

Année universitaire 2021-2022
07/01/2022



Projet Tutoré : Analyse des projections climatiques régionalisées de la plaine de la Limagne

DANS LE CADRE DE L'ETUDE DE L'IMPACT DU
CHANGEMENT CLIMATIQUE SUR LA
PRODUCTION CÉRÉALIÈRE

Rapport final

Etudiants
CHAPUIS Paul
DAO Tuan Linh
MCGUIGAN Scott
MOREL Alexis

Tuteur
GARAT Philippe
Commanditaires
ALLARD Vincent (INRAE)
GREGOIRE Gérard (LJK)



Table des matières

Remerciements	2
Résumé de l'étude	3
Français	3
Anglais	3
Glossaire	4
1. Introduction	5
Contexte	5
Objectifs	6
2. Présentation des données brutes	8
2.1. Présentation graphique à l'échelle globale	8
2.2. Présentation à l'échelle locale	13
3. Séries chronologiques	18
3.1 Filtrage linéaire	18
3.2 La présentation des données brutes avec les quantiles	19
4. Résumés statistiques des données et Tests d'hypothèses	21
5. Interpolation spatiale	25
6. Classification automatique de courbe	35
Conclusion	40
Annexe	41
1. Cahier des charges	41
2. Diagramme de GANTT	44
3. Indicateurs météorologiques de Clermont-Ferrand 1971-2000	45
4. Comptes rendus de réunion	47
5. Classification automatique de courbe	50

Remerciements

Nous tenons à remercier Mr ALLARD Vincent et Mr GREGOIRE Gérard, qui sont nos commanditaires et qui nous ont permis de pouvoir travailler sur ce projet.

Nous tenons à exprimer notre profonde reconnaissance à Mr GARAT Philippe qui a été notre tuteur, pour sa disponibilité et son aide.

Et finalement nous remercions les organismes responsables du logiciel Rstudio qui nous permet la mise en œuvre de notre projet.

Résumé de l'étude

Français

Le but de notre projet est la création d'outils de décision qui permettront de comprendre l'impact futur des contraintes climatiques sur la production céréalière de la plaine de Limagne

Pour répondre à notre problématique, nous utilisons un jeu de donnée de prévisions climatiques de 2006 à 2100 sur un maillage de 45 points spatiaux équidistants situés dans la plaine de Limagne.

Grâce au logiciel Rstudio, nous avons réalisé différentes fonctions : l'une permettant de créer des résumés statistiques des différents indicateurs disponibles (température, pluviométrie et vitesse du vent) et des graphiques comparatifs de nos indicateurs avec des données passées. Une fonction permettant l'interpolation spatiale selon différente méthode et indicateurs afin de pouvoir commencer à regrouper les points qui sont similaires.

Pour la suite du projet, nous avons aussi réalisé des tests de conformité entre les données réelles et les prévisions, et une fonction permettant la classification automatique grâce à une ACP afin de pouvoir plus facilement regrouper nos points spatiaux ensembles.

Des techniques de filtrage linéaire (moyenne mobile) complètent l'étude afin d'extraire les tendances et saisonnalités des séries chronologiques.

Anglais

The purpose of our project is to create decisional tools that will help to understand the future impact of climatic constraints on cereal production in the Limagne plain.

To solve our problem, we used a data set of climate forecasts from 2006 to 2100, which was divided into 45 points located in the Limagne plain (all these points are separated and each point is 8 km apart). Thanks to RStudio software, we created different functions :

The first one is calculating statistical summaries of the different indicators and is creating graph comparing our indicators with past data.

We also created a function that is using spatial interpolation to create maps in order to group all the similar points.

We also did conformity tests, and linear filtering. By using an automatic classification method PCA - Principal Component Analysis, we can group our spatial points more easily.

Glossaire

Rstudio : Logiciel de programmation utilisé pour le traitement de données et l'analyse statistique.

Interpolation spatiale : L'interpolation spatiale est le processus d'utilisation des points avec des valeurs connues pour des valeurs estimées à d'autres points inconnus. Par exemple, pour faire une carte, L'interpolation spatiale peut estimer la température à des endroits sans données enregistrées en utilisant des relevés de températures connus dans des stations météo à proximité

Analyse en composantes principales : Méthode permettant d'explorer des jeux de données multidimensionnels constitués de variables quantitatives.

Classification ascendante hiérarchique : La classification ascendante hiérarchique est dite "ascendante" car elle part d'une situation où tous les individus sont seuls dans une classe, puis sont rassemblés en classes de plus en plus grandes.

Méthode de Ward : La méthode de Ward permet de regrouper deux classes d'une partition pour obtenir une partition plus agrégée.

Dendrogramme : Graphique illustrant l'arrangement des groupes générés par une CAH.

1. Introduction

Contexte

Dans un monde où la croissance démographique est de plus en plus importante, la production agricole notamment céréalière est essentielle si on veut arriver à subvenir aux besoins de tous.

C'est un enjeu de taille avec le changement climatique, car certaines parcelles risquent de ne plus être cultivables si les conditions n'y sont plus favorables. Le problème est d'ordre mondial, lors de la COP26 qui s'est déroulée à Glasgow du 1 au 13 novembre 2021, plus de 196 pays se sont réunis afin de parvenir à un consensus sur les actions à mener pour faire face aux changements climatiques, mais il est possible qu'elles soient trop insuffisantes et qu'elles arrivent trop tard.

C'est pourquoi l'institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) de Clermont Ferrand nous a demandé d'étudier les modèles climatiques liés à la plaine de Limagne, proche de Clermont Ferrand afin d'arriver à comprendre l'impact futur des contraintes climatiques à la fois sur la variabilité génétique des céréales et aussi sur la qualité de la production céréalière de la plaine de Limagne.

Pour cela, nous travaillons sur un jeu de données contenant des indicateurs météorologiques (température, pluviométrie et vitesse du vent) sur 45 points spatiaux séparés de 8km entre eux répartis dans la plaine de Limagne.

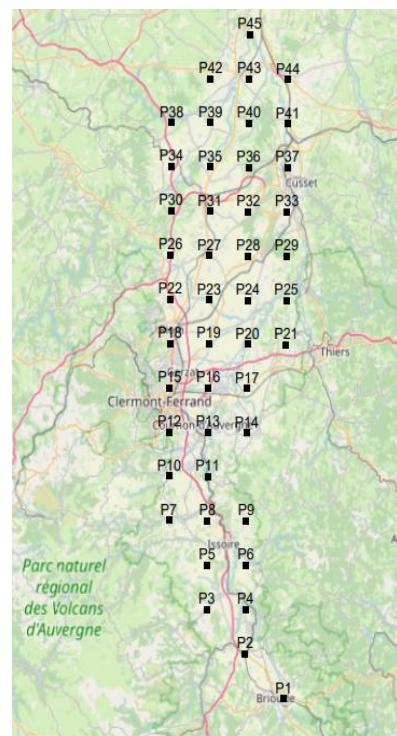


Figure 1 - Carte de la plaine de Limagne avec les 45 points spatiaux

Les données sont des prévisions de l'année 2006 à l'année 2100. Les données ont été créées à partir de scénarios d'émissions des combustibles fossiles (les RCP) et d'une douzaine de modèles climatiques régionaux (les RCM) qui sont déduits de modèles climatiques globaux (les GCM) à l'échelle de la planète. Il y a plusieurs modèles différents (américains, japonais, français, etc.), le consensus scientifique est qu'il faut utiliser tous ces modèles (pas un seul) pour quantifier la part d'incertitude liée au modèle.

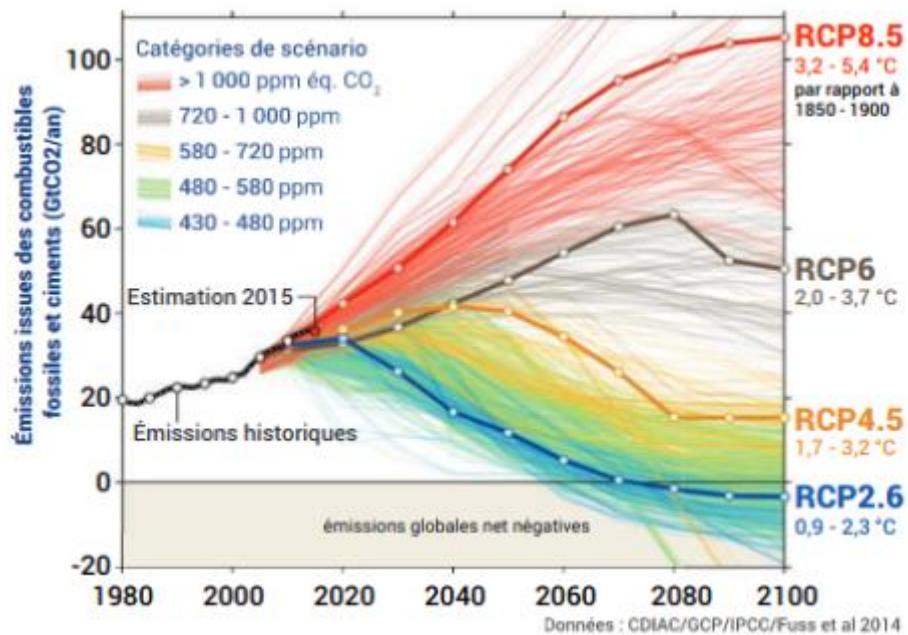


Figure 2 - Graphique des scénarios d'émissions des combustibles fossiles

Objectifs

Avec ces données, nous avons donc plusieurs objectifs :

1. Étudier la variabilité temporelle des données : déterminer une tendance haussière ou baissière des températures pour pouvoir mieux appréhender la production céréalière. Et par la suite calculer les corrélations sérielles, ce qui nous permettra de voir si les valeurs de nos variables sont indépendantes entre chaque jour, et si elles ne le sont pas, de quelle manière.
2. Comparer les prévisions que nous avons avec les données réelles : nos données de prévision vont de 2006 à 2100 et nos données climatiques de référence récoltées par Météo France vont de 1971 à 2000. On peut donc regarder comment nos variables ont évolué après plusieurs années. On pourrait également comparer nos données de prévision à des données réelles de la même année pour évaluer notre modèle de prévision.

3. Étudier la variabilité spatiale des données : détecter la présence ou non de microclimats ou une opposition entre le Nord et le Sud de la Plaine. Puis nous pourrons voir si les microclimats évoluent de la même manière.
4. Regrouper les points spatiaux similaires entre eux, ceux qui ont des trajectoires climatiques similaires et même ceux qui sont statistiquement proches. Il s'agira également de rechercher les variables les plus pertinentes pour atteindre cette classification.

2. Présentation des données brutes

Dans cette partie, nous allons expliquer le jeu de données fourni et quelles informations nous avons pu en tirer.

Notre jeu de données comprend 3 variables, la température, la pluviométrie et la vitesse du vent. Elles sont expliquées par plusieurs variables, pour la température on a la température minimale, maximale et moyenne à 2 mètres, elles sont toutes exprimées en degrés Celsius. Pour la pluviométrie, on a des précipitations totales et solides en millimètres. Et également l'humidité spécifique (en kilogramme d'eau par kilogramme d'air humide). Concernant le vent, on a la vitesse du vent à 10 mètres (en mètre par seconde).

Nom codé	Nom complet	Unité de mesure
tasminadjust	Température minimale	°C
tasmaxAdjust	Température maximale	°C
tasAdjust	Température moyenne à 2 mètres	°C
prtotAdjust	Précipitations totales	Millimètres
prsnAdjust	Précipitations solides	Millimètres
hussAdjust	Humidité spécifique à 2 mètres	Kilogramme(eau) par kilogramme(air humide)
sfcWindAdjust	Vitesse de vent à 10 mètres	Mètre par seconde

Tableau 1 - Tableau des variables de notre jeu de données

Pour cette partie nous utiliserons le point 12 qui correspond approximativement à la ville de Clermont-Ferrand.

2.1. Présentation graphique à l'échelle globale

Dans cette partie à l'échelle globale nous essayerons de mieux comprendre notre modèle prévisionnel et donc de savoir comment nos variables évoluent sur les 100 ans de données prévisionnelles.

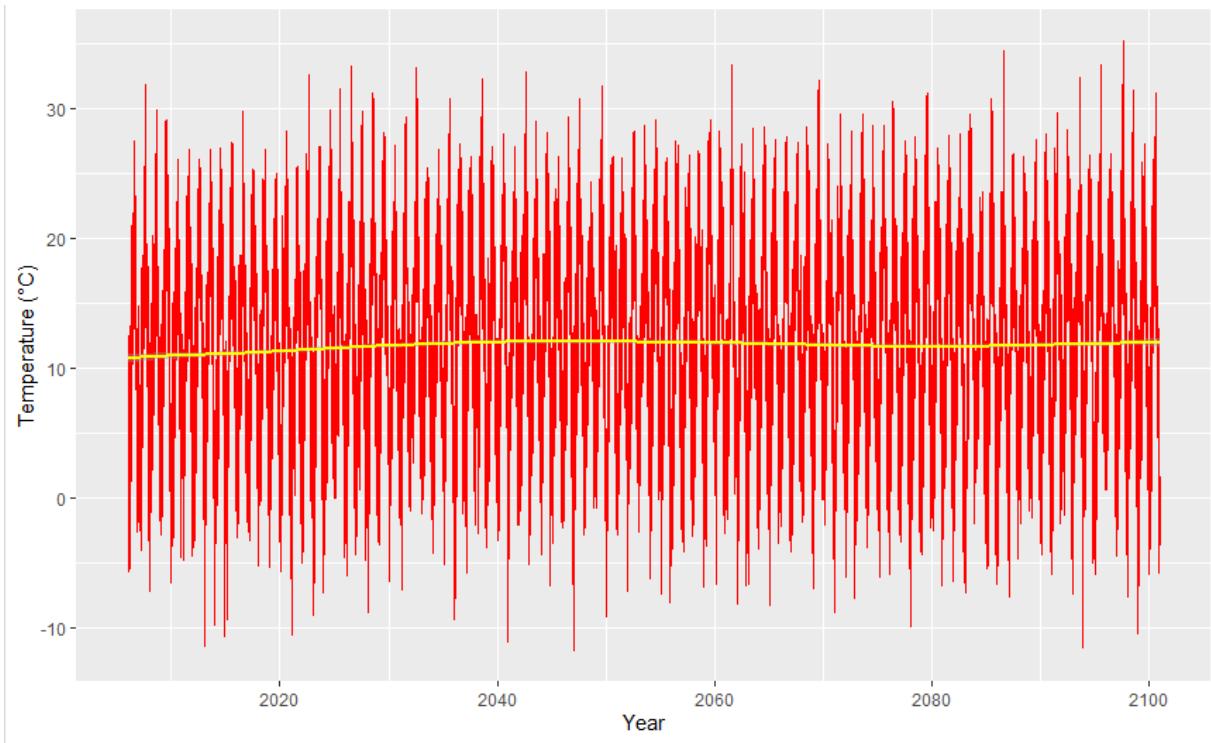


Figure 3 – Graphique représentant la prévision de la température moyenne à Clermont-Ferrand de 2006 à 2100

Ce graphique représente la prévision de la température moyenne à Clermont-Ferrand de 2006 à 2100. En jaune on peut voir la courbe de tendance, on peut voir qu'au cours des années la température moyenne augmente légèrement puis stagne pour finir par augmenter à nouveau. En presque 100 ans, la température moyenne, selon ce modèle de prévision, augmente effectivement de 1.04 degrés en 95 ans.

Au niveau de la variabilité globale des températures, il est difficile de conclure visuellement quoi que ce soit. Des statistiques descriptives et des tests sont nécessaires.



Figure 4 - Graphique de la prévision de la vitesse moyenne du vent sur 100 ans du point P12

Cette fois-ci le graphique représente la vitesse du vent de 2006 à 2100, toujours pour le point P12. On peut voir en première approximation que la courbe de tendance de notre modèle de prévision ne prévoit pas de variation pour la vitesse du vent sur 100 ans. Cela peut paraître paradoxale puisque le dérèglement climatique fait augmenter l'énergie dans l'atmosphère.

Présentation de la pluviométrie moyenne durant 95 ans du point P12

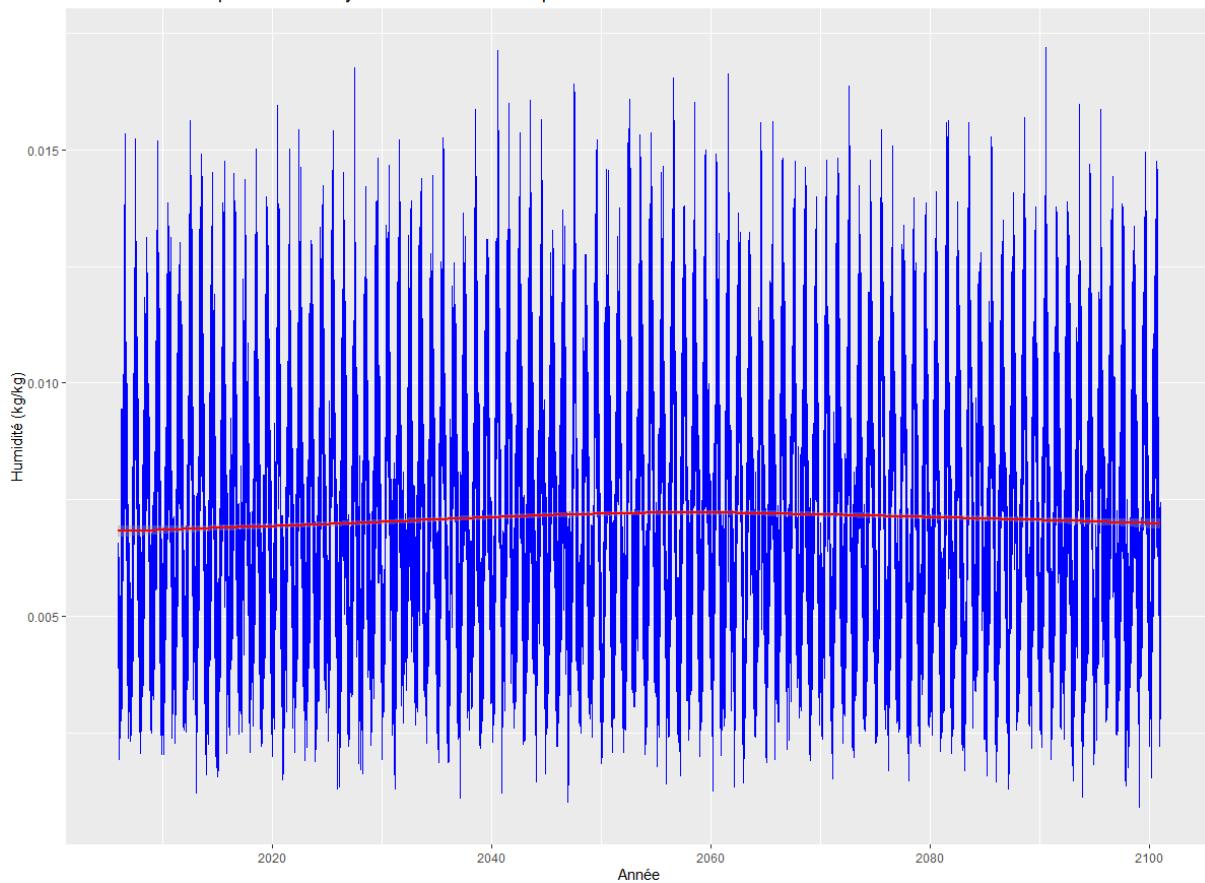


Figure 5-Graphique de la prévision de l'humidité sur 100 ans du point P12

On est encore dans le même contexte, mais ici on représente l'humidité moyenne. Notre premier aperçu est semblable à la vitesse du vent, l'humidité ne variera pas d'ici 100 ans selon nos prévisions. Étant donné qu'un air plus chaud peut contenir plus de vapeur d'eau, l'atmosphère devrait être plus humide. Notre résultat est donc surprenant mais il est en lien avec nos prévisions de la température, qui n'augmente pas significativement sur 100 ans.

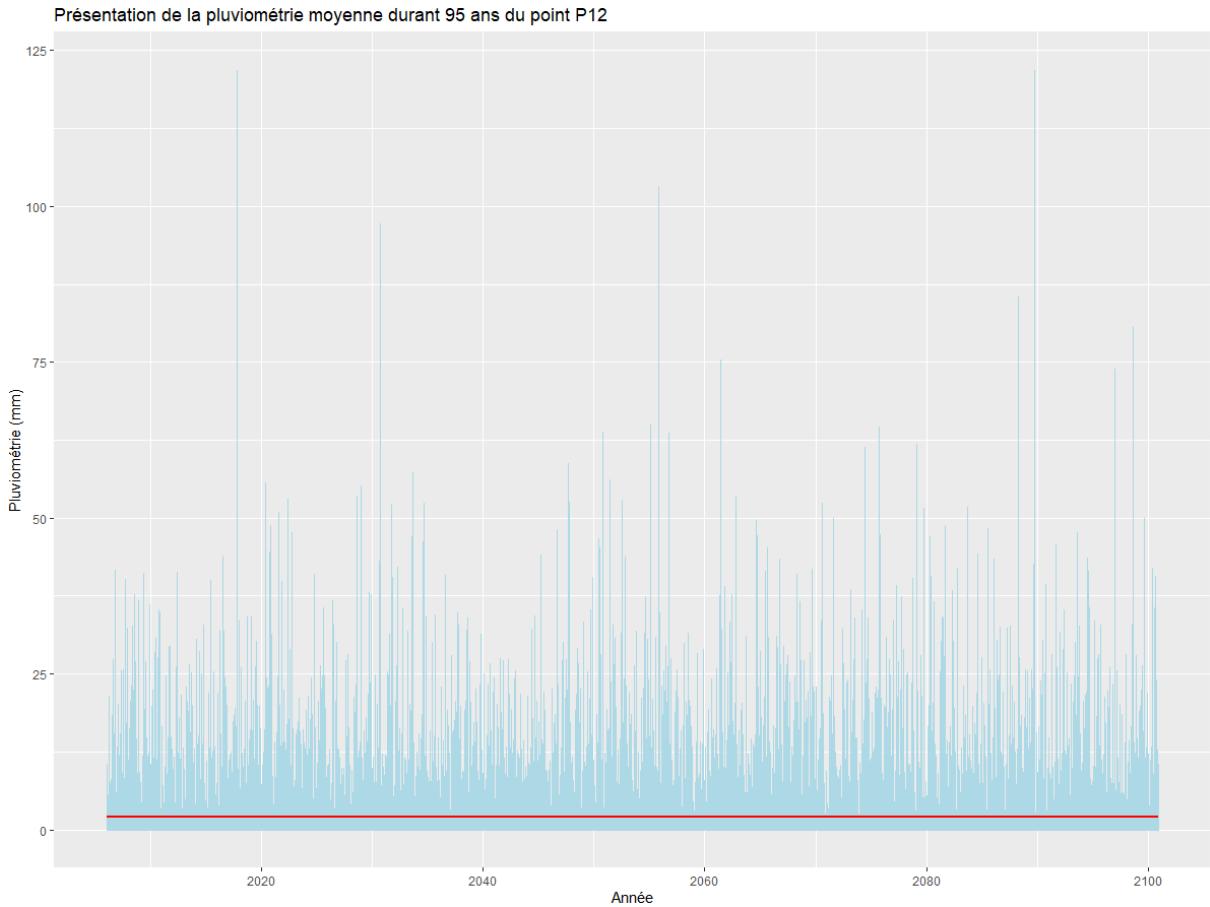


Figure 6 - Graphique de la prévision de l'humidité sur 100 ans du point P12

Sur ce graphique on est encore dans les mêmes conditions, c'est la pluviométrie qui est représentée. On peut encore une fois apercevoir avec la courbe de tendance que la pluviométrie moyenne n'a pas augmenté en 95 ans. Avec des tests de conformité il faudrait confirmer cette impression et il faudra également s'intéresser à la variance que l'on ne peut pas apercevoir de visuel.

Cette partie était présente pour que l'on ai un premier aperçu de nos variables sur les 95 ans de prévision. Et le premier aperçu n'était pas celui escompté puisque les variables ne semblent pas avoir augmenté mis à part la température qui a augmenté faiblement. Il faudra donc vérifier cela avec des tests de conformité et également analysé la variance.

2.2. Présentation à l'échelle locale

Dans cette partie, nous avons créé deux fonctions en utilisant le package “GGPLOT2” et la fonction “facet_wrap”, l'une est dédiée aux 3 types de température et nommée func_temp qui prend en entrée 3 paramètres:

- Les données brutes
- Le point spatial (il y en a 45)
- L'année choisie (de 2006 à 2100)

Cette fonction sert à visualiser 3 types de température pour chaque mois en fonction d'une année choisie.

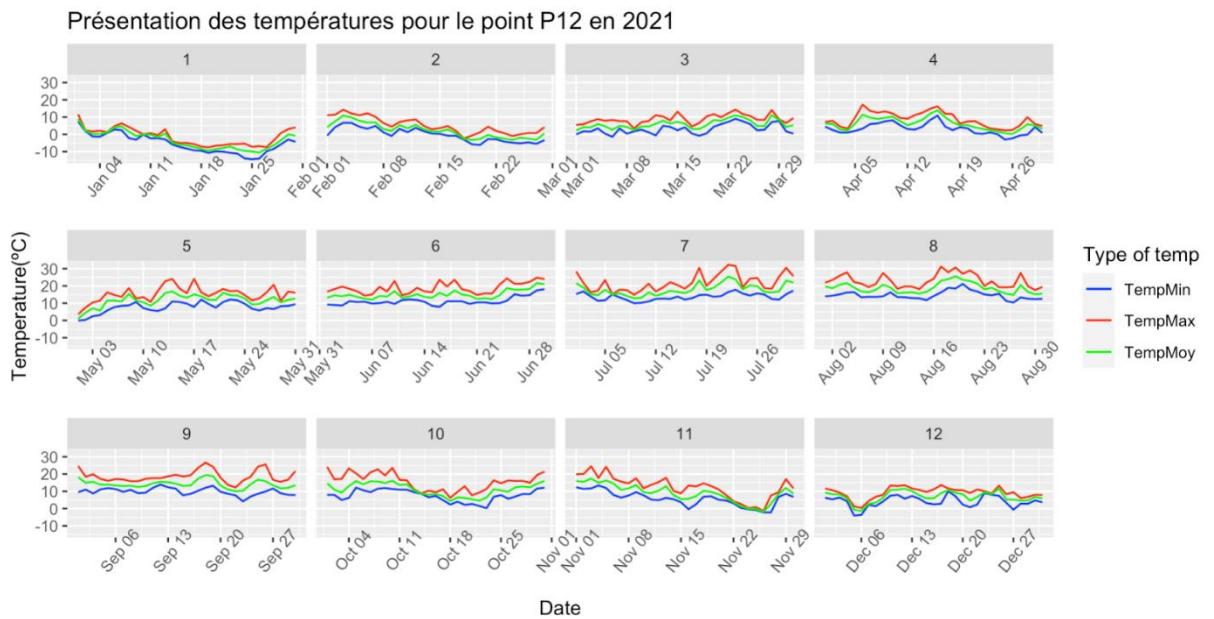


Figure 7 - Graphique des températures pour le point 12 en 2021

Par exemple, ce graphique combine 12 graphiques, chacun contient 3 courbes de températures : la température maximale (la ligne rouge), la température moyenne (la ligne verte) et la température minimale (la ligne bleue). Nous remarquons que l'évolution est non linéaire pour les 3 types de température pour l'année 2021, il y a du « bruit » et donc de l'aléatoire dans la création de nos prévisions.

L'autre est nommée func_facet qui prend en entrée 4 paramètres :

- Les données brutes
- La variable choisie (il y en a 7)
- Le point spatial (il y en a 45)
- L'année choisie (de 2006 à 2100)

Le but de cette fonction est de visualiser les données brutes (la variable choisie) pour chaque mois en fonction d'une année choisie.

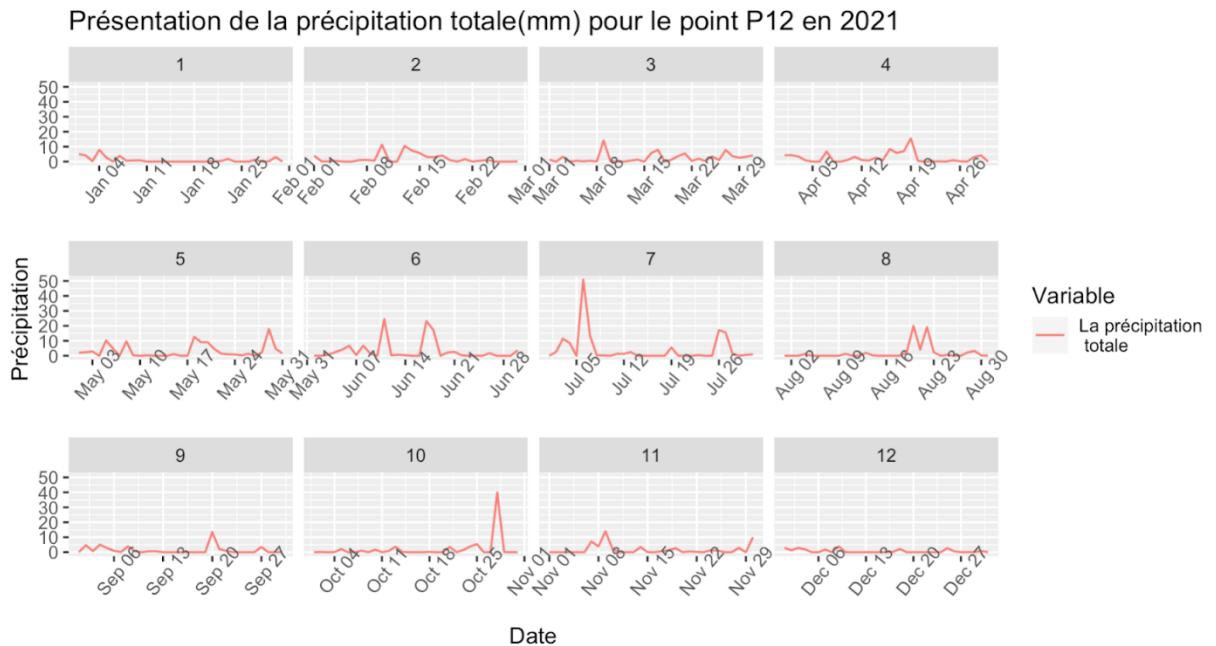
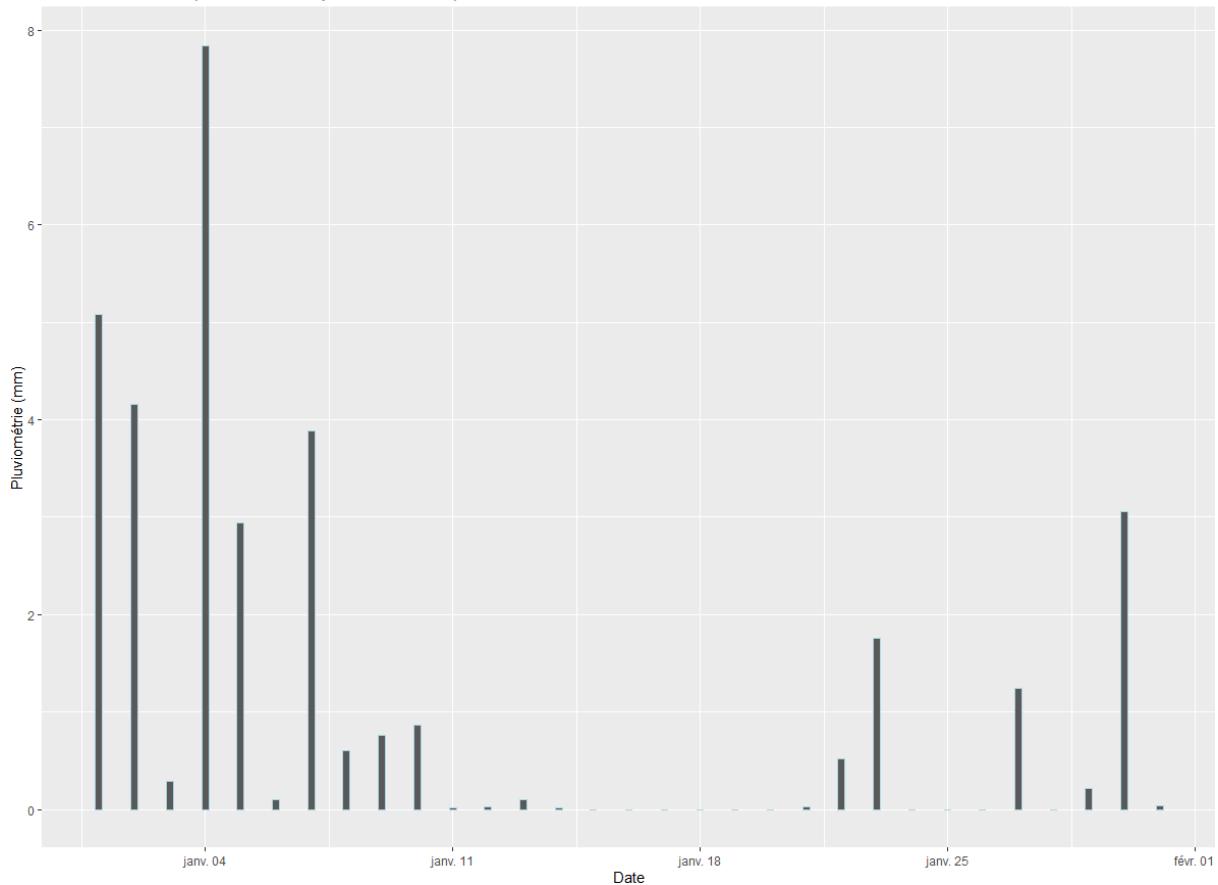


Figure 8 - Graphique de la prévision de la précipitation totale de l'année 2021 pour le point P12

Par exemple, ce graphique représente la précipitation totale pour le point P12 en 2021. Nous remarquons que l'évolution de cette variable n'est pas forte parce qu'elle est présentée sur une année. La précipitation des mois juillet, octobre est la plus grande.

Présentation de la pluviométrie en janvier 2021 du point P12



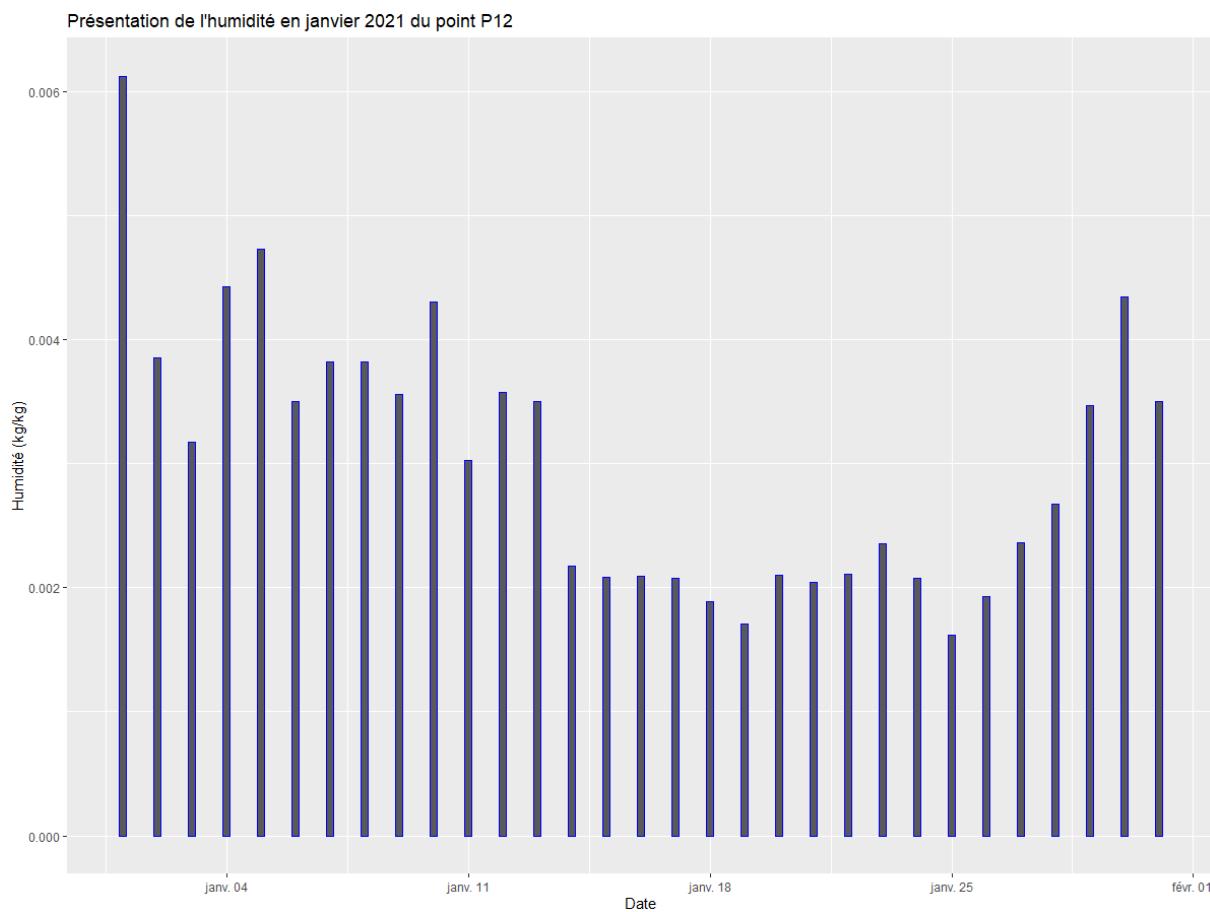


Figure 10 - Graphique de la prévision de l'humidité en janvier 2021 du point P12

Ce graphique présente l'humidité du mois de janvier en 2021, durant ce mois on n'observe pas des variations très importantes, les jours semblent liés entre eux concernant cette variable.

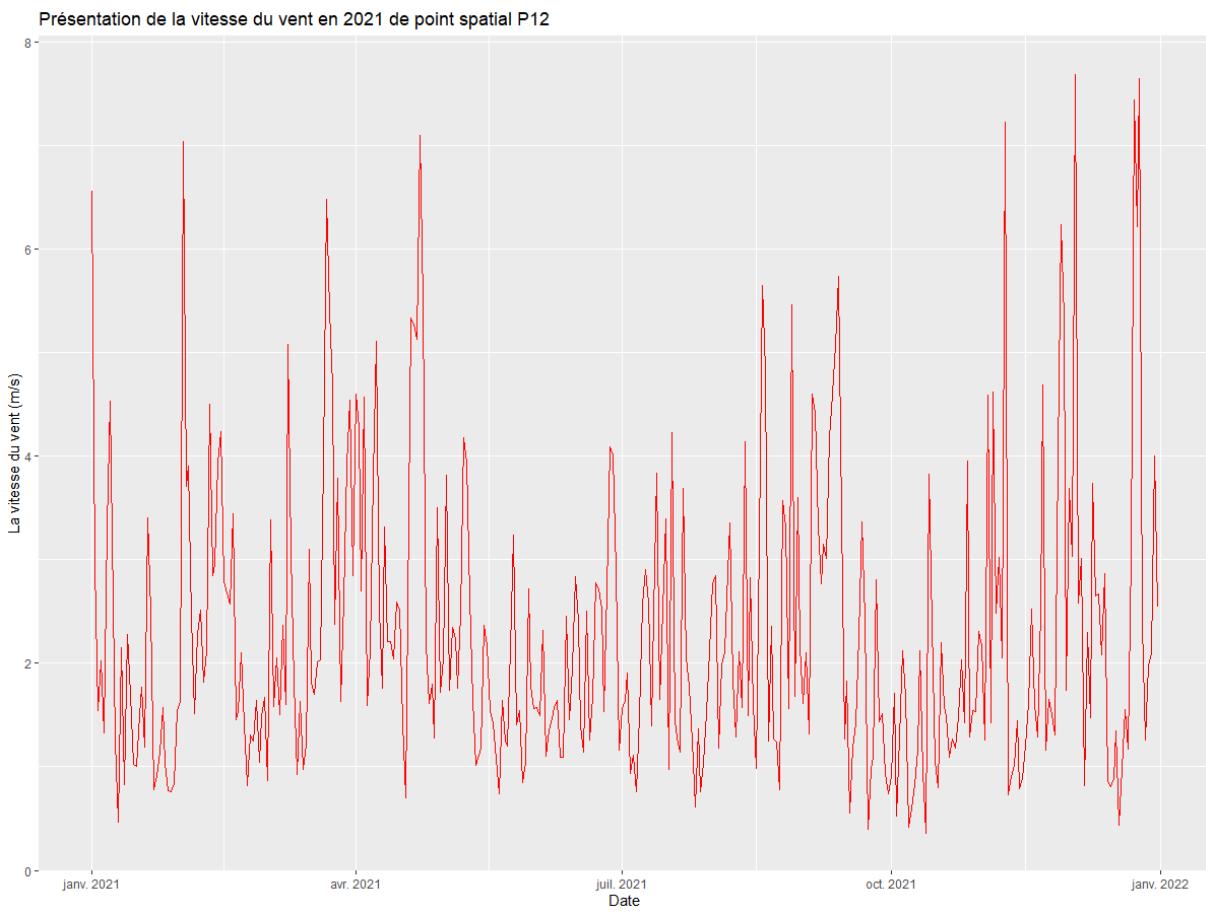


Figure 11 - Graphique de la prévision de la vitesse du vent en 2021 du point P12

Ce graphique présente l'évolution de la vitesse du vent pour l'année 2021, on constate que la vitesse du vent est plus importante en hiver est moins importante en été. On peut également voir que l'évolution n'est pas linéaire.

Tous ces graphiques à l'échelle locale nous permettent de mieux comprendre notre modèle de prévision. Et cette partie nous permet de dire qu'il y a de l'aléa dans nos prévisions. Et ils nous ont fait nous questionner sur le lien qu'il pouvait avoir entre chaque jour selon les variables.

3. Séries chronologiques

3.1 Filtrage linéaire

Dans cette partie, nous avons créé deux fonctions nommées “Filtrage_1” avec la base de R et “Filtrage_2” avec le package GGPLOT2. Les deux fonctions prennent en entrée 7 paramètres:

- Les données brutes
- La variable choisie dans les données brutes (il y en a 7)
- Le point spatial (il y en a 45)
- La date de début
- La date de fin
- Moyenne mobile d'ordre 1
- Moyenne mobile d'ordre 2

Le but est de visualiser les données brutes (la variable choisie) avec la tendance, la saisonnalité et les résidus en fonction des dates, du point spatial et les différentes moyennes mobiles arithmétiques. À la fin de la fonction, nous pouvons afficher les données filtrées par la date de début, la date de fin et le point spatial dans une nouvelle fenêtre de R.

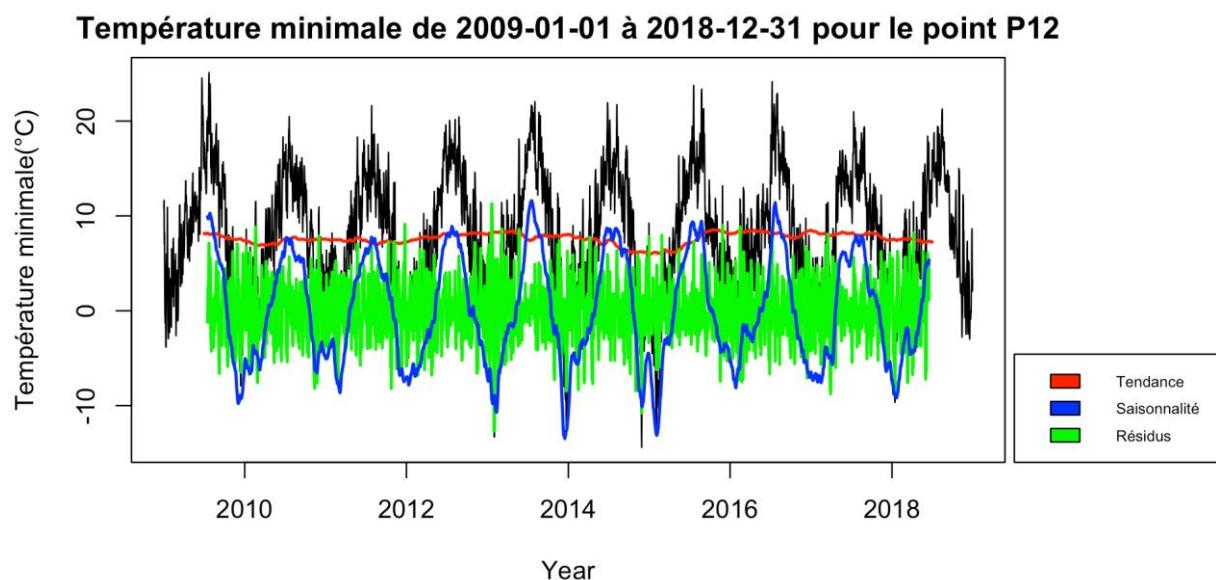


Figure 12 - Graphique de la température minimale de 01/01/2009 à 31/12/2018 pour le point spatial 12
(La version avec la base de R)

Température moyenne(°C) de 2009-01-01 à 2018-12-31
pour le point P12

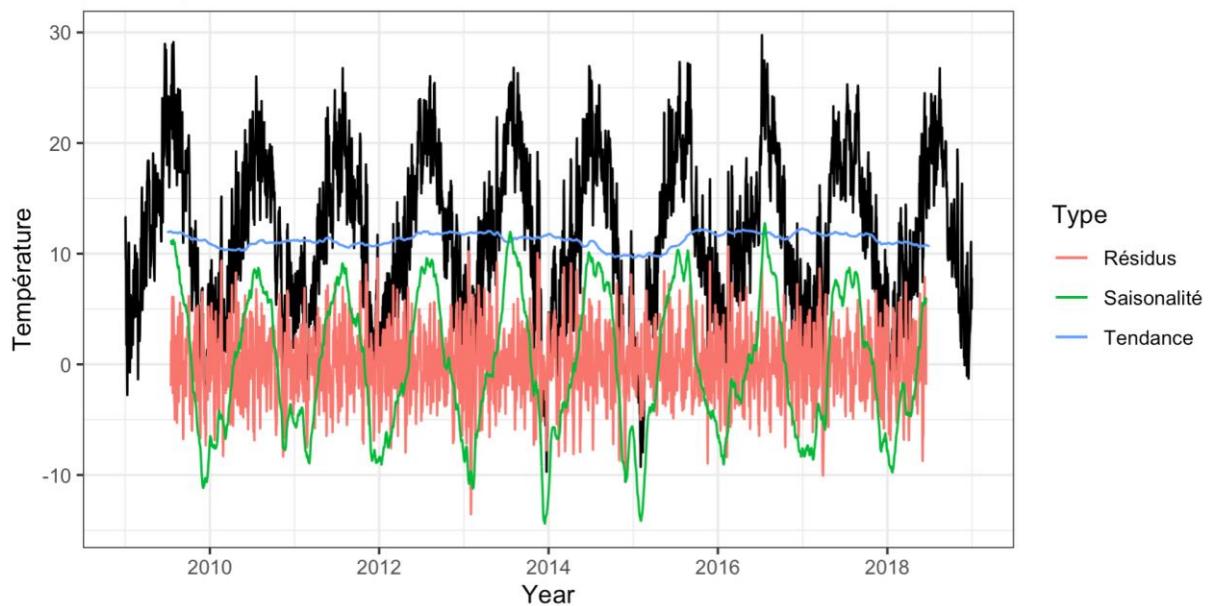


Figure 13 - Graphique de la température minimale de 01/01/2009 à 31/12/2018 pour le point spatial 12
(La version avec le package GGPLOT2)

Une série chronologique comprend 3 composantes: la saisonnalité, la tendance et les résidus. Ce graphique présente les données brutes (la variable tasminAdjust - la température moyenne) en noir, la courbe de tendance en rouge, la courbe de saisonnalité en bleu, la courbe des résidus en vert. Ici la tendance n'est pas linéaire, elle est assez stable. La saisonnalité correspond à la répétition d'un phénomène à intervalle de temps (de 01/01/2009 à 31/12/2018). Après l'élimination de la tendance et la saisonnalité, nous obtenons les résidus qui correspondent à des fluctuations irrégulières.

3.2 La présentation des données brutes avec les quantiles

Pour cette partie, nous avons créé une fonction nommée “func_quantile” qui prend en entrée 5 paramètres:

- Les données brutes
- La variable choisie dans les données brutes (il y en a 7)
- La date de début
- La date de fin
- Le point spatial (il y en a 45)

Le but est de visualiser les données brutes (la variable choisie) en ajoutant 5 courbes des quantiles (avec les probabilités personnalisées). À la fin de la fonction, nous pouvons afficher

les données filtrées par la date de début, la date de fin et le point spatial dans une nouvelle fenêtre de R.

Température moyenne($^{\circ}$ C) de 2021-02-01 à 2022-12-31 pour le point P12

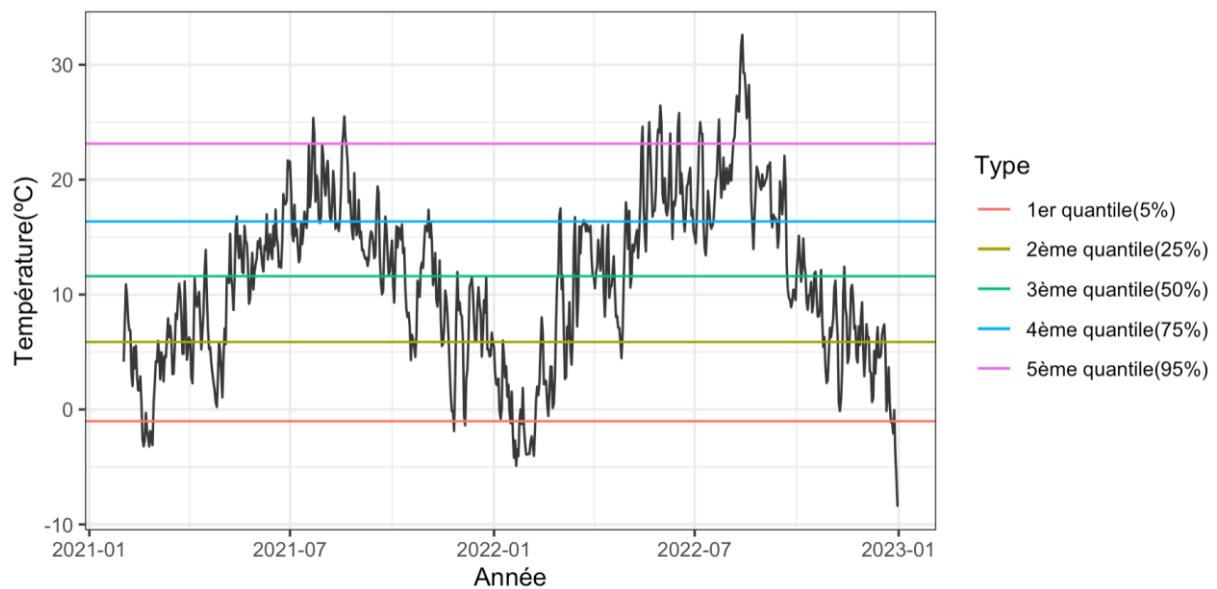


Figure 13 - Graphique de la température moyenne de 01/02/2021 à 31/12/2022 pour le point spatial 12

Ce graphique représente la température moyenne de 01/02/2021 à 31/12/2022. Les quantiles sont les courbes horizontales. Par exemple, si nous calculons le quantile de la température moyenne avec la probabilité de 5% et nous obtenons 0.676, c'est-à-dire environ 5% de toutes les températures moyennes ont des valeurs de -0.13 $^{\circ}$ C ou moins. Nous pouvons voir clairement l'évolution de la température moyenne des mois et les comparer avec les courbes des quantiles.

4. Résumés statistiques des données et Tests d'hypothèses

Dans cette partie, nous avons créé une fonction nommée GraphComparatifs qui en choisissant différents paramètres qui sont : l'année de début, l'année de fin, la variable étudiée et les points spatiaux que l'on veut comparer, permet d'avoir différents résultats.

Elle permet dans un 1^{er} temps de tracer un tableau avec les différentes moyennes de la statistique climatique choisie en fonction du mois et du point spatial comme on peut le voir dans le Tableau 2. Nous avons aussi la ligne « Moyennepassé » qui contient les données sur la température moyenne que nous avons sur la plaine de Limagne de 1971 à 2000 que nous avons mis en annexe. Nous avons des données sur la température moyenne, la pluviométrie et la vitesse du vent. Il est possible de changer ces données afin d'avoir la possibilité de comparer notre échantillon avec d'autre données passées.

Ensuite nous avons calculé les écart-types pour chaque point et n qui est la taille totale de l'échantillon : ici la somme des jours de chaque mois de 2006 à 2035.

	01	02	03	04	05	06	07	08	09	10	11	12
mean.P1	3.28	3.93	7.30	9.83	14.19	17.14	21.27	20.99	16.13	11.47	7.37	4.51
mean.P12	3.16	3.96	7.23	9.76	14.10	16.86	21.02	20.75	16.04	11.25	7.27	4.57
mean.P45	3.66	4.59	8.18	10.67	15.09	17.75	21.86	21.51	16.74	11.71	7.61	4.83
Moyennepassé	-0.30	0.50	2.20	4.10	8.20	11.20	13.40	13.20	10.30	7.10	2.60	0.70
sd.P1	4.48	4.29	4.45	3.91	4.24	3.40	3.65	3.71	3.92	3.84	4.64	4.15
sd.P12	4.61	4.47	4.58	3.95	4.08	3.43	3.86	3.91	3.61	3.72	4.64	4.29
sd.P45	4.51	4.35	4.34	3.85	3.95	3.30	3.82	3.81	3.67	3.76	4.58	4.39
n	899.00	819.00	899.00	870.00	899.00	870.00	899.00	899.00	870.00	899.00	870.00	899.00

Tableau 2-Tableau des températures moyennes pour chaque mois des points 1,12,45 de l'année 2006 à 2035

Ici par exemple, nous avons calculé les températures moyennes pour chaque mois des points spatiaux 1, 12 et 45 de l'année 2006 à 2035. Nous choisissons le point 1 car c'est celui situé le plus au Sud, le 12 car c'est celui situé à Clermont Ferrand et le 45 car c'est celui le plus au Nord. Cela pourrait nous permettre de déterminer s'il y a une opposition Nord-Sud.

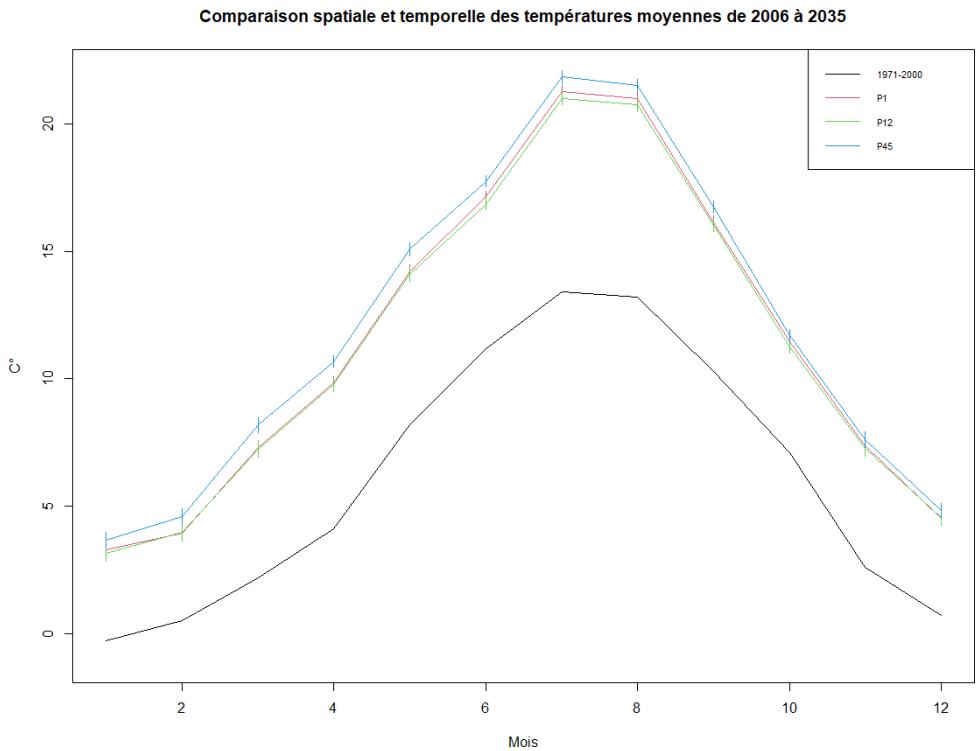


Figure 14 -Graphique de la moyenne des températures moyennes des points 1,12 et 45 et des données passées.

Ensuite notre fonction trace un graphique de la variable moyenne pour chaque mois des points choisi et nous permet de comparer avec les données passées que nous avons. Nous avons calculé des barres d'erreurs à partir de la formule des intervalles de confiance d'un test de Student avec alpha = 5% afin de voir si l'écart entre les données passées et notre échantillon pouvait être expliqué.

Sur ce graphique, nous avons décidé de comparer les températures moyennes qui ont été relevées de 1971 à 2000 sur la Plaine de Limagne avec notre jeu de données (toujours de 2006 à 2035) en prenant encore une fois, un point situé au Nord de la Plaine et un situé au Sud. On peut voir que les températures ont beaucoup augmenté sur nos prévisions avec presque 5 degrés Celsius de différence au cours de l'année. il y a peu de variation entre la température moyenne du Nord et du Sud

Et pour finir, si la variable possède des données passées. Notre fonction nous fournit un tableau de p-valeur d'un test de conformité entre notre échantillon testé et nos données passées.

	1	2	3	4	5	6	7	8	9	10	11	12
P1	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	3.47e-03
P2	0.00e+00	4.90e-09	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	0.00e+00
P3	1.27e-14	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	0.00e+00
P4	2.44e-09	0.00e+00	3.79e-01	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	2.67e-01	0.00e+00
P5	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	0.00e+00
P6	0.00e+00	0.00e+00	0.00e+00	2.12e-01	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	0.00e+00
P7	0.00e+00	6.27e-02	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	0.00e+00
P8	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	0.00e+00	0.00e+00	0.00e+00
P9	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0	0	0	0	1.65e-09	0.00e+00	0.00e+00

Tableau 3-Tableau des pvaleurs pour les 9 premiers points

Comme c'est visible sur le graphique, les pvaleurs sont proche de 0 car l'écart entre nos valeurs est trop élevé.

Nous avons calculé les moyennes pour toutes les variables et nous avons décidé de tracer des graphiques sur les variables dont nous avions des données passées afin de pouvoir faire un comparatif.

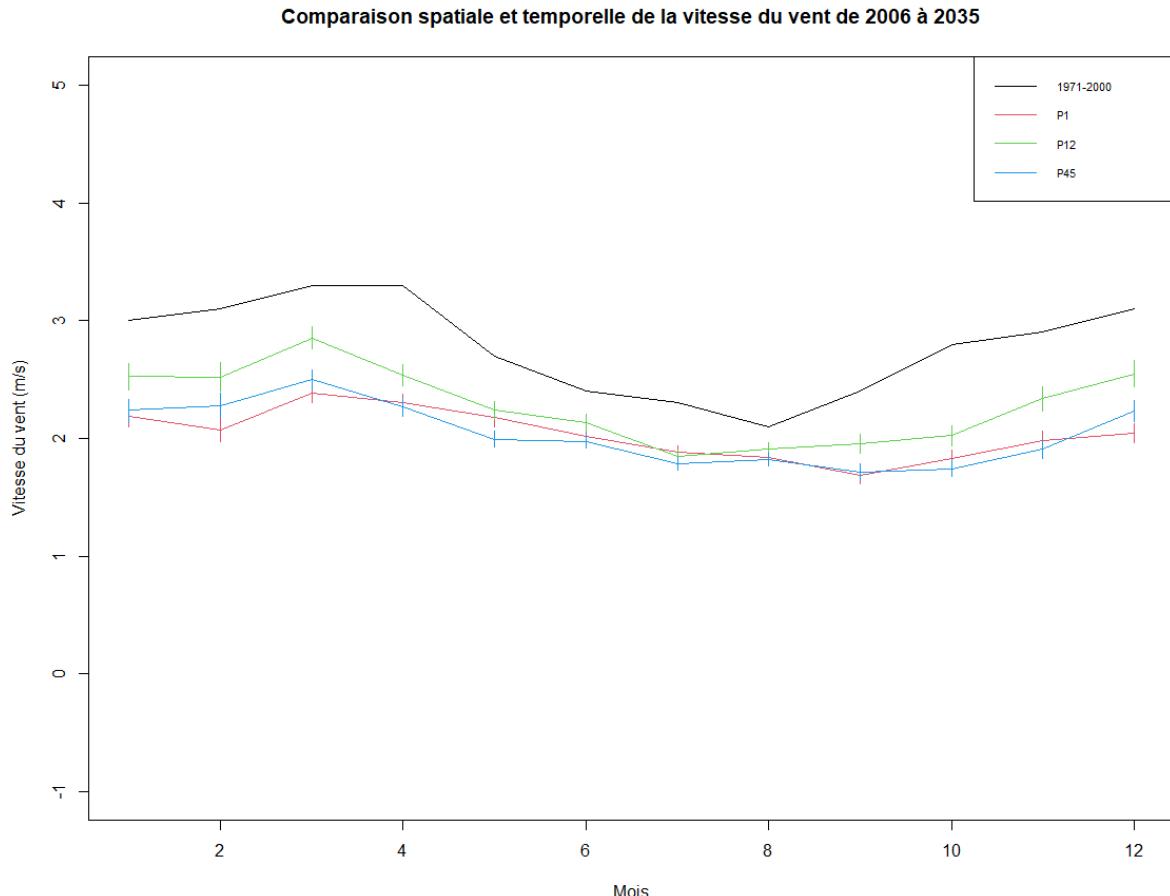


Figure 15 - Graphique de la moyenne des vitesses du vent des points 1,12 et 45 et de la moyenne des vitesses du vent de Clermont Ferrand de 1971 à 2000

Sur ce graphique, on ne voit pas de réelles différences de vitesse du vent entre le Nord et le Sud, mais on peut voir que le vent a plus de vitesse à Clermont-Ferrand tout au long de l'année. La vitesse du vent était globalement plus élevé de 1971 à 2000.

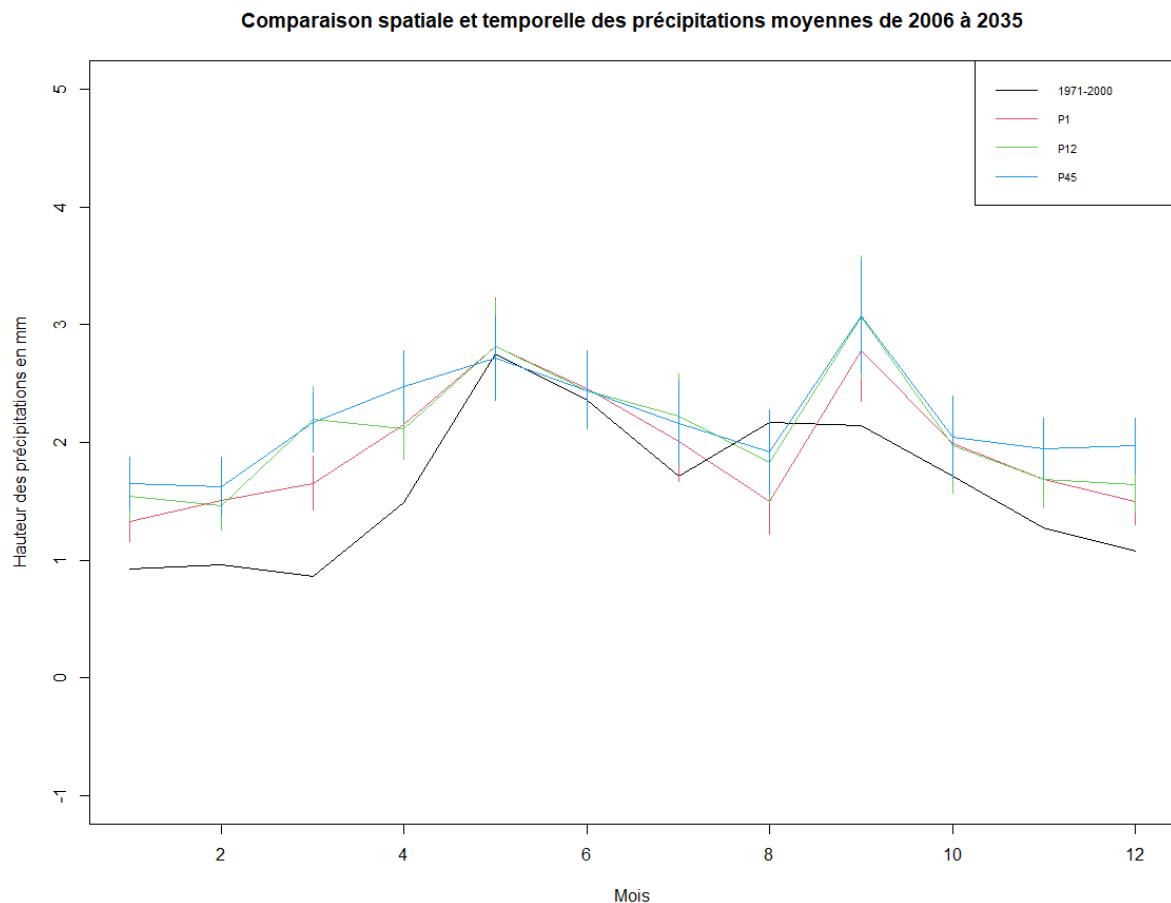


Figure 16 - Graphique de la moyenne des hauteurs de précipitation des points 1,12 et 45

Ici sur ce graphique, on arrive à discerner plus de différences entre le point 1 et le point 45. En effet, en août le point 1 a une hauteur de précipitation de 1.5mm tandis que les 2 autres points à une hauteur de précipitation de 2 mm en moyenne.

5. Interpolation spatiale

Nous voulons désormais être capable de faire des cartes de notre plaine, pour que l'on puisse avoir un aperçu visuel des microclimats qui peuvent être présent sur la plaine. Cependant nous n'avons pas des valeurs pour toute notre plaine, ce qui nous donnerait une carte très peu visuelle. Il va donc falloir généraliser nos variables à tout le domaine d'étude et pour cela nous allons utiliser la méthode d'interpolation spatiale.

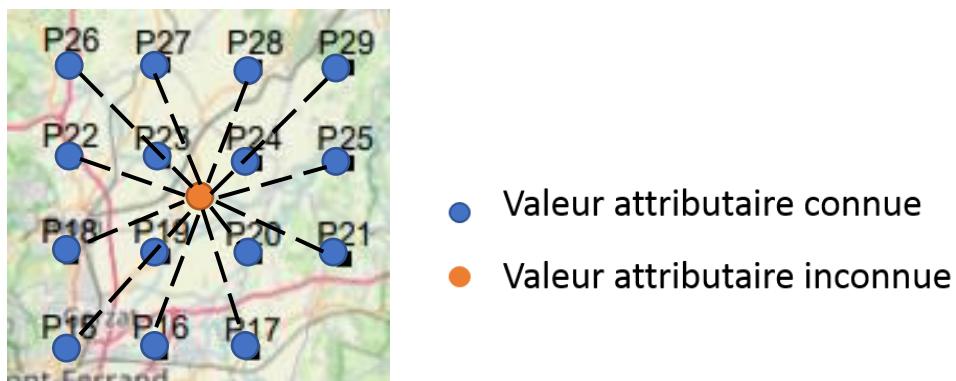


Figure 17 – Présentation de la méthode d'interpolation spatiale

La méthode d'interpolation spatiale consiste à utiliser la valeur de point connu pour créer une valeur à des points qui n'aurait pas de valeurs. C'est une méthode qui est utilisée dans la météo tous les jours et qui peut être utile à tout le monde. Exemple, si vous habitez dans un village et que vous voulez connaître la météo alors qu'il n'y a pas de centre météorologique là où vous vous situez. Cela est possible grâce à cette méthode, car en utilisant les données récoltées aux alentours il est possible de calculer la météo à un point précis.

Dans notre cas, on a 45 points dans notre plaine où des valeurs sont enregistrées. Pour expliquer simplement le principe qui se cache derrière le calcul de nouveaux points. Plus la distance entre le point où l'on veut calculer sa valeur et un point connu est grande moins la valeur de celui-ci aura d'impact sur la valeur du nouveau point.

Pour calculer la valeur de nos nouveaux points il y a plusieurs méthodes. Mais nous allons d'abord nous concentrer sur le profil inverse distance. La valeur de notre nouveau point sera

calculée selon l'inverse de la distance avec les points déjà connu, avec également un paramètre p qui est dans le calcul.

Ce paramètre p est un paramètre de lissage que l'on doit choisir et qui fait varier notre résultat, ce paramètre va faire varier le poids d'un point dans le calcul d'une valeur.

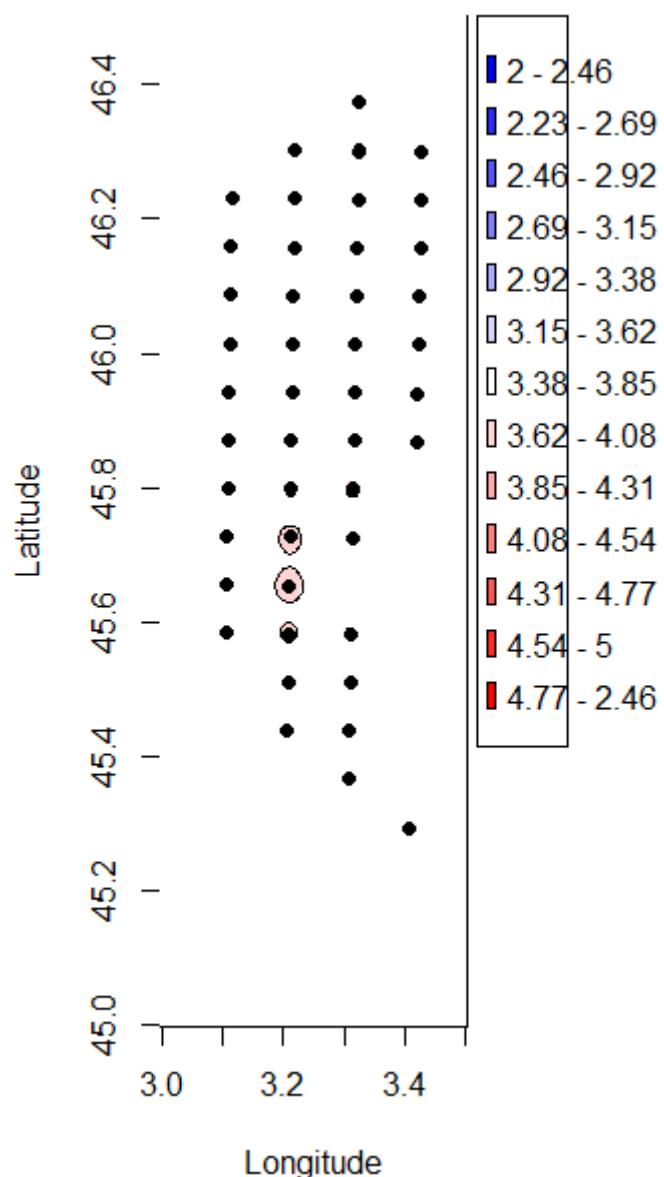


Figure 18 - Graphique de la température moyenne en janvier avec $p=0.6$ (faible) selon la méthode d'inverse de la distance

Si l'on utilise une valeur trop faible, on obtient une carte trop lisse. Et donc avec un paramètre faible les points très éloigné au nouveau calculé ont beaucoup de poids dans le calcul de celui-ci, c'est ce phénomène qui nous donne une carte trop lisse.

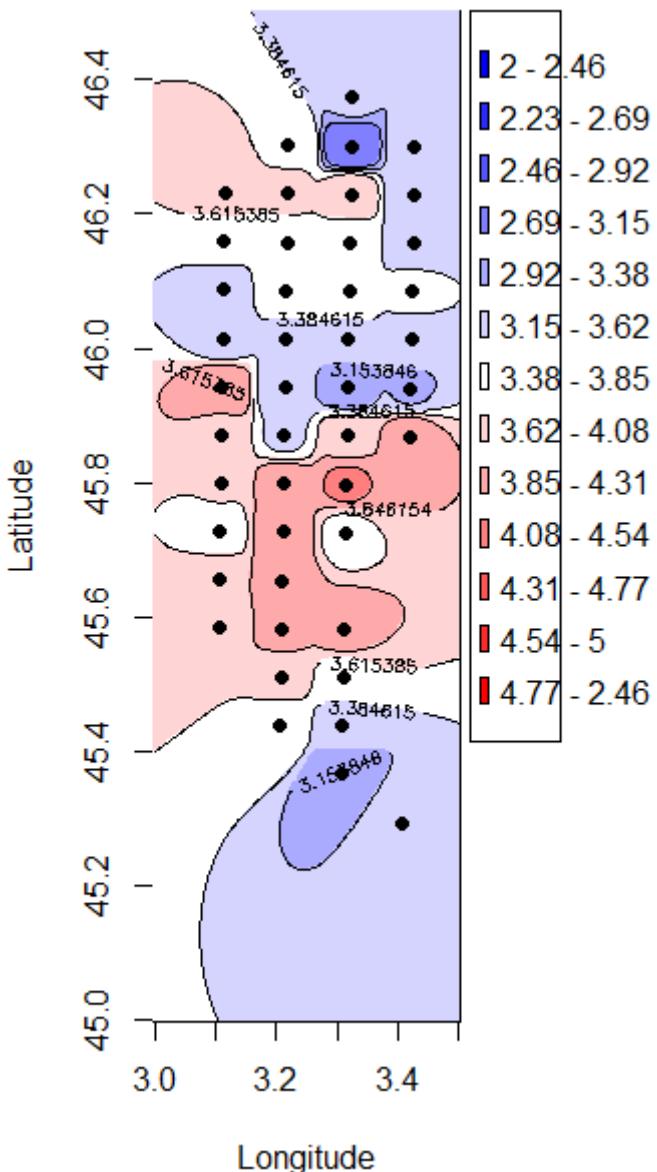


Figure 19 - Graphique de la température moyenne en janvier avec $p=6$ (fort) selon la méthode d'inverse de la distance

Si on utilise une valeur trop grande, on peut voir qu'autour de certains points on obtient des formes rectangulaires ce qui ne correspond pas à la réalité et il y a trop de variables.

Il va donc falloir calculer la valeur optimale de p . Il faut savoir que lorsque l'on utilise la méthode d'interpolation spatiale même les points qui ont déjà des valeurs connus vont être recalculés grâce à la valeur des autres points. Donc pour avoir la meilleure valeur de p possible, on va comparer la valeur de nos points d'après nos prévisions et celle avec l'interpolation spatiale. Ce qui nous donnera un écart quadratique moyen pour une valeur de p donné. On conservera la valeur de p qui à l'écart quadratique moyen le plus faible.

Calcul de la valeur optimal de p

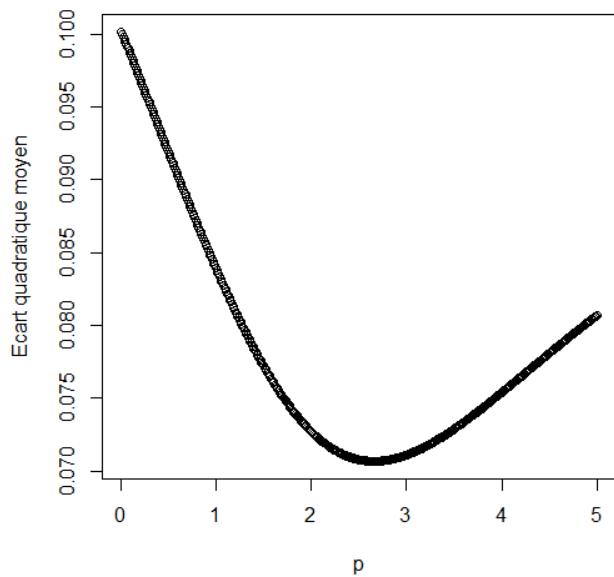


Figure 20 - Courbe représentant le calcul de la valeur optimale de p selon l'inverse de la distance

Donc pour notre exemple la valeur de p qui a l'écart quadratique moyen le plus faible et donc qui convient le mieux vaut 2.7.

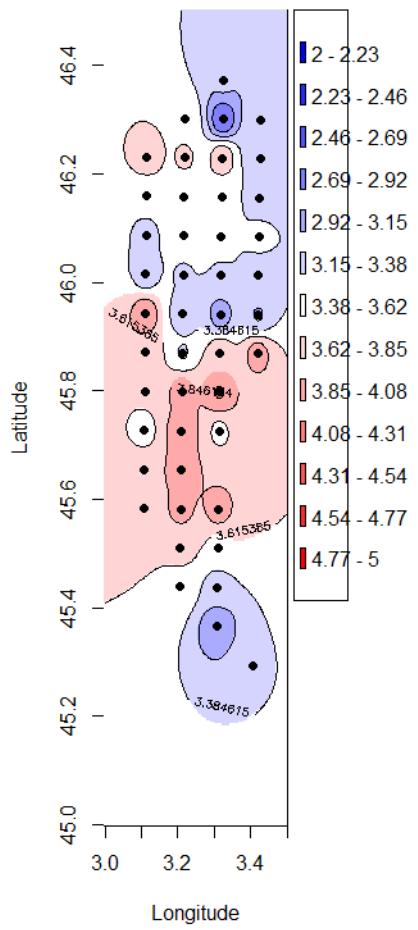


Figure 21 - Graphique de la température moyenne en janvier avec $p=2.7$ (optimal) selon la méthode d'inverse de la distance

On a donc la meilleure carte possible. On peut désormais voir plusieurs microclimats. On peut également voir qu'il n'y a pas de différence de température entre nord et sud. Les températures les plus élevées se trouvent au milieu de la plaine.

Maintenant que l'on arrive à optimiser une méthode de calcul d'interpolation spatiale, qui est celle de l'inverse de la distance.

On va ajouter deux méthodes à notre modèle pour que l'on ait le meilleur résultat possible.

La première calcule la nouvelle valeur en suivant le profil gaussien et l'autre le profil exponentiel. Ces deux modèles comprennent encore un paramètre p que l'on doit choisir. Pour trouver sa valeur la plus optimisé on utilise la même méthode, c'est-à-dire garder la valeur de p pour laquelle l'écart quadratique moyen est le plus faible.

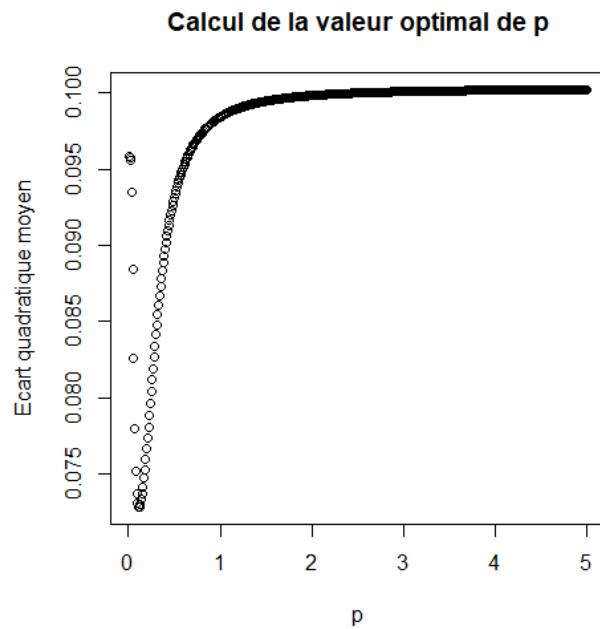


Figure 22 - Courbe représentant le calcul de la valeur optimale de p selon le profil gaussien

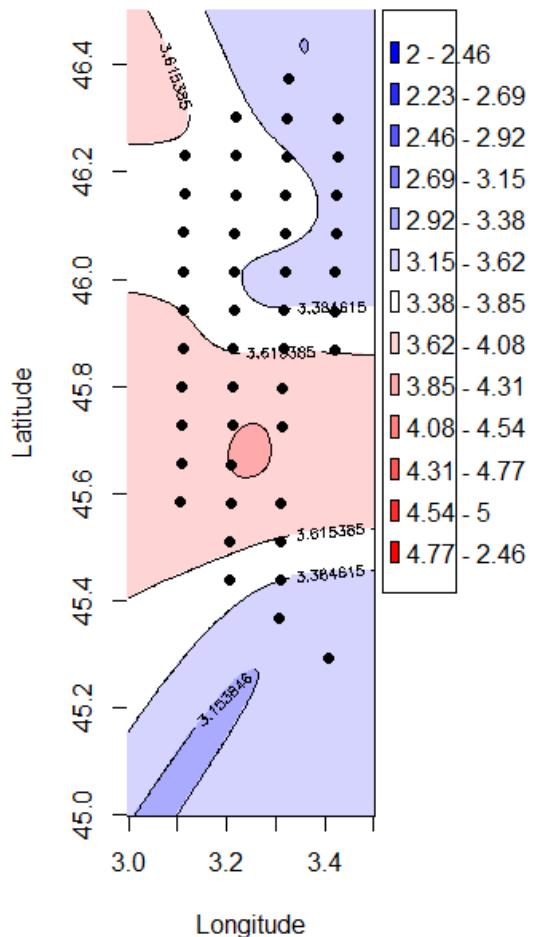


Figure 23 - Graphique de la température moyenne en janvier avec $p=0.11$ (optimal) selon le profil gaussien

Avec le profil gaussien, on obtient un résultat vraiment différent avec des formes de microclimats vraiment différentes de la méthode de l'inverse de la distance.

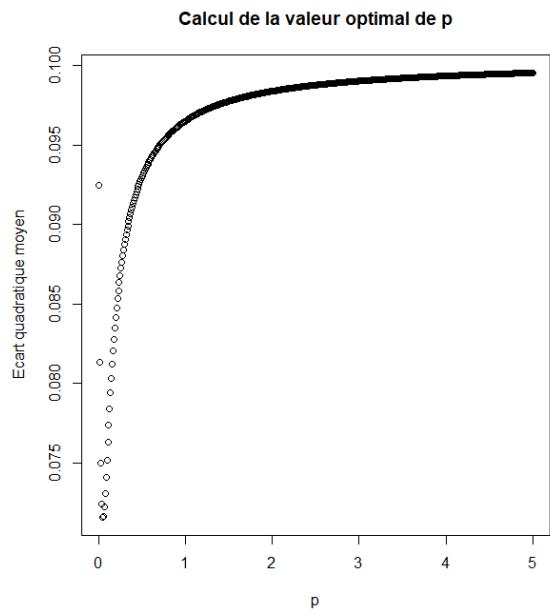


Figure 24 - Courbe représentant le calcul de la valeur optimale de p selon le profil exponentiel

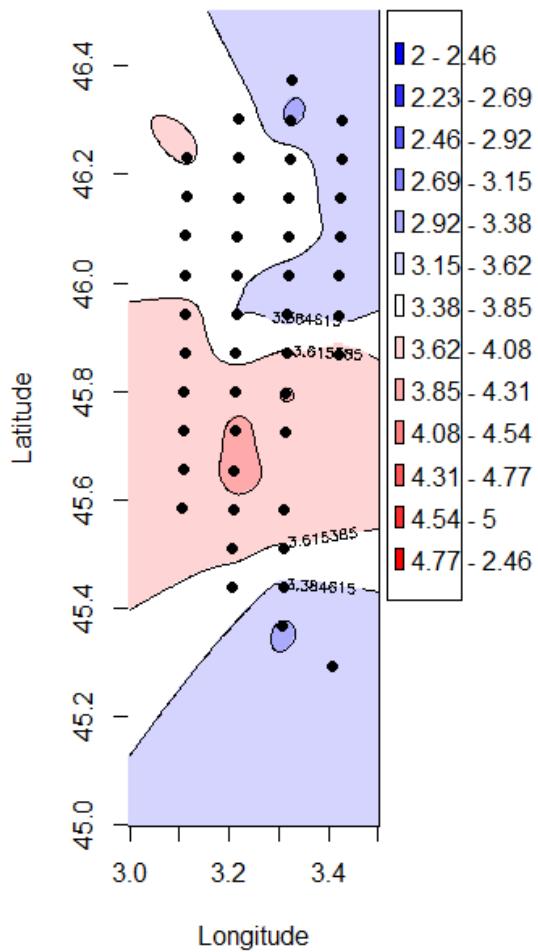


Figure 25 - Graphique de la température moyenne en janvier avec $p=0.11$ (optimal) selon le profil exponentiel

Nos deux nouveaux modèles se ressemblent on a une carte avec le profil exponentiel qui ressemble plus à celle suivant le profil gaussien que celle avec l'inverse de la distance.

Maintenant qu'on a optimisé chacun de nos modèles, on veut utiliser le plus performant pour notre interpolation spatiale. Pour cela lors de la création de notre carte, nous allons tester nos 3 modèles et comparer l'écart quadratique moyen le plus faible de chacun. Et le programme choisira automatiquement celui avec la plus faible valeur.

Maintenant que l'on peut avoir les meilleures cartes possibles, nous allons créer des fonctions simples, que notre commanditaire pourra utiliser pour réaliser des cartes.

La première fonction nous permettra de faire une carte de n'importe quelle statistique climatique, avec un mois donné et où les données seront agrégées par mois de 2006 à 2035. Elle s'appelle `Carte.mois` et a en arguments d'entrée :

- `statc` : la statistique climatique que l'on veut sur notre carte
- `mois` : le mois spécifique
- `nb` : en combien de classe veut-on notre légende, il faut choisir un nombre impair pour que l'on ait du blanc au milieu

Notre deuxième fonction consiste elle à faire une carte de n'importe quelle statistique climatique cette fois-ci non agrégées où l'on choisira le mois et l'année. Cette fonction s'appelle `Carte.annee` et a en argument d'entrée :

- `statc` : la statistique climatique que l'on veut sur notre carte
- `année` : l'année spécifique
- `mois` : le mois spécifique
- `nb` : en combien de classe veut-on notre légende, il faut choisir un nombre impair pour que l'on ait du blanc au milieu

Cette fonction a deux alternatives très similaires `Carte.annee.pdf` qui enregistre donc la carte dans un format pdf et `Carte.annee.jpg` qui elle l'enregistre dans un format jpeg.

Grâce à cette deuxième fonction, avec la simple création d'une boucle où l'on choisit une année de début et de fin, on peut créer des cartes de la même statistique climatique, du même mois avec juste les années qui évoluent. Et en créant un GIF sur internet avec toutes ces images, on peut avoir une visualisation très claire d'une statistique climatique qui évoluent au cours des années.

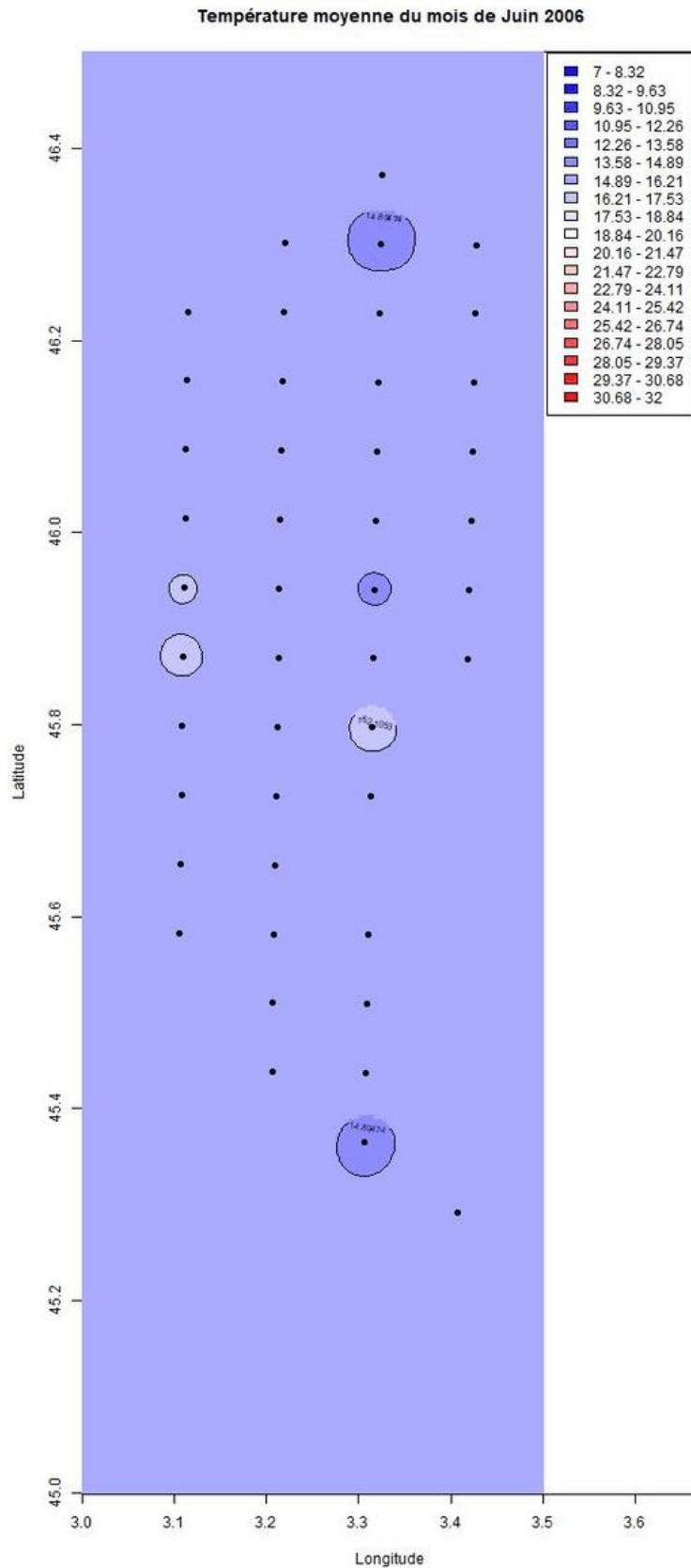


Figure 26 - GIF de l'évolution de la température moyenne du mois de Juin de 2006 à 2100.

Voici un GIF de l'évolution de la température moyenne du mois de Juin de 2006 à 2100.

6. Classification automatique de courbe

Pour la suite de notre projet, nous avons trouvé pertinent, au fur et à mesure des analyses des statistiques climatiques, de regrouper nos points spatiaux en classes afin de bien visualiser les différences entre différents groupes de points spatiaux. Le but est donc d'expliquer les phénomènes météorologiques de différentes zones de la plaine de la Limagne.

Tout d'abord nous avons corrigé les données en enlevant à chaque observation, la moyenne de l'individu afin de procéder à une ACP normée. Nous utilisons la méthode de l'ACP normée ici car nous avons des statistiques climatiques qui s'expriment avec des unités assez différentes.

De plus, nous avons procédé à une CAH (Classification ascendante hiérarchique) pour laquelle nous avons choisis comme méthode d'agrégation, la méthode de Ward.

Ensuite, nous avons décidé de fixer à 5, le nombre de classe que doit retenir cette classification. Ce choix est libre mais si nous prenons moins ou plus de groupes, nous nous éloignons de l'objectif principal.

Pour finir, nous traçons le dendrogramme des différentes statistiques climatiques et nous représentons les classes sous forme de courbes afin de visualiser les tendances des différents groupes et d'identifier les proximités entre les groupes.

Pour se rendre compte géographiquement des regroupements des points spatiaux, nous avons coloré la carte de la plaine avec des couleurs en fonction des classes.

Nous avons décidé de représenter seulement les températures moyennes car c'est la statistique climatique qui nous semblait-être la plus pertinente pour nos analyses. Nous pouvons ici comparer les classifications des points spatiaux selon deux agrégations des données (Années et mois).

Nous verrons dans un premier temps, les regroupements de nos points spatiaux des données agrégées par années. Et pour finir, nous verrons ceux des données agrégées par mois.

Températures moyennes agrégées par années :

Dendrogramme des points spatiaux classifiés par les températures moyennes agrégées par années

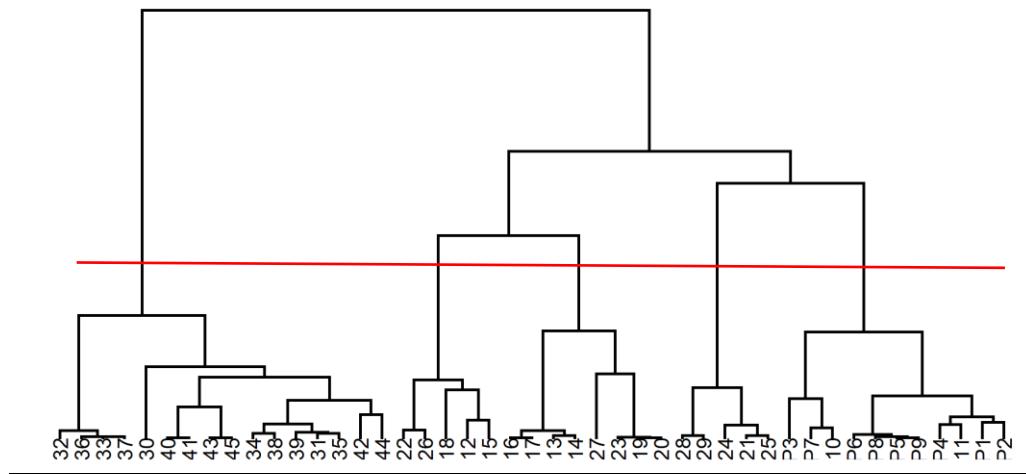


Figure 27 – Dendrogramme des points spatiaux classifiés par les températures moyennes agrégées par années

Nous pouvons voir sur la figure ..., le dendrogramme des points spatiaux classifiés par les températures moyenne agrégées par années. Ce dendrogramme nous permet de constater que les classes ne sont pas forcément constituées d'effectifs homogènes. En effet, mettons-nous au niveau de la ligne rouge qui nous indique que nous avons coupé le dendrogramme en 5 classes. Nous voyons que le premier groupe possède 16 individus tandis que le groupe qui suit n'en possède que 5.

Le problème du dendrogramme est qu'il ne fait que renseigner sur les individus regroupés mais ne permet pas de voir les différences de température entre les groupes.

C'est pourquoi nous allons maintenant vous présenter cela sous forme de courbes colorées de 5 façons différentes.

Classification des courbes des différentes classes des températures moyennes agrégées par années

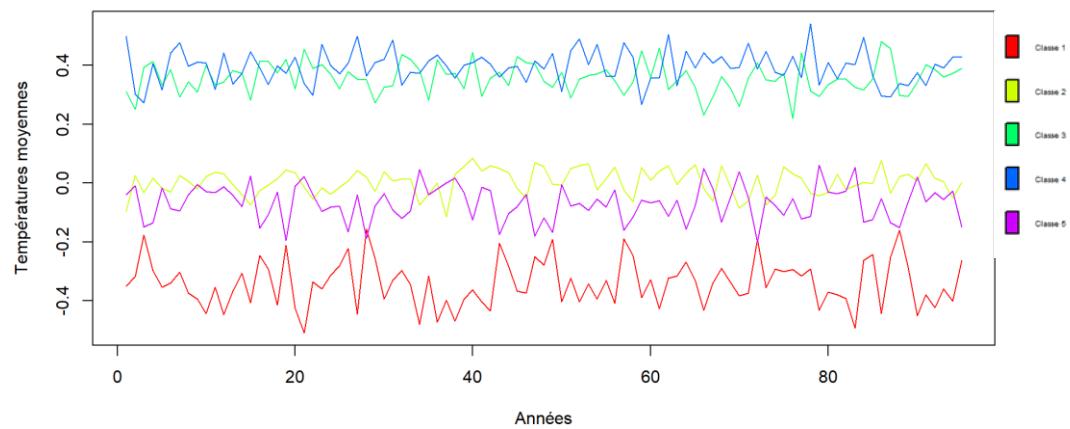


Figure 28 – Classification des courbes des différentes classes des températures moyennes agrégées par année

Nous remarquons ici les différentes classes d'individus et leurs positions sur le graphique. Les courbes se trouvant en bas du graphique sont les individus qui ont été classés selon des températures moyennes basses tandis que les courbes se trouvant en haut du graphique sont les individus qui ont été à l'inverse classés selon des températures moyennes hautes.

On remarque que les classes 3 et 4 semblent être très proche du point de vue des températures moyennes. Mais on voit bien que la classe 1 semble sortir de lot. En effet, c'est la plus isolée de toutes les classes.

Pour finir, on remarque que pour chaque classe, la tendance sur les 95 ans semble être assez aléatoire.

Nous verrons pour finir, une vision encore plus précise des classes en vous représentant chaque point spatial avec leurs couleurs de classes associées.

Carte de la plaine de la Limagne des points spatiaux classifiés par les températures moyennes agrégées par années

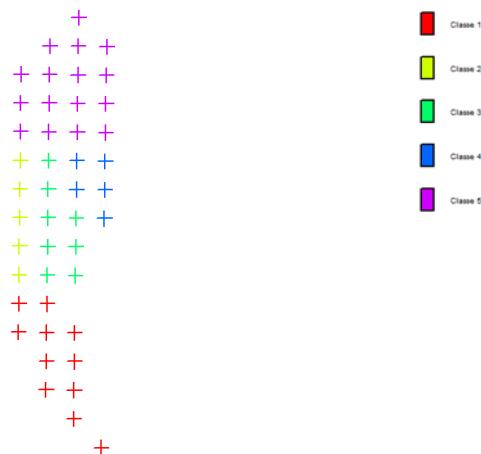


Figure 29 –Carte de la plaine de la Limagne des points spatiaux classifiées par les températures moyennes agrégées par année

Sur cette carte, on retrouve bien les classes avec le plus et le moins d'individus. En effet, la classe qui possède le plus d'individus correspond aux points spatiaux les plus au nord (Classe 5) et celle qui possède le moins d'individus correspond aux points spatiaux légèrement en dessous (Classe 4).

De plus, en analysant en même temps les courbes et la carte, on remarque que les nord et le sud possèdent des températures moyenne assez similaires alors qu'elles sont les plus éloignées.

A l'inverse, les courbes des classes 3 et 4 sont très proche ce qui est probablement expliqué par la proximité des zones.

Températures moyennes agrégées par mois :

Dendrogramme des points spatiaux classifiés par les températures moyennes agrégées par mois

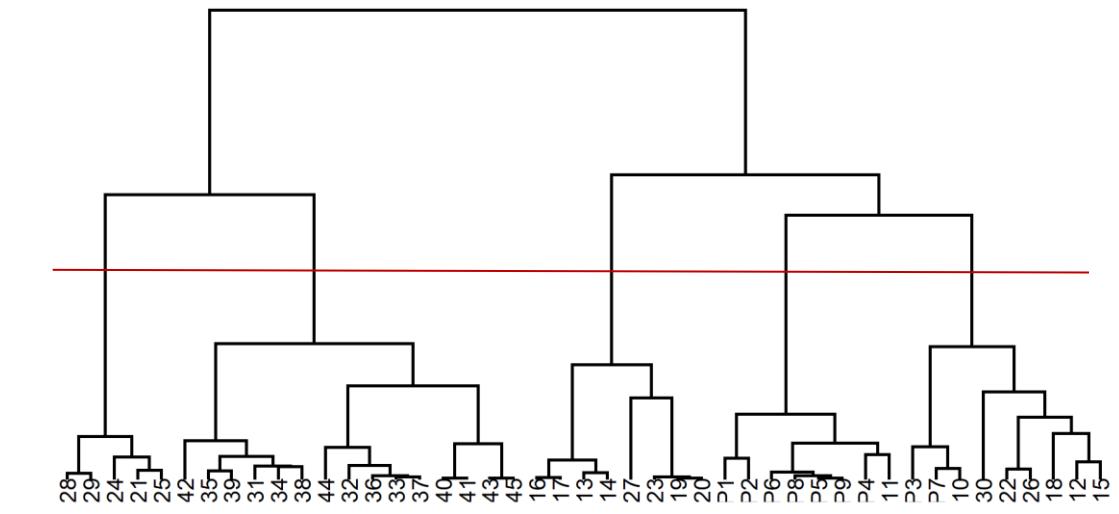


Figure 30 – Dendrogramme des points spatiaux classifiés par les températures moyennes agrégées par mois

Comme pour les données agrégées par années, on remarque une distribution assez hétérogène des individus dans les classes. On remarque que ce dendrogramme est quasiment identique à celui de la figure ...

Nous allons maintenant représenter chaque classe sur une période de 1140 mois.

Classification des courbes des différentes classes des températures moyennes agrégées par mois

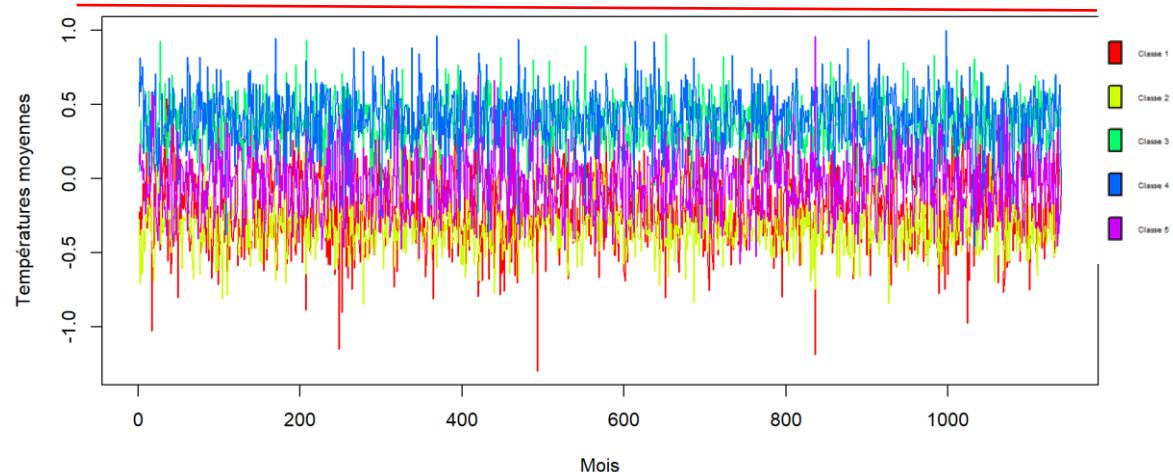


Figure 31 – Classification des courbes des différentes classes des températures moyennes agrégées par mois

Ici, il est un peu plus difficile d'analyser clairement les classes mais nous retrouvons à peu près la même position des courbes soit les classes 3 et 4 vers les températures moyennes plus hautes et les autres vers les températures moyennes plus basses.

De plus, on remarque de très forts piques qui se répètent sur plusieurs périodes pour la classe 1. On pourrait donc se demander pourquoi de tels piques ?

Carte de la plaine de la Limagne des points spatiaux classifiés par les températures moyennes agrégées par mois

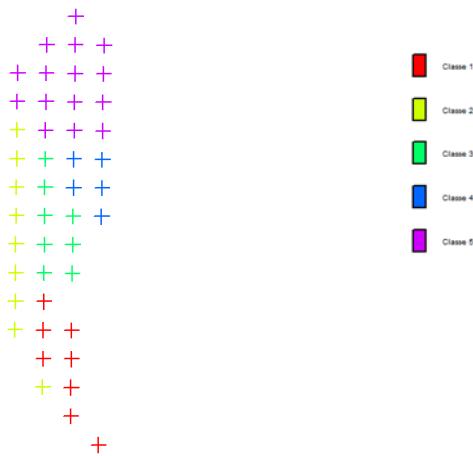


Figure 31 –Carte de la plaine de la Limagne des points spatiaux classifiés par les températures moyennes agrégées par mois

On remarque ici que cette carte n'est pas tout à fait identique à celle des données agrégées par années. En effet, nous avons plus de points spatiaux appartenant à la classe 2 et moins de points de la classe 1. Or, nous pouvons voir que les points appartenant aux classes 3 et 4 n'ont pas changé de disposition.

Pour parler du point correspondant à Clermont-Ferrand, on remarque qu'il appartient à la classe 2. On peut donc dire qu'avec les données agrégées par année, les températures moyenne de Clermont-Ferrand se situent entre les maximales et les minimales. Or, ce n'est pas le cas pour les données agrégées par mois. En effet, cette classe se situe au niveau des températures minimales au même niveau que la classe 1.

Vous pouvez retrouver les représentations de toutes les autres statistiques climatiques en Annexe. En effet, compte tenu du nombre de statistiques climatiques, nous ne pouvons pas toutes les présenter.

Conclusion

Face à la complexité de produire des interprétations pertinentes suite à nos analyses, nous avons préféré ne pas se lancer dans des interprétations absurdes. Nous laissons donc la partie interprétation de nos résultats à des experts du sujet.

Malgré cela, ce projet nous a permis de créer des outils d'aide à la décision pour l'INRAE, nous a permis de découvrir une nouvelle notion mathématique telle que l'interpolation spatiale, nous a permis d'appliquer différentes notions abordées à STID. En effet, nous avons utilisé les notions de tests d'hypothèses, filtrages linéaires et de Data Mining.

Nous avons été ravis de travailler sur des données de ce type qui nous ont permis de nous retrouver dans un contexte professionnel. Un contexte qui nous forme pour la suite du fait des objectifs fixés par les commanditaires, des deadlines...

Nous sommes satisfait du rendu final qui contient ce rapport et le code R qui permet d'utiliser les outils décrits.

Annexe

1. Cahier des charges

Partie 1 : Présentation du projet

1 – Acteurs

Le tableau 1 ci-dessous répertorie les acteurs du projet, ainsi que leur rôle dans celui-ci et leur contact.

Prénom Nom	Rôle	Contact
GREGOIRE Gérard	Commanditaire	/
ALLARD Vincent	Commanditaire	/
GARAT Philippe	Tuteur	philippe.garat@univ-grenoble-alpes.fr
CHAPUIS Paul	Maitre d'œuvre	paul.chapuis@etu.univ-grenoble-alpes.fr
DAO Tuan Linh	Maitre d'oeuvre	tuan.Dao@etu.univ-grenoble-alpes.fr
MCGUIGAN Scott	Maitre d'œuvre	Scott.Mcguigan@etu.univ-grenoble-alpes.fr
MOREL Alexis	Maitre d'œuvre	alexis.morel1@etu.univ-grenoble-alpes.fr

2 – Contexte

Avec le changement climatique, les parcelles qui aujourd’hui permettent de faire pousser des céréales ne seront peut-être pas les mêmes demain.

C'est pourquoi l'institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) nous a demandé d'étudier les modèles climatiques liés à la plaine de Limagne, proche de Clermont Ferrand afin d'arriver à comprendre l'impact futur des contraintes climatiques à la fois sur la variabilité génétique des céréales et aussi sur la qualité de la production céréalière de la plaine de Limagne.

3 – Objectifs

Avec ces données, nous avons donc plusieurs objectifs :

- Regrouper les points spatiaux similaires entre eux, ceux qui ont des trajectoires climatiques similaires et même ceux qui sont statistiquement proches. Cela nous permettrait de comprendre plus facilement l'évolution de nos variables et par la suite leur impact.
- Étudier la variabilité temporelle des données : déterminer une tendance haussière ou baissière des températures pour pouvoir mieux appréhender la production céréalière. Et par la suite calculer les corrélations sérielles, ce qui nous permettra de voir si les valeurs de nos variables sont indépendantes entre chaque jour, et si elles ne le sont pas, de quelle manière.
- Étudier la variabilité spatiale des données : détecter la présence ou non de microclimats ou une opposition entre le Nord et le Sud de la Plaine. Puis nous pourrons voir si les microclimats évoluent de la même manière.
- Comparer les prévisions que nous avons avec les données réelles : nos données de prévision vont de 2006 à 2100 et nos données climatiques de référence récoltées par Météo France vont de 1971 à 2000. On peut donc regarder comment nos variables ont évolué après plusieurs années. On pourrait également comparer nos données de prévision à des données réelles de la même année pour évaluer notre modèle de prévision.

4 – Contraintes

Dans ce projet, nous pouvons rencontrer certaines contraintes de temps, de qualité ou encore budgétaire. Ci-dessous, une liste des principales contraintes de ce projet :

- Situation sanitaire : en cause du covid, absence d'un ou plusieurs élèves, néanmoins ce projet est en télétravail
- Problèmes matériels : problèmes avec le logiciel Rstudio, perte de documents, problème avec le téléchargement du package rJava
- Contraintes de moyens : prendre en compte le temps que nous avons afin d'ajuster jusqu'où nous voulons traiter le sujet.
- La qualité : Le projet doit répondre à la problématique initiale afin de pouvoir satisfaire les commanditaires et répondre à leur besoin.

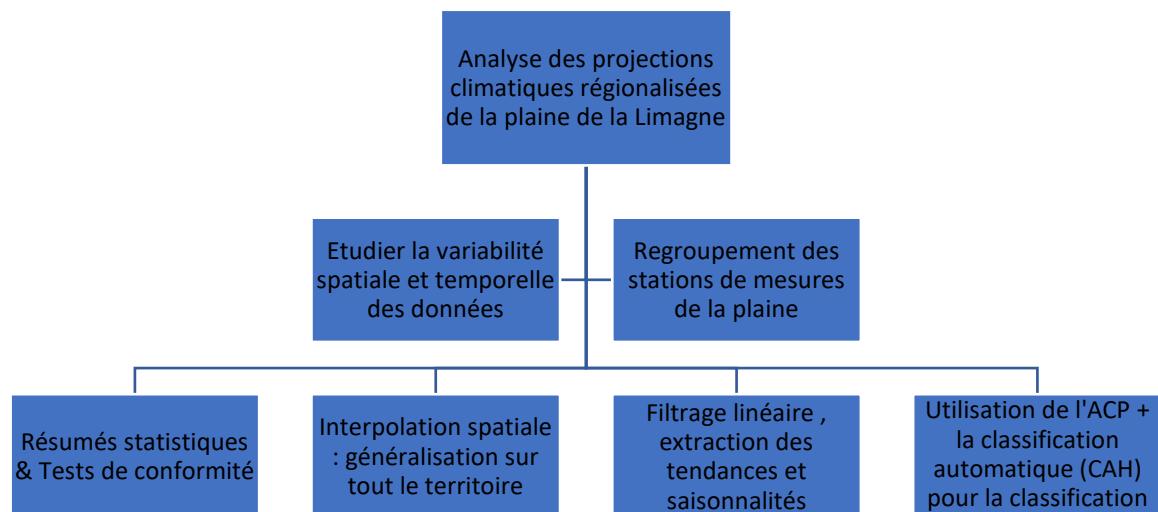
5 – Livrables

- Rapport intermédiaire : 06/01/2022
- Rapport final : Fin mars
- Code R : Fin mars

Partie 2 : Mise en œuvre du projet

1- Découpage du projet

Nous avons découpé ce projet en 2 sous-projets. Plusieurs tâches sont à réaliser pour ces 2 sous-projets. Nous avons donc établi un organigramme technique du projet (Figure 1) pour visualiser toutes les tâches à réaliser par la suite.



2-Outils utilisés

Ce projet nécessite l'utilisation de plusieurs logiciels et sites. Premièrement, les données nous ont été transmises sous forme d'un fichier texte (.txt) par les commanditaires. Ces données ont par la suite été traité sur RStudio. Les indicateurs, les graphiques et les résultats ont et seront traité intégralement sur RStudio. Enfin, nous partagerons et mettrons en commun nos résultats et nos documents grâce à un dossier partagé sur Chamilo

3-Cartographie des risques

Le projet comporte quelques risques plus ou moins probables ayant une certaine gravité sur le bon déroulement de celui-ci. Sur la figure 2, on retrouve chaque risque énoncé ci-

dessous. Certains risques sont suivis d'une flèche et d'une solution apportée pour éviter de grosses conséquences sur le projet.

1. Les prévisions soient mauvaises
2. Ne pas rendre les documents à temps. Travailler à plusieurs sur une tâche pour être sûr qu'elle soit faite.
3. Nouveau confinement.
4. Perte des documents. Enregistrer les documents à plusieurs endroits.
5. Abandon d'un étudiant au cours du projet

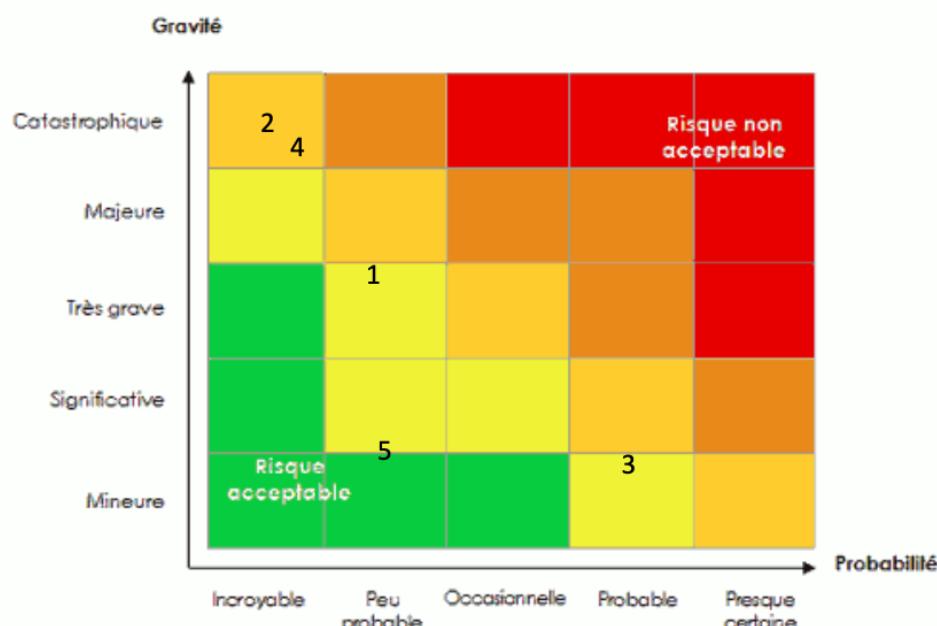
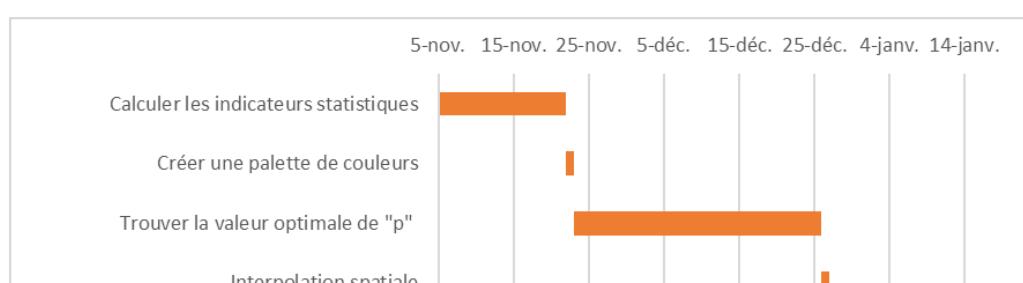


Figure 2 : Cartographie des risques

On remarque que les risques sont plus proches des « risques acceptables », il est donc très envisageable d'atteindre les objectifs du projet dans les temps.

2. Diagramme de GANTT

Travaux	Date début	Durée	Date de fin
Calculer les indicateurs statistiques	5-nov.	17	22/11/2021
Créer une palette de couleurs	22-nov.	1	23/11/2021
Trouver la valeur optimale de "p"	23-nov.	33	26/12/2021
Interpolation spatiale	26-déc.	1	27/12/2021
Le diaporama PowerPoint	27-déc.	7	03/01/2022
La présentation des données brutes, rapport intermédiaire	3-janv.	3	06/01/2022
Soutenance intermédiaire	6-janv.	1	07/01/2022



3. Indicateurs météorologiques de Clermont-Ferrand 1971-2000

CLERMONT-FERRAND

Puy-de-Dôme (63)

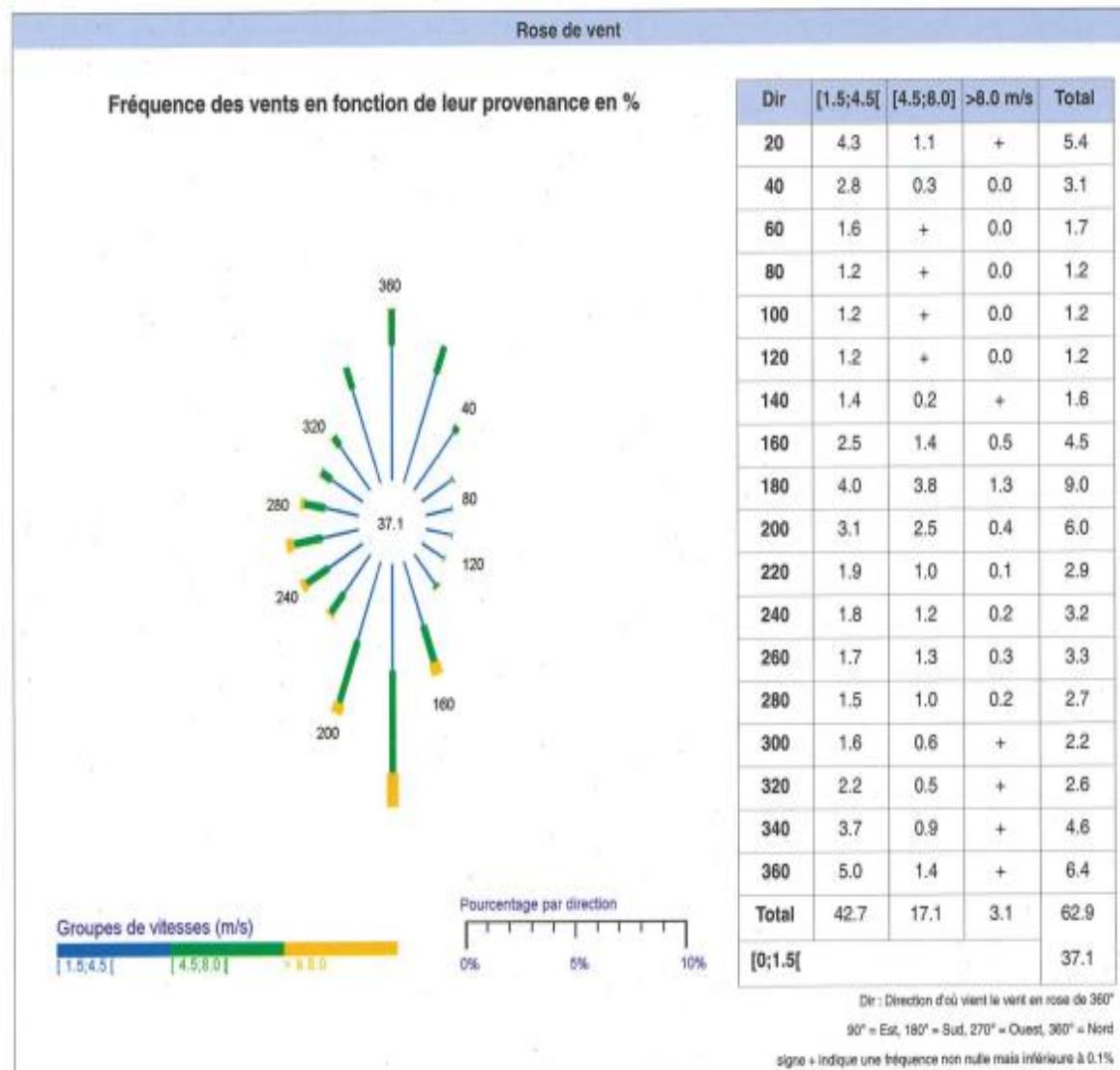
Températures	Janv.	Févr.	Mars	Avril	Mai	Juin	Juill.	Août	Sept.	Oct.	Nov.	Déc.	Année
Température minimale (°C)													
Q80	1.8	2.4	3.4	5.3	9.7	12.1	14.4	14.5	11.4	8.5	4.8	2.4	6.8
MOY	-0.3	0.5	2.2	4.1	8.2	11.2	13.4	13.2	10.3	7.1	2.6	0.7	6.1
Q20	-2.0	-1.8	0.6	2.8	6.8	10.0	12.1	12.1	9.2	5.7	0.8	-0.7	5.4
Température moyenne (°C)													
Q80	5.9	7.1	9.0	10.7	15.5	18.0	20.5	21.0	17.7	13.3	8.8	6.0	12.0
MOY	3.5	4.9	7.3	9.5	13.7	16.8	19.7	19.5	16.2	12.0	6.8	4.4	11.2
Q20	1.9	3.0	5.8	8.4	12.0	15.7	18.3	18.3	14.9	10.9	5.0	2.9	10.6
Température maximale (°C)													
Q80	9.8	12.1	14.9	16.4	21.6	23.7	27.3	27.6	24.5	18.6	13.1	9.8	17.1
MOY	7.3	9.3	12.5	14.8	19.2	22.5	25.9	25.8	22.2	16.9	11.0	8.2	16.3
Q20	5.3	6.5	10.5	13.0	17.1	21.1	24.2	24.2	20.0	15.7	9.0	6.4	15.7
Température la plus basse (°C)													
Date	-23.1	-15.7	-10.7	-5.2	-3.6	1.0	4.4	2.4	-1.0	-9.2	-11.8	-14.2	-23.1
	05-1971	10-1986	06-1971	11-1976	01-1978	02-1973	22-1980	24-1983	22-1977	29-1997	23-1993	03-1980	1971
Température la plus élevée (°C)													
Date	19.9	23.7	26.6	28.8	31.6	35.5	40.7	39.5	36.8	29.0	23.0	21.3	40.7
	10-1991	20-1998	25-1981	24-1992	14-1992	23-1998	31-1983	16-1974	16-1987	01-1985	02-1981	16-1989	1993
Nombre moyen de jours avec													
Tn <= -5°C	5.9	3.7	2.2	0.1	0.0	0.0	0.0	0.0	0.0	0.1	1.8	3.7	17.6
Tn <= 0°C	15.2	12.5	9.4	4.8	0.5	0.0	0.0	0.0	0.1	1.7	8.8	14.4	67.5
Tx <= 0°C	2.6	1.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	2.3	7.3
Tx >= 25°C	0.0	0.0	0.1	0.5	4.0	9.1	17.6	17.2	8.5	1.5	0.0	0.0	58.6
Tx >= 30°C	0.0	0.0	0.0	0.0	0.3	1.7	6.1	6.2	1.4	0.0	0.0	0.0	15.6
Tx >= 35°C	0.0	0.0	0.0	0.0	0.0	0.1	0.6	0.7	0.1	0.0	0.0	0.0	1.5

Précipitations	Janv.	Févr.	Mars	Avril	Mai	Juin	Juill.	Août	Sept.	Oct.	Nov.	Déc.	Année
Hauteur de précipitations (mm)													
Q80	41.3	39.3	36.5	70.5	121.5	106.5	75.5	101.3	99.8	75.6	59.9	47.5	733.7
MOY	28.7	26.8	26.6	44.6	85.3	70.6	53.0	67.2	64.3	53.1	38.1	33.3	591.6
Q20	9.2	12.5	11.5	14.9	62.4	41.6	30.7	33.4	27.8	24.6	18.7	14.2	485.4
Hauteur maximale quotidienne de précipitations (mm)													
Date	49.0	25.2	26.9	60.8	41.7	57.6	51.2	51.3	54.2	66.6	57.2	31.2	66.6
	06-1982	03-1994	23-1977	29-1998	14-1983	26-1990	19-1979	30-1974	21-1992	10-1979	04-1994	08-1972	1979
Nombre moyen de jours avec													
Rr >= 1 mm	6.3	6.2	6.1	8.7	11.5	8.7	8.4	7.6	7.1	7.7	6.8	6.4	89.6
Rr >= 5 mm	1.4	1.7	1.5	2.6	5.3	3.7	3.1	3.8	3.7	3.1	2.1	2.0	33.9
Rr >= 10 mm	0.4	0.4	0.5	1.0	2.6	2.2	1.9	2.1	1.8	1.3	0.9	0.8	15.9

Insolation	Janv.	Févr.	Mars	Avril	Mai	Juin	Juill.	Août	Sept.	Oct.	Nov.	Déc.	Année
Durée d'insolation (heures)													
Q80	93.4	148.7	184.0	198.8	238.4	241.9	265.6	277.4	222.2	166.9	101.6	86.2	2035.2
MOY	85.5	113.7	162.5	173.6	203.5	206.1	245.9	248.8	182.5	121.9	85.3	69.2	1898.0
Q20	72.9	63.5	129.9	125.8	165.0	163.0	222.7	216.6	126.7	89.8	61.6	55.2	1748.2
Nombre moyen de jours avec fraction d'insolation													
= 0 %	9.5	5.4	3.3	2.6	2.1	1.3	0.6	0.4	2.1	5.0	7.1	8.1	47.5
<= 20 %	17.5	12.3	10.7	9.4	8.5	8.4	6.3	4.3	7.6	13.3	15.9	18.5	132.7
>= 80 %	4.6	5.1	7.3	5.5	5.7	5.4	7.6	9.3	7.8	5.1	3.2	2.8	69.4

Phénomènes	Janv.	Févr.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.	Année
Nombre moyen de jours avec													
Brouillard	4.7	2.8	1.1	0.6	0.6	0.2	0.2	0.3	1.6	3.5	4.4	4.8	24.7
Orage	0.0	0.1	0.1	1.3	5.2	5.5	5.4	6.4	2.9	1.1	0.3	0.2	28.5
Grêle	0.0	0.1	0.0	0.1	0.3	0.3	0.2	0.2	0.1	0.0	0.0	0.0	1.3
Neige	-	4.2	2.6	1.6	0.1	0.0	0.0	0.0	0.0	0.0	2.1	3.6	-

Vent	Janv.	Févr.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.	Année
Rafale maximale de vent (m/s)													
Date	27.0	30.0	32.0	28.0	23.0	28.0	27.0	26.0	28.0	29.0	36.0	44.0	44.0
	17-1995	03-1990	24-1988	07-1998	11-1983	28-1990	13-1995	21-1994	25-1982	10-1987	07-1982	27-1999	1999
Vitesse du vent moyenné sur 10 mn (moyenne en m/s)													
	3.0	3.1	3.3	3.3	2.7	2.4	2.3	2.1	2.4	2.8	2.9	3.1	2.8
Nombre moyen de jours avec rafales													
$\geq 16 \text{ m/s}$	6.3	4.9	5.5	5.4	2.9	2.3	2.0	1.9	2.3	5.3	4.7	6.4	49.8
$\geq 28 \text{ m/s}$	0.0	0.2	0.2	0.1	0.0	0.1	0.0	0.0	0.1	0.1	0.5	0.4	1.5
16m/s = 56 km/h, 28m/s = 100 km/h													



4. Comptes rendus de réunion



PROJET TUTORÉ : Analyse des projections climatiques régionalisées de la plaine de la Limagne

COMPTE-RENDU DE REUNION DU JEUDI 5 NOVEMBRE 2021

- Date : 05/11/2021
- Heure de début : 15h00
- Heure de fin : 16h00

Ordre du jour : Questionnements à propos du début du projet et résolution de problèmes

Présents en présentiel :	Excusés :
Tuteur : M. Garat Et les maîtres d'œuvre : M. Alexis Morel M. Scott Mc Guigan M. Paul Chapuis M. Tuan Linh Dao	/

Ouverture de la séance

M. Garat ouvre la séance à 15h00.

La réunion s'est déroulée sur 3 parties. Une première partie sur une explication globale du code R fourni au préalable par notre tuteur. Une deuxième partie porte sur la compréhension et la résolution de problèmes liés à ce code. Et pour finir, une partie sur l'explication des statistiques climatiques et la répartition des premières tâches à réaliser.

1. Explication du code R

M. Garat nous a entièrement repris le code R en nous expliquant l'utilité et le principe de fonctionnement de plusieurs fonctions telle que « strsplit » ...

Il nous a pour finir expliqué à quoi correspondait le « lambert93 » et le « wgs84 ».

2. Compréhension et résolution de problème

M. Garat a essayé de nous aider face à un problème que tous les maîtres d'œuvre ont rencontré. Ce problème est lié au package « Openstreetmap » de R qui ne fonctionne pas. Il nous a prévenu qu'il essaiera de corriger ce problème

3. Explication des statistiques climatiques et tâches à réaliser

M. Garat nous a expliqué et présenté les différentes statistiques climatiques que nous avons dans le fichier de données avant de nous donner un premier travail à réaliser. Nous devons donc avant la prochaine réunion calculer des indicateurs statistiques sur une période de 30 ans, afin de comparer aux statistiques fournies (période 1971-2000) de la station météo de Clermont-Ferrand. Chaque maître d'œuvre va devoir prendre un point qui sont radicalement opposés aux autres et les comparer entre eux afin d'en conclure les phénomènes qui entrent en jeu.

COMPTE-RENDU DE REUNION DU LUNDI 22 NOVEMBRE 2021

- Date : 22/11/2021
- Heure de début : 17h00
- Heure de fin : 18h15

Ordre du jour : Explication de la cartographie par interpolation spatiale

Présents en présentiel :	Excusés :
Tuteur : M. Garat	M. Scott Mc Guigan
Et les maîtres d'œuvre : M. Alexis Morel M. Paul Chapuis M. Tuan Linh Dao	

Ouverture de la séance

M. Garat ouvre la séance à 17h00.

La réunion s'est déroulée sur 2 parties. Une première partie sur une présentation générale du code R que nous avons programmé afin de calculer les différents indicateurs statistique liés aux statistiques climatiques. Puis, une deuxième partie sur l'explication et la présentation d'un code R permettant de faire une cartographie par interpolation spatiale.

1. Validation du code R pour le calcul des indicateurs statistiques des statistiques climatiques

M. Garat nous a validé le code R qui nous a permis de calculer les indicateurs statistiques des différentes données climatiques. Il nous explique que ces indicateurs statistiques (ex : Moyenne) pourront être réutilisées lors de cartographie par interpolation spatiale.

2. Explication et présentation d'un code R pour une cartographie par interpolation spatiale

M. Garat nous a expliqué dans les grandes lignes à quoi servait et comment réaliser une cartographie par interpolation spatiale avant de nous présenter un code R permettant de la réaliser. Après avoir réalisé une cartographie par interpolation spatiale, M. Garat nous explique et nous présente une variable importante qui est la variable « p » et qui va permettre de lisser au mieux les courbes d'interpolation. Cette variable va jouer un rôle important dans la qualité de représentation de la cartographie.

Le prochain travail est donc de créer une boucle permettant de trouver la bonne valeur de ce « p » pour avoir la meilleure qualité de représentation. De plus, il faut réussir à créer une palette de couleurs adaptée aux conditions climatiques.

- Date : 04/01/2022
- Heure de début : 9h00
- Heure de fin : 10h30

Ordre du jour :

Présents en présentiel :	Excusés :
Tuteur : M. Garat Et les maîtres d'œuvre : M. Alexis Morel M. Paul Chapuis M. Tuan Linh Dao M. Scott Mc Guigan	

Ouverture de la séance

M. Garat ouvre la séance à 9h00.

M. Garat nous demande de faire un récapitulatif de toute nos missions depuis le début du projet.

- Calcul d'indicateurs statistiques pour 45 points sur une durée de 100 ans.

1. Visualisation des données brutes

Il nous demande de faire des graphiques sur les données brutes pour savoir si les données sont lisses ou si elles ont du bruit afin de voir si elles ont de tendance haussière ou baissière.
Représenter la température sur 1 mois ou 1 an.

Représenter la pluviométrie en diagramme en bâton sur 1 semaine.

Calculer les corrélations sérielles

2. Objectifs du projet tutoré

Regrouper les points similaires (typologie) par classification automatique qui sera vue en cours de DataMining.

Etudier la variabilité temporelle

Etudier la variabilité spatiale : 8km entre chaque point, y'a-t-il des microclimats ? (opposition nord-sud), caractériser ces micros-climats en termes de pluviométrie.

3. Mise en œuvre / Réalisation prévue

Résumés statistiques

Test statistique de conformité (Rentrer les données météorologiques de la plaine de Limagne de 1971 à 2000.)

Interpolation spatiale pour 1 jour donné

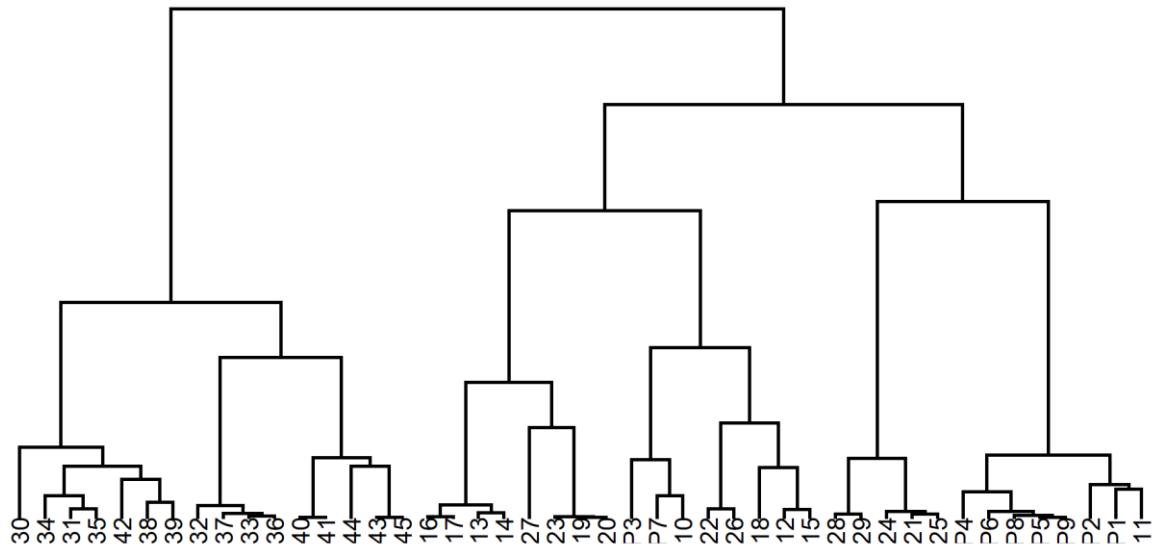
Classification de courbes (cours Data Mining)

5. Classification automatique de courbe

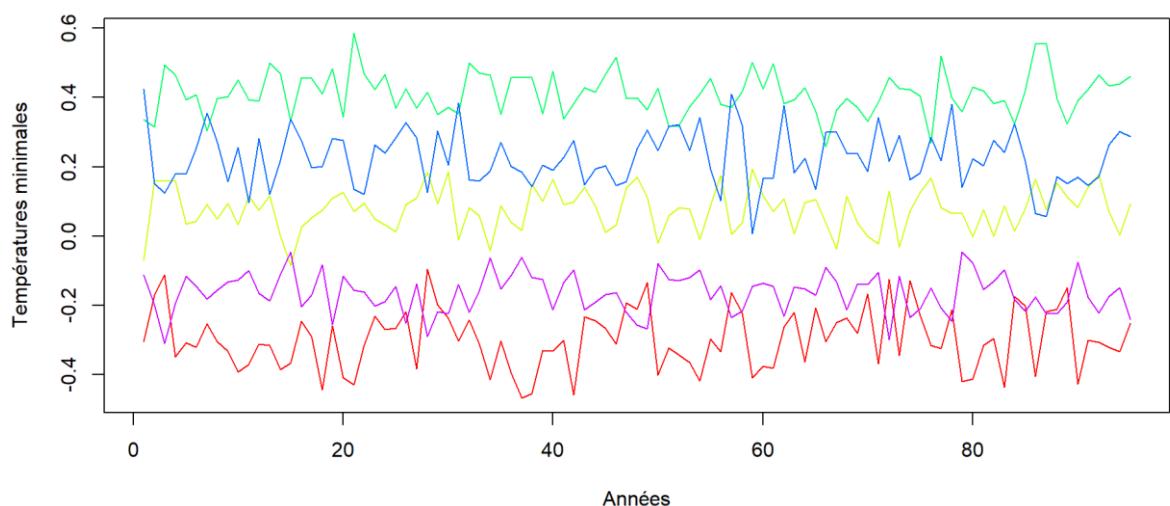
Données agrégées par années

Températures minimales :

Dendrogramme des points spatiaux classifiés par les températures minimales agrégées par années



Classification des courbes des différentes classes des températures minimales agrégées par années

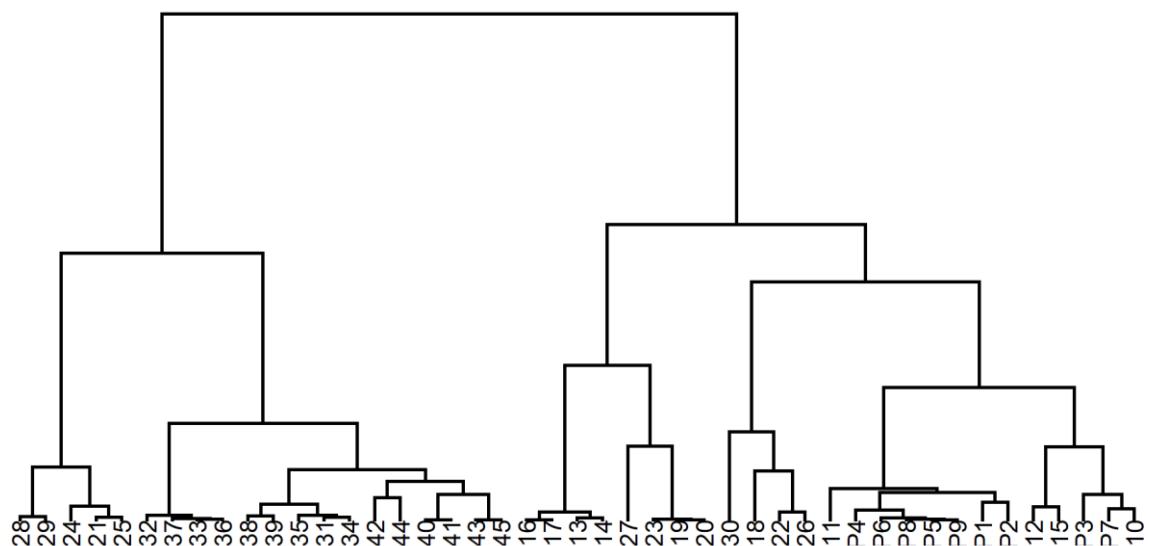


Carte de la plaine de la Limagne des points spatiaux classifiés par les températures minimales agrégées par années

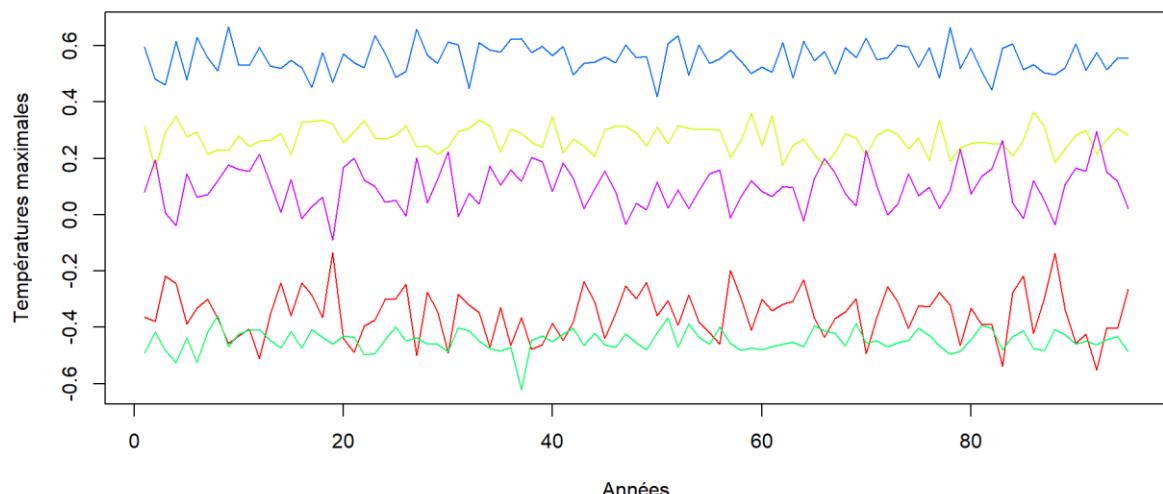


Températures maximales :

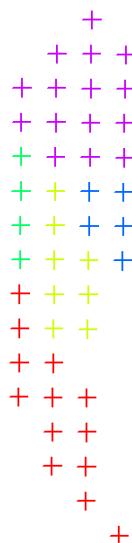
Dendrogramme des points spatiaux classifiés par les températures maximales agrégées par années



Classification des courbes des différentes classes des températures maximales agrégées par années

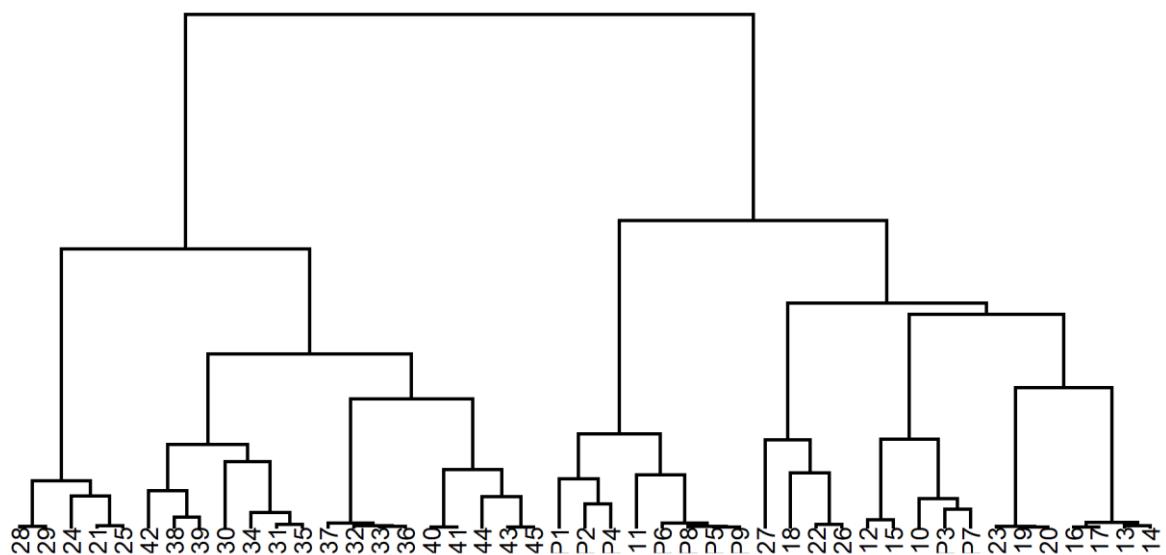


Carte de la plaine de la Limagne des points spatiaux classifiés par les températures maximales agrégées par années

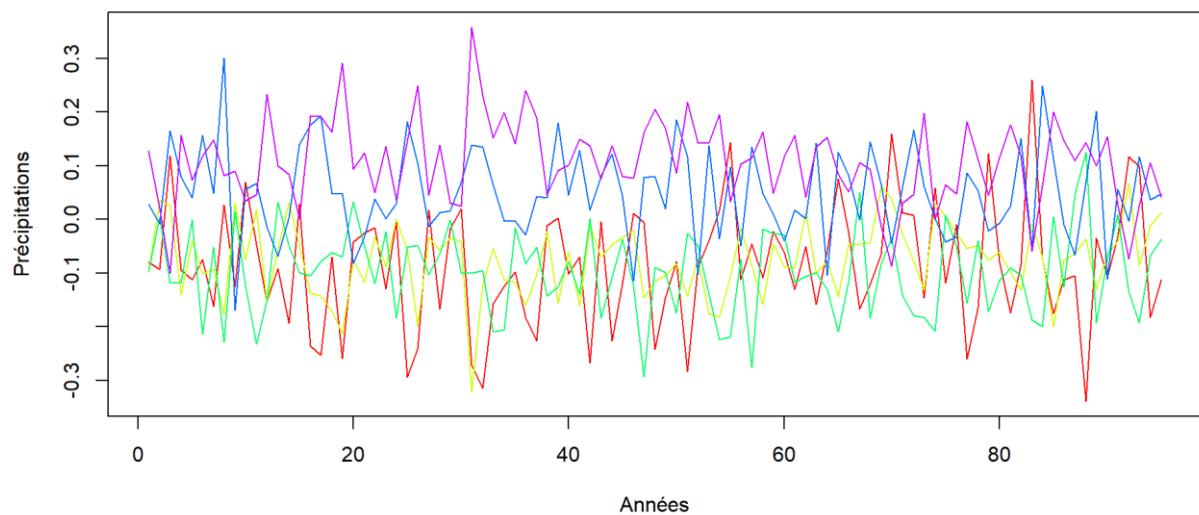


Précipitations :

Dendrogramme des points spatiaux classifiés par les précipitations agrégées par années



Classification des courbes des différentes classes des précipitations agrégées par années

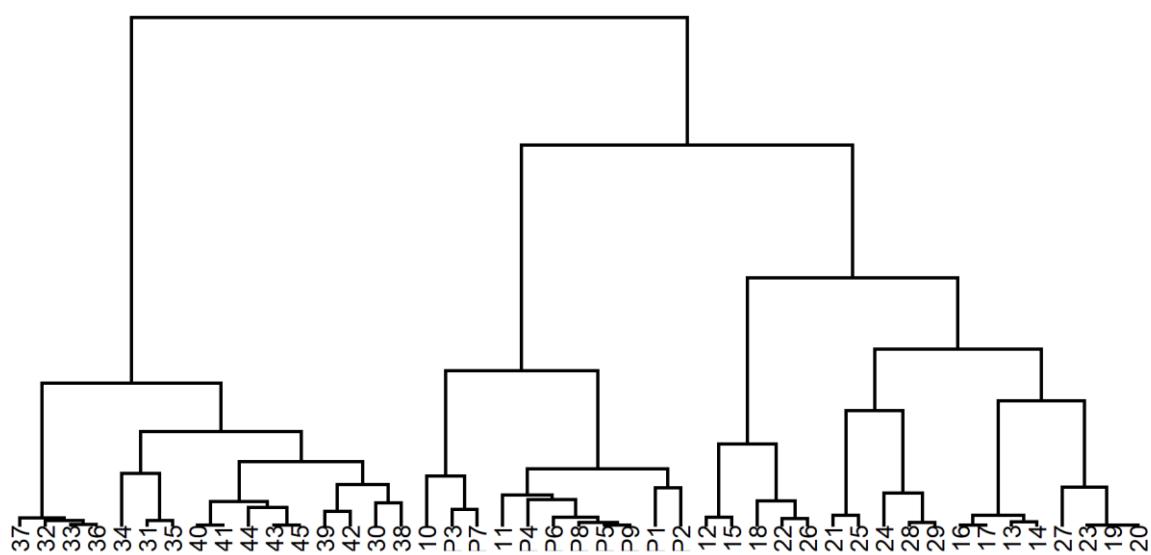


Carte de la plaine de la Limagne des points spatiaux classifiés par les précipitations agrégées par années

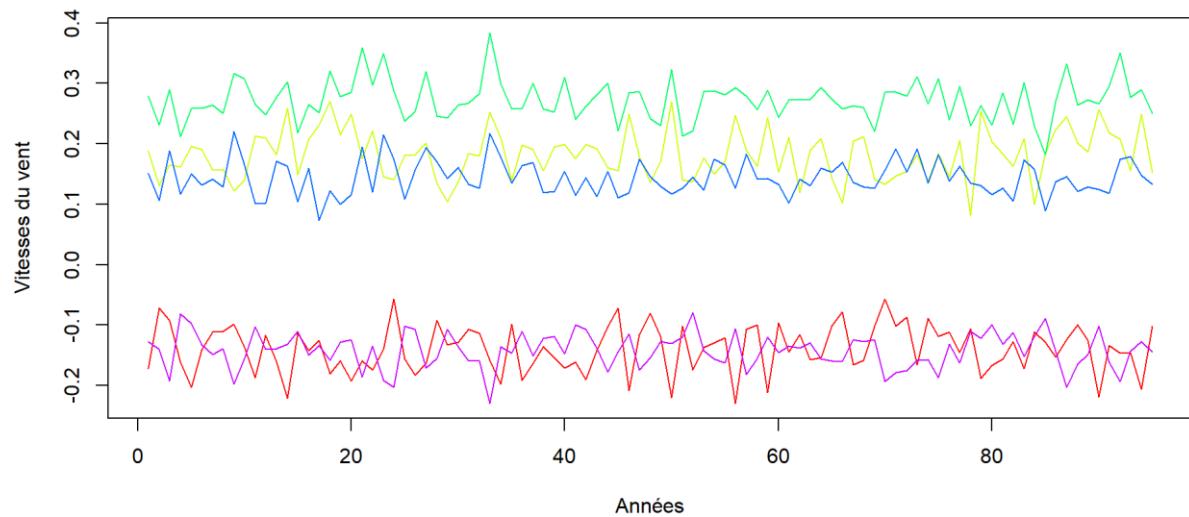


Vitesses du vent :

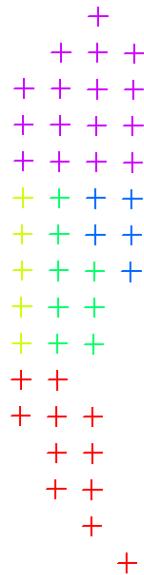
Dendrogramme des points spatiaux classifiés par les vitesses du vent agrégées par années



Classification des courbes des différentes classes des vitesses du vent agrégées par années

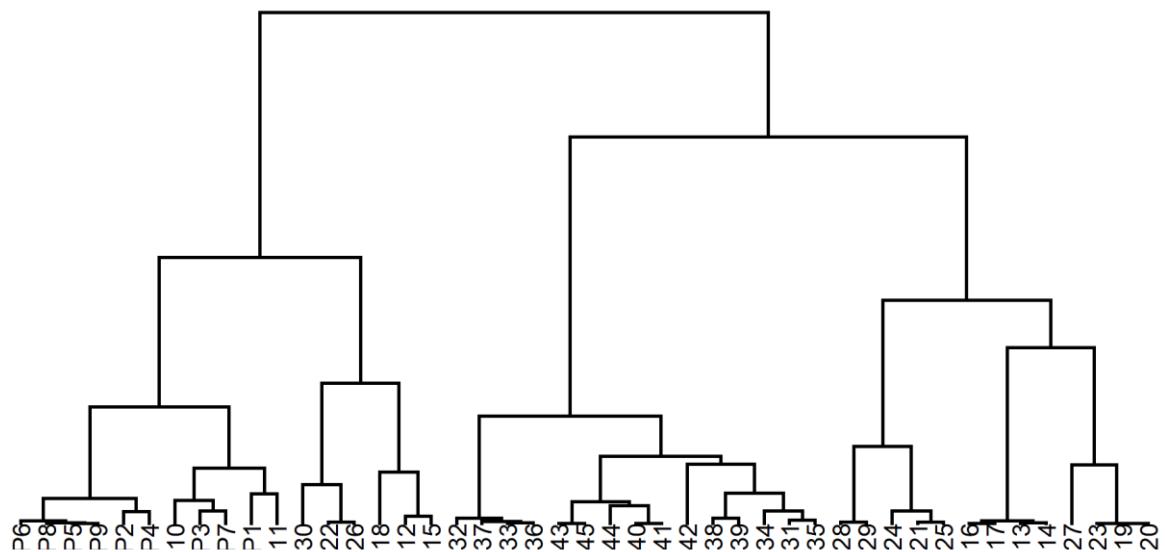


Carte de la plaine de la Limagne des points spatiaux classifiés par les vitesses du vent agrégées par années

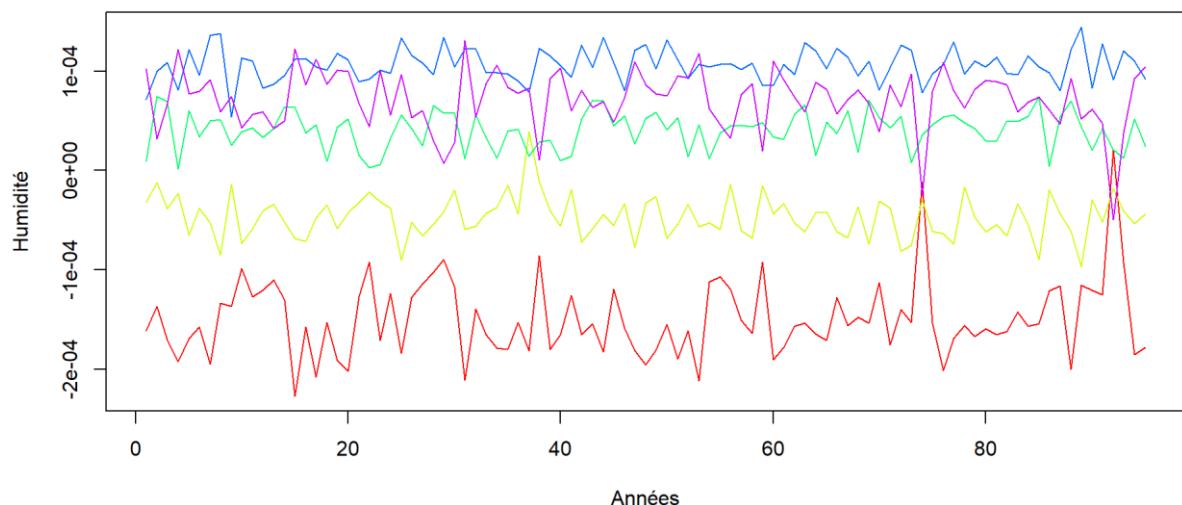


Humidités :

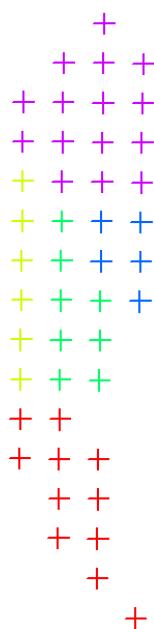
Dendrogramme des points spatiaux classifiés par l'humidités agrégées par années



Classification des courbes des différentes classes d'humidités agrégées par années



Carte de la plaine de la Limagne des points spatiaux classifiés par l'humidités agrégées par années



+

+++

+++

+++

+++

+++

+++

+++

++

++

++

++

++

++

++

++

++

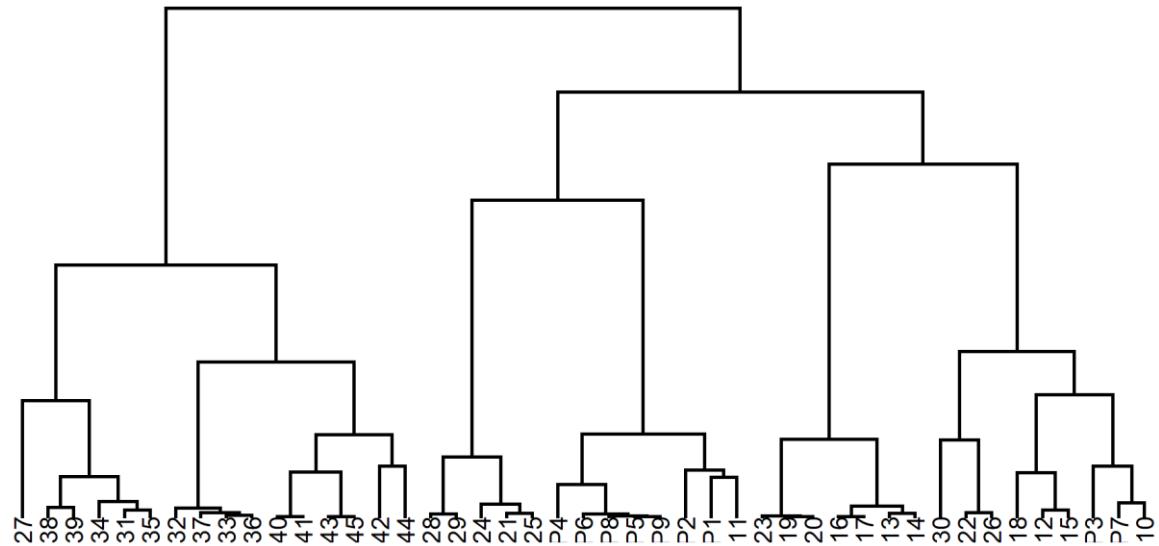
++

++

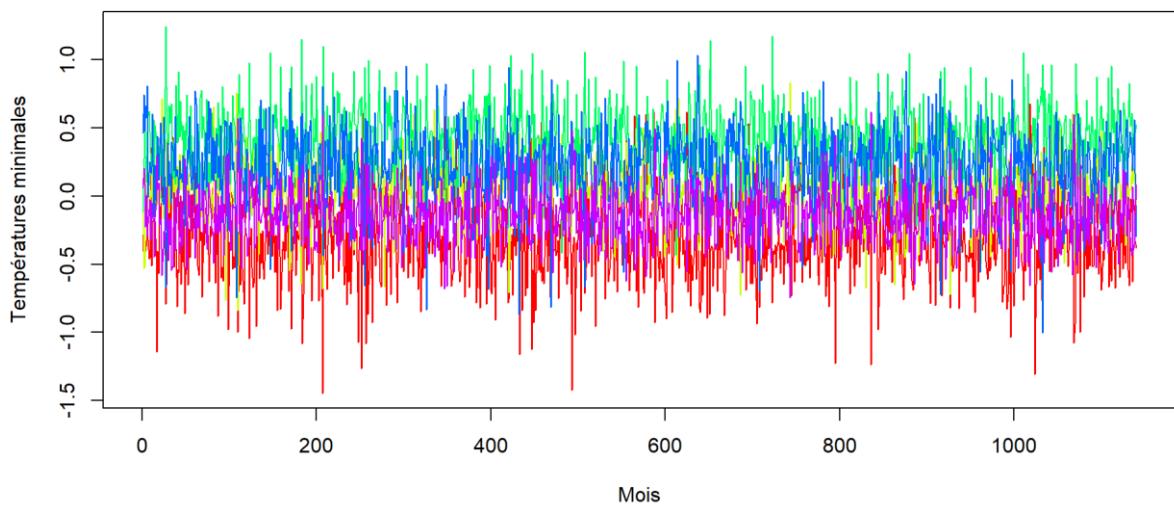
Données agrégées par mois

Températures minimales :

Dendrogramme des points spatiaux classifiés par les température minimales agrégées par mois



Classification des courbes des différentes classes des températures minimales agrégées par mois

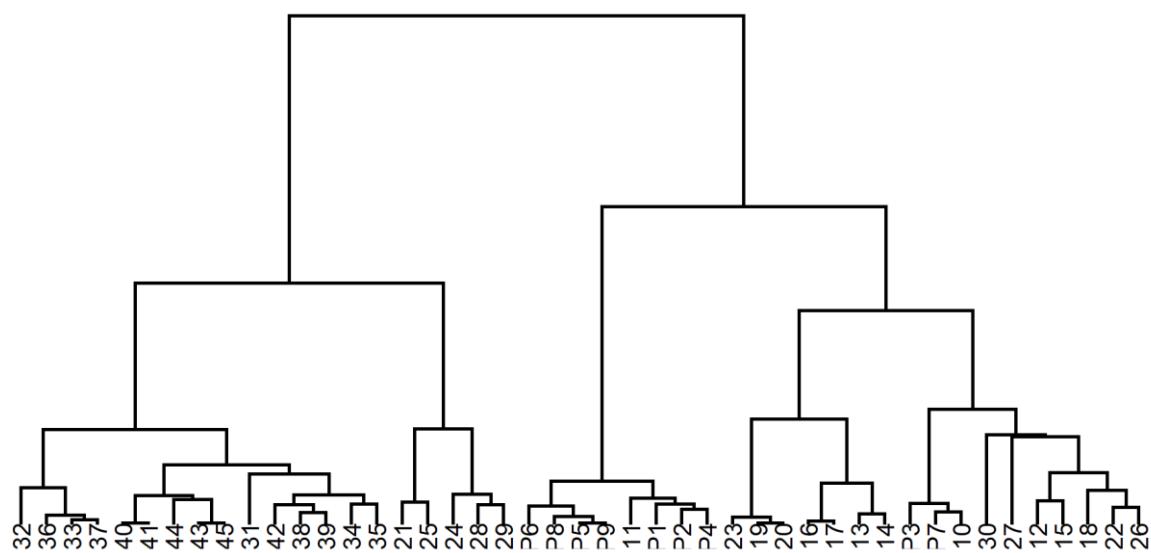


Carte de la plaine de la Limagne des points spatiaux classifiés par les températures minimales agrégées par mois

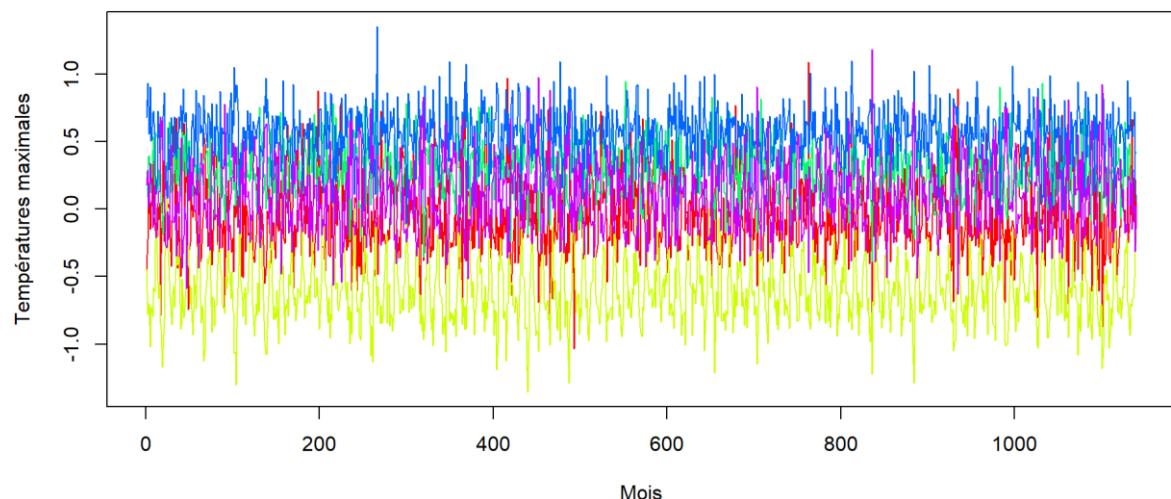


Températures maximales :

Dendrogramme des points spatiaux classifiés par les température maximales agrégées par mois



Classification des courbes des différentes classes des températures maximales agrégées par mois

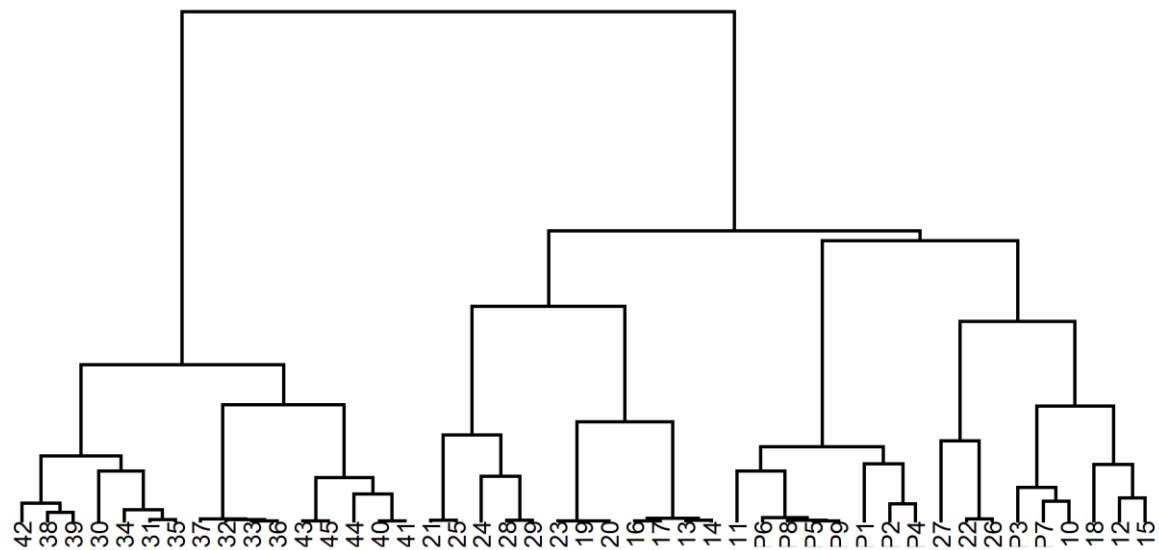


Carte de la plaine de la Limagne des points spatiaux classifiés par les températures maximales agrégées par mois

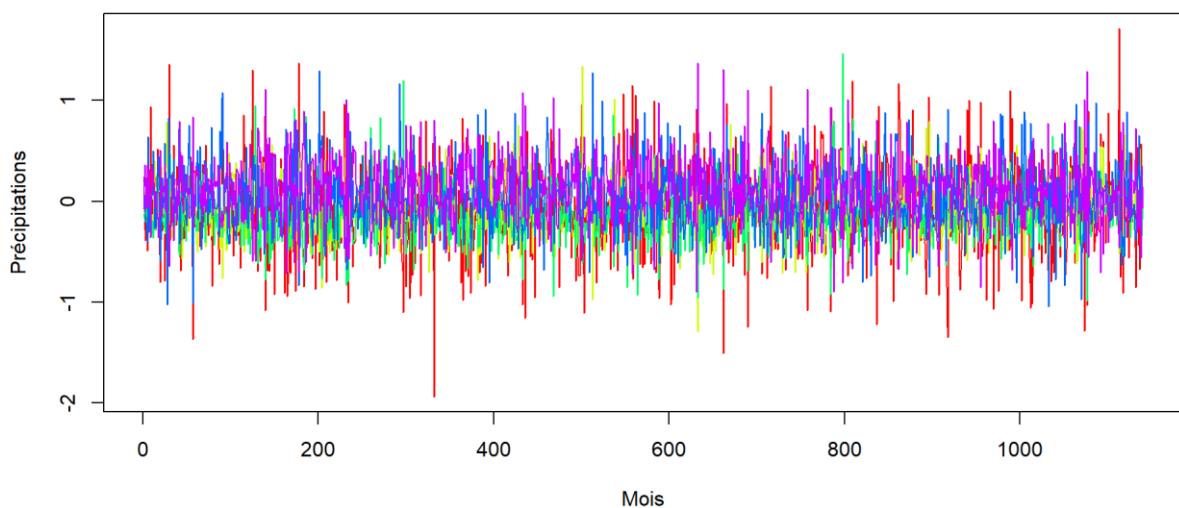


Précipitations :

Dendrogramme des points spatiaux classifiés par les précipitations agrégées par mois



Classification des courbes des différentes classes des précipitations agrégées par mois

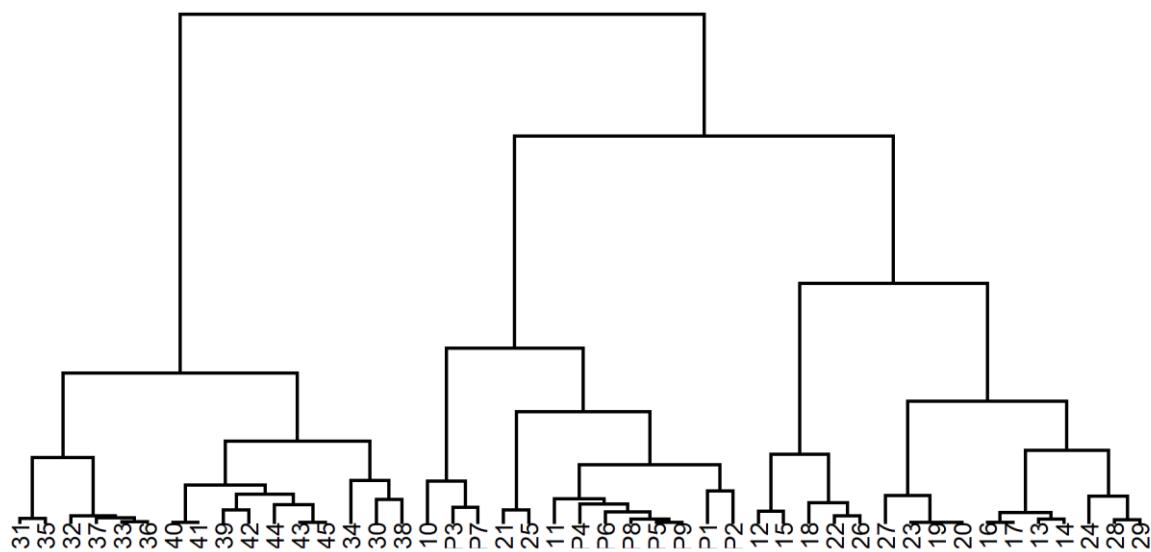


Carte de la plaine de la Limagne des points spatiaux classifiés par les précipitations agrégées par mois

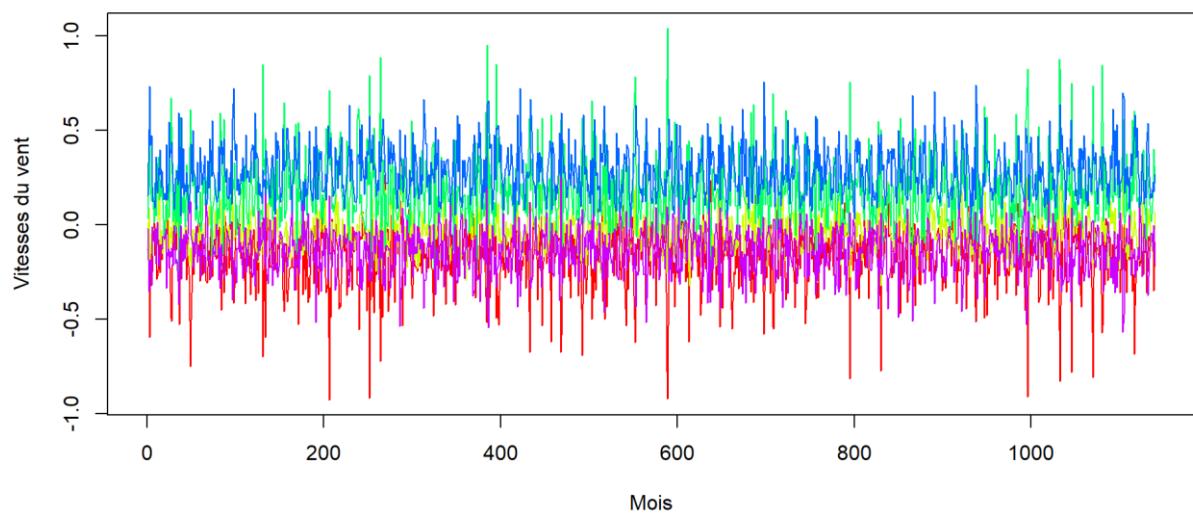


Vitesses du vent :

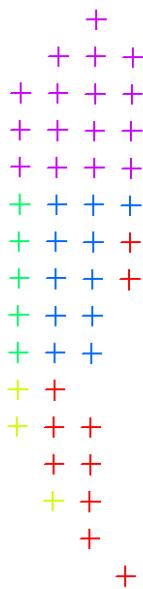
Dendrogramme des points spatiaux classifiés par les vitesses du vent agrégées par mois



Classification des courbes des différentes classes des vitesses du vent agrégées par mois

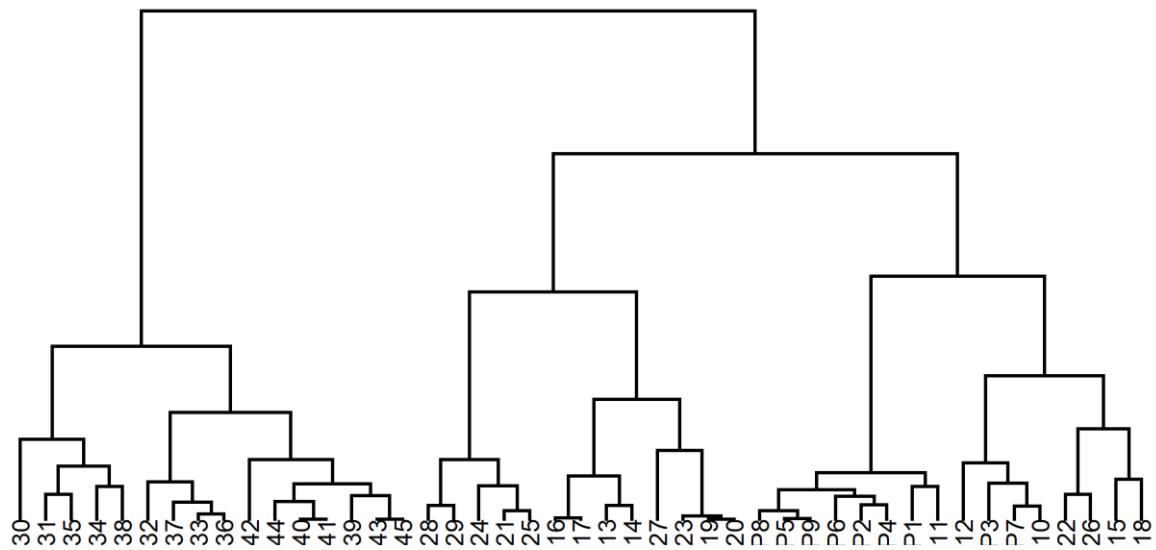


Carte de la plaine de la Limagne des points spatiaux classifiés par les vitesses du vent agrégées par mois

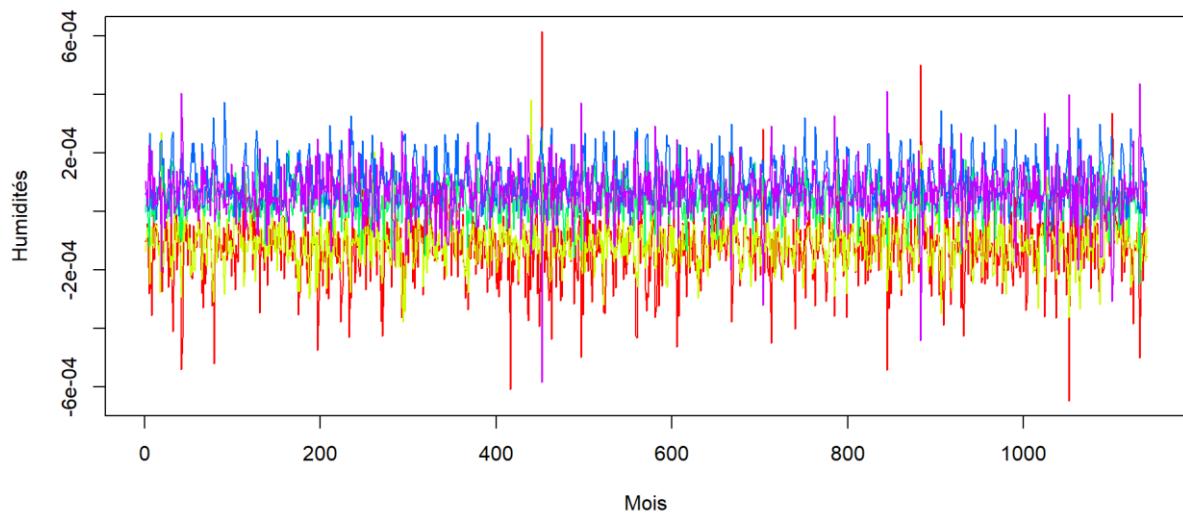


Humidités :

Dendrogramme des points spatiaux classifiés par l'humidités agrégées par mois



Classification des courbes des différentes classes d'humidités agrégées par mois



Carte de la plaine de la Limagne des points spatiaux classifiés par l'humidités du vent agrégées par mois

