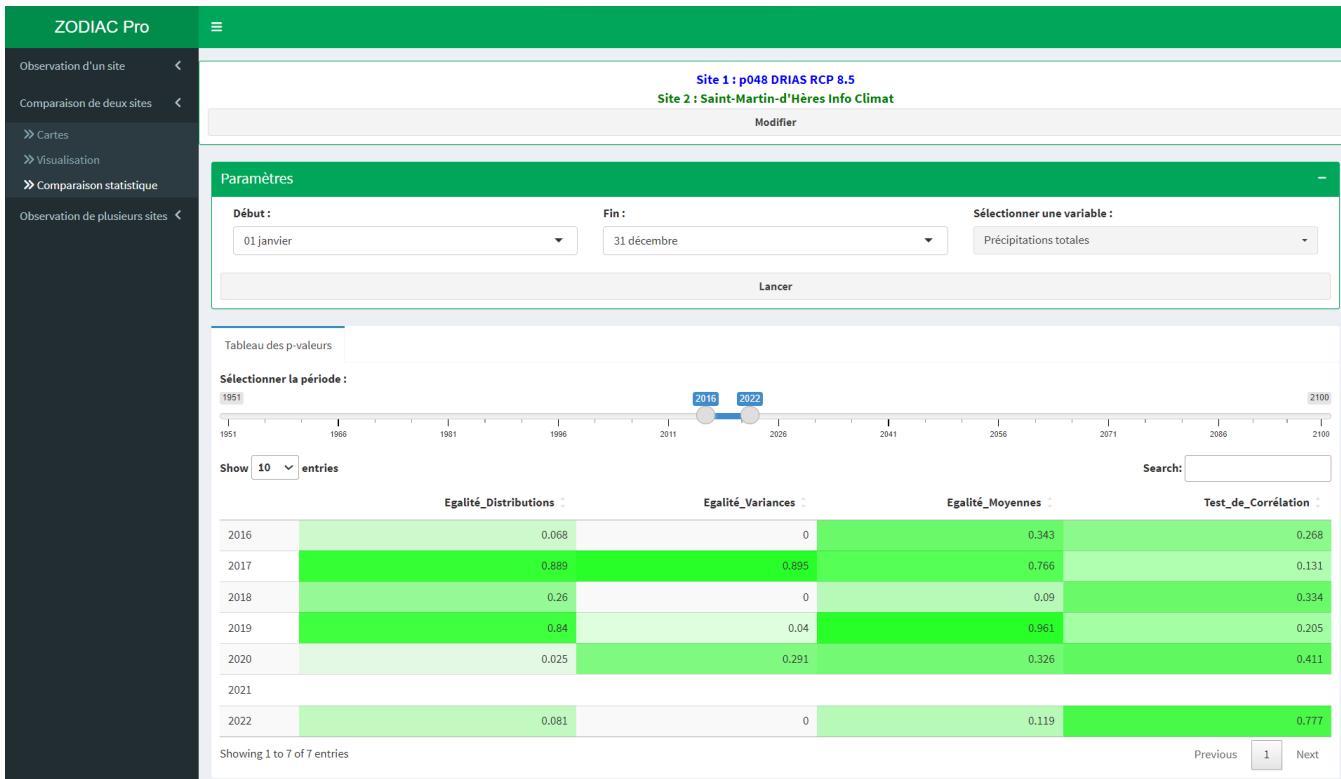
Visualisation

Comparaison des p-valeurs

Commençons avec la section *cartes*. Nous gardons le même principe que pour l'observation d'un site. Nous avons ici deux cartes. Chaque carte permet de sélectionner un jeu de données ainsi qu'un point qui deviendra rouge sur la carte lorsqu'il est choisi. Le site 1 sera identifié par la couleur bleue, et le site 2 par du vert. En haut de page, nous résumons les sélections, si aucun choix n'est réalisé un message rouge apparaît, il stipulera de faire un choix. Nous pouvons maintenant accéder aux observations, soit avec les boutons en bas de page, ou bien avec le menu à gauche.



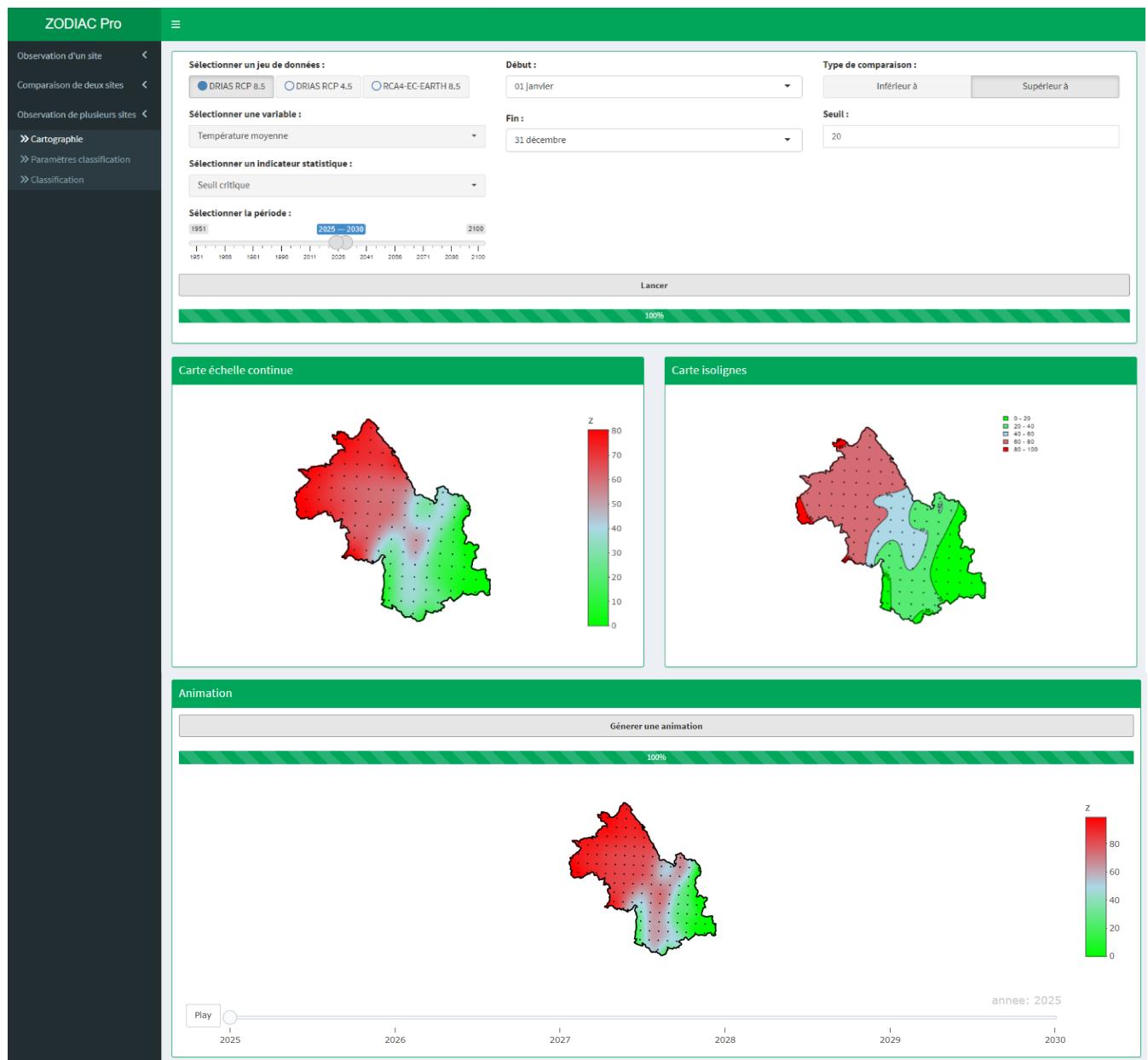
La section *visualisation* est exactement la même que pour l'observation d'un site. En haut de page, nous rappelons les paramètres sélectionnés grâce aux cartes, nous pouvons retourner à la page de sélection grâce au bouton "Modifier". Nous retrouvons ensuite quatre panels, les deux premiers sont les graphiques de série chronologique et de distribution. Les deux autres correspondent aux données de la sélection, présentées exactement comme pour les données pour un site.



La section *comparaison statistique* permet l'affiche du tableau des tests statistiques. Les deux jeux de données étant déjà définis, il ne manque plus qu'à choisir une date de début et de fin, et une variable pour la comparaison. En appuyant sur "Lancer", nous affichons le tableau avec les résultats des tests. Une ligne vide signifie qu'il n'y avait pas assez de données conformes pour réaliser les tests.

C. Observation de plusieurs sites

Pour finir, il ne nous reste plus qu'à découvrir la partie sur le zonage géographique et nous aurons fait le tour des fonctionnalités disponibles dans l'application ZODIAC Pro.



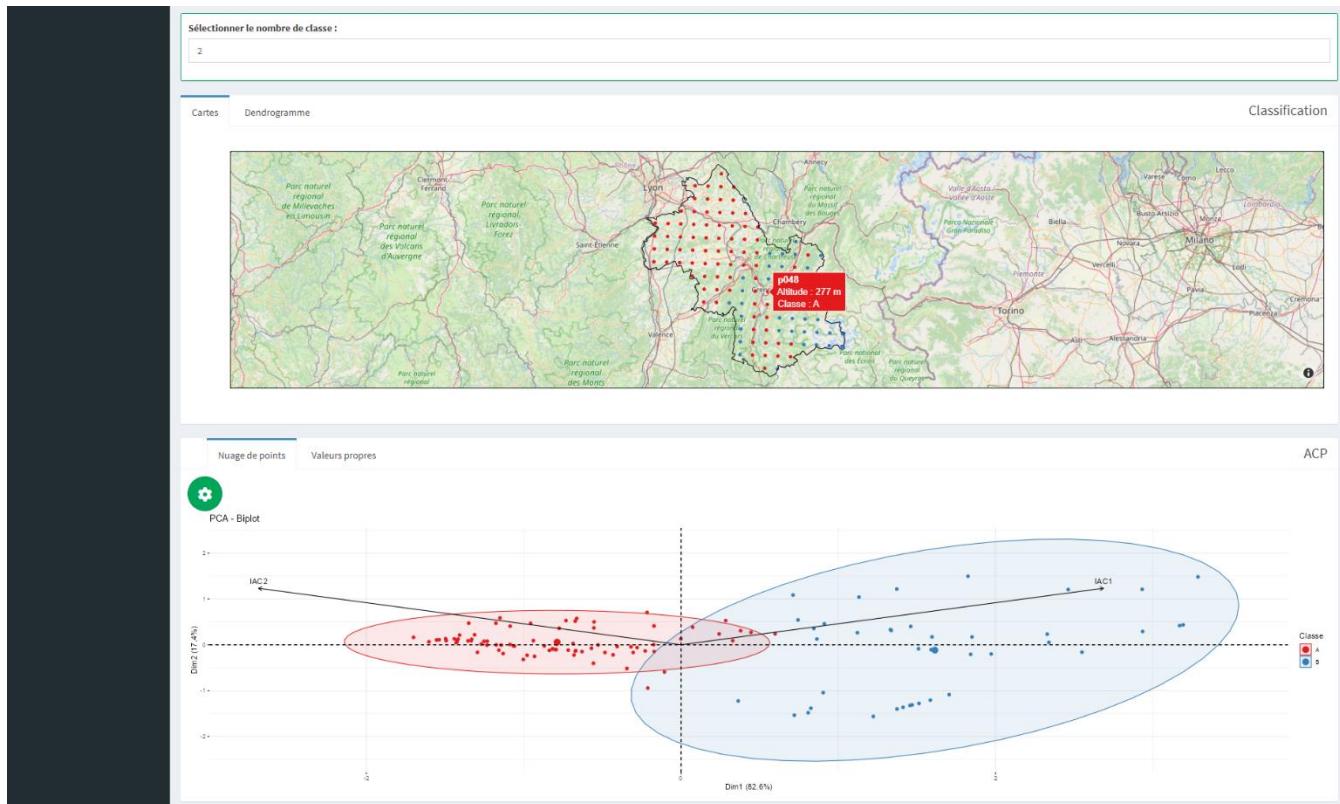
Voici la section *cartographie* dédiée à l'interpolation spatiale. En tête de page, nous trouvons les divers paramètres à entrer, les principaux étant : le jeu de données, la variable, l'IAC et la période. Des paramètres additionnels sont inclus en fonction de l'indicateur statistique choisi. Dans cette illustration, par exemple, le seuil critique est sélectionné, ce qui dévoile les paramètres liés à cette catégorie d'IAC. Ensuite, les deux cartes d'interpolation sont affichées. Nous ajoutons la possibilité de générer une carte animée. Cette animation produit une carte à échelle continue pour chaque année appartenant à la période choisie. Attention, plus la période englobe d'années, plus la génération de l'animation est longue.

IAC	Variable	Indicateur stat	Début	Fin	Type de comparaison	Seuil	Cible	écartage
IAC1	Precipitations totales	Somme	01 janvier	31 décembre				
IAC2	Température maximale	Seuil critique	01 janvier	31 décembre	Supérieur à	25		

Voici la page dédiée à la configuration des paramètres de classification. En haut de page, les principaux paramètres sont affichés, comprenant la sélection du jeu de données, de la période et du nombre d'IAC à calculer. La section de sélection des IAC s'adapte alors à ce nombre. Dans l'exemple donné, une classification basée sur deux IAC a été choisie, mais ce nombre pourrait être plus élevé. En bas de la page, un tableau récapitulatif des choix effectués est présent, accompagné d'un bouton "Lancer" qui redirige vers la section de classification et déclenche les calculs.

ID	classe	IAC2	IAC1
1	A	44	1120.554
2	B	14.2	1253.425
3	B	10.4	1605.109
4	A	58.4	1035.928
5	A	65.9	1010.762
6	A	38.7	1150.242
7	A	49.5	1102.41
8	B	12.5	1580.134
9	A	58.3	1061.124
10	A	59.5	1056.943

Pour la dernière page de l'application, nous la présenterons en deux parties. La partie haute dans un premier temps. Ici, nous retrouvons deux tableaux. Tout d'abord, nous retrouvons le tableau récapitulatif des paramètres des IAC sélectionnés, accompagné d'un bouton "Modifier" pour retourner à la page précédente. Juste au-dessus, nous affichons le jeu de données et la période choisis. Le deuxième tableau affiche les résultats du calcul des IAC pour chaque point, avec leurs classes d'appartenance.



Le bas de page quant à lui affiche les différents résultats. Nous avons inclus la possibilité de déterminer le nombre de classes à générer, suivi de l'affichage des cartes générées à partir de la classification. Lorsque nous survolons un point, des informations supplémentaires s'affichent. De plus, il est possible de visualiser le dendrogramme issu de la CAH. En dessous, le biplot de l'ACP est présenté, accompagné d'un bouton en forme d'engrenage permettant de modifier les axes à afficher. En changeant de panel, nous obtenons une visualisation des pourcentages de variance expliquée.

Nous venons de présenter l'ensemble des fonctionnalités offertes par l'application ZODIAC Pro. Notre travail touche à sa conclusion. Il est temps de faire un récapitulatif de tout ce que nous avons accompli pendant ce projet.

V. CONCLUSION

Après trois mois de travail, nous arrivons au terme de ce stage. Pour rappel, l'objectif était de concevoir le prototype d'une application web pour la chambre d'agriculture de l'Isère, afin de sensibiliser le milieu agricole aux changements climatiques, et ainsi de prendre de meilleures décisions pour leurs productions. Afin d'élaborer une synthèse complète, nous allons structurer notre conclusion en trois parties. Dans un premier temps, nous procéderons à une synthèse de toutes les réalisations techniques accomplies. Par la suite, nous proposerons des pistes d'améliorations envisagées pour ZODIAC Pro. Enfin, nous clôturerons en dressant un bilan personnel de cette expérience.

1. Récapitulation des réalisations

Revenons sur l'ensemble des réalisations du projet. Notre travail, s'est découpé en trois parties. La première consistait à récupérer l'ensemble des données nécessaires, que ce soit les simulations DRIAS ou les données réelles récupérées sur Infoclimat. Une fois toutes les données en main, il était nécessaire de les transformer afin d'obtenir un format commun entre nos données, et ainsi de les rendre comparables. La seconde partie traitait des différents IAC et des analyses statistiques mis en œuvre. Nous avons commencé par voir comment sont définies, calculées et visualisées les IAC. Par la suite, nous nous sommes intéressé à deux types de traitements statistiques. La comparaison statistique par des graphiques et des tests d'homogénéité, et le zonage géographique avec nos cartes d'interpolation et de classification complété par un biplot résultant d'une ACP. Pour finir, nous nous sommes focalisés sur l'application ZODIAC Pro, en élaborant un guide d'utilisation qui recense l'ensemble des fonctionnalités de l'application. Le travail réalisé constitue la plus grosse partie pour l'application ZODIAC Pro, mais il ne s'agit pas de la version finale. En effet, ce n'est qu'une première version de l'application, certains points peuvent être soumis à des corrections et améliorations. Voyons quelques idées proposées, mais non-disponibles dans la version actuelle.

2. Perspectives d'amélioration

Commençons avec les données. Actuellement, l'application se limite à un seul modèle DRIAS. L'inclusion de nouveau modèle, comme le propose l'application CANARI, est une perspective pertinente pour l'application. Ceci permettrait d'avoir davantage d'appréciation sur les changements climatiques. Pour les données réelles, nous pouvons trouver une nouvelle source avec des informations plus complètes, mais ceci n'est pas une priorité majeure.

Les principaux changements à prévoir concernent les IAC et les analyses statistiques. Tout d'abord pour les IAC, nous avons vu trois catégories d'IAC que nous avons nous-même définis, mais certaines catégories d'IAC reste à inclure dans l'application. Par exemple, un IAC mélangeant le seuil critique et la quantité critique existe. Il consiste à introduire de la variabilité d'une année à l'autre pour les dates de début et de fin du seuil critique en utilisant la quantité critique. Pour mieux comprendre, nous pouvons effectuer des mesures de la quantité critique sur plusieurs années, ce qui nous fournit une date pour chaque année. Ensuite, pour les mêmes années, nous appliquons le seuil critique en ajustant la période de plus ou moins 10 jours par rapport à la date obtenue grâce à la quantité critique. Voici un des IAC indisponible dans la version actuelle de ZODIAC Pro. Il y a également des changements à apporter pour le calcul et la visualisation des IAC. Nous avons vu que dans la partie de zonage géographique que

nous pouvons réaliser le calcul des IAC grâce à la moyenne et à la somme. Cette option doit également être ajoutée pour l'observation d'un site. La moyenne représente l'indicateur statistique le plus reconnu et compréhensible pour un utilisateur de l'application. La somme, quant à elle, est importante pour les agriculteurs, qui l'utilisent principalement pour le calcul des précipitations dans une période. Concernant les analyses statistiques, quelques améliorations sont à prévoir notamment pour le zonage géographique. Actuellement, nous pouvons réaliser le calcul des IAC pour le seuil critique avec le nombre de jours, mais le calcul avec les cumuls n'est pas possible. Cette option est donc à mettre en place. Pour la classification, particulièrement, nous nous limitons à une seule méthode pour classer nos points. Il serait intéressant d'inclure d'autres méthodes de classification, comme par exemple les kmean, et ainsi permettre à l'utilisateur de choisir sa méthode favorite.

Pour les changements sur l'application, tout dépend des demandes de la chambre d'agriculture suivant leur ressenti lorsqu'ils l'utiliseront. Plus généralement, si l'application répond à leurs attentes, ils pourront la proposer à d'autres chambres d'agriculture françaises. Ainsi, l'application ZODIAC Pro doit pouvoir étendre son champ d'action à de nouveaux départements, et ne plus se focaliser sur le département Isérois.

3. Bilan du stage

Mon expérience de stage a été extrêmement bénéfique et enrichissante. Les enseignements que j'ai acquis et les compétences que j'ai développées ont répondu à mes attentes. Ce stage s'est avéré être une opportunité enrichissante, me permettant de plonger dans le domaine de l'agriculture tout en améliorant mes compétences en développement R et Rshiny. Les moments passés au sein du laboratoire ont été très agréables, grâce à l'environnement chaleureux et à l'équipe accueillante qui m'ont entouré. Je suis reconnaissant d'avoir été choisi pour cette expérience qui a non seulement renforcé mes compétences professionnelles, mais m'a également permis de cultiver de nouvelles relations. Je suis convaincu que l'application ZODIAC Pro sera un outil précieux pour la chambre d'agriculture, apportant des solutions concrètes et innovantes faces au défi climatique auquel elle est confrontée.

VI. ANNEXES

1. Fonction de scraping

```
```{r}
station = "STATIC0030"
data = data.frame()
date_1 = as.Date("1951-01-01")
date_2 = 0
while (date_2 < as.Date("2023-12-31")) {
 api_key = "KKEp4TqrjHTA1$Rnftd6RjapmkPPGFivk8NfpufDhG0MveyixeFW"
 date_2 = date_1 + 7
 stationID = station
 url <- paste0("https://www.infoclimat.fr/opendata/?method=get&format=csv&stations[1]", stationID,
 "&start=", date_1, "&end=", date_2, "&token=", api_key)
 donnees <- read.csv(url, sep = ";", skip = 4)[-1,]
 data = rbind(data, donnees)
 date_1 = date_2+1
}
```

```

2. Extrait des données

A. Données DRIAS

| Date | annee | mois | ID | Longitude | Latitude | Altitude | tasAdjust | tasminAdjust | tasmaxAdjust | prtotAdjust | prsnAdjust | hussAdjust | sfcWindAdjust |
|------------|-------|---------|------|-----------|----------|----------|-----------|--------------|--------------|-------------|------------|------------|---------------|
| 1951-01-01 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -6.16 | -9.43 | -2.05 | 0.290 | 0.290 | 2.528 | 3.794 |
| 1951-01-02 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -6.39 | -10.46 | -4.13 | 0.000 | 0.000 | 1.865 | 1.694 |
| 1951-01-03 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -6.39 | -10.00 | -3.70 | 0.000 | 0.000 | 1.747 | 0.462 |
| 1951-01-04 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -6.11 | -10.10 | -4.19 | 0.000 | 0.000 | 1.772 | 0.615 |
| 1951-01-05 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -7.20 | -10.78 | -4.04 | 0.000 | 0.000 | 1.632 | 1.424 |
| 1951-01-06 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -7.53 | -10.72 | -4.20 | 0.000 | 0.000 | 1.504 | 1.463 |
| 1951-01-07 | 1951 | janvier | p001 | 5.7202 | 44.7419 | 1158 | -7.35 | -11.08 | -3.42 | 0.000 | 0.000 | 1.414 | 0.848 |

| rsdsAdjust | rldsAdjust | evpsblpotAdjust | Bilan_hydrique |
|------------|------------|-----------------|----------------|
| 23.882 | 267.230 | 0.156 | 0.134 |
| 87.721 | 204.169 | 0.259 | -0.259 |
| 86.518 | 208.433 | 0.199 | -0.199 |
| 87.157 | 205.709 | 0.207 | -0.207 |
| 89.564 | 197.721 | 0.311 | -0.311 |
| 89.997 | 197.961 | 0.354 | -0.354 |
| 89.425 | 207.223 | 0.285 | -0.285 |

B. Données Infoclimat

| Date | annee | mois | ID | tasAdjust | tasminAdjust | tasmaxAdjust | prtotAdjust | sfcWindAdjust |
|------------|-------|---------|-------|-----------|--------------|--------------|-------------|---------------|
| 2018-07-16 | 2018 | juillet | 00014 | 22.928 | 15.9 | 30.9 | 6122.2 | 0.667 |
| 2016-07-07 | 2016 | juillet | 00014 | 22.829 | -17.7 | 32.3 | 5120.4 | 1.033 |
| 2018-06-07 | 2018 | juin | 00014 | 19.045 | 15.9 | 26.8 | 5113.4 | 0.167 |
| 2018-05-09 | 2018 | mai | 00014 | 17.617 | 13.1 | 25.7 | 4254.2 | 0.733 |
| 2007-08-29 | 2007 | août | 00014 | 20.000 | 14.7 | 23.4 | 449.5 | NA |
| 2007-08-07 | 2007 | août | 00014 | 19.609 | 16.2 | 23.4 | 240.1 | 4.944 |
| 2016-04-05 | 2016 | avril | 000DV | 11.821 | 9.9 | 13.3 | 172.0 | 3.375 |
| 2007-06-06 | 2007 | juin | 00014 | 19.231 | 15.8 | 25.2 | 167.3 | NA |
| 2007-02-12 | 2007 | février | 00009 | 7.694 | 4.8 | 11.1 | 166.6 | 7.560 |

3. Fonction de traitement des données Infoclimat

```

```{r}
Transformation des données infoclimat
transform_infoclimat = function(donnees) {
 df_final = data.frame()
 st_id = unique(donnees$station_id)

 donnees$dh_utc = as.Date(donnees$dh_utc)
 donnees$pluie_1h = as.numeric(donnees$pluie_1h)
 donnees$pluie_3h = as.numeric(donnees$pluie_3h)
 donnees$vent_moyen = as.numeric(donnees$vent_moyen)
 donnees$temperature = as.numeric(donnees$temperature)

 NB_tot = table(donnees$dh_utc)

 NB_val_temp = tapply(donnees$temperature,
 INDEX = donnees$dh_utc,
 FUN = function(x) sum(!is.na(x)))
 NB_val_pluie = tapply(donnees$pluie_1h,
 INDEX = donnees$dh_utc,
 FUN = function(x) sum(!is.na(x)))

 NB_val_vent = tapply(donnees$vent_moyen,
 INDEX = donnees$dh_utc,
 FUN = function(x) sum(!is.na(x)))

 # Création des variables pertinentes
 temp_max = aggregate(temperature~dh_utc+station_id, donnees,FUN = max)
 temp_min = aggregate(temperature~dh_utc+station_id, donnees,FUN = min)
 temp_moy = aggregate(temperature~dh_utc+station_id, donnees,FUN = mean)
 #pluie_3h = aggregate(pluie_3h~dh_utc+station_id, donnees,FUN = sum)

 pluie_1h =aggregate(pluie_1h~dh_utc+station_id, donnees,FUN = sum)
 vent = aggregate(vent_moyen~dh_utc+station_id, donnees,FUN = mean)

 # Création des variables pertinentes
 temp_max = aggregate(temperature~dh_utc+station_id, donnees,FUN = max)
 temp_min = aggregate(temperature~dh_utc+station_id, donnees,FUN = min)
 temp_moy = aggregate(temperature~dh_utc+station_id, donnees,FUN = mean)
 #pluie_3h = aggregate(pluie_3h~dh_utc+station_id, donnees,FUN = sum)

 pluie_1h =aggregate(pluie_1h~dh_utc+station_id, donnees,FUN = sum)
 vent = aggregate(vent_moyen~dh_utc+station_id, donnees,FUN = mean)

 temp_max[unique(which(NB_tot < 12 | NB_val_temp < 0.50*NB_tot)), 3] <-NA|
 temp_min[unique(which(NB_tot < 12 | NB_val_temp < 0.50*NB_tot)), 3] <-NA|
 temp_moy[unique(which(NB_tot < 12 | NB_val_temp < 0.50*NB_tot)), 3] <-NA|
 pluie_1h[unique(which(NB_tot < 12 | NB_val_pluie < 0.50*NB_tot)), 3] <-NA|
 vent[unique(which(NB_tot < 12 | NB_val_vent < 0.50*NB_tot)), 3] <-NA

 # Joindre
 df_meteo = left_join(temp_moy,temp_max,by = c("dh_utc","station_id"))
 df_meteo = left_join(df_meteo,temp_min,by = c("dh_utc","station_id"))
 #df_meteo = left_join(df_meteo,pluie_3h,by = c("dh_utc","station_id"))
 df_meteo = left_join(df_meteo,pluie_1h,by = c("dh_utc","station_id"))
 df_meteo = left_join(df_meteo,vent,by = c("dh_utc","station_id"))

 # Crées année et mois
 df_meteo = transform(df_meteo , annee = format(as.Date(df_meteo$dh_utc), "%Y"))
 df_meteo = transform(df_meteo , mois = format(as.Date(df_meteo$dh_utc), "%B"))
 df_meteo$annee = as.numeric(df_meteo$annee)

 # Réorganisation
 colnames(df_meteo) = c("Date","ID","tasAdjust", "tasmaxAdjust", "tasminAdjust",
 "prtotAdjust","sfcWindAdjust", "annee","mois")
 df_meteo = df_meteo %>% relocate(Date, annee, mois, ID,
 tasAdjust, tasminAdjust,tasmaxAdjust ,
 prtotAdjust,sfcWindAdjust)

 df_final = rbind(df_final,df_meteo)

 # Supprimer les années = 2023
 df_final = subset(df_final, annee != 2023)

 # Date entre 2005 et 2023
 Date <- data.frame(Date = seq.Date(min(df_final$date),max(df_final$date),
 by = "day"))

 df_final = merge(Date, df_final, all = TRUE)

 # Arrondie
 df_final$tasAdjust = round(df_final$tasAdjust, 3)
 df_final$sfcWindAdjust = round(df_final$sfcWindAdjust, 3)

 df_final$ID = st_id

 # Remet les années et mois
 df_final = transform(df_final , annee = format(as.Date(df_final$date), "%Y"))
 df_final = transform(df_final , mois = format(as.Date(df_final$date), "%B"))
 df_final$annee = as.numeric(df_final$annee)

 return(df_final)
}
```

```

4. Fonction de calcul des IAC

```

# fonctions calcul des Indicateurs Agro Climatiques (IAC) --

nbj_seuil<-function(x,comparateur,seuil){
  if(comparateur=="Supérieur à") nbj<-sum(x>seuil)
  if(comparateur=="Inférieur à") nbj<-sum(x<seuil)
  return(nbj)
}

nbj_quant<-function(x,comparateur,seuil,ecretage,cible){
  #ecretage
  if(comparateur=="supérieur à") x[x>ecretage]<-ecretage
  if(comparateur=="Inférieur à") x[x<ecretage]<-ecretage
  #écart au seuil
  y<-x-seuil
  if(comparateur=="supérieur à") y[y<0]<-0
  if(comparateur=="Inférieur à") y[y>0]<-0

  #calcul nb_j jusqu'à atteindre la cible
  if(comparateur=="supérieur à") nbj<-which(cumsum(y)>cible)
  if(comparateur=="Inférieur à") nbj<-which(cumsum(y)<=cible)
  if(length(nbj)>0) return(nbj[1]) else return(366)
}

calcul_iac<-function(df_name,ID,period,varlabel,stat,date_debut,date_fin
                      ,seuil=NULL,comparateur=NULL,depart,ecretage=NULL,cible=NULL,aggregation=TRUE){

  list_var_labels<- c("Température moyenne", "Température minimale", "Température maximale",
                      "Précipitations totales", "Chute de neige à grande échelle",
                      "Humidité", "Vitesse du vent", "Rayonnement visible", "Rayonnement infra-rouge",
                      "Evapotranspiration", "Bilan hydrique")

  df<-dico_data[[df_name]]

  list_var_names<-colnames(df)[9:19]

  varname<- list_var_names[list_var_labels==varlabel]

  if(!is.null(date_debut))if(nchar(date_debut)>0)
    date_debut<- substr(as.Date(date_debut,format="%d %B"),6,10)
  if(is.null(date_fin))if(nchar(date_fin)>0)
    date_fin<- substr(as.Date(date_fin,format="%d %B"),6,10)
  if(is.null(depart))if(nchar(depart)>0)
    depart<- substr(as.Date(depart,format="%d %B"),6,10)

  df_iac<-data.frame()

  if(stat == "Quantité critique") {
    date_debut<-depart
    date_fin <-depart
  }

  for(i in 1:length(period)){
    an<-period[i]
    cat('an= ',i,'\n')
    if(date_debut >= date_fin) {
      select1<-which(df$date %in% seq.Date(from=as.Date(paste(an,date_debut,sep="-")),
                                              to=as.Date(paste(an,"12-31",sep="-")),by=1))
      select2<-which(df$date %in% seq.Date(from=as.Date(paste(an+1,"01-01",sep="-")),
                                              to=as.Date(paste(an+1,date_fin,sep="-"))),by=1))
      data<-df[c(select1,select2),c("ID",varname)]
    } else {
      select<-which(df$date %in% seq.Date(from=as.Date(paste(an,date_debut,sep="-")),
                                              to=as.Date(paste(an,date_fin,sep="-"))),by=1))
      data<-df[select,c("ID",varname)]
    }

    if(stat=="Moyenne") iac<-aggregate(~ID,data=data,FUN=mean,na.rm=TRUE)
    if(stat=="Somme") iac<-aggregate(~ID,data=data,FUN=sum,na.rm=TRUE)
    if(stat=="Q05") iac<-aggregate(~ID,data=data,FUN=quantile,probs=0.05,na.rm=TRUE)
    if(stat=="Q25") iac<-aggregate(~ID,data=data,FUN=quantile,probs=0.25,na.rm=TRUE)
    if(stat=="Q50 (Médiane)") iac<-aggregate(~ID,data=data,FUN=quantile,probs=0.5,na.rm=TRUE)
    if(stat=="Q75") iac<-aggregate(~ID,data=data,FUN=quantile,probs=0.75,na.rm=TRUE)
    if(stat=="Q95") iac<-aggregate(~ID,data=data,FUN=quantile,probs=0.95,na.rm=TRUE)
    if(stat=="Seuil critique") iac<-aggregate(~ID,data=data,FUN=nbj_seuil,comparateur=comparateur,seuil=seuil)
    if(stat=="Quantité critique") iac<-aggregate(~ID,data=data,FUN=nbj_quant,comparateur=comparateur,seuil=seuil,ecretage=ecretage,cible=cible)

    df_iac<-rbind.data.frame(df_iac,cbind.data.frame(annee=an, ID=iac[,1], IAC=iac[,2]))
  }

  if(aggregation==TRUE) {
    df_iac<-aggregate(IAC~ID,FUN=mean,data=df_iac)
    df_iac$IAC<-round(df_iac$IAC,3)
    rownames(df_iac)<-df_iac$ID
  }
  return(df_iac)
}

```

5. Fonction des tests statistiques

```

# ----- Tableau Test stat -----
output$tableau = renderDataTable({
  annee_inf = input$annee_comp[1]
  annee_sup = input$annee_comp[2]
  input$valide3
  isolate({
    data_1 = data_compare_1()
    data_1 = data_1[c("annee", "Date", df_var[input$var_pval, 1])]

    data_2 = data_compare_2()
    data_2 = data_2[c("annee", "Date", df_var[input$var_pval, 1])]

    data_1 = filter(data_1, annee %in% c(annee_inf:annee_sup) )
    data_2 = filter(data_2, annee %in% c(annee_inf:annee_sup) )

    df_compare = data.frame()
    pv_ks = c()
    pv_var = c()
    pv_moy = c()
    pv_cor = c()

    for ( an in c(input$annee_comp[1]:input$annee_comp[2]) ) {
      ech1 = filter(data_1, annee == an)
      ech2 = filter(data_2, annee == an)

      ech1 = ech1[[ c(df_var[input$var_pval, 1]) ]]
      ech2 = ech2[[ c(df_var[input$var_pval, 1]) ]]

      if (length(ech1) != length(ech2)){
        nb_mank = max(length(ech1), length(ech2)) - min(length(ech1), length(ech2))
        if (length(ech1) < length(ech2)) {
          ech1 = c(rep(NA, time = nb_mank) , ech1)
        }
        else{
          ech2 = c(rep(NA, time = nb_mank) , ech2)
        }
      }

      # Pour gerer les échantillons avec bcp de NA
      min_val = length(ech1)/2

      if ( sum(is.na(ech1)) >= min_val || sum(is.na(ech2)) >= min_val ) {
        pv_ks = c(pv_ks, NA)
        pv_var = c(pv_var, NA)
        pv_moy = c(pv_moy, NA)
        pv_cor= c(pv_cor, NA)
      }

      else {
        # Récupère les pvalues
        pv_ks = c(pv_ks, ks.test(ech1, ech2)$p.value)
        pv_var = c(pv_var, var.test(ech1, ech2)$p.value)
        pv_moy = c(pv_moy, t.test(ech1,ech2)$p.value)
        pv_cor= c(pv_cor, cor.test(ech1,ech2, alternative = "greater")$p.value)
      }
    }
  } # Fin
  df_compare = data.frame(c(input$annee_comp[1]:input$annee_comp[2]),
    round(pv_ks,3), round(pv_var,3),
    round(pv_moy,3), round(pv_cor,3))

  colnames(df_compare) = c("Année", "Égalité_Distributions", "Égalité_Variances",
    "Égalité_Moyennes", "Test_de_Corrélation")
}

rownames(df_compare) = df_compare$Année
df_compare = df_compare[,-1]

# Mettre de la couleur
breaks <- quantile(df_compare, probs = seq(.05, .95, .05), na.rm = TRUE)
colors <- round(seq(255, 40, length.out = length(breaks) + 1), 0) %>%
  {paste0("rgb", " ", " ", 255, " ", " ")}

df_compare = datatable(df_compare) %>% formatStyle(names(df_compare),
  backgroundColor = styleInterval(breaks, colors))

return(df_compare)
})
}

```

6. Fonction d'interpolation

```
# Fonction cartographie -----
kernel = function(d,a,type) {
  w<-NULL
  if(type=="exp") w<-exp(-abs(d/a))
  if(type=="gauss") w<-exp(- (d/a)^2)
  return(w)
}

interp.gwr = function(x,y,z,grid.x,grid.y,a,type="gauss",degree=1){
  n <- length(z)
  MAP <- matrix(NA,length(grid.x),length(grid.y))
  for(i in 1:length(grid.x))for (j in 1:length(grid.y)){
    x0 <- grid.x[i]
    y0 <- grid.y[j]

    dist<-sqrt((x-x0)^2+(y-y0)^2)
    w<-kernel(dist,a,type)
    W<-diag(w)
    X<- cbind(rep(1,n),x,y)
    if(degree==0) X<-X[,1]
    A<- t(X)%% W %% X
    B<- t(X)%% W %% z
    coef<-solve(A,B)
    if( degree==0) MAP[i,j] <- coef[1]
    if( degree==1)
      MAP[i,j] <- coef[1] + coef[2]*x0 + coef[3]*y0
  }
  MAP[MAP>max(z)] = max(z)
  MAP[MAP<min(z)] = min(z)
  return(MAP)
}
```