



# Projet industriel : Méthode Optimal Fingerprint et Extensions

Projet industriel MAIN4 - Rapport

*Référent : H. Durand*

*Membres de l'équipe :  
Sara Grissia, Chadha Hassine, Elliott Le Dauphin, Francis Jégou et Mohamed Ghali Es  
Sadel*

Le 26 Mars 2025

# Table des Matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Thématique	3
1.2	Problématique	3
1.3	Définitions	3
1.4	Preliminaire Statistique	5
<b>2</b>	<b>Théorie de la méthode Optimal Fingerprint</b>	<b>6</b>
2.1	Un Problème de Signal et de Bruit	6
2.2	Un Problème de Régression Linéaire	7
2.3	Démarche	8
2.3.1	Du théorique au pratique	8
2.3.2	Gestion de la taille des données :	8
<b>3</b>	<b>Approche non linéaire pour la détection de changement climatique</b>	<b>9</b>
3.1	Méthode à kernel et SVM	9
3.2	Modèle Bayésien : les Processus Gaussiens	12
<b>4</b>	<b>Application de la méthode Optimal Fingerprint</b>	<b>14</b>
4.1	Présentation du code	14
4.1.1	Environnement de travail	14
4.1.2	Implémentation de l'algorithme	14
4.1.3	Calibrage du test	15
4.2	Application sur les données et résultats	15
<b>5</b>	<b>Application des méthodes non linéaires</b>	<b>17</b>
5.1	Séparation des données	17
5.1.1	Explication des méthodes	17
5.1.2	Implémentation des Méthodes	17
5.1.3	Pourquoi privilégier la méthode avec <code>gpytorch</code> ?	18
5.2	SVM et test statistique fréquentiste	18
5.2.1	Explication de la méthode	18
5.2.2	Implémentation de la méthode	18
5.2.3	Conclusion	20
5.3	Processus Gaussiens, SVM et test statistique bayésien	20
5.3.1	Explication de la démarche	20
5.3.2	Implémentation de la méthode	20
5.3.3	Conclusion	21
<b>6</b>	<b>Résultats expérimentaux</b>	<b>21</b>
6.1	Présentation des jeu de données	21
6.1.1	Présentation du jeu de données sur les températures	21
6.1.2	Présentation du jeu de données sur les précipitations maximum	21
6.2	Résultats de l'Optimal Fingerprint	21
6.3	Résultats des approches non linéaires	23
6.3.1	Calibration et puissance du test pour différent seuil $\alpha$	23
6.3.2	Observations globales	23
6.4	Observations Amérique du Nord et Europe	24
6.5	Résultats du test bayésien	25
6.6	Précipitations extrêmes	26
<b>7</b>	<b>Réflexion sur les aspects RSE</b>	<b>27</b>

<b>8 Conclusion :</b>	<b>28</b>
<b>9 Bibliographie</b>	<b>29</b>

# 1 Introduction

## 1.1 Thématique

Le changement climatique constitue l'une des problématiques majeures auxquelles notre société est confrontée aujourd'hui. En effet, il engendre des conséquences graves, telles que des catastrophes liées aux événements climatiques extrêmes, des zones devenant inhabitées et inhabitables, provoquant des migrations massives, des conflits géopolitiques, ainsi que des famines et des crises alimentaires (Trisos et al., 2022). Depuis plusieurs décennies, la question de l'impact des activités humaines sur le climat fait l'objet de nombreuses études. Selon le dernier rapport du GIEC, publié en Mars 2023, « les émissions de gaz à effet de serre dues aux activités humaines ont réchauffé le climat à un rythme sans précédent : la température de la surface du globe s'est élevée d' 1,1 °C par rapport à la période pré-industrielle ». Le réchauffement climatique, causé en grande partie par l'émission de gaz à effet de serre, est désormais un phénomène bien documenté. Il est ainsi possible de relier directement l'augmentation des températures globales à ces émissions. Cependant, il reste plus complexe d'attribuer les changements climatiques régionaux, comme la fréquence et l'intensité des événements extrêmes, à des causes spécifiques, en raison de la variabilité naturelle du climat et des incertitudes inhérentes aux modèles.

Dans ce contexte, les méthodes statistiques jouent un rôle clé dans la détection et l'attribution des phénomènes climatiques. La méthode de l'Optimal Fingerprint (OF) est l'une des approches couramment utilisées pour analyser les signaux de changement climatique. Elle permet d'isoler l'impact des forçages externes (comme les gaz à effet de serre) de la variabilité naturelle du climat, souvent qualifiée de « bruit ». Il correspond à la variabilité climatique naturelle, englobant les fluctuations aléatoires du système climatique indépendantes des forçages anthropiques. L'objectif est donc de détecter un « signal » dans les observations climatiques tout en minimisant l'effet de ce bruit. Cependant, malgré ses avantages, la méthode Optimal Fingerprint présente des limites, notamment lorsqu'il s'agit de traiter des données à des échelles régionales ou de détecter des événements climatiques extrêmes.

L'objectif de ce projet est de comprendre et d'améliorer la méthode OF, en l'intégrant à des techniques récentes telles que le machine learning ou l'inférence causale, pour permettre une détection plus fine et plus précise des changements climatiques. L'application de ces approches pourrait ouvrir de nouvelles perspectives, notamment en termes de détection régionale ou de mesure du « time of emergence », un concept qui permet de mieux comprendre quand un signal de changement climatique devient détectable.

## 1.2 Problématique

Nous nous proposons donc dans ce rapport de répondre à la problématique suivante : Peut-on détecter qu'il existe un changement climatique ? Celui-ci est-il dû à l'activité humaine ?

### **Sous-problématique : Compréhension et amélioration de la méthode Optimal Fingerprint**

- Qu'est-ce que la méthode Optimal Fingerprint (OF), et comment s'applique-t-elle aux données climatiques ?
- Quels sont les résultats obtenus par cette méthode dans le passé, et quelles en sont les limites (notamment pour les précipitations, la détection à l'échelle locale/régionale) ?
- Comment peut-on améliorer cette approche en utilisant des techniques récentes de machine learning ou d'inférence causale ?
- Quels nouveaux cas d'usage peuvent être imaginés pour ces extensions (détection régionale, “time of emergence”, etc.) ?

## 1.3 Définitions

Nous commençons par définir les termes et concepts importants et récurrents de notre rapport afin d'en simplifier la compréhension.

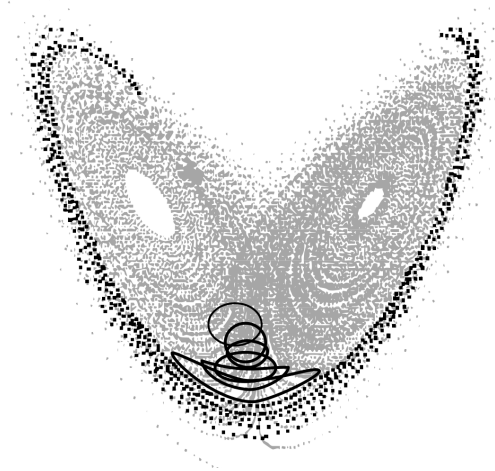


FIGURE 1 – Exemple de système chaotique (proposé par [Lorenz, 1963](#)). Une légère perturbation des conditions initiales (cercle central) peut entraîner une divergence rapide et importante de l'évolution du système (pointillés). De manière analogue, le système climatique présente un comportement chaotique : de faibles variations de pression atmosphérique en un point donné peuvent provoquer d'importants changements de température ou de précipitations quelques jours ou semaines plus tard.

— **Détection :**

La détection vise à démontrer qu'un changement climatique s'est produit, au sens statistique, sans pour autant chercher à en identifier la cause.

— **Forçages externes :**

La notion de forçages externes représente les facteurs qui sont extérieurs au système climatique (c'est à dire qu'ils l'affectent sans être affecté par celui-ci en retour), comme par exemple les variations orbitales ou les variations solaires qui sont des mécanismes naturels, ou les perturbations dites anthropiques (liées à l'activité humaine) telles que les émissions de gaz à effet de serre. Ce sont ces forçages qui sont à l'origine du signal utilisé dans l'optimal fingerprint.

— **Variabilité interne du climat :**

La variabilité interne désigne les fluctuations naturelles du système climatique en l'absence de forçages externes, comme les activités humaines ou les variations de l'ensoleillement. Elle est une composante essentielle des systèmes climatiques, mais elle rend la détection du changement climatique particulièrement difficile, car elle agit comme un "bruit" dans les modèles, notamment dans la méthode de l'optimal fingerprint.

Comme cela a été reconnu très tôt par Edward Lorenz ([Lorenz, 1963](#)), le système climatique est chaotique. Cela signifie que des très faible perturbation peuvent entraîner de fort changement même sur une courte période de temps. C'est le fameux effet papillon (voir figure 1). Cela rend donc la variabilité interne du climat très difficile à prédire sur des longues périodes temporelles et elle est donc modélisée comme un bruit stochastique. Cette variabilité découle en grande partie des processus physiques du climat, qu'ils soient thermodynamiques, chimiques ou biologiques. Les mécanismes thermodynamiques incluent la balance énergétique, les mouvements des océans, et la circulation des masses d'air, qui influencent les températures et les précipitations. Les processus chimiques, comme la décomposition des gaz dans l'atmosphère et l'effet d'albédo (variation de la réflexion de la lumière par la surface terrestre), jouent également un rôle significatif. Enfin, les processus biologiques, tels que la capture de CO<sub>2</sub> par la végétation, ajoutent une couche supplémentaire de complexité.

Plus la variabilité interne est grande, plus il est difficile de détecter les changements dus aux forçages externes, car ces fluctuations naturelles masquent les signaux anthropiques. Cette variabilité est également difficile à modéliser, car elle ne peut pas être directement observée. Le climat actuel est une

combinaison de variabilité interne et de forçage externe, et démêler ces deux composantes constitue un défi majeur pour les climatologues.

— **Modèles climatiques :**

Ces modèles, basés sur la résolution numérique des équations physiques, chimiques et biologiques permettent de réaliser des projections climatiques, avec ou sans forçage externe ([Warner, 2010](#)). Ils sont cruciaux pour certaines méthodes de détection et d'attribution, qui s'appuient non seulement sur des observations mais aussi sur des simulations numériques. Ils servent notamment à prédire les réponses du climat aux forçages externes, sous forme de variations spatiales, temporelles, ou les deux.

— **Méthodes de détection :**

Les méthodes de détection reposent principalement sur l'analyse statistique des données climatiques, qu'elles proviennent d'observations directes ou de modèles climatiques ([Hegerl and Zwiers, 2011](#)). Ces méthodes sont utilisées pour identifier des changements significatifs dans les phénomènes climatiques, qu'il s'agisse d'événements extrêmes ou de tendances à long terme.

- L'approche la plus courante consiste à utiliser des tests d'hypothèse. Dans cette méthode, un événement climatique extrême est d'abord défini, et sa distribution est estimée dans un climat sans influence anthropique, en utilisant des modèles ou des données historiques. Ensuite, si la fréquence ou l'intensité de cet événement dévie de manière significative par rapport à cette distribution, cela peut suggérer l'existence d'un changement, ce qui conduit au rejet de l'hypothèse nulle (absence de changement).
- En parallèle, il existe des méthodes permettant de détecter des tendances climatiques à long terme, comme la méthode de l'Optimal Fingerprint. Cette méthode se distingue des tests d'événements extrêmes en ce qu'elle cherche à isoler un signal de changement climatique dans des séries temporelles en séparant les forçages externes (par exemple, l'effet des gaz à effet de serre) de la variabilité interne du climat. L'Optimal Fingerprint utilise des techniques statistiques pour modéliser ce signal tout en prenant en compte le bruit naturel, permettant ainsi de détecter des tendances climatiques de manière plus précise et fiable.

## 1.4 Préliminaire Statistique

Les tests statistiques jouent un rôle fondamental dans l'analyse des données en permettant de prendre des décisions rationnelles face à l'incertitude. Ils sont utilisés pour évaluer si une hypothèse formulée sur un phénomène observé peut être validée ou rejetée à partir d'un échantillon de données. Dans cette section, nous présentons les concepts clés nécessaires à la compréhension des tests statistiques utilisés dans ce rapport. Pour une description plus détaillée de ces concepts, nous renvoyons le lecteur à [Wasserman \(2013\)](#).

1. **Test statistique :** Un test statistique est une procédure méthodologique qui permet de choisir entre deux hypothèses en fonction des données disponibles. Par exemple, on peut vouloir tester si une intervention a un effet mesurable sur une population ou si une tendance observée est due au hasard. Le test repose sur un cadre probabiliste et utilise un échantillon représentatif des données pour répondre à une problématique spécifique. L'objectif est de prendre une décision, en minimisant le risque d'erreurs.
2. **Hypothèse nulle  $H_0$  :** L'hypothèse nulle représente l'état initial ou le modèle par défaut que l'on cherche à tester. Par exemple, dans une étude médicale,  $H_0$  pourrait stipuler qu'un traitement n'a aucun effet sur les patients. Elle est formulée pour être réfutée ou confirmée à l'aide des données. Rejeter  $H_0$  revient à affirmer que les données contredisent l'hypothèse par défaut.
3. **Hypothèse alternative  $H_1$  :** L'hypothèse alternative est celle que l'on accepte si les données fournissent suffisamment de preuves contre  $H_0$ . Elle représente une situation où  $H_0$  est invalide, traduisant généralement un effet significatif ou une différence mesurable. Par exemple, si  $H_0$  affirme que la moyenne d'une population est égale à une valeur donnée,  $H_1$  pourrait suggérer qu'elle est différente (ou supérieure, selon le cas étudié).

4. **Erreur de type 1** : L'erreur de type 1 survient lorsque l'on rejette  $H_0$  alors qu'elle est en réalité vraie. Cela correspond à une fausse alarme. Par exemple, conclure qu'un nouveau médicament est efficace alors qu'il ne l'est pas est une erreur de type 1. La probabilité associée à cette erreur est notée  $\alpha$ , et on fixe souvent cette probabilité à 5 % dans les tests statistiques courants.
5. **Erreur de type 2** : L'erreur de type 2 se produit lorsque l'on conserve  $H_0$  alors qu'elle est en réalité fausse. Cela revient à manquer un effet réel. Par exemple, ne pas détecter une différence significative entre deux traitements lorsqu'elle existe constitue une erreur de type 2. La probabilité associée à cette erreur est notée  $\beta$ , et son complément  $(1 - \beta)$  est directement lié à la puissance d'un test.
6. **Puissance d'un test** : La puissance d'un test est définie comme la probabilité de rejeter  $H_0$  lorsque  $H_1$  est vraie. Elle reflète donc la capacité du test à détecter un effet réel ou une différence significative lorsque celle-ci existe. Une puissance élevée est souhaitable, car elle réduit le risque d'erreur de type 2. Par exemple, dans une expérience, une puissance de 80 % indique qu'il y a 80 % de chances de détecter une différence significative si elle existe réellement. La puissance dépend de plusieurs facteurs, tels que la taille de l'échantillon, l'amplitude de l'effet et le seuil de signification  $\alpha$ .

## 2 Théorie de la méthode Optimal Fingerprint

### 2.1 Un Problème de Signal et de Bruit

Dans cette section, nous présentons des notions utilisées dans l'article *The detection of climate change* (W.Zwiers, 1999) qui propose une description simple de la méthode de l'optimal fingerprint. L'idée principale est l'observation d'une quantité  $T(x, t)$ . Ici,  $x$  représente la localisation et  $t$  le temps. Dans notre cas, cette quantité correspond à la température ou aux précipitations et est décrite par l'équation  $T(x, t) = X + N$ . Dans cette formule,  $X$  est un signal déterministe. Il représente la réponse du climat à un forçage externe, comme les gaz à effet de serre ou l'activité volcanique.  $N$ , quant à lui, est un bruit aléatoire. Il est modélisé comme un processus stochastique de moyenne nulle et traduit la variabilité naturelle du climat.

Le problème de la détection du changement climatique consiste, d'une part, à identifier la présence du signal  $X$  dans les observations  $T$  et, d'autre part, à attribuer tout ou partie du signal détecté aux activités humaines. Klaus Hasselmann a proposé une méthode pour répondre à ce problème. Elle repose sur certaines hypothèses, notamment que  $T(x, t)$  est composé de variables indépendantes et identiquement distribuées (i.i.d) et que  $N$  suit une loi gaussienne.

Pour séparer le signal  $X$  du bruit  $N$ , les chercheurs utilisent un filtre spatio-temporel. Bell (1982) propose une variable de détection basée sur une moyenne pondérée spatio-temporelle, exprimée par l'équation suivante :

$$A_t = \sum_{x=1}^l \sum_{\tau=1}^m w(x, \tau) T(x, t - \tau + 1), \quad \text{avec} \quad \sum_{x=1}^l \sum_{\tau=1}^m w(x, \tau) = 1. \quad (1)$$

Ce filtre peut également être exprimé en notation matricielle-vecteur sous la forme :

$$A_t = \mathbf{w}^T \mathbf{T}_t. \quad (2)$$

Le filtre optimal  $\mathbf{w}^*$  est défini comme celui qui maximise le rapport signal/bruit donné par :

$$\gamma^2(\mathbf{w}) = \frac{\mathbb{E}(A_t)^2}{\text{Var}(A_t)} \quad \text{ou} \quad \gamma^2(\mathbf{w}) = \frac{(\mathbf{w}^T \mathbf{X}_t)^2}{\mathbf{w}^T \Sigma_{N_t} \mathbf{w}}, \quad (3)$$

où  $\mathbf{w}^T$  désigne la transposée du vecteur des poids  $\mathbf{w}$ ,  $\mathbf{X}_t$  le vecteur du signal au temps  $t$ , et  $\Sigma_{N_t}$  la matrice de covariance du bruit  $N$ . Une illustration de la direction optimal, aussi appelée **optimal fingerprint**, est donnée pour un exemple en deux dimension dans la figure 2.

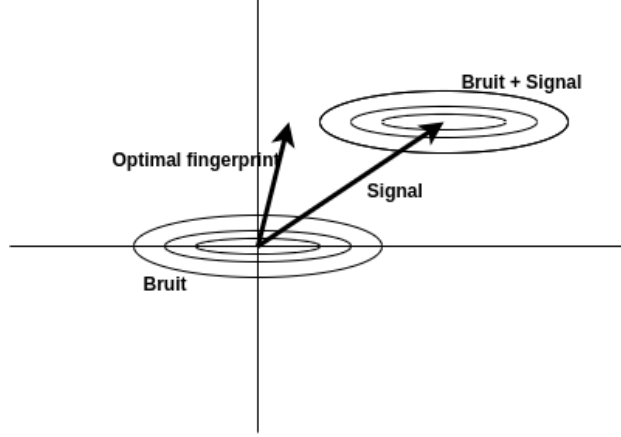


FIGURE 2 – Un exemple de la méthode de l’optimal fingerprint pour un problème de détection en deux dimensions. La distribution du bruit n’étant pas isotrope, le signal est partiellement aligné avec celui-ci, ce qui en fait une direction sous-optimale pour projeter les données observées. En revanche, l’optimal fingerprint maximise le rapport signal/bruit, ce qui en fait la direction optimale pour distinguer les distributions avec et sans signal.

Pour déterminer l’existence du signal  $X$ , un test statistique est réalisé. L’hypothèse nulle  $H_0$  stipule que  $T = N$  (absence de signal), tandis que l’hypothèse alternative  $H_1$  suggère que  $T = X + N$  (existence du signal). La statistique de test  $Z_t$  est alors calculée. L’hypothèse  $H_0$  est rejetée si  $Z_t$  dépasse un seuil critique fixé à 1,96 (correspondant à un test à 5 % d’erreur pour une gaussienne centrée réduite). La formule utilisée est la suivante :

$$Z_t = \frac{\mathbf{X}_t^T \Sigma_{N_t}^{-1} \mathbf{T}_t}{\sqrt{\mathbf{X}_t^T \Sigma_{N_t}^{-1} \mathbf{X}_t}} = \frac{\mathbf{X}_t^T \Sigma_{N_t}^{-1} \mathbf{T}_t}{\gamma} \geq 1.96. \quad (4)$$

Ici ,  $Z_t$  suit une gaussienne centrée car c’est une somme de gaussienne centrée et elle réduite en la normalisant avec  $\gamma$ .

## 2.2 Un Problème de Régression Linéaire

Il a été reconnu quelques années plus tard que la formulation originalement proposé par Hasselmann (Hasselmann, 1993) et décrite par Zwiers avait une formulation équivalente. Allen décrit la théorie de l’optimal fingerprint (OF) dans son article *Estimating signal amplitudes in optimal fingerprinting* M. R. Allen (2003) où il présente ce concept comme un modèle de régression linéaire :

$$y = \sum_{i=1}^m x_i \beta_i + v = X\beta + v, \quad (5)$$

Ici,  $y$  est le vecteur des observations climatiques.  $X$  est la matrice des *fingerprints* simulés par un modèle climatique. Chaque colonne de  $X$  représente un *pattern* de réponse à un forçage spécifique. Le vecteur  $\beta$  contient les amplitudes des *fingerprints* à estimer. Enfin,  $v$  est le bruit climatique, c’est-à-dire la variabilité interne du système climatique qui n’est pas expliquée par les forçages. Pour chaque  $i$ , on teste l’hypothèse  $H_0 : \beta_i = 0$ . Cela permet d’évaluer si la variable  $x_i$  influence significativement les observations  $y$ . Si l’intervalle de confiance pour  $\beta_i$  exclut zéro, on rejette  $H_0$ . Cela signifie que le forçage  $x_i$  a probablement contribué au changement climatique observé. En conclusion, cette méthode permet d’identifier les forçages climatiques ayant un impact significatif sur le climat mesuré. Allen prend également en compte l’incertitude d’échantillonnage dans les simulations grâce à l’approche *total least squares* (TLS) (Golub and Van Loan, 1980). En



revanche, Zwiers utilise une méthode basée sur les *ordinary least squares* (OLS), qui suppose que le bruit provient uniquement des observations. Cette différence mise à part, les deux méthodes sont équivalentes lorsqu'il s'agit de détecter un changement climatique.

## 2.3 Démarche

### 2.3.1 Du théorique au pratique

Les deux méthodes ont une puissance optimale mais selon différentes hypothèses : erreurs de mesures ou non. Pour des raisons de simplicité, nous avons choisi l'approche de Zwiers tout en reprenant des notions abordés par Allen dans le calcul.

En effet, si les hypothèses sont correctes, la puissance du test est optimale avec cette méthode.

On considère  $T$  un vecteur aléatoire tel que ses colonnes représentent l'espace et ses lignes représentent le temps.

Nous utilisons les formules et notations suivantes dans notre code :

1. **Calcul de  $X$  et  $C_N$**  : qui représentent respectivement le signal et la matrice de covariance

Pour calculer  $X$  et  $C_N$ , on va avoir besoin d'introduire la saisonnalité qui va représenter la température moyenne mensuelle observée dans le passé pour pouvoir ensuite calculer les anomalies du passé (resp. présent) qui est égale à la différence entre la saisonnalité et les températures du passé (resp. présent).

Désormais, on peut calculer  $X$  et  $C_N$  :

- $X$  : est la différence entre les anomalies du présent et celles du passé,
- $C_N$  : est la matrice de covariance des anomalies du passé (qui est équivalente à la matrice de covariance du bruit  $\Sigma_{N_t}$  dans l'article).

2. **Calcul de  $\beta$**  : Le vecteur des amplitudes à estimer, donné par la formule suivante :

$$\beta = C_N^{-1} X \quad (6)$$

avec :

- $X$  : Le signal,
- $C_N$  : La matrice de covariance.

3. **Hypothèses du test statistique :**

$H_0$  : "aucun changement climatique " :  $Z_t \leq 2$

$H_1$  : "changement climatique observé" :  $Z_t > 2$

Nous effectuons le test à un niveau de confiance de 5%.

4. **Calcul de  $Z$**  : La statistique de test avec la formule suivante :

$$Z = \frac{Y(t)\beta}{\gamma} \quad (7)$$

Pour cela, nous devons calculer  $\gamma$  :

$$\gamma^2 = X^T C_N^{-1} X \quad (8)$$

### 2.3.2 Gestion de la taille des données :

Vu la taille des données sur lesquelles nous avons travaillé, nous avons eu recours à ce qu'on appelle **PCA** (Principal Components Analysis).

L'idée principale de cette approche est de réduire la dimensionnalité des données tout en conservant au maximum l'information essentielle, en se basant sur les directions où la variance est la plus élevée.

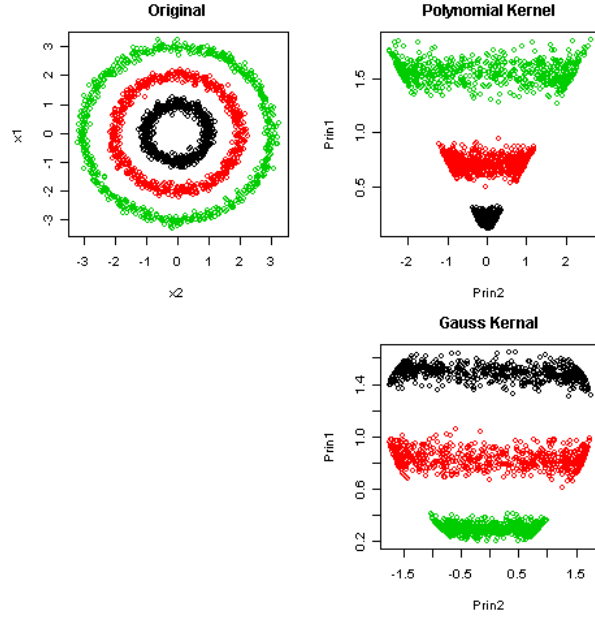


FIGURE 3 – Illustration du fonctionnement des méthodes à noyaux pour noyau polynomial (Polynomial Kernel) et gaussien (Gauss Kernel). Alors que les données ne sont pas linéairement séparables dans leur espace original en 2 dimensions, en les projetant dans des espaces multidimensionnels, elles deviennent linéairement séparables.

Nous passons alors de données contenant plusieurs variables corrélées à un jeu de variables non corrélées appelées **composantes principales**. Cette étape a été cruciale pour l'obtention de résultats valides.

En effet, la méthode nécessite l'inversion de la matrice de covariance  $C_N$  mais celle-ci est en général mal conditionnée car les données sont en grande dimension. En réduisant la dimension des données nous obtenons donc une estimation de la matrice de covariance qui est inversible.

### 3 Approche non linéaire pour la détection de changement climatique

Les méthodes non linéaires sont particulièrement adaptées à l'étude du climat, car celui-ci est régi par des équations physiques hautement non linéaires. Les interactions complexes entre l'atmosphère, les océans et la biosphère rendent souvent inopérantes les approches linéaires classiques. Une modélisation plus fidèle de ces dynamiques nécessite donc l'utilisation de techniques non linéaires capables de capturer ces phénomènes complexes.

#### 3.1 Méthode à kernel et SVM

Les machines à vecteurs de support (SVMs) sont largement utilisées en apprentissage automatique en raison de leur fondement théorique solide et de leur capacité à généraliser efficacement sur des données inconnues. Elles sont particulièrement adaptées aux problèmes de grande dimensionnalité et ont montré leur efficacité dans de nombreuses applications, y compris en climatologie, où elles permettent d'identifier des motifs complexes et d'améliorer la détection des changements climatiques.

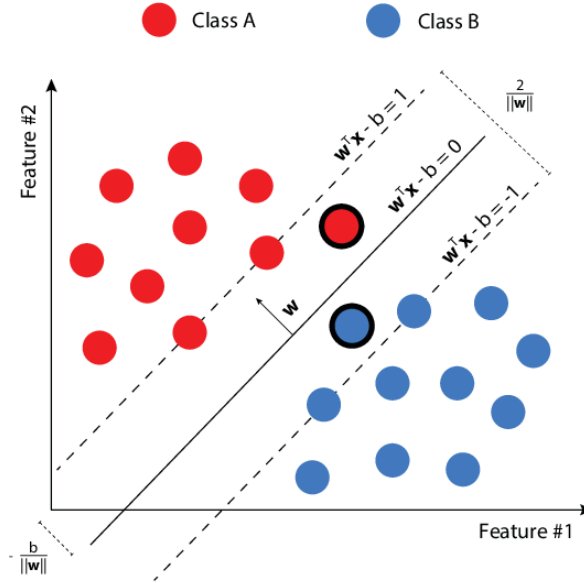


FIGURE 4 – Exemple de Machine à Vecteurs de Support pour un problème de classification en deux dimensions. Les lignes en pointillées définissent la variable d'écart par à l'hyperplan de décision (ligne pleine). Les points proche de l'hyperplan de décision à la variable d'écart près ne pénalise pas la fonction de perte.

a) **Kernel Methods ( méthodes à noyau ) :**

- Les méthodes à noyaux représentent une classe fondamentale d'algorithmes en apprentissage automatique permettant de traiter des problèmes non-linéaires complexes. Leur principe fondamental repose sur la transformation implicite des données dans un espace de caractéristiques de dimension supérieure, où les problèmes deviennent linéairement séparables.
- **Principe fondamental :** La fonction noyau  $K(x, y)$  calcule le produit scalaire  $\langle \phi(x), \phi(y) \rangle$  dans l'espace transformé sans nécessiter le calcul explicite de la transformation  $\phi$ . Cette propriété, connue sous le nom d'*astuce du noyau* (kernel trick), confère aux méthodes à noyaux leur efficacité computationnelle remarquable. Cette approche est illustré par la figure 3.
- **Fonctions noyaux principales :**
  - **Noyau linéaire :**  $K(x, y) = \langle x, y \rangle$  Application : Problèmes linéairement séparables
  - **Noyau polynomial :**  $K(x, y) = (\langle x, y \rangle + c)^d$  Paramètres :  $c$  (constante),  $d$  (degré) Application : Capture des interactions polynomiales entre caractéristiques
  - **Noyau gaussien (RBF) :**  $K(x, y) = \exp(-\gamma \|x - y\|^2)$  Paramètre :  $\gamma$  (inverse de la largeur du noyau) Application : Modélisation de relations complexes non-linéaires
  - **Noyau sigmoïde :**  $K(x, y) = \tanh(\alpha \langle x, y \rangle + \beta)$  Paramètres :  $\alpha, \beta$  (constantes) Application : Alternative au noyau RBF pour certains problèmes
- **Propriétés des fonctions noyaux :** Pour qu'une fonction  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  soit un noyau valide, elle doit satisfaire les conditions de Mercer :
  - (a) Symétrie :  $K(x, y) = K(y, x)$  pour tout  $x, y \in \mathcal{X}$
  - (b) Semi-définie positive :  $\sum_{i=1}^n \sum_{j=1}^n c_i c_j K(x_i, x_j) \geq 0$  pour tout  $n \in \mathbb{N}$ ,  $x_1, \dots, x_n \subset \mathcal{X}$  et  $c_1, \dots, c_n \subset \mathbb{R}$

b) **Machines à Vecteurs de Support (SVM) :**

- Les SVM constituent une approche puissante d'apprentissage supervisé pour la classification et la régression, fondée sur la théorie de l'apprentissage statistique (Cortes and Vapnik, 1995).

- **Principes théoriques :** L'objectif des SVM est de déterminer l'hyperplan optimal qui maximise la marge entre les classes tout en minimisant l'erreur de classification. Cette optimisation est formulée comme un problème d'optimisation quadratique sous contraintes illustré par la figure 4.
- **Formulation mathématique :** Pour un problème de classification binaire avec données d'entraînement  $(x_1, y_1), \dots, (x_n, y_n)$ , où  $x_i \in \mathbb{R}^d$  et  $y_i \in \{-1, 1\}$ , l'objectif est de résoudre :

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i \\ \text{sous contraintes} \quad & y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n \\ & \xi_i \geq 0 \quad \forall i = 1, \dots, n \end{aligned} \tag{9}$$

où :

- $w$  : vecteur normal à l'hyperplan
- $b$  : terme de biais
- $\xi_i$  : variables d'écart pour les points mal classés
- $C$  : paramètre de régularisation contrôlant l'équilibre entre maximisation de la marge et minimisation de l'erreur
- **Formulation duale :** La formulation duale du problème SVM s'exprime comme :

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{sous contraintes} \quad & 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \tag{10}$$

où  $\alpha_i$  sont les multiplicateurs de Lagrange. La fonction de décision devient :

$$f(x) = \text{sgn} \left( \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \tag{11}$$

- **Importance des Hyperparamètres dans les SVM :** Les performances d'un SVM dépendent fortement des hyperparamètres sélectionnés. Deux hyperparamètres essentiels à ajuster sont :
  - **Paramètre  $C$  :** contrôle l'équilibre entre la maximisation de la marge et la minimisation des erreurs de classification. Une valeur faible de  $C$  permet plus d'erreurs de classification (favorisant la généralisation), tandis qu'une valeur élevée réduit les erreurs mais peut conduire au surapprentissage.
  - **Paramètre gamma ( $\gamma$ ) pour le noyau RBF :** détermine la portée de l'influence d'un seul point de données. Un  $\gamma$  élevé conduit à une frontière de décision plus complexe et à un potentiel surapprentissage, alors qu'un  $\gamma$  faible entraîne une frontière de décision plus lisse et peut améliorer la généralisation.

Ces hyperparamètres sont généralement choisis par des techniques de validation croisée afin de maximiser la capacité de généralisation du modèle.

## 3.2 Modèle Bayésien : les Processus Gaussiens

### a. Inférence bayésienne

L'inférence bayésienne est une méthode statistique qui interprète la probabilité comme une mesure de croyance ou de certitude, contrairement à la statistique fréquentiste, qui considère la probabilité comme une fréquence d'occurrence à long terme, basée sur des paramètres fixes. Dans le cadre bayésien, on commence avec des croyances initiales appelées *a priori*, que l'on met à jour à l'aide de données observées pour obtenir des croyances *a posteriori*. Cette mise à jour repose sur le **théorème de Bayes** (voir [Jaynes, 1968](#)), qui permet de calculer les probabilités à postériori en fonction des nouvelles informations ou données  $D$ . La formule du théorème de Bayes est la suivante :

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}$$

- $P(H|D)$  : probabilité **a posteriori** du paramètre (ou hypothèse)  $H$  étant donné les données  $D$  (croyance mise à jour après l'observation des données)
- $P(H)$  : probabilité **a priori** de  $H$  (croyance initiale avant l'observation des données)
- $P(D|H)$  : **vraisemblance**, c'est-à-dire la probabilité d'observer les données  $D$  si l'hypothèse  $H$  est vraie
- $P(D)$  : probabilité **marginale** des données (facteur de normalisation pour s'assurer que la somme des probabilités est égale à 1)

### b. Processus Gaussiens

Un **processus gaussien** est un ensemble de variables aléatoires indexées par le temps ou l'espace, dont toute sous-collection finie suit une distribution normale multivariée.

**Définition formelle :** Un processus stochastique  $\{X_t\}_{t \in T}$  est dit gaussien si, pour toute collection finie de temps  $t_1, t_2, \dots, t_n$  dans  $T$ , le vecteur aléatoire  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  suit une loi normale multivariée.

#### Caractéristiques principales :

- **Fonction moyenne**  $m(t)$  : Elle représente l'espérance mathématique du processus à l'instant  $t$ , définie par  $m(t) = \mathbb{E}[X_t]$ .
- **Fonction de covariance**  $k(t, t')$  : Elle mesure la dépendance linéaire entre les valeurs du processus aux instants  $t$  et  $t'$ , donnée par  $k(t, t') = \text{Cov}(X_t, X_{t'})$ .

Ces deux fonctions déterminent entièrement le comportement d'un processus gaussien.

En pratique, les processus gaussiens permettent :

- D'estimer la valeur moyenne d'une fonction inconnue,
- De quantifier l'incertitude de cette estimation à l'aide de la variance prédite,
- De sélectionner intelligemment des données pour l'entraînement.

Concrètement, on a pour chaque instant  $t$

$$f(t) \sim \mathcal{N}(m(t), k(t, t')).$$

En présence de nouvelles observations  $\mathbf{x}_*$ , on obtient la distribution postérieure :

$$f_* | \mathbf{x}_*, \mathbf{X}, \mathbf{y} \sim \mathcal{N}(\mu_*, \sigma_*^2),$$

où :

$$\mu_* = m(\mathbf{x}_*) + k_*^T (\mathbf{K} + \sigma^2 I)^{-1} (\mathbf{y} - \mathbf{m}),$$

$$\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - k_*^T (\mathbf{K} + \sigma^2 I)^{-1} k_*.$$

Ici  $\mathbf{m}$  et  $\mathbf{K}$  sont respectivement l'espérance et la matrice de covariance à priori. Ainsi, les processus gaussien sont une méthode d'inférence statistique bayésienne : partant d'une distribution à priori, on met à jour la distribution en présence de nouvelles observations  $\mathbf{x}_*$  et nous obtenons la distribution à postériori  $f_* | \mathbf{x}_*, \mathbf{X}, \mathbf{y}$ .

### c. Test bayésien

Contrairement au test statistique fréquentiste qui repose sur une valeur seuil (liée à la probabilité jugée acceptable de faire une erreur de type I), le **test bayésien** évalue la probabilité des hypothèses directement à l'aide des données, en calculant le **facteur de Bayes (Bayes Factor ou BF, Kass and Raftery, 1995)**.

Le facteur de Bayes permet de comparer deux hypothèses  $H_0$  et  $H_1$  en mesurant la vraisemblance des données sous chaque hypothèse :

$$BF = \frac{P(D | H_1)}{P(D | H_0)}$$

où :

- $P(D | H_1)$  est la probabilité des données  $D$  si  $H_1$  est vraie,
- $P(D | H_0)$  est la probabilité des données  $D$  si  $H_0$  est vraie.

En faisant l'hypothèse que la probabilité à priori de  $H_1$  est la même que celle de  $H_2$  nous avons par la formule de Bayes que

$$BF = \frac{P(D | H_1) P(H_0)}{P(D | H_0) P(H_1)} = \frac{P(H_1 | D)}{P(H_0 | D)}. \quad (12)$$

Ainsi les facteurs de Bayes permettent d'estimer à quel point l'hypothèse  $H_1$  est plus (ou moins) probable que l'hypothèse  $H_0$  étant données les nouvelles données  $D$ . Ils permettent donc de quantifier le degré auquel les nouvelles données soutiennent l'une ou l'autre des hypothèses. Si  $BF > 1$ , les données soutiennent davantage l'hypothèse  $H_1$  et si  $BF < 1$ , elles soutiennent  $H_0$ .

**Interprétation avec l'échelle de Jeffreys** Pour interpréter la valeur du facteur de Bayes, **Harold Jeffreys** a proposé une échelle qualitative, connue sous le nom de *tableau de Jeffreys* (Jeffreys, 1939), qui associe des intervalles de valeurs de  $BF$  à des niveaux de preuve :

BF (rapport $H_1/H_0$ )	Interprétation
$< 1$	Preuve en faveur de $H_0$ (hypothèse nulle)
$1 - 3$	Évidence faible pour $H_1$
$3 - 10$	Évidence modérée pour $H_1$
$10 - 30$	Évidence forte
$30 - 100$	Évidence très forte
$> 100$	Évidence décisive pour $H_1$

Cette méthode permet non seulement de détecter un phénomène, mais aussi d'en évaluer la **force de manière continue**, contrairement à une simple décision binaire (rejeter ou non  $H_0$ ).

## 4 Application de la méthode Optimal Fingerprint

### 4.1 Présentation du code

#### 4.1.1 Environnement de travail

- Nous avons travaillé tout au long du semestre sur Google Collab de manière interactive.
- Les librairies Python auxquelles nous avons eu recours sont principalement :
  - `numpy` : gestion efficace des calculs numériques sur des tableaux multidimensionnels.
  - `matplotlib` : visualisation graphique des données, création de graphiques et figures.
  - `pandas` : traitement et analyse des données structurées sous forme de tableaux et séries temporelles.
  - `xarray` : manipulation efficace de données multidimensionnelles, particulièrement adaptée aux jeux de données climatiques (`netCDF`).
  - `cftime` : gestion précise des dates et des calendriers spécifiques utilisés dans les jeux de données climatiques.
  - `cartopy` : création de cartes géospatiales et représentation géographique des résultats.
  - `os` : gestion du système de fichiers, accès aux dossiers et fichiers sur l'environnement de travail.
  - `google.colab` : utilisation des fonctionnalités spécifiques à l'environnement Google Colab, notamment l'intégration avec Google Drive.

#### 4.1.2 Implémentation de l'algorithme

- a) **Traitement des données** : pour chaque simulation, les étapes suivantes sont appliquées
- Nous avons réduit la granularité spatiale. Les données ont été rééchantillonnées à une résolution spatiale plus faible de l'ordre de mille points. Pour cela, nous avons implémenté une fonction `reduce_granularity` qui utilise `coarsen` de la bibliothèque `xarray`.
  - Ensuite, nous avons séparé les données en deux périodes : passées de 1880 à 1950 et présentes de 1951 à 2022.
  - Avec les données passées, nous avons réussi à calculer les saisonnalités (moyennes mensuelles pour chaque mois). Cette étape a été réalisée à l'aide de la fonction `groupby` sur l'axe `time.month` suivie du calcul de la moyenne.
  - Après, nous avons calculé les anomalies passées ainsi que les anomalies présentes. Pour ce faire, les saisonnalités calculées ont été retranchées des données.
  - Pour finir, nous avons regroupé les anomalies par année puis avons calculé la moyenne des valeurs pour chaque groupe d'une année donnée.

⇒ Ce traitement nous a permis d'obtenir des anomalies climatiques qu'on utilisera pour calculer le signal  $X$  ainsi que le bruit  $C_N$ .

b) **Calcul de beta** :

- Nous avons implémenté une fonction `compute_covariance_and_signal` qui renvoie le signal ainsi que la matrice de covariance. Le signal est la différence entre les moyennes des anomalies passées et présentes.
- La fonction `calculate_beta` retourne alors le vecteur de dimension  $(lat*lon,1)$  aussi appelé détecteur optimal  $\beta$

c) **Statistique de test** :

- Nous avons commencé par implémenter la fonction `calculate_gamma`.
- Ensuite, nous avons écrit la fonction `calculate_Z` qui retourne notre statistique de test sous forme de série temporelle.

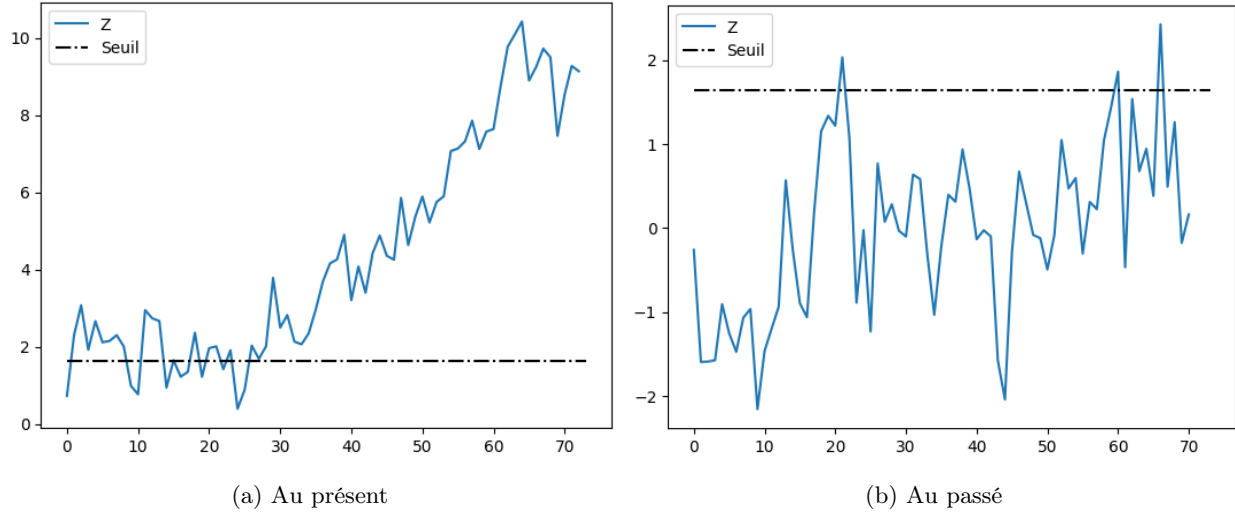


FIGURE 5 – Détection de changement climatique pour (a) des données récentes (après 1940) et passées (avant 1940). On observe que le test contrôle bien l’erreur de première espèce (taux de détection du changement climatique pour les données passées inférieur au seuil de 5%) et à une puissance relativement élevée pour les années les plus récentes.

⇒ Maintenant que nous avons implémenté les fonctions nécessaires pour faire le test, nous allons passer au calibrage de ce dernier ainsi qu’à l’analyse des résultats obtenus.

#### 4.1.3 Calibrage du test

Le but de cette partie est principalement de tester la fiabilité de notre algorithme.

Pour ce faire, nous l’avons entraîné sur le jeu de données en utilisant une approche de validation croisée : un fichier parmi les 50 a été utilisé comme fichier test, tandis que les 49 restants servaient à calculer les valeurs de beta.

Ce processus a été répété pour chaque fichier, ce qui nous a permis d’obtenir deux graphes par fichier [figure 5](#) : le premier représentant les valeurs de Z au passé [figure 5b](#) et le second au présent [figure 5a](#).

Les résultats similaires obtenus pour tous les fichiers montrent que notre algorithme détecte efficacement le changement climatique.

Concernant le choix du nombre de composantes pour la PCA, nous avons d’abord utilisé une valeur aléatoire, puis, pour garantir les meilleurs résultats, nous avons calculé la puissance du test et l’erreur de première espèce en faisant varier ce paramètre.

L’objectif était d’obtenir une puissance maximale tout en minimisant l’erreur de première espèce. En effet, on veut que cette erreur soit inférieur ou égale à  $\alpha$ .

Après analyse des résultats, nous avons déterminé que le choix optimal était  $n=70$ .

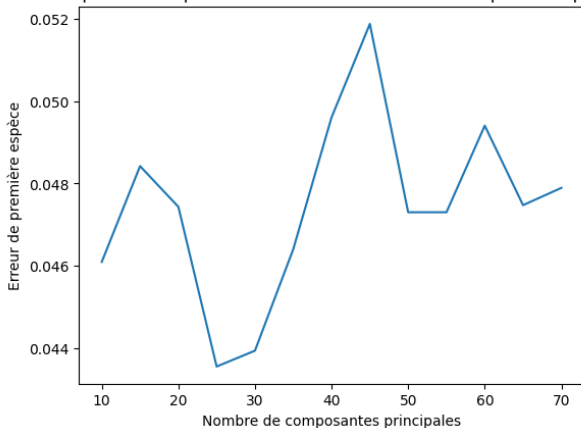
## 4.2 Application sur les données et résultats

a) Données :

Pour cette partie, nous avons travaillé sur un jeu de données d’observations fournies par notre tuteur. Ces données sont de ré-analyse ([Hersbach et al., 2020](#)), 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.

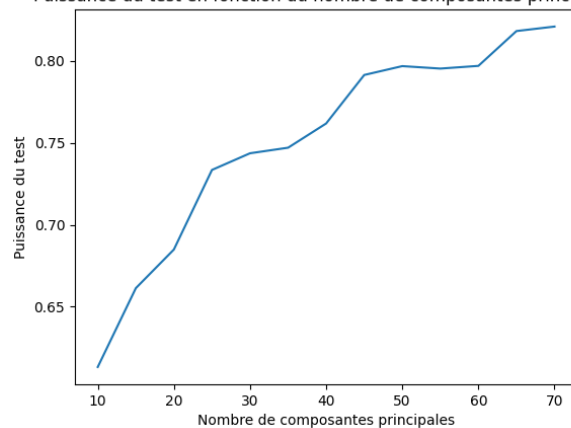


Erreur de première espèce en fonction du nombre de composantes principales



(a) Erreur première espèce en fonction de  $n$

Puissance du test en fonction du nombre de composantes principales



(b) Puissance du test en fonction de  $n$

FIGURE 6 – Impact de  $n$  sur les performances

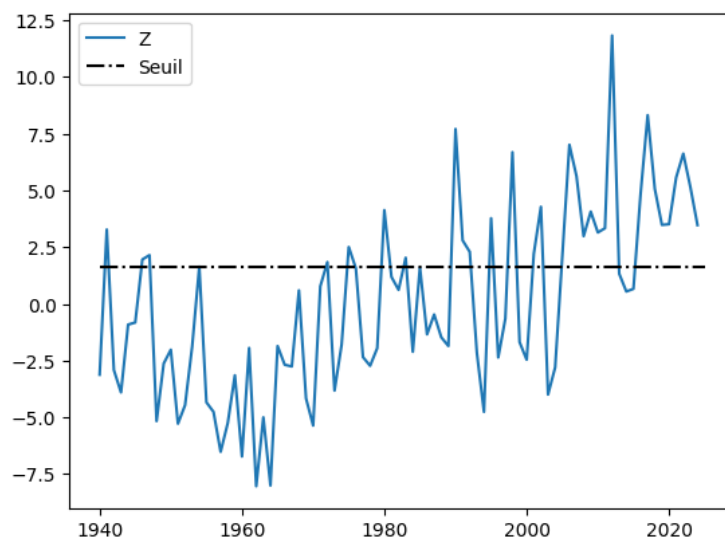


FIGURE 7 – Valeur de la statistique observée  $Z$  après avoir exécuté notre algorithme de l'Optimal Fingerprint

## 5 Application des méthodes non linéaires

L’hypothèse de linéarité est souvent utilisée en modélisation, mais elle ne reflète pas toujours fidèlement la complexité du monde réel. Les phénomènes climatiques sont influencés par de nombreux facteurs, souvent de manière non linéaire. Il est donc essentiel d’explorer des approches qui ne reposent pas uniquement sur des relations linéaires afin d’améliorer la précision des modèles de détection du changement climatique.

Dans cette partie nous allons donc voir comment implémenter de nouvelles méthodes de détection du changement climatique en utilisant cette fois des approches non linéaires. Ces méthodes vont se baser sur les Machines à Vecteurs de Support (SVM) et les Processus Gaussiens (GP) afin d’obtenir des résultats plus précis et plus fiables pour améliorer la robustesse de nos analyses.

### 5.1 Séparation des données

Dans cette première partie, nous allons explorer différentes approches pour séparer nos données en trois sous-ensembles :

- **Données d’entraînement** : utilisées pour ajuster les paramètres du modèle.
- **Données de validation** : utilisées pour ajuster les hyperparamètres et éviter le surajustement.
- **Données de test** : utilisées pour évaluer les performances finales du modèle.

L’objectif est de réussir à séparer de manière plus intelligente afin d’optimiser l’entraînement de nos modèles. Une séparation efficace des données permet d’obtenir des modèles précis et robustes.

Nous allons comparer deux méthodes de séparation :

1. Séparation **aléatoire** : les données sont divisées de manière purement aléatoire en 3 sous ensembles.
2. Séparation **basée sur les GP** : cette méthode permet d’optimiser la sélection des données d’entraînement en minimisant la variance des points sélectionnés pour l’entraînement.

#### 5.1.1 Explication des méthodes

##### 1. Séparation aléatoire

La séparation aléatoire est la plus simple et la plus courante. Cette méthode est rapide et facile à mettre en œuvre, mais elle ne prend pas en compte l’incertitude ou la fiabilité associée aux données.

##### 2. Séparation basée sur les Processus Gaussiens

L’utilisation des GP permet une sélection plus intelligente des données. L’idée est de prédire l’incertitude des données à l’aide d’un GP et de classer les points en fonction de leur variance. On divise ensuite les données en trois groupes :

- Les points les plus **sûrs** (faible variance) pour l’entraînement.
- Les points avec une variance **intermédiaire** pour la validation.
- Les points les plus **incertains** (variance élevée) pour le test.

Cette méthode permet à l’ensemble des données d’entraînement d’être le plus qualitatif possible, ce qui améliore la robustesse du modèle final.

#### 5.1.2 Implémentation des Méthodes

Pour la première méthode, qui sépare de manière aléatoire les données, nous allons tout simplement utiliser deux fois la fonction `train_test_split` de la librairie `sklearn`, qui permet de séparer un jeu de données en deux. Nous allons d’abord séparer les données une première fois pour obtenir les données d’entraînement,

puis une seconde fois pour séparer les données de test et celles de validation.

Pour la seconde méthode, qui utilise un processus gaussien, il va falloir suivre les étapes suivantes :

#### 1. Entraînement du Modèle GP

Nous commençons par :

- Charger et préparer les données.
- Séparer les données en train/test de manière aléatoire dans un premier temps.
- Initialiser un processus gaussien pour modéliser l'incertitude des données.
- Utiliser `sklearn` avec `GaussianProcessRegressor` ou `gpytorch` pour une approche plus flexible.

#### 2. Prédiction de l'incertitude

Avec le GP, nous pourrions prédire la variance pour l'ensemble des données et normaliser les valeurs de variance afin d'obtenir une échelle cohérente.

#### 3. Séparation des données en fonction de l'incertitude

Les quantiles permettent de classer les données en trois groupes :

- Entraînement : données avec la plus faible incertitude.
- Validation : données avec une incertitude moyenne.
- Test : données les plus incertaines.

Ces ensembles sont ensuite stockés pour l'entraînement du modèle final.

### 5.1.3 Pourquoi privilégier la méthode avec `gpytorch` ?

L'utilisation de `gpytorch` est préférable à celle de `sklearn` car :

- Elle permet de **mieux comprendre** l'optimisation et la modélisation.
- Elle offre un **meilleur contrôle** sur le choix du noyau et l'optimisation des hyperparamètres.
- Elle est **optimisée** pour gérer efficacement des ensembles de données volumineux.

Grâce à cette approche, nous obtenons des modèles plus précis et robustes.

## 5.2 SVM et test statistique fréquentiste

### 5.2.1 Explication de la méthode

Dans cette partie, nous allons exploiter la puissance des machines à vecteurs de support (SVM) combinée à une approche statistique fréquentiste pour la détection du changement climatique.

L'objectif est de procéder en deux étapes :

1. Entraîner un **SVM** sur des données de simulation pour détecter le changement climatique.
2. Utiliser une **courbe ROC** pour analyser les probabilités fournies par le modèle et déterminer un seuil de décision optimal qui contrôle le niveau  $\alpha$  du test .

Cette approche permet de profiter des capacités de classification des SVM tout en les intégrant dans un cadre statistique rigoureux pour réaliser un test d'hypothèse fiable.

### 5.2.2 Implémentation de la méthode

#### 1. Validation croisée et entraînement du modèle

Pour obtenir un modèle précis et robuste, il est essentiel de :

- Minimiser l'incertitude des données d'entraînement.

- Trouver les meilleurs hyperparamètres pour éviter le sous-apprentissage et le sur-apprentissage.

### Choix des hyperparamètres et validation croisée

Nous allons utiliser la validation croisée pour sélectionner les meilleurs paramètres du SVM de manière à maximiser la précision du modèle. L'optimisation se fera en faisant varier les hyperparamètres suivants :

- $\gamma$  : paramètre de la fonction noyau qui contrôle l'influence d'un seul point d'entraînement. Il prendra les valeurs 0.1, 1 et 10.
- $C$  : paramètre de régularisation, qui équilibre la complexité du modèle et la minimisation des erreurs. Il prendra également les valeurs 0.1, 1 et 10.
- Degré du noyau : comme nous utilisons un noyau polynomial, nous allons tester les degrés 2, 3 et 4.

Nous choisissons un noyau polynomial car nous cherchons une séparation non linéaire des données.

### Mise en œuvre avec GridSearchCV

Nous avons choisi la validation croisée k-fold, avec  $k=5$ . Pour la mettre en place, nous utilisons la fonction `GridSearchCV` de la librairie `sklearn`, qui permet de tester toutes les combinaisons des hyperparamètres.

Le principe :

- (a) **Diviser** les données en 5 parties.
- (b) **Entraîner** 5 fois le modèle sur 4 parties et tester sur la 5e.
- (c) **Moyenner** les performances pour obtenir une évaluation globale.
- (d) **Répéter** l'opération pour chaque combinaison d'hyperparamètres pour avoir une précision moyenne pour chaque combinaison.
- (e) **Sélectionner** le modèle avec la meilleure performance.

Cette méthode garantit que nous obtenons un SVM optimal capable de généraliser efficacement sur de nouvelles données.

## 2. Détermination du seuil et test statistique

Une fois le modèle SVM entraîné, il est capable de détecter un changement climatique. Cependant, nous souhaitons aller au-delà d'une simple classification en intégrant une approche statistique rigoureuse.

### Lien entre SVM et test d'hypothèse

Les SVM classifient les données en estimant une probabilité d'appartenance à la classe du changement climatique. Par défaut, un SVM attribue un point à une classe si sa probabilité d'appartenance est supérieure à 0.5.

Cependant, cette approche ne tient pas compte les erreurs de type I (faux positifs) et de type II (faux négatifs).

Nous voulons donc transformer cette classification en test d'hypothèse en définissant un seuil de décision optimisé pour minimiser les erreurs.

### Utilisation de la courbe ROC pour ajuster le seuil

Nous allons :

- (a) Extraire les probabilités des prédictions SVM.
- (b) Tracer une courbe ROC (Receiver Operating Characteristic) qui montre le compromis entre le taux de vrais positifs et le taux de faux positifs pour différents seuils.
- (c) Choisir le seuil optimal qui garantit un niveau de test  $\alpha$  souhaité.

### Pourquoi utiliser la courbe ROC ?

- Elle permet de contrôler l’erreur de type I en ajustant le seuil de décision.
- Elle nous donne la puissance du test, un indicateur clé pour évaluer la performance du modèle.

Cette approche nous permet donc de combiner l’apprentissage machine (SVM) avec une rigueur statistique pour améliorer la fiabilité de la détection du changement climatique.

#### 5.2.3 Conclusion

Nous avons ainsi construit une méthode complète en combinant SVM et test statistique :

1. Optimisation des hyperparamètres avec GridSearchCV pour obtenir le meilleur modèle possible.
2. Détermination d’un seuil optimal avec la courbe ROC pour garantir un bon compromis entre erreur et puissance du test.

Cette approche permet d’obtenir des résultats plus robustes que l’utilisation isolée d’un SVM ou d’un simple test statistique.

### 5.3 Processus Gaussiens, SVM et test statistique bayésien

Dans cette dernière partie, nous allons implémenter une méthode de détection du changement climatique en utilisant un test statistique bayésien au lieu d’un test fréquentiste. Pour mettre une telle méthode en œuvre, nous aurons besoin de combiner les SVM et les GP avec les statistiques bayésiennes.

#### 5.3.1 Explication de la démarche

Ce test repose sur les statistiques bayésiennes, notamment sur le calcul d’une loi *a posteriori* à l’aide d’un processus gaussien. L’objectif est d’estimer les facteurs de Bayes pour chaque point de données afin de déterminer si un changement climatique est détecté ou non.

Contrairement aux tests fréquentistes, qui reposent sur des seuils fixes de significativité (*p-value*), l’approche bayésienne permet d’obtenir une mesure de plausibilité du changement climatique en comparant les modèles à l’aide du facteur de Bayes.

#### 5.3.2 Implémentation de la méthode

1. **Estimation des probabilités issues du modèle SVM** Nous utilisons d’abord les probabilités de classification fournies par le SVM entraîné précédemment. Ces probabilités représentent le degré de confiance du modèle dans la classification des données climatiques.
2. **Apprentissage du processus gaussien** Un processus gaussien (GP) est ensuite entraîné sur ces probabilités afin de modéliser leur distribution et d’estimer l’incertitude associée aux prédictions. Pour entraîner le processus gaussien, nous utilisons la fonction `GaussianProcessClassifier` de la librairie `sklearn` qui permet de classer les données en utilisant un GP tout en indiquant à quel point il est sûr de sa prédiction ce qui va nous permettre de faire l’étape suivante.
3. **Calcul du facteur de Bayes** À partir de la sortie du processus gaussien, nous pouvons calculer les facteurs de Bayes pour chaque point de données. Le facteur de Bayes permet de comparer l’hypothèse  $H_0$  (absence de changement climatique) et l’hypothèse  $H_1$  (présence d’un changement climatique). Pour le calculer, on a l’équation (12). Le processus gaussien permet d’obtenir une distribution de probabilité pour chaque point, ce qui permet d’évaluer les vraisemblances  $P(D|H_1)$  et  $P(D|H_0)$  de manière plus robuste et adaptative et ainsi calculer ce facteur.
4. **Interprétation avec l’échelle de Jeffreys** Enfin, nous utilisons l’échelle de Jeffreys (3.2) pour interpréter les valeurs du facteur de Bayes et déterminer si le changement climatique est détecté de manière significative.

### 5.3.3 Conclusion

Grâce à cette méthode, nous obtenons une quantification plus rigoureuse de la détection du changement climatique. Le facteur de Bayes nous permet non seulement de détecter un changement, mais aussi d'évaluer la force de cette détection avec plus de précision qu'un simple test fréquentiste.

## 6 Résultats expérimentaux

### 6.1 Présentation des jeu de données

#### 6.1.1 Présentation du jeu de données sur les températures

Notre client nous a fourni un jeu de données contenant 50 fichiers de type NetCDF (.nc) représentant différentes simulations climatologiques générées par le modèle climatique Community Earth System Model Version 2 (CESM2) (Danabasoglu et al., 2020).

Tous les fichiers sont multidimensionnels de dimension (time,lat,lon,bnds). La période couverte va du **16 janvier 1880** au **16 décembre 2022** soit 1716 mois.

Les variables contenues sont :

- **tas** : variable principale représentant les températures en Kelvin à la surface terrestre de dimension (time,lat,lon).
- **time\_bnds** : définit les bornes temporelles pour chaque point dans la dimension *time*.

On nous a aussi fourni les observations des température dans le monde par mois sur la même période de temps. CE sont des données ERA5 produite par le Centre européen pour les prévisions météorologiques à moyen terme (ECMWF) dans le cadre du programme Copernicus.

#### 6.1.2 Présentation du jeu de données sur les précipitations maximum

Ce jeu de données, fourni par notre client, contient également 50 fichiers provenant du modèle CESM2. Cette fois, les données représentent les précipitations maximales par mois sur une période allant de 1880 à 2100, même si nous allons nous intéresser à la même période que précédemment.

Nous n'avons pas pu avoir accès aux données de précipitations dans les temps et nous gardons l'application de nos méthodes nonlinéaires à celles-ci pour des travaux futurs.

### 6.2 Résultats de l'Optimal Fingerprint

Les résultats obtenus à l'aide de l'algorithme d'Optimal Fingerprint serviront de référence pour évaluer les performances des méthodes non linéaires telles que les SVMs et les processus gaussiens. Cela permettra de comparer leur capacité à détecter les signaux de changement climatique et à capturer des relations complexes au sein des données.

Nous passons maintenant à l'application directe de l'algorithme d'Optimal Fingerprint. Nous avons commencé par entraîner le modèle sur les 50 fichiers de simulations et ainsi calculer la matrice de covariance, le signal et beta.

Sur la carte du signal [figure 8b](#), on observe particulièrement un changement climatique au pôle Nord. On observe aussi aucun signal négatif, il peut y avoir des endroits avec un signal faible mais jamais négatif ce qui indique que le réchauffement climatique s'observe partout dans le monde mais à des intensités différentes.

La carte de beta (voir [figure 8a](#)), quant à elle, représente une balance entre les endroits où le signal est fort et où le bruit est faible. Les régions où l'on va pouvoir observer le plus sûrement un changement climatique

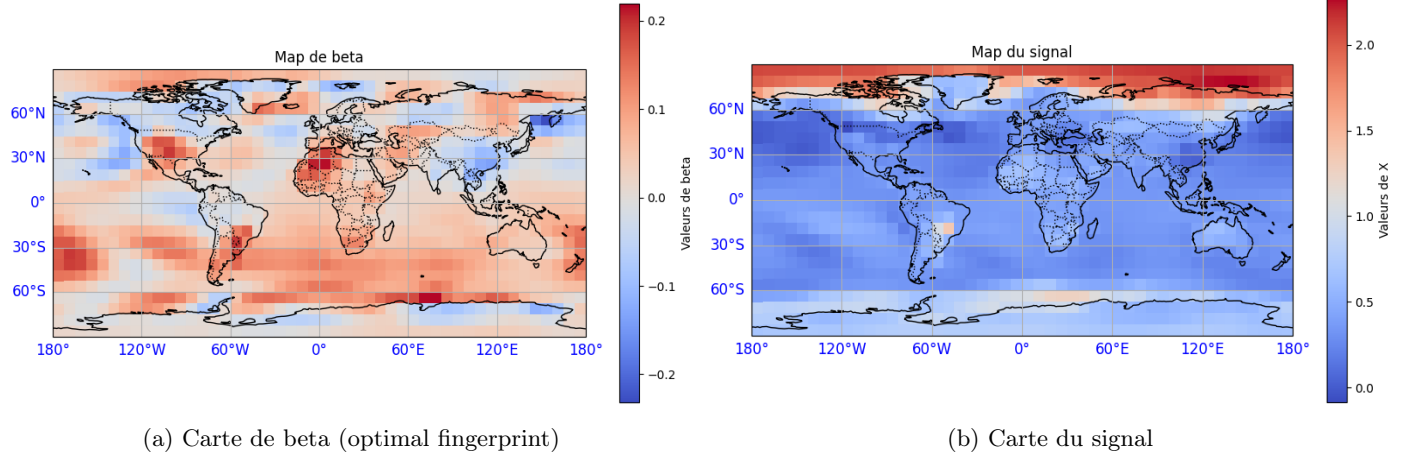


FIGURE 8 – Cartes représentatives des résultats

sont l'Afrique du Nord, l'Amérique du Sud (au niveau de l'Uruguay) et aussi au milieu de l'océan Pacifique.

Nous avons réussi après à calculer la statistique de test pour valider ou non notre hypothèse  $H_0$  (pas de changement climatique observé).

D'après le graphe obtenu, [figure 7](#), nous observons que de 1940 à 1980, l'hypothèse nulle  $H_0$  n'est pas rejetée, ce qui signifie que nous ne détectons pas de changement climatique significatif. Cependant, il est important de noter que ne pas rejeter l'hypothèse nulle ne signifie pas l'accepter. Cela indique simplement que les données ne fournissent pas suffisamment de preuves pour conclure à un changement climatique, mais cela ne signifie pas pour autant qu'il n'existe pas.

Ensuite, des années 1990 à 2010, il y a de plus en plus d'années où la statistique observée dépasse le seuil, environ une année sur deux.

Enfin, de 2010 à nos jours, tous les ans on rejette  $H_0$  et on peut affirmer avec plus de 95% de confiance, qu'il y a un changement climatique dû aux activités humaines. De plus, sur cette période, on aurait pu choisir un niveau beaucoup plus faible pour notre test statistique comme 0.1% et tout de même rejeter  $H_0$  et pouvoir affirmer qu'il y a un changement climatique car le nouveau seuil serait de 3.3 alors que notre statistique observée est supérieure à 3.5. Le niveau de confiance que nous avons dans le fait qu'il y a un changement climatique est donc très élevé.

On obtient également une puissance de test supérieure à 0,80 (exactement 0.824), ce qui est très satisfaisant. Cela confirme la robustesse et la précision de nos résultats avec cette méthode.

Pour conclure, grâce à la méthode de l'Optimal Fingerprint, nous avons réussi à montrer qu'il y avait un réchauffement climatique dû aux activités humaines et que notre niveau de confiance dans ce résultat est très élevé.

Cependant, nous constatons que cette détection intervient relativement tard par rapport aux connaissances actuelles, qui suggèrent que le changement climatique était déjà détectable dans les années 1980-1990. Cela soulève la question de la sensibilité des méthodes utilisées. Nous proposons donc d'explorer une approche non linéaire, telle que les SVMs ou les processus gaussiens, afin d'évaluer si elle permettrait d'identifier les signaux du changement climatique plus précocement.

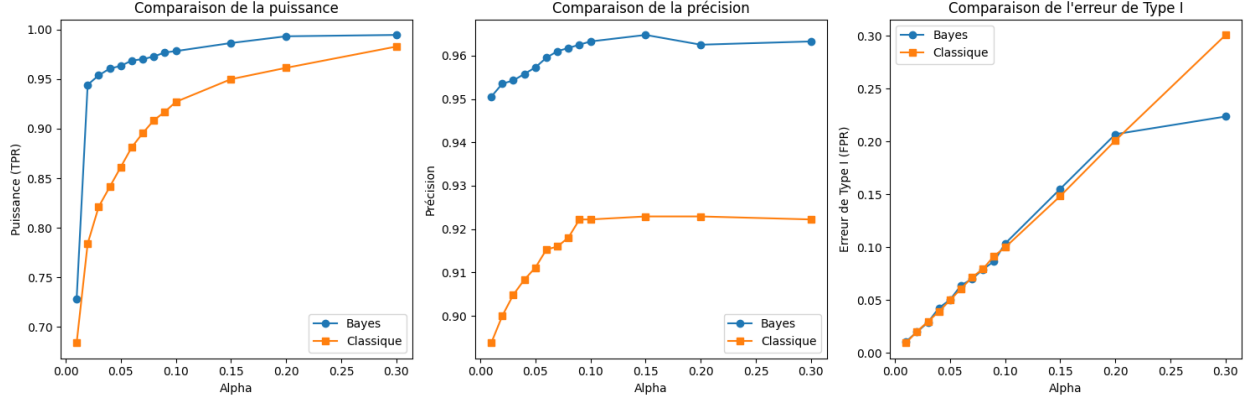


FIGURE 9 – Comparaison de la séparation classique et celle basée sur les processus gaussiens pour l'entraînement des modèles

## 6.3 Résultats des approches non linéaires

### 6.3.1 Calibration et puissance du test pour différent seuil $\alpha$

Nous allons maintenant comparer les deux méthodes de séparation des données (la séparation aléatoire et la celle utilisant un GP) pour voir laquelle permet d'avoir les modèles les plus précis et robustes. Pour ce faire nous allons regarder :

- La précision du modèle (erreur moyenne).
  - La puissance statistique (capacité à détecter un effet réel).
  - L'erreur de type I (faux positifs).
- pour différents niveaux de test  $\alpha$ .

La première chose que l'on remarque est qu'avec cette approche non linéaire, on obtient pour les deux méthodes de séparation des données, une puissance de test supérieure à celle de l'optimal fingerprint qui était de 0.824. On voit sur la [figure 9](#) que la séparation basée sur les processus gaussiens offre une puissance de test supérieure ou égale à 0.95 pour toute valeur de  $\alpha$ . De plus, ces modèles sont dotés de précision supérieure à environ 5% comparé à la séparation des données "classique".

On en conclut donc que les modèles entraînés grâce à cette séparation seront plus précis et plus fiables, ce qui justifie son utilisation systématique dans la suite du projet.

Nous utiliserons maintenant uniquement des modèles entraînés sur des données dont la séparation est basée sur les processus gaussiens.

### 6.3.2 Observations globales

Utilisons maintenant nos données d'observations globales pour détecter s'il y a un changement climatique ou non, au fil des années, à l'aide de notre modèle prédictif.

On remarque, sur la [figure 10](#), que l'on commence à détecter le changement climatique aux alentours de 1990. Le résultat est plus ou moins en accord avec les observations climatiques où on a constaté la croissance du réchauffement aux années 80. En comparaison avec la méthode optimal fingerprint qui détecte le réchauffement climatique uniquement à partir de 2010. On est donc capable de détecter beaucoup plus tôt le changement grâce à l'approche non linéaire.



Graphique de la détection du changement climatique en fonction des années grace au svm non linéaire

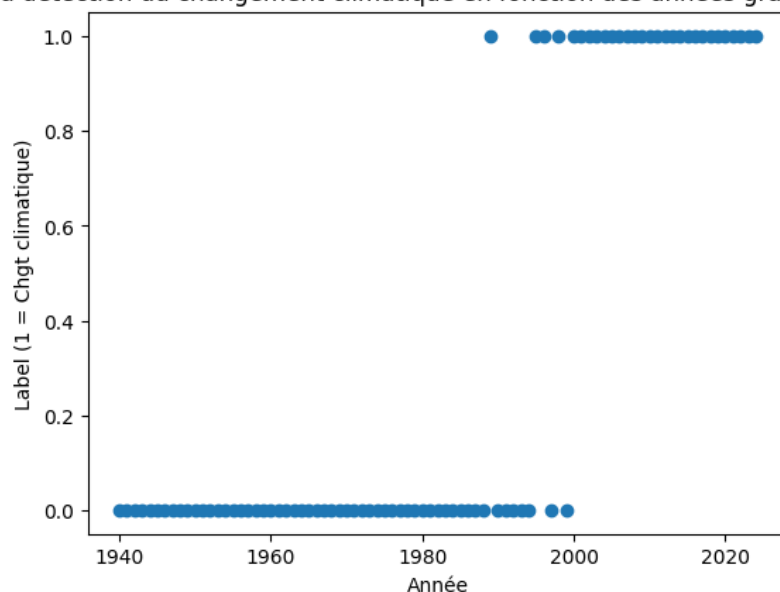


FIGURE 10 – Détection du changement climatique en fonction des années

## 6.4 Observations Amérique du Nord et Europe

Focalisons nous maintenant sur une région du monde, on va ici prendre l'exemple de l'Europe et l'Amérique du Nord. Notons d'abord que comme nous allons tester sur des plus petites régions, la détection sera plus difficile à cette échelle spatiale. Cela est dû au bruit qui sera plus fort, en raison des variations locales (orographie, effets de microclimat, influences régionales) qui introduisent des fluctuations naturelles plus marquées, sans que le signal climatique soit plus fort, ce qui nous donnera un rapport signal/bruit plus faible. Pour ce faire, nous allons procéder comme ceci :

- Déterminer les coordonnées de la région à étudier ( longitudes et latitudes )
- Sélectionner les données correspondants aux coordonnées de la région.
- Entraîner le modèle à l'aide de ces données, comme expliqués précédemment, à l'aide de la méthode processus gaussiens.
- Appliquer le modèle sur les observations de ces mêmes régions

Sur la [figure 11a](#), on a deux des anomalies de deux régions, l'Amérique du Nord et l'Europe. On obtient ces cartes en calculant les valeurs moyennes de la température des données du passé, puis on les compare avec les nouvelles valeurs. Ce qui nous donne des anomalies fortes, si la différence est significative ou encore aucune anomalie, si la valeur de la différence est identique.

La carte sur la [figure 11a](#) couvre toute l'Amérique du Nord, en bleu, sont marqués les anomalies faibles ou négatives, tandis qu'en rouge, ce sont les anomalies élevées. On observe au Nord-Est de Canada, au Québec, de grandes anomalies signifiant un changement significatif. On remarque aussi côté ouest des Etats-Unis, en Californie, quelques zones orangées signifiant des anomalies légères. Beaucoup de bleu est présent dans les zones maritimes ou au sud-est des Etats-Unis, on interprète cela comme des valeurs normales ou en-dessous de la moyenne attendue.

L'autre carte ([figure 11b](#)) couvre toute l'Europe, avec la même légende. On y remarque des zones rouges en Europe de l'est, dans les Balkans, à la mer Noire ou encore en Turquie, ce qui est très probablement liées à un événement climatique extrême, type vague de chaleur ou sécheresse. Quant à elle, L'Europe centrale présente des anomalies modérées à fortes. Tandis qu'en France, Royaume-Uni, Espagne, on observe peu ou pas d'anomalies.

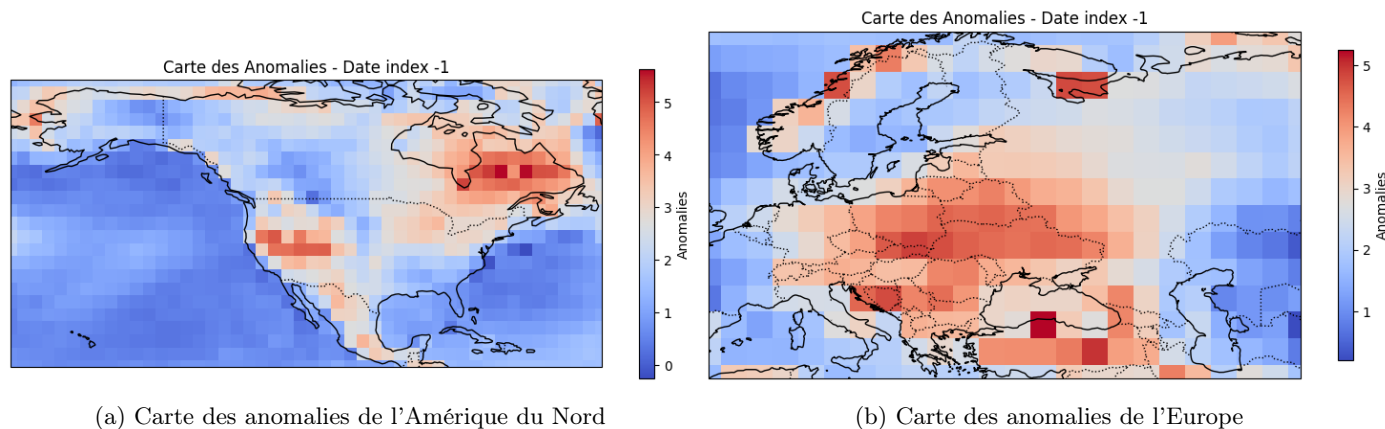


FIGURE 11 – Cartes représentatives des observations régionales.

Graphique de la détection du changement climatique en fonction des années grace au svm non linéaire et à un test bayésien

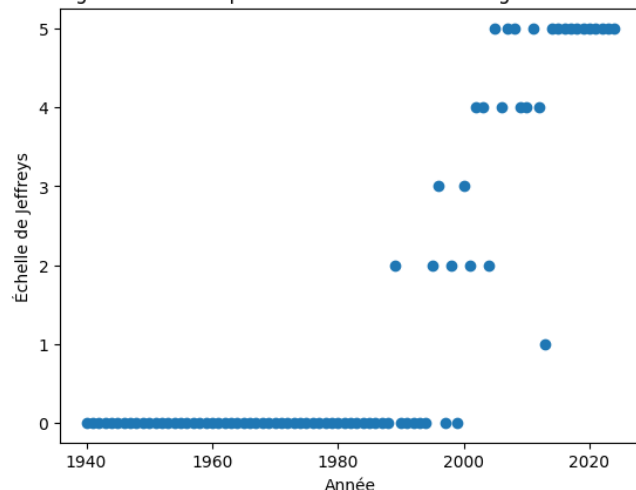


FIGURE 12 – Détection globale du changement climatique en fonction des années avec la méthode bayésienne

En Europe, les anomalies sont plus concentrées sur une bande géographique tandis qu'en Amérique du Nord, elles sont plus diffuses et réparties.

## 6.5 Résultats du test bayésien

- Facteurs de Bayes pour les observations (Globales/Europes/Precipitations) Comme pour la partie précédente, nous allons nous intéresser aux observations globales mais cette fois-ci, avec un modèle de détection du changement climatique utilisant les facteurs de Bayes. On remarque sur ce graphe (figure 12) qu'on détecte progressivement le changement climatique à partir de 1990. L'avantage de cette méthode est qu'on peut quantifier la "confiance" que l'on a quand à la présence ou non du changement climatique comparés à la figure 10 où l'on a uniquement une réponse binaire.

Qu'en est-il des observations régionales pour cette méthode ?

On remarque que l'on obtient pour les observations régionales tout comme les globales, une détection progressive du changement climatique ce qui est intéressant à interpréter. Sur la figure 13a, on peut voir qu'à partir de 1950, on détecte du changement climatique de manière certaine selon l'échelle

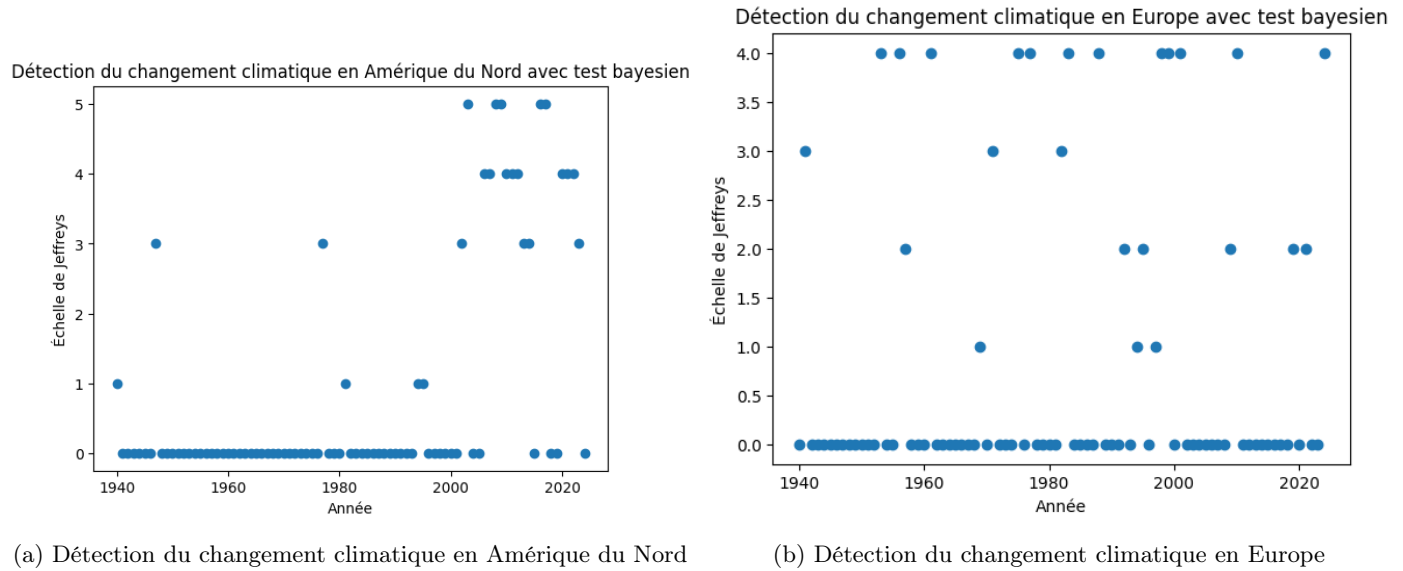


FIGURE 13 – Détection du changement en fonction des années avec la méthode bayésienne et des observations régionales

de Jeffreys, cela peut s'interpréter comme les conséquences de la révolution industrielle ainsi eu lieu aux Etats-Unis lors de la fin du 19ème siècle. A partir des années, on est certains de la présence du changement climatique. Alors que sur la [figure 13b](#), on ne converge pas vers un changement climatique certains à l'aide de ce modèle. Cela peut s'expliquer par le fait que l'Europe est une petite région où il y a de fortes disparités de détection du changement climatique, comme vue sur la carte des anomalies de l'Europe ([figure 11b](#)).

On peut donc conclure que ce modèle est plus intéressant pour apporter des réponses nuancées avec un indice de confiance que nous fournit l'échelle de Jeffreys tandis que celui basé sur les processus gaussiens est plus adapté à une réponse binaire, de présence ou non, de changement climatique.

## 6.6 Précipitations extrêmes

Afin d'évaluer l'efficacité de nos modèles au-delà des données habituelles de température, nous avons décidé de tester leur performance sur des données de précipitations extrêmes, reconnues pour leur complexité et leur forte variabilité. Les précipitations représentent en effet un cas typique de phénomène climatique extrême, généralement plus difficile à modéliser en raison de leur nature non linéaire et leur caractère souvent imprévisible.

Les deux approches (SVM et Optimal Fingerprint) sont entraînées et calibrées sur 49 des 50 fichiers et la puissance est ensuite testé sur le dernier fichier. Ainsi nous obtenons pour chacune des approches 50 valeurs de puissance.

Nous pouvons observer sur la [figure 14a](#) que l'approche basée sur les SVMs à une puissance bien plus élevée et est donc capable de détecter le changement climatique des données de précipitations de manière plus précise avec des résultats plus fiables.

Cette différence notable souligne le potentiel considérable des SVMs pour modéliser efficacement les données climatiques extrêmes, particulièrement complexes.

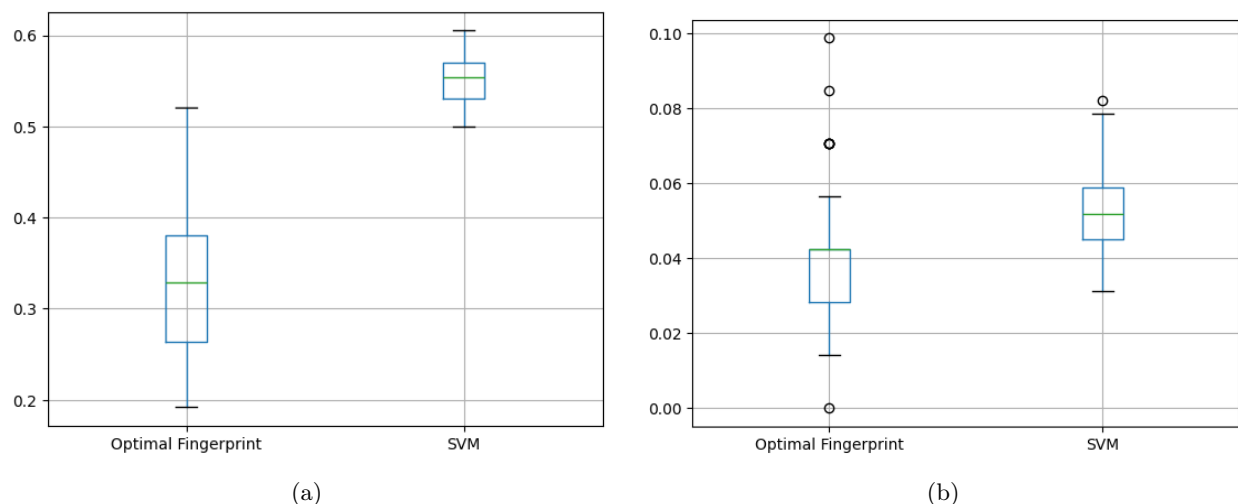


FIGURE 14 – (a) Puissance des tests linéaire (Optimal Fingerprint) et non linéaire (SVM). Les boxplots représentent la distribution des puissances pour chaque fichier de simulation de test. On voit que la puissance du test non linéaire est bien plus élevée. (b) Erreur de type I pour méthode linéaire et non linéaire. La méthode nonlinéaire à un contrôle de l'erreur de type I légèrement moins bon.

## 7 Réflexion sur les aspects RSE

Tout d'abord il est important de noter que notre projet ne s'inscrit pas totalement dans la définition d'un projet industriel, nous ne travaillons pas pour une entreprise mais pour un chercheur dans une université. En conséquence, notre produit final ne représente pas un bien de consommation ou de production mais bien un rapport scientifique, répondant à une problématique.

### — Impact environnemental :

Impact positif : Nous expliquons les méthodes que nous utilisons et tentons de verbaliser un maximum les résultats pour qu'ils soient compréhensibles. Le but est que nos lecteurs prennent conscience des enjeux climatiques, que les individus et entreprises prennent des décisions pour limiter cet impact.

Impact négatif : exécuter nos algorithmes nécessite une quantité non négligeable d'électricité, cela produit donc des gaz à effet de serre. Nous tentons de limiter cet impact en réalisant nos simulations tests sur de petits échantillons.

### — Impact éthique :

Impact positif : nous faisons avancer la recherche en présentant des résultats honnêtes, sans biais. Nos résultats viennent s'ajouter à d'autres pour appuyer la thèse d'un changement dû à l'Homme.

Impact négatif : on dépense beaucoup d'énergie tout en critiquant la consommation d'énergies fossiles.

### — Impact sociétal :

Impact positif : nos résultats pourraient être un signal d'alarme pour nos lecteurs, qui prendraient enfin conscience de l'ampleur du problème. Cela pourrait les amener à changer leurs habitudes et prendre des décisions en faveur de l'environnement.

Impact négatif : si nos résultats sont très négatifs, cela pourrait déclencher un sentiment de fatalisme chez nos lecteurs qui pourraient penser qu'il n'y a rien à faire, et c'est bien le danger, ne rien faire est la pire des solutions. Si nos résultats sont plus mitigés ils pourraient être déformés par certains de nos lecteurs qui seraient confortés dans leur pensée climatosceptique.

Pour conclure ce paragraphe, il faut noter que la manière dont les résultats sont communiqués est presque aussi importante que les résultats eux-mêmes. Nous avons fait un travail rigoureux pour utiliser les bons

mots, aux bons endroits, afin d'éviter toute mauvaise interprétation.

Les chercheurs en sciences climatiques publient des papiers de recherche qui peuvent se distinguer en 2 grandes catégories : ceux qui donnent un état de la situation telle qu'elle est ; et ceux qui fournissent également des solutions. Les deux visions sont respectables, la première vise à ne pas évoquer de sujet politique pour ne pas être décrédibilisée. La seconde tente d'exposer aux lecteurs et décideurs politiques des améliorations, au risque d'être rejetés par une partie des lecteurs qui le verraient comme un message politique qu'on tente de leur imposer.

Pour notre part, nous avons choisi de rester dans le cadre scientifique strict, en laissant les décideurs politiques et les lecteurs tirer leurs propres conclusions sur les actions à mener.

## 8 Conclusion :

En résumé, le projet a permis d'explorer et de démontrer la pertinence de la méthode Optimal Fingerprint pour détecter les signaux du changement climatique, en s'appuyant sur des données climatiques et des techniques d'analyse statistique robustes. Les résultats obtenus confirment avec un haut degré de confiance l'impact des activités humaines sur le réchauffement climatique, tout en soulignant les variations géographiques et l'évolution temporelle des signaux détectés. De plus, on a observé que les méthodes non linéaires permettent une meilleure détection du changement climatique et que ces méthodes pourraient être explorées plus en profondeur afin de voir si elles peuvent aider à la détection de changement climatique à des échelles locales ou sur des variables climatiques très chaotiques telles que les précipitations ou la circulation atmosphérique, qui représentent deux des principaux défis actuels en détection et attribution (D&A).

De plus, les méthodes non linéaires ont montré un potentiel prometteur pour une détection plus précoce et plus fine du changement climatique. Leur capacité à capturer des relations complexes suggère qu'elles pourraient être explorées plus en profondeur, notamment pour améliorer la détection à des échelles locales ou sur des variables climatiques très chaotiques, comme les précipitations ou la circulation atmosphérique. Ces aspects représentent aujourd'hui des défis majeurs en détection et attribution.

Par ailleurs, une réflexion approfondie sur les aspects RSE du projet a mis en lumière l'importance de la rigueur scientifique et de la transparence dans la communication des résultats. L'objectif de sensibiliser à l'urgence climatique tout en minimisant les impacts environnementaux et éthiques liés à la conduite du projet a guidé les décisions méthodologiques. Bien que des limites existent, le travail contribue à une meilleure compréhension des enjeux climatiques, encourageant ainsi des actions collectives et responsables pour un avenir durable.

## 9 Bibliographie

### Références

- Thomas L Bell. Optimal weighting of data to detect climatic change : Application to the carbon dioxide problem. *Journal of Geophysical Research : Oceans*, 87(C13) :11161–11170, 1982.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20 :273–297, 1995.
- Gokhan Danabasoglu, J-F Lamarque, J Bacmeister, DA Bailey, AK DuVivier, Jim Edwards, LK Emmons, John Fasullo, R Garcia, Andrew Gettelman, et al. The community earth system model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, 12(2) :e2019MS001916, 2020.
- Gene H Golub and Charles F Van Loan. An analysis of the total least squares problem. *SIAM journal on numerical analysis*, 17(6) :883–893, 1980.
- Klaus Hasselmann. Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate*, 6(10) :1957–1971, 1993.
- Gabriele Hegerl and Francis Zwiers. Use of models in detection and attribution of climate change. *Wiley interdisciplinary reviews : climate change*, 2(4) :570–591, 2011.
- Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730) :1999–2049, 2020.
- Edwin T Jaynes. Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3) :227–241, 1968.
- Harold Jeffreys. *Theory of Probability*. Clarendon Press, Oxford, England, 1939.
- Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90 (430) :773–795, 1995.
- Edward N Lorenz. Deterministic nonperiodic flow. *Journal of atmospheric sciences*, 20(2) :130–141, 1963.
- P. A. Stott M. R. Allen. Estimating signal amplitudes in optimal fingerprinting, part i : theory. 2003.
- C.H Trisos, I.O. Adelekan, E. Totin, A. Ayanlade, J. Efitre, A. Gemed, K. Kalaba, C. Lennard, C. Masao, Y. Mgaya, G. Ngaruiya, D. Olago, N.P. Simpson, and S. Zakieldean. *Africa*, pages 1285–1455. Cambridge University Press, Cambridge, UK and New York, USA, 2022. ISBN 9781009325844. doi : 10.1017/9781009325844.011.1286.
- Thomas Tomkins Warner. *Numerical weather and climate prediction*. cambridge university press, 2010.
- Larry Wasserman. *All of statistics : a concise course in statistical inference*. Springer Science & Business Media, 2013.
- Francis W.Zwiers. The detection of climate change. 1999.

Vous pouvez accéder au code associé à ce projet via ce lien :  
[https://github.com/Ghali062/Projet\\_Industriel\\_OF.git](https://github.com/Ghali062/Projet_Industriel_OF.git).