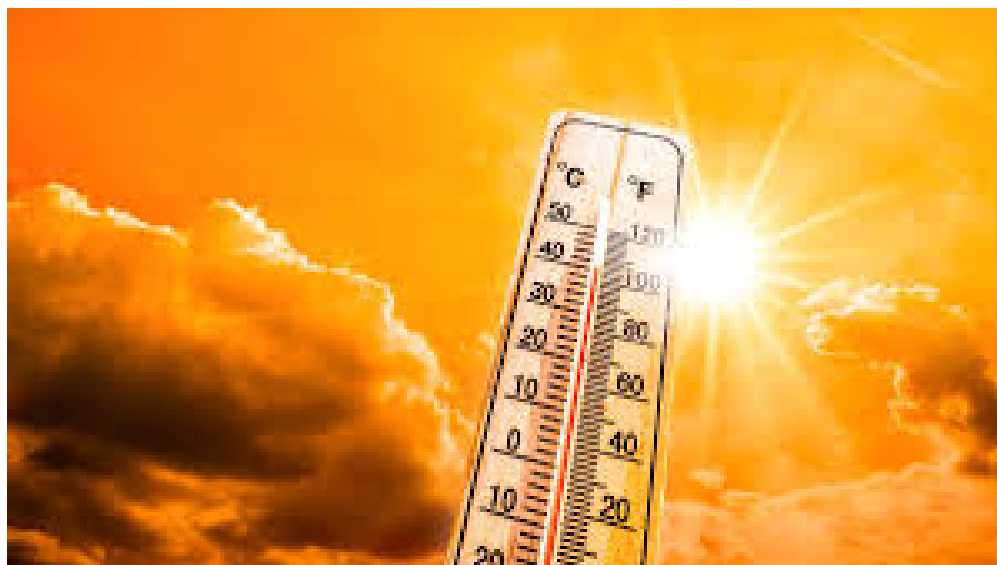


# Apprentissage statistique pour la Détection et Attribution de changement climatique : exemple de la vague de chaleur de septembre 2023

Projet d'initiation MAIN - Rapport final

*Rafaëlle Lacraz   Sara Grissia   Céline Sakr   Cassandra Martin*

*Référent : H. Durand*



Le 20 mai 2024

# Table des Matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Contexte . . . . .	3
1.1.1	Concepts clés de D&A . . . . .	3
1.1.2	Méthodes de D&A . . . . .	3
<b>2</b>	<b>Cahier des charges</b>	<b>8</b>
2.1	Présentation du projet . . . . .	8
2.1.1	Définition du Problème . . . . .	8
2.1.2	Objectifs . . . . .	9
2.1.3	Périmètre . . . . .	9
2.2	Organisation . . . . .	9
2.3	Besoins et contraintes . . . . .	10
<b>3</b>	<b>Résultats expérimentaux</b>	<b>11</b>
3.1	Caractéristiques de la vague de chaleur de septembre 2023 . . . . .	11
3.2	Attribution de la vague de chaleur de septembre de 2023 . . . . .	13
3.2.1	Analyse des résultats . . . . .	13
<b>4</b>	<b>Ouverture</b>	<b>19</b>
<b>5</b>	<b>Conclusion</b>	<b>22</b>
5.1	Bilan du projet et difficultés rencontrées . . . . .	22
5.2	Utilisation d'IA générative . . . . .	23
<b>6</b>	<b>Remerciements</b>	<b>23</b>
<b>7</b>	<b>Bibliographie</b>	<b>24</b>

# 1 Introduction

De nos jours, nous assistons de plus en plus fréquemment à des phénomènes climatiques extrêmes. Cela peut être des canicules (vagues et pics de chaleur), des incendies ou des feux de forêt, des inondations ou des pluies torrentielles, des cyclones, ou encore des ouragans. Ces événements peuvent persister aussi bien pendant plusieurs semaines ou des mois, que sur de très courtes périodes mais avec une intensité plus importante.

D'après le GIEC<sup>1</sup>, des preuves se sont accumulées ces dernières années concernant le changement de fréquence et d'intensité de ces événements climatiques extrêmes dans de nombreuses régions du monde. En effet, d'après leur dernier rapport [1], il est presque certain ("virtually certain") qu'il y ait eu une augmentation des vagues de chaleur et une diminution des vagues de froid extrêmes en Europe, et il est de plus en plus probable que ces phénomènes soient attribuables à l'activité humaine.

En parallèle, nous sommes témoins d'un réchauffement climatique qui prend de plus en plus d'ampleur dans notre quotidien. D'après le WWA<sup>2</sup>, les températures ont haussé de 1,1°C en moyenne depuis 1850 [2]. Cet accroissement est lié aux émissions massives de gaz à effet de serre. De plus, ce dérèglement climatique, qui ne cesse de s'accélérer et de s'accroître, est responsable de catastrophes humaines et matérielles [3].

Ici, on se propose d'étudier un événement climatique extrême récent : la vague de chaleur de septembre 2023 en France. Nous l'avons choisi car c'est un phénomène actuel qui nous a directement concernés, car il a, à ce jour, été peu étudié, et car il n'est pas encore très bien compris. D'après Météo-France (service officiel de la météorologie et de la climatologie en France depuis 1993), le mois de septembre 2023 a pulvérisé les records de chaleur sur l'ensemble du territoire métropolitain [4]. Ce mois s'est caractérisé par des températures très élevées presque continuellement sur l'ensemble de la France, et plus fortement dans le centre. La température moyenne nationale mensuelle était de 21,1°C, soit 3,6°C au dessus de la normale (1991-2020)<sup>3</sup>. Le précédent record était de 19,5°C en 1949. Pour résumer, ce mois de septembre est au premier rang des mois de septembre les plus chauds enregistrés depuis le début du XX<sup>e</sup> siècle, et plus globalement, l'année 2023 a été la seconde année la plus chaude du XX<sup>e</sup> siècle en France métropolitaine derrière 2022.

Pour étudier ce phénomène, nous allons utiliser des méthodes de détection et attribution. Dans un premier temps, une définition de ce domaine de recherche sera donnée, puis certaines de ses méthodes les plus communes seront abordées.

---

1. Le GIEC (en anglais IPCC) est un groupe d'experts intergouvernemental sur l'évolution du climat créé en 1988. Les rapports du GIEC, par leur effort de synthèse de la recherche actuelle, évaluent l'état des connaissances sur l'évolution du climat, ses causes et ses impacts, afin de trouver des solutions viables.

2. Le WWA (World Weather Attribution) est un consortium international de scientifiques fondé en 2014. Ses membres utilisent des observations et modèles climatiques pour attribuer des événements extrêmes au réchauffement climatique.

3. La normale est une moyenne arithmétique calculée sur une période de 30 ans. Elle caractérise le climat de cette période et sert de référence.

## 1.1 Contexte

### 1.1.1 Concepts clés de D&A

La détection et attribution (D&A) des changements climatiques consiste en la mise en évidence statistique des changements climatiques et l'étude de leurs causes. La détection cherche à démontrer que le climat a changé au sens statistique, sans donner de raison pour ce changement, tandis que l'attribution établit les causes les plus probables de ce changement. Le changement climatique peut être étudié à différentes échelles temporelles et spatiales. Ainsi, certaines recherches se concentrent sur l'évolution globale du climat sur de longues périodes, tandis que d'autres visent à caractériser des événements extrêmes dans des régions très localisées. Pour pouvoir comprendre l'application des méthodes de D&A qui seront décrites plus tard, il faut se familiariser avec plusieurs concepts clés qui sont indispensables :

- La notion de forçages externes qui sont les facteurs qui sont extérieur au système climatique (c'est à dire qu'ils l'affectent sans être affecté par celui en retour), comme par exemple les variations orbitales ou les variations solaires qui sont des mécanismes naturels, ou les perturbations dites anthropiques telles que les émissions de gaz à effet de serre, c'est à dire liées à l'activité humaine.
- La variabilité interne du climat, qui est la variabilité du système climatique en l'absence des forçages externes.
- Les modèles climatiques, basés sur la résolution numérique des équations de la physique, qui vont servir à faire des projections climatiques avec ou sans forçages externes, et sont donc importants pour certaines méthodes de D&A qui vont se baser sur ces projections, et pas seulement sur des observations. Ils sont donc souvent utilisés pour prédire les réponses aux forçages externes, sous la forme de cycles de variation de l'espace, du temps ou des deux.

### 1.1.2 Méthodes de D&A

Dans cette partie, nous allons brièvement parler des différentes méthodes fréquemment utilisées dans la détection et attribution des événements extrêmes.

Les méthodes de détection reposent généralement sur l'analyse statistique de données climatiques, qu'il s'agisse d'observations ou de données provenant de modèles climatiques. L'approche la plus courante repose sur des tests d'hypothèse. Après avoir défini un événement climatique extrême, on estime sa distribution dans un climat sans influence anthropogénique (en utilisant des modèles climatiques ou des données historiques) et on détecte un changement dans la fréquence ou l'intensité de cet événement si celui-ci s'éloigne suffisamment de cette distribution (rejet de l'hypothèse nulle).

Nous retrouvons aussi l'utilisation de méthodes d'*analyse des séries temporelles* dans le

but d'identifier des variations à long terme à partir de données collectées sur une période de temps.

Une fois qu'un changement a été détecté concernant un événement extrême, nous cherchons à en déterminer les causes, c'est-à-dire à attribuer ce changement à des facteurs externes tels que l'activité humaine, la variabilité climatique interne ou d'autres types de forçages naturels (comme les radiations solaires, par exemple). Pour cela, nous avons recours aux méthodes d'attribution ci-dessous (voir par exemple [5]) :

**Méthodes reposant sur l'observation des données** : Ces méthodes se basent principalement sur des observations pour identifier les signes anthropiques. Nous retrouvons diverses approches, qu'on ne citera pas toutes.

- *Méthodes utilisant des schémas spatiaux pour séparer la variabilité interne et les forçages externes [6]* : La filtration des variations dans le schéma d'anomalies sur terre et en mer de signes opposés, connu sous le nom de schéma "Cold Ocean Warm Land" (COWL), permet d'identifier une tendance résiduelle de température à grande échelle. Cette tendance est moins influencée par la variabilité dynamique hivernale, offrant ainsi une réflexion plus claire des réponses aux forçages. Ce schéma se forme car les dynamiques hivernales fortes, par exemple, associées à des vents d'ouest plus forts ou plus faibles que la moyenne, déplacent plus ou moins de chaleur de la couche mélangée de l'océan vers la terre, entraînant des anomalies positives ou négatives de la température moyenne dans l'hémisphère nord. Les schémas spatiaux observés de changement peuvent aussi servir à examiner les données à la recherche des mécanismes physiques sous-jacents à un changement climatique.

Cependant, seules des simulations de modèles seraient en mesure de confirmer de manière fiable si les changements hypothésisés peuvent expliquer les résultats obtenus.

- *Méthodes de séparation du signal et du bruit basées sur les échelles de temps* : Les chercheurs utilisent ces méthodes pour différencier le forçage externe à long terme, tel que les émissions de gaz à effet de serre, et les changements dynamiques à court terme en se concentrant sur les différences d'échelle de temps. Cette approche vise à identifier les tendances à long terme dans les séries temporelles souvent dominées par des dynamiques à court terme. Elle suppose que les changements des agents externes, en particulier l'augmentation des gaz à effet de serre, conduisent à des schémas statistiquement distincts des schémas à plus court terme enregistrés en raison de la variabilité historique générée au sein du système climatique. C'est par exemple sur cette hypothèse que repose la méthode *Low-Frequency Component Analysis* (voir [7]).

Cependant, la variabilité climatique interne est présente sur toutes les échelles de temps, y compris à long terme, et donc la séparation du changement anthropique sera incomplète.

- *Analyses basées sur les analogues* : Une autre approche consiste à prendre en compte les caractéristiques dynamiques de la circulation observée pour conditionner les résultats sur les conditions climatiques présentes lors d'un certain événement. Les analogues de circulation ont été développés dans le but d'estimer les conditions climatiques passées sous une circulation à grande échelle similaire à celle observée aujourd'hui. Cette méthode offre une manière d'explorer l'impact du changement climatique sur les événements inhabituels. Du point de vue statistique, les températures analogues sont considérées comme des répliques aléatoires de la température du jour conditionné par la circulation atmosphérique (qui pourra par exemple être estimé en utilisant les niveaux de pression atmosphérique), ce qui permet de déterminer les distributions de probabilité de la variabilité de la température influencée par la circulation atmosphérique. Une fois que les distributions de température conditionnées par la circulation sont obtenues, il est possible de calculer la probabilité de l'événement étudié.
- *Approches empiriques* : Appliquée directement aux observations, cette méthode a été utilisée pour estimer comment le changement climatique affecte la probabilité ou les intervalles de retour de certaines classes d'événements. Les chances de températures record peuvent être liées à l'augmentation des températures moyennes. Les estimations résultantes indiquent qu'un réchauffement à long terme a entraîné une augmentation de cinq fois du nombre d'extrêmes mensuels de température record dans le monde, ce qui est largement cohérent avec les estimations utilisant les modèles climatiques. Les approches empiriques offrent un moyen de vérifier les résultats des modèles climatiques et sont souvent valorisées par les utilisateurs qui doutent de la fiabilité de ces modèles, même si elles reposent également sur des hypothèses. De plus, elles peuvent être appliquées à des événements qui ne sont pas encore bien pris en compte par les modèles climatiques utilisés pour l'attribution, comme les événements convectifs extrêmes en été.

**Méthodes avec modèles climatiques** [8] Dans cette partie, nous aborderons les deux méthodes les plus couramment utilisées qui font appel aux modèles climatiques. Pour comprendre les méthodes, il faut se familiariser avec la notion de Fraction of Attributable Risk (FAR), également connu sous le nom de fraction du risque attribuable. Le concept a été appliqué pour la première fois en 2004 dans une analyse de la canicule européenne de 2003 (voir [9]). Cette étude a fourni un lien direct entre le changement climatique anthropique et un événement climatique extrême individuel. Pour parvenir à ce résultat, la probabilité ( $P_1$ ) d'un été exceptionnellement chaud dans une région européenne particulière a été comparée à sa probabilité ( $P_0$ ) si les influences anthropiques sur le climat avaient été absentes.

L'étude a conclu que l'influence humaine avait très probablement (probabilité  $> 90\%$ ) plus que doublé la probabilité d'un été exceptionnellement chaud. Par conséquent, après avoir calculé les probabilités de l'événement en présence et en l'absence de changement climatique anthropique,  $P_1$  et  $P_0$ , les résultats peuvent être exprimés sous forme de FAR, calculée comme  $FAR = 1 - P_0/P_1$  où un FAR de plus de 0.5, comme dans ce cas, indique que sa probabilité a plus que doublé.

Les estimations de  $P_1$  et  $P_2$  reposent respectivement sur la simulation de larges ensembles dans les conditions actuelles (avec influence anthropogénique) ou contrafactuelles (sans influence anthropogénique), c'est-à-dire en perturbant les conditions initiales de chaque simulation afin d'obtenir une distribution sur la probabilité d'occurrence de chaque événement. Contrairement à l'attribution de tendances globales (par exemple, le réchauffement terrestre global), l'estimation des distributions  $P_1$  et  $P_0$  nécessite un grand nombre de simulations pour obtenir des estimations robustes.

Nous décrivons maintenant les principales méthodes utilisées.

- *Approche des modèles couplés* : Ces modèles de circulation générale couplés (GCM), de complexité croissante, qui incluent souvent non seulement l'atmosphère, l'océan et les terres, mais aussi les processus biologiques et chimiques, fournissent les simulations les plus complètes du système climatique.

Il s'agirait alors de regrouper des données à partir d'ensembles de simulations multi-modèles avec et sans influences anthropiques, ce qui génère de grands échantillons de la variable climatique pertinente (la température dans notre cas).

La distribution de la variable dans le monde "réel"  $P_1$  et le monde "naturel" contrafactuel  $P_0$  sans influence humaine sur le climat peut ainsi être construite, à partir de laquelle des estimations de la fraction du risque attribuable (FAR) pour l'événement étudié sont obtenues.

Cette méthode d'attribution est utilisée pour fournir des évaluations rapides, disponibles dès qu'un événement extrême est observé. La probabilité changeante des extrêmes est estimée par rapport à des seuils prédéfinis.

- *Approches des modèles forcés par les températures de surface de la mer* : Une autre façon de conditionner les résultats sur les aspects des conditions climatiques présentes au moment de l'événement est de prescrire les anomalies observées de température de surface de la mer (TSM) dans un modèle climatique atmosphérique uniquement.

Ainsi, de nombreuses études d'attribution d'événements contrastent les simulations de modèles de circulation générale atmosphérique seulement (AGCM) représentant le monde "réel" incluant l'évolution observée des TSM avec des simulations du monde contrafactuel "naturel", un "monde qui aurait pu être", s'il n'y avait eu aucune influence humaine sur le climat.

Comme avec les approches des modèles couplés, cette méthodologie nécessite également la disponibilité de ensembles de données de modèles suffisamment importants pour simuler les statistiques des événements en question, et repose sur la capacité du modèle à simuler de manière fiable les conditions climatiques générant l'événement

extrême.

Prescrire les TSM dans un AGCM plutôt que d'utiliser des modèles couplés peut réduire les biais du modèle et permettre de simuler plus de membres d'ensemble car ils sont moins coûteux à exécuter, ce qui peut potentiellement conduire à une meilleure représentation des événements extrêmes et un meilleur rapport signal-bruit.

Cependant, cette approche ne représente pas le couplage atmosphère-océan et pourrait donc conduire à une moins bonne représentation des événements extrêmes fortement affectés par ce couplage.

**Critères d'évaluation des modèles :** L'attribution des événements extrêmes dépend fortement de la qualité des simulations des modèles climatiques, ce qui nécessite une validation approfondie pour assurer la crédibilité des résultats. Il est essentiel que ces modèles soient capables de reproduire fidèlement une gamme diversifiée d'extrêmes, adaptée à des événements spécifiques.

Cette validation implique généralement une comparaison minutieuse entre les distributions d'événements simulées par le modèle et les données observationnelles. Des tests statistiques peuvent être utilisés pour évaluer l'accord entre les deux sur une période spécifique, tandis que la cohérence de la variabilité simulée avec les observations est également examinée. Ces évaluations sont cruciales pour déterminer la pertinence du modèle dans l'analyse des événements extrêmes, fournissant ainsi des informations précieuses pour les analyses de détection et d'attribution.

De plus, une évaluation approfondie implique également une comparaison entre les données des modèles et les observations pour garantir la cohérence des différentes distributions d'événements extrêmes. La fiabilité des données observationnelles est indispensable pour attribuer de manière fiable l'influence humaine sur les événements de température extrême. Cependant, les limitations dans les observations et les modèles peuvent compromettre les attributions pour les événements de précipitations extrêmes. Ainsi, l'amélioration continue des méthodes d'évaluation renforcera l'utilité des produits d'attribution pour les parties prenantes.



## 2 Cahier des charges

### 2.1 Présentation du projet

Dans le cadre de ce projet nous allons étudier la vague de chaleur de 2023, à travers des recherches sur l’attribution des événements extrêmes. Le mois de septembre 2023 a été le plus chaud jamais enregistré en France, à cause d’une canicule tardive, en effet la température de ce mois de septembre a été d’environ  $3,5^{\circ}\text{C}$  au dessus des moyennes (voir [10]).

Ce genre d’évènement est étudié grâce à la science de l’attribution, qui correspond à l’étude des rapports de causalité entre le dérèglement climatique dû aux activités humaines et les événements climatiques, dans ce notre cas, extrêmes. Ces études sont importantes pour plusieurs domaines : en politique cela permet de guider les prises de décisions ou encore les scientifiques pour expliquer d’autres phénomènes. La science de l’attribution peut aussi affecter les litiges liés au dérèglement climatique en permettant des poursuites contre les entreprises pour avoir causé le changement climatique ou contre les gouvernements pour avoir échoué à le prévenir. Nous pouvons aussi citer d’autres personnes directement impactées telles que les écologistes, les économistes et le grand public sont tous concernés par la vague de chaleur et son attribution, car elle a des implications majeures pour la santé publique, l’agriculture, l’économie et l’environnement.

Actuellement, les scientifiques ont une compréhension solide des mécanismes généraux de la vague de chaleur et de son lien avec le changement climatique. Cependant, l’attribution précise des événements individuels à des facteurs spécifiques reste un défi, bien que des progrès significatifs aient été réalisés grâce à des techniques avancées de modélisation et d’analyse statistique. Les lacunes persistent dans notre capacité à attribuer avec certitude des événements extrêmes à des causes spécifiques, en particulier à l’échelle locale. De plus, il reste des incertitudes quant à l’ampleur exacte de l’impact du changement climatique sur la fréquence et l’intensité des vagues de chaleur.

Malgré tout cette recherche vaut la peine d’être entreprise, elle est cruciale car elle fournit des informations essentielles pour informer les politiques du changement climatique. Comprendre comment les vagues de chaleur sont liées au dérèglement climatique peut également aider à préparer les sociétés à faire face à de telles crises à l’avenir. En outre, cela permettrait de mieux éduquer le public sur les impacts du changement climatique et de promouvoir des actions visant à réduire les émissions de gaz à effet de serre.

#### 2.1.1 Définition du Problème

Les principales problématiques sont :

- Quelles sont les causes de la vague de chaleur ?
- L’activité humaine est-elle une cause ?
- Si oui, dans quelle mesure ?

Ces questions pourront évoluer au long du projet.

## 2.1.2 Objectifs

Lors de ce projet nous allons effectuer une revue de littérature afin de développer une compréhension des phénomènes climatiques ayant engendré la vague de chaleur de l'été 2023 et développer une compréhension des méthodes et concepts d'attribution des phénomènes climatiques extrêmes. De plus, notre objectif est d'utiliser des données climatiques (simulations et/ou observations) et appliquer les méthodologies étudiées pour chercher à répondre aux questions de recherche. Pour finir, nous voulons rapporter les résultats obtenus à travers un rapport constituée tout au long du projet et une présentation orale finale.

## 2.1.3 Périmètre

Nous pouvons définir plusieurs limites pour ce projet, nous nous concentrons exclusivement sur la vague de chaleur survenue en France en 2023. Les analyses et les conclusions seront spécifiques à cette région géographique. L'étude se limite au mois de septembre 2023, les données et les analyses seront basées sur cette période précise.

Le projet utilisera principalement des simulations climatiques pour évaluer l'attribution de la vague de chaleur de 2023 en France. Plus précisément, la méthode de Fraction of Attributable Risk (FAR) sera employée dans un premier temps pour estimer la contribution du changement climatique à l'intensité et à la fréquence de l'événement.

Bien que la méthode principale repose sur des simulations climatiques, le projet reste ouvert à l'intégration de données d'observations pour valider et compléter les résultats obtenus par les simulations.

## 2.2 Organisation

Pour notre organisation nous utilisons un diagramme de Gantt pour se répartir les différentes tâches. Pour la partie informatique de ce projet nous utilisons des notebooks sur Google Colab pour que monsieur Durand puissent suivre notre avancée.

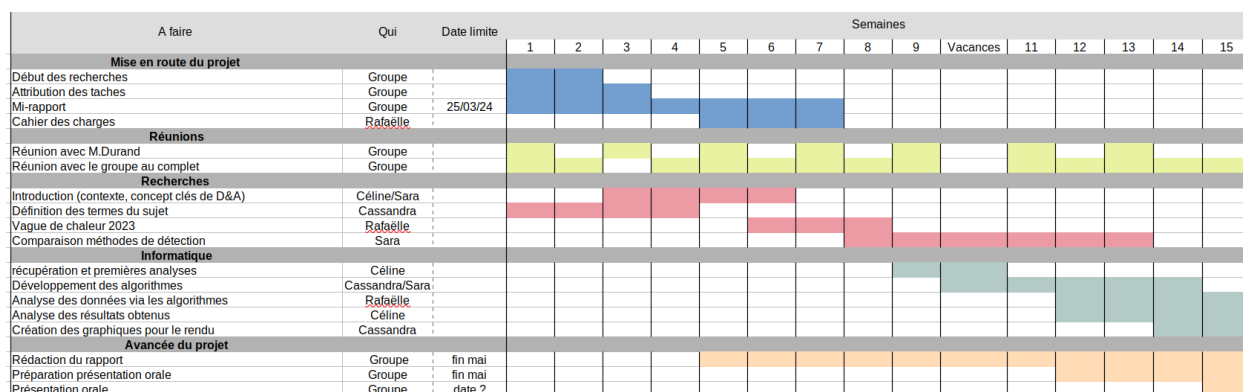


FIGURE 1 – Diagramme de Gantt

## 2.3 Besoins et contraintes

Pour chaque tâches du diagramme précédent nous avons imaginé des besoins et des contraintes présentés dans le tableau suivant.

Objectifs	Besoins	Contraintes	Ressources
Rapport écrit et bibliographie	Écrire le rendu en LaTeX	Partage d'un document en LaTeX	Utilisation de Overleaf
Développement des algorithmes	Connaissances en python	Partage des algorithmes	Utilisation de Google Collabs
Analyse des données via les algorithmes	Données climatiques et Algorithmes de D&A	Temps et accès aux données	Données de simulation climatique fournies, Données d'observation en libre accès
Analyse des résultats obtenus et production des graphiques	Utilisation des sources bibliographiques et des données accumulées tout au long du projet		Données de simulation climatique, comparaison avec d'autres études
Présentation orale	Graphiques résultats		

TABLE 1 – Besoins et contraintes

### 3 Résultats expérimentaux

Pour commencer cette partie informatique du projet, nous avons extrait un 1er jeu de données depuis Copernicus ([11]), une plateforme d'accès libre à des données climatiques.

#### 3.1 Caractéristiques de la vague de chaleur de septembre 2023

Nous avons commencé par travailler sur un jeu de données regroupant les températures maximales annuelles de 1950 à 2023 en France.

Ce premier jeu nous a permis de nous familiariser avec le langage python ainsi qu'avec la détection et la modélisation. Ces données sont de ré-analyse, c'est à dire qu'elles combinent des observations satellites avec des simulations climatiques qui permettent de combler les données manquantes et d'obtenir un ensemble plus homogène et approprié à l'analyse statistique.

Nous avons voulu représenter ces températures sur une carte de France afin de regarder la cohérence des données.

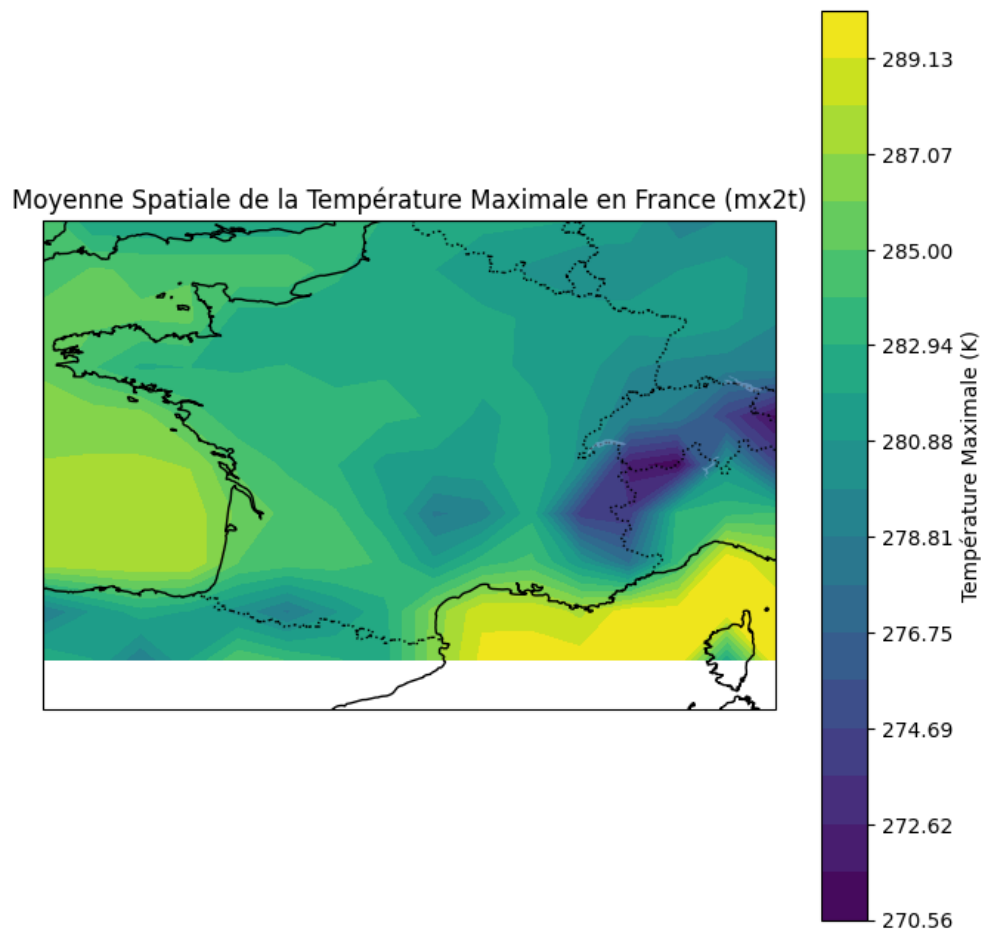


FIGURE 2 – Moyenne spatiale de la température maximale en France

Cette carte semble logique quant aux zones les plus chaudes et les plus froides en France. En effet nous pouvons observer que les zones de mers et d'océan sont plus chaudes que les zones terrestres à cause de la grande capacité thermique de l'eau. De plus, les zones bleues sont les plus froides, il s'agit des zones montagneuses, tel que les Alpes, les Pyrénées ou encore le Massif Central, car la pression atmosphérique y est plus faible.

Par la suite, nous avons représenté cette même base sur un graphique représentant les températures maximales moyennes par an pour observer les tendances.

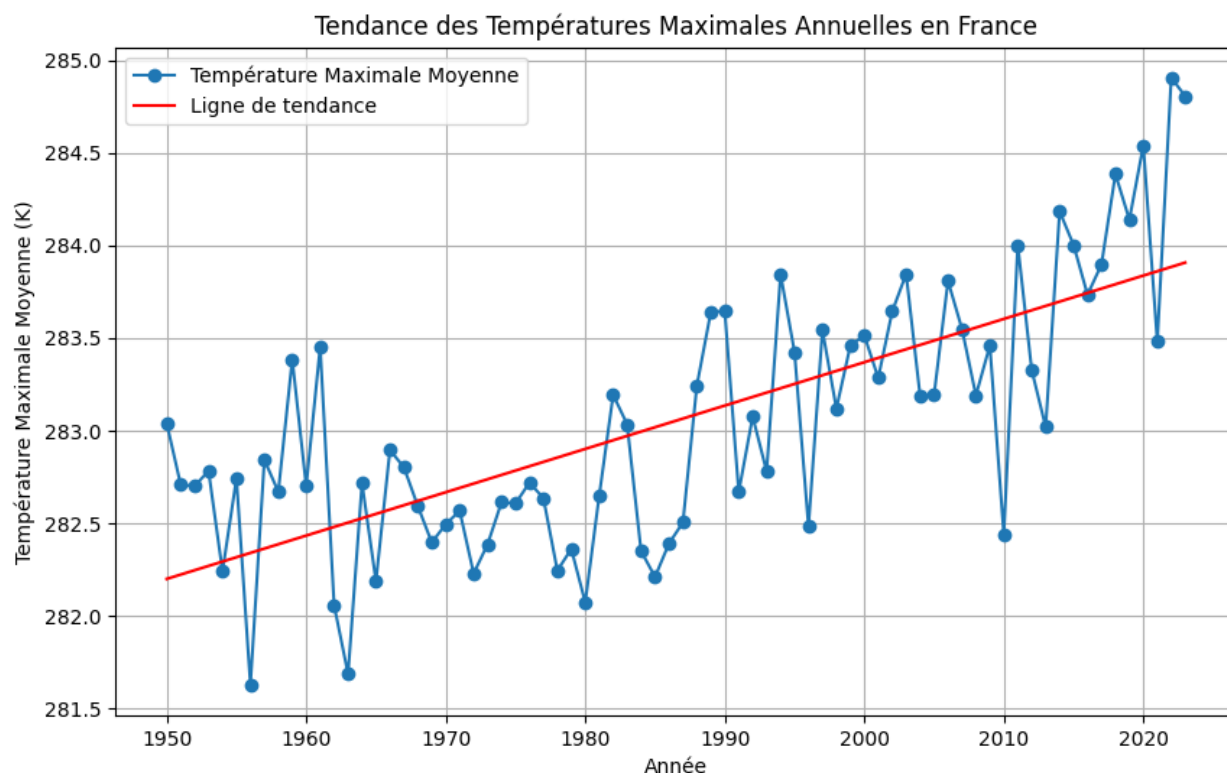


FIGURE 3 – Tendance des températures maximales annuelles en France

Nous pouvons conclure que la température augmente d'année en année et ce de plus en plus rapidement.

Ce jeu de données étant observationnel, une analyse d'attribution requiert des outils méthodologiques statistiques complexes. Une approche utilisant des simulations est dans un premier temps plus accessible, pour découvrir et comprendre le fonctionnement de la détection et attribution.

Pour compléter nos recherches et poursuivre notre projet, notre référent, M.Durand, nous a fourni de nouvelles données que nous traiterons dans la partie suivante.

## 3.2 Attribution de la vague de chaleur de septembre de 2023

Après avoir bien pris en main le premier jeu de données, il a été possible d'étudier de nouvelles données pour analyser la vague de chaleur de septembre 2023.

Pour ce faire, nous avons décidé d'exploiter une méthode avec modèle climatique (1.1.2) puisque cette méthode est idéale pour établir un lien entre cette vague de chaleur et l'influence anthropique sur le climat.

Nous avons donc utilisé pour cette étude deux jeux de données semblables : le premier avec des simulations historiques sur la période 1850-1900 pour caractériser l'état du climat sans influence humaine, et le second avec des simulations selon le scénario de projection climatique pour caractériser l'état du climat sous influence humaine.

Ce modèle de projection suit une trajectoire définie par le dernier rapport du GIEC sous le scénario SSP2-4.5. Ce scénario représente une projection considérée comme la plus probable, prévoyant un réchauffement climatique compris entre 2 et 4,5 degrés Celsius au-dessus des niveaux pré-industriels ([12]). Ces données représentent les températures moyennes mensuelles. En les comparant, nous pouvons étudier l'impact de l'influence humaine.

### 3.2.1 Analyse des résultats

Dans un premier temps nous avons voulu regarder la probabilité d'avoir des températures dépassant le seuil de canicule en septembre entre 1850 et 1900.

Nous avons fixé ce seuil à 295 Kelvin soit 21,85 degrés celsius, ce seuil peut paraître trop élevé pour des moyennes mensuelles du mois de septembre cependant nous avons fait ce choix car il est utilisé dans d'autres études de ce genre.

```
# Calcul de la probabilité de tomber sur des jours avec température maximale dépassant le seuil de canicule
p0 = days_above_threshold / total_days

# Affichage de la probabilité
print("La probabilité d'avoir des températures mensuelles en septembre dépassant le seuil de canicule est :", p0)

La probabilité d'avoir des températures mensuelles en septembre dépassant le seuil de canicule est : 0.02430971128608924
```

FIGURE 4 – Probabilité d'avoir des des températures dépassant le seuil de canicule

Nous pouvons voir que la probabilité qu'il y ait une canicule en septembre est de 0.024, il y a donc très peu de chance qu'une vague de chaleur se produise entre 1850 et 1900 donc avant l'arrivée de l'ère industrielle.

Ensuite, nous avons fait cette même analyse avec les données allant de 2020 à 2026 à l'aide d'un *mask* dans notre fonction.

```
# Filtrage des données pour les mois de septembre des années 2020 à 2026
mask_september = (ds['time'].dt.year >= 2020) & (ds['time'].dt.year <= 2026) & (ds['time'].dt.month == 9)
```

FIGURE 5 – Variable *Mask* utilisée pour filtrer

Nous avons affiché l'éventualité qu'il y ait une canicule en septembre sur cette période. Nos résultats sont cohérents car la probabilité est de 0.75. Cette dernière est donc beaucoup plus importante que sans influence humaine (avant l'ère industrielle).

```
# Affichage des résultats
print(f'Nombre total de jours en septembre de 2020 à 2026: {total_days_sept}')
print(f'Nombre de jours de septembre de 2020 à 2026 avec température maximale dépassant le seuil de canicule: {days_above_threshold_sept}')
print(f'Probabilité de dépasser le seuil de canicule en septembre de 2020 à 2026: {prob_sept:.2f}')

Nombre total de jours en septembre de 2020 à 2026: 224
Nombre de jours de septembre de 2020 à 2026 avec température maximale dépassant le seuil de canicule: 167
Probabilité de dépasser le seuil de canicule en septembre de 2020 à 2026: 0.75
```

FIGURE 6 – Probabilités pour les mois de septembre de 2020 à 2026

De plus, la fraction de risque attribuable est de 0.97, et d'après la catégorisation du GIEC, une FAR comprise entre 0.9 et 1 veut dire qu'il est très probable que la canicule soit causée par l'influence humaine. En effet, si le risque attribuable est égal à 1 alors l'évènement est uniquement dû à l'influence humaine.

```
far = 1-(p0/prob_sept)

print("Fraction de Risque Attribuable (FAR) :", far)

Fraction de Risque Attribuable (FAR) : 0.9673929621072815
```

FIGURE 7 – Fraction risque attribuable pour 2020-2026

Ensuite, nous nous sommes focalisées sur le mois de septembre 2023 pour essayer de voir si la probabilité d'avoir une canicule était proche ou pas de la probabilité trouvée pour tous les mois de septembre groupés.

L'approche était similaire à celle utilisée précédemment, avec quelques modifications. La probabilité d'une canicule trouvée est de 0.78.

Nous pouvons remarquer que c'est supérieur à la probabilité sur toute la période 2020-2026, ce qui nous amène à faire la même réflexion que le paragraphe précédent par rapport à la catégorisation du GIEC.

Cette analyse nous permet de supposer que le mois de septembre 2023 que nous cherchons à étudier est en effet un mois exceptionnel en terme de températures moyennes.

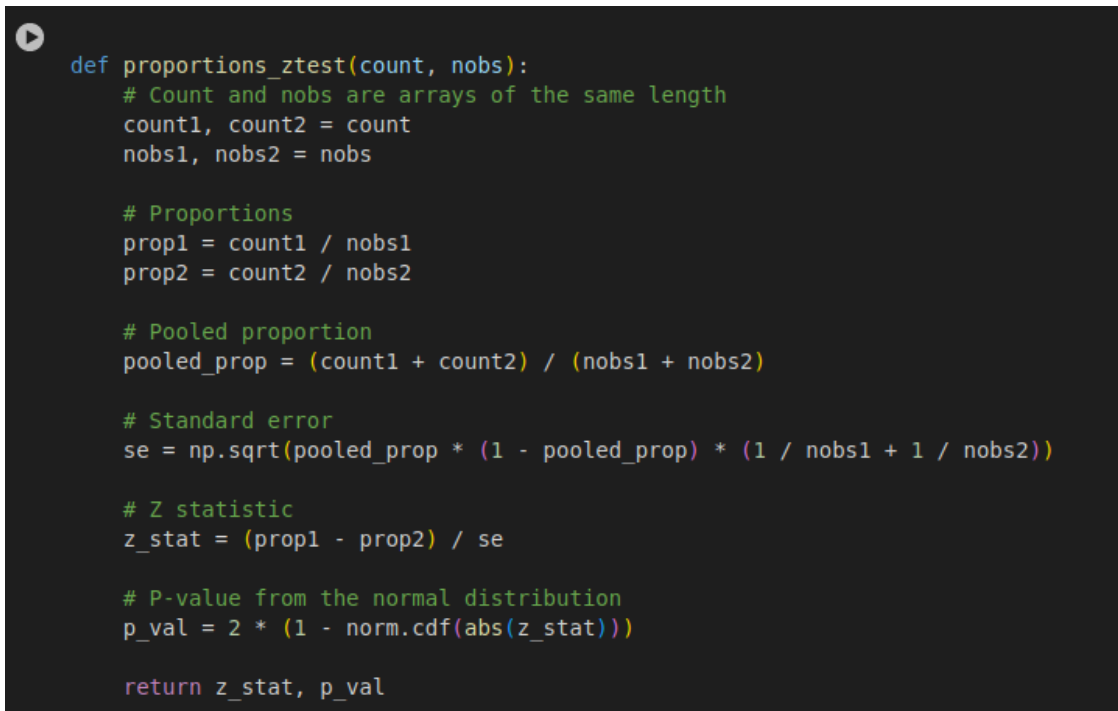
```
# Affichage des résultats
print(f'Nombre de jours de septembre 2023 avec température maximale dépassant le seuil de canicule: {days_above_threshold_sept_2023}')
print(f'Probabilité de dépasser le seuil de canicule en septembre 2023: {prob_sept_2023:.2f}')

Nombre de jours de septembre 2023 avec température maximale dépassant le seuil de canicule: 25
Probabilité de dépasser le seuil de canicule en septembre 2023: 0.78
```

FIGURE 8 – Probabilité pour les jours du mois de septembre 2023

Nous avons cherché à déterminer si les températures maximales enregistrées en septembre 2023 dépassent significativement un seuil de canicule par rapport aux températures enregistrées en septembre des autres années comprises entre 2020 et 2026. Pour pouvoir valider notre hypothèse, nous avons utilisé un test de proportion, pour évaluer l'importance de la différence observée, ce qui nous permettra de discuter de la présence d'une vague de chaleur exceptionnelle en septembre 2023 et de l'influence humaine sur les événements climatiques extrêmes.

Pour cela, nous avons calculé la probabilité de tomber sur des jours dépassant le seuil pour septembre 2023 puis la probabilité combinée pour les autres années. Ensuite, nous avons implémenté le test qui compare les proportions de jours dépassant le seuil de canicule entre septembre 2023 et les autres années. La signification statistique de la différence est déterminée par la p-value obtenue du test de proportion pour l'hypothèse nulle  $H_0 : p_0 = p_1$ .



```
def proportions_ztest(count, nobs):  
    # Count and nobs are arrays of the same length  
    count1, count2 = count  
    nobs1, nobs2 = nobs  
  
    # Proportions  
    prop1 = count1 / nobs1  
    prop2 = count2 / nobs2  
  
    # Pooled proportion  
    pooled_prop = (count1 + count2) / (nobs1 + nobs2)  
  
    # Standard error  
    se = np.sqrt(pooled_prop * (1 - pooled_prop) * (1 / nobs1 + 1 / nobs2))  
  
    # Z statistic  
    z_stat = (prop1 - prop2) / se  
  
    # P-value from the normal distribution  
    p_val = 2 * (1 - norm.cdf(abs(z_stat)))  
  
    return z_stat, p_val
```

FIGURE 9 – Fonction test de proportion

La fonction *proportions ztest* calcule la statistique z et la p-value pour un test de proportion. Une fois ces deux valeurs calculées, si la p-value est inférieure à 0,05, la différence entre les probabilités est considérée comme statistiquement significative au seuil 5%.



```
if p_val < 0.05:
    print('La différence est statistiquement significative.')
else:
    print('La différence n\'est pas statistiquement significative.')

⇒ Probabilité de dépasser 295 K en septembre 2023: 0.78
   Probabilité de dépasser 295 K pour les autres années: 0.08
   Z-statistic: 14.32
   P-value: 0.00
   La différence est statistiquement significative.
```

FIGURE 10 – Résultats test

Pour souligner l'importance de septembre 2023, nous avons aussi recalculé notre FAR en utilisant la probabilité de dépasser le seuil en septembre de 2020 à 2026 sans prendre en compte l'année 2023.

```
far2 = 1-(p0/prob_combined_other_years)

print("Fraction de Risque Attribuable (FAR) :", far2)
print("FAR plus petite que si on avait pris en compte l'année 2023")

⇒ Fraction de Risque Attribuable (FAR) : 0.7028364753161042
   FAR plus petite que si on avait pris en compte l'année 2023
```

FIGURE 11 – FAR pour septembre 2023

Nous pouvons alors soutenir l'hypothèse d'une vague de chaleur exceptionnelle en septembre 2023, possiblement influencée par des facteurs anthropiques.

Pour appuyer encore plus notre hypothèse, nous avons visualisé les densités de probabilité avec et sans influence humaine.

Nous avons tracé les courbes sur 3 périodes :

- 1850-1900 : passé
- 2020-2026 : présent
- 2090-2099 : futur

Nous avons aussi ajouté une ligne verticale qui représente le seuil de canicule.

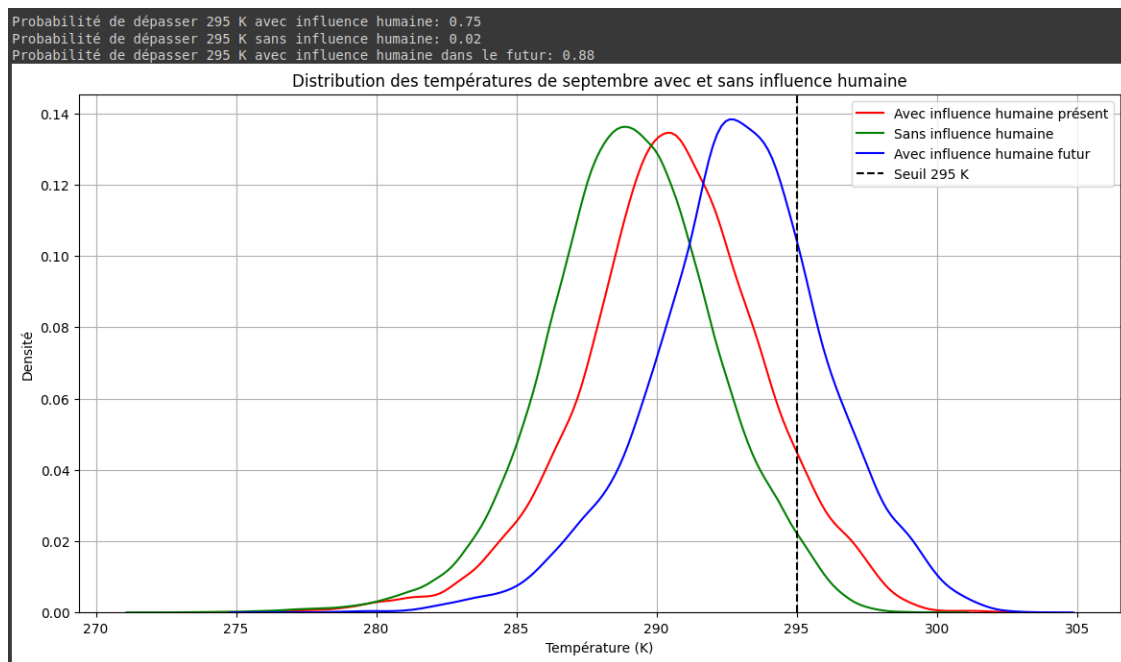


FIGURE 12 – Densités de probabilités

Sur ce graphique nous pouvons voir la distribution des températures du mois de septembre avec influence humaine sur le présent et le futur mais aussi sans influence humaine. Pour représenter cette distribution on trace les densités en fonction de la température, le maximum d'une courbe représente la température qui a le plus de chance de se produire en septembre.

La courbe verte, qui représente la distribution sans l'influence humaine (de 1850 à 1900), ne croise le seuil de 295°K que sur une petite partie il y a donc peu de chance qu'un épisode caniculaire ait lieu.

La courbe rouge correspond à la distribution avec influence humaine pour les années 2020-2026, et donc le présent. Nous pouvons voir que le phénomène de canicule a plus de chance de se produire que précédemment.

La courbe bleue quant à elle représente une distribution avec influence humaine dans le futur, ce ne sont donc que des suppositions faites à partir des tendances actuelles. Le maximum de densité est de plus en plus proche du seuil de canicule ce qui laisse entendre qu'à l'avenir les vagues de chaleur seront beaucoup moins rares au mois de septembre.

Enfin, nous avons fourni une interprétation causale qui vise à mettre en lumière l'impact significatif de l'influence humaine sur les extrêmes climatiques.

Pour se faire, nous avons calculé :

- la probabilité de causalité suffisante (PS) : évalue la probabilité qu'un événement ait pu se produire uniquement en raison de l'influence humaine.
- la probabilité de causalité nécessaire (PN) : évalue la probabilité que l'influence humaine ait été nécessaire pour que l'événement se produise. C'est équivalent à la Frac-

- tion de Risque Attribuable (FAR).
- la probabilité de causalité nécessaire et suffisante (PNS) : évalue la probabilité qu'un événement n'aurait pas pu se produire sans l'influence humaine et que cette influence était suffisante pour le déclencher.

```
[11]
# Calculer PS, PN et PNS
PS = 1 - (1 - prob_sept) / (1 - p0)
PN = 1 - (p0 / prob_sept)
PNS = prob_sept - p0

# Affichage des résultats
print(f'Probabilité de dépasser {seuil_canicule} K avec influence humaine (p1): {prob_sept:.2f}')
print(f'Probabilité de dépasser {seuil_canicule} K sans influence humaine (p0): {p0:.2f}')
print(f'Probability of Sufficient Causation (PS): {PS:.2f}')
print(f'Probability of Necessary Causation (PN): {PN:.2f}')
print(f'Probability of Necessary and Sufficient Causation (PNS): {PNS:.2f}')
```

```
⇒ Probabilité de dépasser 295 K avec influence humaine (p1): 0.75
Probabilité de dépasser 295 K sans influence humaine (p0): 0.02
Probability of Sufficient Causation (PS): 0.74
Probability of Necessary Causation (PN): 0.97
Probability of Necessary and Sufficient Causation (PNS): 0.72
```

FIGURE 13 – Valeurs de probabilités de causalité

Avec les résultats obtenus, nous pouvons conclure qu'il y a une forte probabilité que les températures extrêmes ne se seraient pas produites sans l'influence humaine, qu'il est très probable que l'influence humaine soit une cause nécessaire pour que les températures dépassent le seuil et que non seulement l'influence humaine était nécessaire, mais qu'elle était aussi probablement suffisante pour provoquer les températures extrêmes.

La haute probabilité de causalité suffisante (PS), la quasi-certitude de causalité nécessaire (PN), et la probabilité substantielle de causalité nécessaire et suffisante (PNS) soulignent que l'influence humaine a joué un rôle crucial dans l'augmentation des températures extrêmes.

## 4 Ouverture

Après avoir étudié la vague de chaleur de septembre 2023 et démontré que c'était un événement exceptionnel qui était pour grande partie causée par l'Homme, nous nous sommes intéressées à l'évolution de l'impact anthropique sur la température moyenne mais cette fois sur des années entières sans se concentrer sur un mois ou événement spécifique. Pour cela, nous avons exploité deux nouvelles bases de données :

- "piControl-multi-model-France" qui contient les températures journalières sur la période de 1850 à 1900 en France, et donc sans l'influence humaine puisqu'on étudie ici la période pré révolution industrielle.
- "ssp245-multi-model-France" qui contient les températures journalières sur la période de 2015 à 2099 en France, avec un fichier par année.

Nous avons aussi augmenté le seuil à partir duquel la température était considérée caniculaire, le faisant passer de  $295^{\circ}\text{K}$  à  $308.15^{\circ}\text{K}$ , soit  $35^{\circ}\text{C}$ . Nous avons tout d'abord compté le nombre de jours au dessus du seuil de canicule afin de calculer la probabilité de tomber sur des jours avec une température maximale dépassant ce seuil. Pour les données contenant l'influence humaine, nous avons seulement pris en compte les données sur les trois périodes mentionnées précédemment pour pouvoir observer l'évolution de ces dernières. La période de retour, qui est le temps moyen pour qu'un événement similaire se produise dans les conditions des périodes étudiées, a aussi été calculée. Nos résultats sont représentés dans le tableau suivant :

Influence anthropique	Probabilités	Période de retour (année)
Sans activité humaine	0.0148	67,57
Avec activité humaine (2015-2019)	0,0332	30,12
Avec activité humaine (2020-2026)	0.0365	27,40
Avec activité humaine (2090-2099)	0.0739	13,53

TABLE 2 – Probabilité qu'une journée présente une température maximale dépassant le seuil de canicule.

Comme on peut le voir, la probabilité d'atteindre des températures caniculaires augmente au fil des années, pour atteindre 0,0739 en prévision pour 2090, ce qui peut paraître très faible, mais en regardant la période de retour on peut se rendre compte de la fréquence bien plus élevée d'événements de chaleur extrême, de nos jours ou dans le futur. Même la toute petite différence entre les deux premières périodes avec activité humaine est importante à prendre en considération puisqu'une telle augmentation n'est pas attendue en si peu de temps.

À l'aide des probabilités calculées, nous avons pu trouver les FAR pour les trois périodes étudiées avec l'activité de l'Homme pour montrer l'évolution de l'influence de l'Homme sur les événements de chaleur extrême en France.

Les résultats de nos calculs donnent des FAR très proches pour 2015-2019 et 2020-2026 qui sont à 0.55 et 0.58. Ces valeurs sont considérables puisque, par exemple dans le deuxième cas cela signifie que l'on peut attribuer près de 58% du risque de températures caniculaires à des facteurs anthropogéniques. Une interprétation causale de cela est qu'il y a une probabilité de 58% que des facteurs anthropogéniques soient une cause nécessaire de cet événement. Pour la période 2090-2099, les prévisions annoncent une FAR de 0.8 qui est bien plus élevée que les années précédentes et cette valeur annonce un futur inquiétant.

Afin de pouvoir illustrer nos résultats et les comparer plus facilement de façon plus concrète, nous avons décidé de calculer puis tracer des fonctions de densité de probabilité. Sur les trois prochains graphes, les densités de probabilité associées à la période qui précède l'ère industrielle sont représentées en vert, et en rouge on peut voir les densités de probabilités d'événements de chaleur extrême liées aux périodes où l'activité humaine impacte la température.

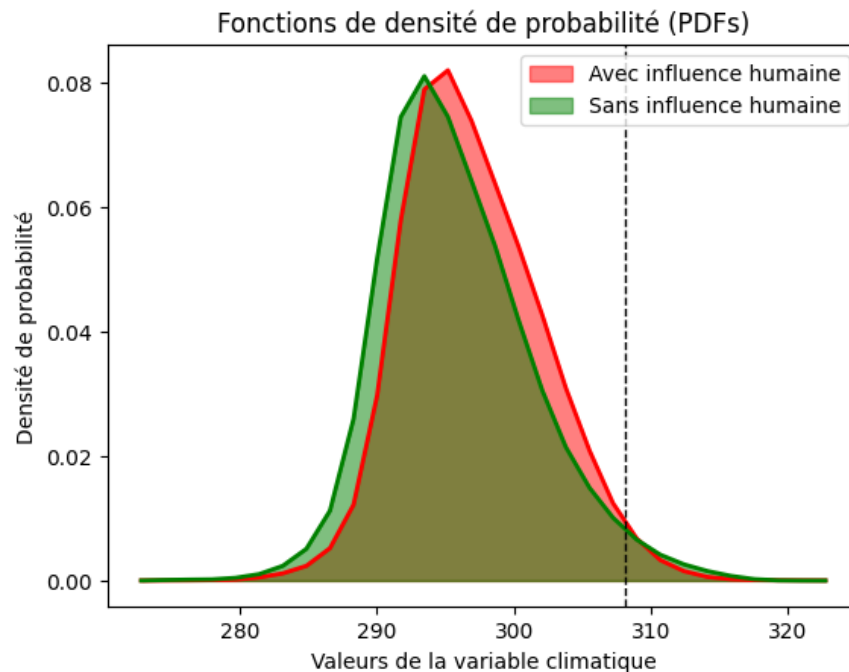


FIGURE 14 – Fonction de densité de probabilité pour les températures de 2015-2019

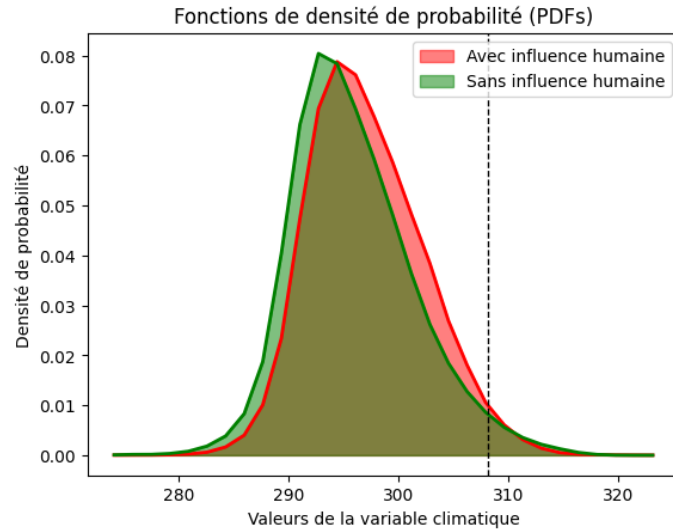


FIGURE 15 – Fonction de densité de probabilité pour les températures de 2020-2026

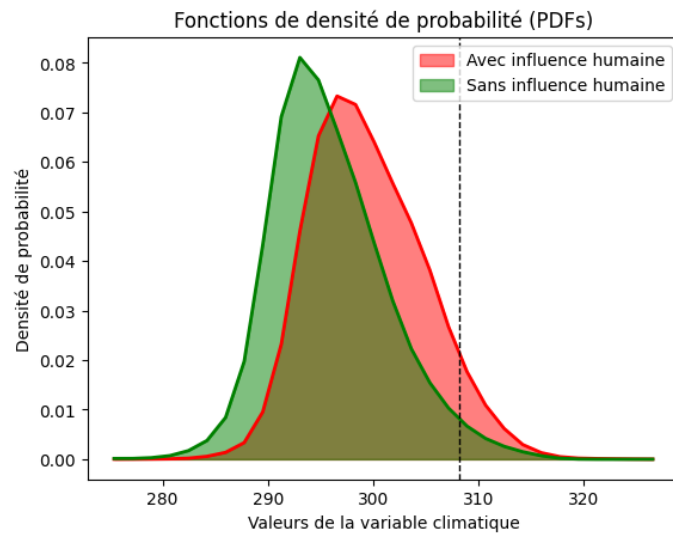


FIGURE 16 – Fonction de densité de probabilité pour les températures de 2090-2099

Nous pouvons très facilement observer l'évolution alarmante de l'influence de l'activité humaine sur l'augmentation de la température. En effet, la probabilité que la température dépasse le seuil de la canicule est déjà un peu plus élevée sur 2020-2026 que 2015-2019, mais nous pouvons observer qu'en restant sur cette lancée, cette probabilité sera encore moins négligeable dans quelques années.

Même si ce ne sont que des prévisions, ces résultats sont inquiétants. On peut donc en conclure que ce qui se produira au niveau du réchauffement climatique sera en grande partie dû à l'homme.

## 5 Conclusion

### 5.1 Bilan du projet et difficultés rencontrées

Grâce à ce projet, nous avons eu l'occasion d'aborder un thème de recherche avec le mode de réflexion qui l'accompagne. En effet, il a été possible de découvrir et expérimenter la démarche à entreprendre lors de travaux de recherche.

Nous avons donc en premier lieu défini un objectif global d'étude qui est la détection et l'attribution de changement climatique, puis avons recentré cette étude autour de la vague de chaleur de septembre 2023. Le cahier des charges ([section 2](#)) a été d'une très grande aide pour cela. Cela a été d'autant plus intéressant que le sujet de notre projet fait référence au phénomène du dérèglement climatique, thème qui est omniprésent à notre époque et que nous avons vécu.

Ensuite, il a été indispensable de bien se renseigner sur la vague de chaleur mais aussi sur toute la théorie liée aux différentes méthodes d'analyse en faisant des recherches.

Enfin, nous avons pu appliquer la théorie sur nos données en découvrant une nouvelle façon d'utiliser le langage de programmation python pour analyser ces données. Cette partie de découverte d'un nouveau mode de travail a été assez complexe car nous n'avons pas souvent l'occasion de l'exploiter. Malgré cette difficulté, cela a été très enrichissant, autant du point de vue des connaissances liées au sujet que de la méthode de recherche.

De plus, la vague de chaleur de septembre 2023 étant un événement récent, il n'a pas été aisé de trouver une approche efficace pour faire notre analyse car ce phénomène n'a pas encore été documenté. Nous avons donc été contraintes de travailler sur plusieurs bases et de changer plusieurs fois de méthodes afin de trouver la plus optimale pour notre étude. Néanmoins, cela rend aussi cette étude plus intéressante car nous obtenons nos propres résultats et notre propre analyse.

D'un autre côté, les données utilisées viennent de simulations climatiques imparfaites et ne permettent donc pas d'avoir des résultats exactes, en revanche, nous pouvons quand même obtenir un ordre de grandeur des probabilités en jeu, et ils peuvent être utiles pour des prises de décision futurs par exemple.

Nous avons aussi été confrontées à quelques difficultés supplémentaires concernant l'écriture du code. En effet, nous avons dû apprendre à utiliser de nouvelles librairies python auxquelles nous n'étions pas habituées, ce qui était compliqué au commencement. De plus, le projet initial était trop ambitieux, c'est pourquoi nous avons rapidement adapté le niveau et revu le projet avec M. Durand. Il nous a ensuite guidées dans notre démarche en nous expliquant les différentes étapes à suivre, ce qui nous a permis d'obtenir les résultats précédents.

Enfin, nous aurions aimé approfondir encore plus notre étude si nous avions eu davantage de temps alloué, bien que celle-ci est déjà satisfaisante. En effet, nous avons réussi à obtenir des résultats intéressants et exploitables pour atteindre les objectifs fixés, mais nous aurions apprécié approfondir en rentrant plus dans les détails.

## 5.2 Utilisation d’IA générative

L’intelligence artificielle a été utilisée de manière modérée dans ce projet. Nous l’avons utilisée lorsque nous avions un problème dans notre code et que nous n’arrivions pas à le résoudre. Tout d’abord, elle nous a été d’une grande utilité pour comprendre comment nous pouvions filtrer nos jeux de données en fonction de l’année ou du mois. Elle nous a aussi aidé pour écrire la fonction de test de probabilité car nous avions un problème pour importer la librairie nécessaire. Ces deux problèmes ont donc été résolus à l’aide de l’intelligence artificielle, tout en gardant un esprit critique quant à la réponse générée. Cela nous a permis de pouvoir continuer notre code en gagnant du temps, tout en vérifiant les réponses générées pour ne pas injecter une information erronée.

## 6 Remerciements

Nous tenons à exprimer notre profonde gratitude à nos encadrants de ce projet, Madame Potop-Butucaru pour avoir organisé et trouvé des personnes disponibles pour nous encadrer.

Nos remerciements les plus sincères vont à Monsieur Durand qui nous a généreusement accordé de son temps précieux. Sa disponibilité, ses conseils et son accompagnement ont permis à notre équipe d’être efficace et productive. Son soutien et ses orientations ont grandement contribué à la qualité et à la rigueur de nos recherches.



## 7 Bibliographie

### Références

- [1] IPCC. [Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change](#). pages 35–115, 2023.
- [2] Ben Clarke and Friederike Otto. [Les évènements météorologiques extrêmes et le changement climatique](#). pages 1–18, 2022.
- [3] Kathryn Conlon Stephanie Herring Kenneth E. Kunkel Jay Lawrimore George Luber Carl Schreck Adam Smith Jesse E. Bell, Claudia Langford Brown and Christopher Uejio. Changes in extreme events and the potential impacts on human health. *Journal of the Air & Waste Management Association*, 68(4) :265–287, 2018.
- [4] Météo-France. [Septembre 2023 : Au 1er rang des mois de septembre les plus chauds depuis 1900](#). 2023.
- [5] Peter A Stott, Nikolaos Christidis, Friederike EL Otto, Ying Sun, Jean-Paul Vanderlinden, Geert Jan van Oldenborgh, Robert Vautard, Hans von Storch, Peter Walton, Pascal Yiou, et al. Attribution of extreme weather and climate-related events. *Wiley Interdisciplinary Reviews : Climate Change*, 7(1) :23–41, 2016.
- [6] Gabriele Hegerl and Francis Zwiers. [Use of models in detection and attribution of climate change](#). COWL, 2011.
- [7] Robert C Wills, Tapio Schneider, John M Wallace, David S Battisti, and Dennis L Hartmann. Disentangling global warming, multidecadal variability, and el niño in pacific temperatures. *Geophysical Research Letters*, 45(5) :2487–2496, 2018.
- [8] Friederike E. L. Otto Ying Sun Jean-Paul Vanderlinden Geert Jan van Oldenborgh Robert Vautard Hans von Storch Peter Walton Pascal Yiou Peter A. Stott, Nikolaos Christidis and Francis W. Zwiers. [ncbi](#). *Attribution of extreme weather and climate-related events*, 2016.
- [9] Peter A Stott, Dáithí A Stone, and Myles R Allen. Human contribution to the european heatwave of 2003. *Nature*, 432(7017) :610–614, 2004.
- [10] info climat. [Indicateur national français des températures](#). 2024.
- [11] [Copernicus](#) .
- [12] Violaine Lepousez Maxime Aboukrat. [Les scénarios SSP](#). *Carbone4*, 2022.