

# MASTER TRIED



---

## Rapport Projet Long

---

### Detection and attribution of climate change A Deep Learning and Variational approach

---

Auteurs

DAHMANI SALAH EDDINE  
XUE XINYU

Tuteurs

M. MEJIA CARLOS  
MME. THIRIA SYLVIE  
M. CONSTANTIN BÔNE

2 avril 2023

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Présentation du sujet et de l'objectif du stage.	3
1.2	Importance de l'IA dans l'étude des changements climatiques	3
1.3	Objectifs du rapport	4
<b>2</b>	<b>Contexte</b>	<b>4</b>
2.1	Présentation du laboratoire	4
2.2	Présentation du contexte des changements climatiques et de l'apprentissage profond	4
<b>3</b>	<b>État de l'art</b>	<b>5</b>
<b>4</b>	<b>Méthodologie</b>	<b>6</b>
4.1	Présentation des données	6
4.2	Procédures de traitement et d'analyse des données.	7
4.2.1	Visualisation des données et suppression des valeurs aberrantes	7
4.2.2	Génération des fichiers All-sim, models et forcings	8
4.2.3	Génération de fichiers pour chaque forçage et filtrage des données pour tous les forçages	8
4.2.4	Génération des données Train/Test/Inversion	9
4.3	Apprentissage direct	10
4.4	Régression linéaire	10
4.5	Apprentissage par inversion (Backward Optimisation)	11
<b>5</b>	<b>Résultats et discussions</b>	<b>12</b>
5.1	Présentation et analyse des résultats obtenus.	13
5.2	Comparaison avec les résultats de la régression linéaire	16
<b>6</b>	<b>Conclusion</b>	<b>17</b>
<b>7</b>	<b>Annexes</b>	<b>18</b>

## Introduction générale

Actuellement, le changement climatique est un véritable problème auquel le monde entier est confronté, suscitant une attention et une recherche croissantes. L'augmentation constante des émissions de gaz à effet de serre peut avoir des conséquences environnementales négatives, telles qu'une hausse de la température globale, une montée du niveau des mers, des événements météorologiques extrêmes plus fréquents, entre autres. Dans cette situation, la détection et l'attribution du changement climatique sont indispensables, car elles permettent de distinguer les facteurs de forçage du changement climatique, d'expliquer les mécanismes du changement climatique, d'identifier les contributions des activités humaines et des facteurs naturels au changement climatique, et de proposer des stratégies d'atténuation et d'adaptation viables. En détail, la détection vise à prouver l'existence du changement climatique et à éliminer sa variabilité interne, qui est causée par les processus intrinsèques du système climatique. L'attribution vise à expliquer et quantifier l'impact des différents facteurs de forçage détectés sur le changement climatique.

Les facteurs de forçage du changement climatique comprennent les activités humaines et les facteurs naturels. Parmi ceux-ci, les activités humaines comprennent principalement les émissions de gaz à effet de serre, les changements d'utilisation des terres, etc., tandis que les facteurs naturels comprennent principalement le rayonnement solaire, les éruptions volcaniques, etc. La force de forçage est la force motrice principale du changement climatique, tandis que la variabilité interne est une variation naturelle du changement climatique.

Cependant, les méthodes de détection et d'attribution du changement climatique actuelles présentent certaines limites. Par exemple, les méthodes actuelles sont généralement basées sur l'hypothèse d'additivité de la force de forçage et ne peuvent pas traiter efficacement les interactions complexes entre les différents facteurs de forçage. Pour cette raison, de nouvelles méthodes ont été proposées, telles que des cadres alternatifs basés sur la prédiction non linéaire, utilisant des réseaux de neurones convolutifs (CNN) et des méthodes d'inversion variationnelle pour effectuer l'attribution du changement climatique. Ces méthodes présentent des avantages pour explorer les structures non linéaires des données de changement climatique et sont susceptibles d'être largement utilisées à l'avenir. Par conséquent, sur cette base, nous devons continuellement améliorer et innover les méthodes pour mieux étudier le changement climatique et relever les défis qu'il pose.

# 1 Introduction

## 1.1 Présentation du sujet et de l'objectif du stage.

Le changement climatique montre une tendance croissante depuis l'industrialisation et a produit des effets irréversibles. Les indicateurs clés de l'atmosphère, des océans, des sols gelés et de la biosphère montrent que le système climatique subit des changements sans précédent. Depuis 1850-1900, la température moyenne à la surface du globe a augmenté de 0,8°C à 1,3°C, et la vitesse de croissance de la température de surface est sans précédent depuis au moins 2000 ans.[1]. Au cours des dernières décennies, les principaux indicateurs du système climatique sont à des niveaux jamais vus auparavant et ont connu des changements à une vitesse sans précédent. L'augmentation de la température a entraîné des événements météorologiques extrêmes, une élévation du niveau de la mer, la fonte des calottes glaciaires polaires, etc. Les activités humaines, en particulier l'utilisation de combustibles fossiles, sont la cause du changement climatique en raison de l'émission de gaz à effet de serre, ce qui intensifie l'effet de serre. Le changement climatique est également influencé par des facteurs naturels tels que les éruptions volcaniques et le rayonnement solaire.

Dans cette situation, la détection et l'attribution du changement climatique sont indispensables, car elles permettent de distinguer les facteurs de forçage du changement climatique, d'expliquer les mécanismes du changement climatique, d'identifier les contributions des activités humaines et des facteurs naturels au changement climatique, et de proposer des stratégies d'atténuation et d'adaptation viables. En détail, la détection vise à prouver l'existence du changement climatique et à éliminer sa variabilité interne, qui est causée par les processus intrinsèques du système climatique. L'attribution vise à expliquer et quantifier l'impact des différents facteurs de forçage détectés sur le changement climatique.

Les facteurs de forçage du changement climatique comprennent les activités humaines et les facteurs naturels. Parmi ceux-ci, les activités humaines comprennent principalement les émissions de gaz à effet de serre, les changements d'utilisation des terres, etc., tandis que les facteurs naturels comprennent principalement le rayonnement solaire, les éruptions volcaniques, etc. La force de forçage est la force motrice principale du changement climatique, tandis que la variabilité interne est une variation naturelle du changement climatique.

Dans ce contexte, l'apprentissage profond apparaît comme une approche potentiellement efficace pour détecter et attribuer les changements climatiques. Les réseaux de neurones convolutifs (CNN) sont capables d'apprendre des relations complexes entre les variables climatiques et les phénomènes atmosphériques, ce qui permet de détecter les changements climatiques à une échelle fine et de les attribuer à des causes spécifiques. En particulier, la technique de l'optimisation inverse est actuellement en cours d'étude pour évaluer son efficacité dans l'estimation des contributions des différentes variables climatiques aux changements observés.

## 1.2 Importance de l'IA dans l'étude des changements climatiques

L'Intelligence Artificielle (IA) joue un rôle en évolution dans l'étude des changements climatiques. Avec la croissance continue de la quantité de données climatiques disponibles, l'IA offre une occasion unique d'analyser de manière efficace et précise ces données.

Les techniques d'apprentissage profond, en particulier, ont montré des résultats prometteurs dans l'identification des tendances et des modèles climatiques complexes à travers des données massives et hétérogènes.[2]. Les algorithmes d'apprentissage profond peuvent apprendre à partir de grandes quantités de données historiques pour aider à prédire les changements climatiques futurs et contribuer ainsi à la prise de décision dans la lutte contre le changement climatique. Non seulement que l'IA permet de détecter les influences humaines sur les changements climatiques et d'attribuer les causes de ces changements à des facteurs spécifiques tels que les émissions de gaz à effet de serre, la fonte de la glace de mer et les changements dans la circulation atmosphérique et océanique, mais elle est également utilisée pour aider à la planification et à la prise de décisions concernant

l'adaptation aux changements climatiques, notamment pour les villes et les communautés côtières confrontées à la montée des eaux et aux tempêtes de plus en plus fréquentes.

### 1.3 Objectifs du rapport

Notre objectif est l'application d'une méthode testée sur le globe à l'étude régionale et de comparer les évolutions du réchauffement sur deux régions distinctes. Ce rapport fournira une vue d'ensemble de la problématique, des méthodes utilisées et des résultats obtenus.

Dans un premier temps, ce rapport passera en revue la littérature sur l'utilisation de l'apprentissage profond dans l'étude des changements climatiques, en mettant l'accent sur la technique de backward optimization. Cette revue de littérature permettra de mieux comprendre l'état de l'art et les enjeux de cette recherche.

Dans un deuxième temps, ce rapport décrira la méthodologie utilisée pour l'analyse des données climatiques. Cette méthodologie comprendra les étapes de préparation des données, de l'entraînement des modèles de réseau de neurones et de la validation croisée pour évaluer les performances des modèles. En même temps, la méthode de régression linéaire sera également présentée pour une comparaison avec la méthode mise en œuvre dans ce stage.

Enfin, les résultats obtenus seront présentés en détail, notamment les conclusions sur les contributions des différentes variables climatiques aux changements de température observés dans l'Atlantique Nord et le Pacifique équatorial. Ce rapport soulignera également les limites de cette étude et les perspectives de recherche futures dans ce domaine mais aussi proposer des pistes pour l'atténuation et l'adaptation face à ces changements.

## 2 Contexte

### 2.1 Présentation du laboratoire

Le Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques (LOCEAN) conduit des études sur les processus physiques et biogéochimiques de l'océan et leur rôle dans le climat en interaction avec les écosystèmes marins. Ses équipes, largement reconnues au niveau international, abordent une large gamme d'échelles de temps et d'espace pour une meilleure compréhension de la dynamique et des variations de l'océan au sein du système climatique ainsi que son évolution présente, passée et future. Elles contribuent aussi au développement de méthodes d'analyses, de modélisation et d'observation, ainsi qu'à l'observation systématique de l'océan, *in situ* ou depuis l'espace.

LOCEAN est également impliqué dans le développement de méthodes d'analyse, de modélisation et d'observation de l'océan, *in situ* ou depuis l'espace. Dans ce cadre, le laboratoire a développé le modèle climatique IPSL-CM6A-LR, un modèle de pointe qui est largement utilisé dans la recherche en climatologie.

### 2.2 Présentation du contexte des changements climatiques et de l'apprentissage profond

Dans ce projet, nous nous intéressons à l'utilisation de l'apprentissage profond et de la technique de backward optimization pour la détection et l'attribution des causes du changement climatique. Nous nous concentrerons sur l'étude des séries temporelles de température historique de l'air dues à la concentration individuelle des différentes types de gaz ou aérosols dans l'atmosphère. Nous distinguons trois types :

- AER : les gaz à effet de serre,
- GHG : les aérosols anthropiques ou produits par l'activité humaine (hors gaz à effet de serre, bien entendu)

- NAT : les aérosols de type naturel (provenant de cendres volcaniques ou dus aux variations solaires)

Ce sont ces trois types de gazes ou aérosols qui peuvent affecter l'atmosphère dans sa capacité à empêcher l'arrivée du rayonnement solaire ou au contraire à retenir la chaleur irradié causant des changements dans la température de l'air dans un sens ou dans l'autre. Nous présentons une méthode de détection-attribution basée sur des CNN entraînées et backward optimization, et nous l'appliquons à des données climatiques pour identifier les facteurs responsables des changements observés. Nous discutons également des limites de cette approche et des perspectives futures pour la détection-attribution des changements climatiques.

### 3 État de l'art

L'attribution et la détection des changements climatiques sont des sujets clés dans le domaine de l'environnement et de la science du climat. Avec la montée des préoccupations sur l'impact des activités humaines sur le changement climatique, la nécessité de détecter les changements climatiques et de les attribuer à des causes spécifiques est devenue plus pressante. Dans cette optique, l'utilisation de l'apprentissage profond pour la détection et l'attribution des changements climatiques est devenue une approche de plus en plus populaire dans la communauté scientifique.

Plusieurs études ont montré que l'apprentissage profond est capable de détecter et d'attribuer les changements climatiques avec une précision accrue. Par exemple, Liu et ses collègues ont utilisé un réseau de neurones convolutifs pour prédire les températures moyennes annuelles dans différentes régions de la Chine, et ont obtenu des résultats très précis [3]. De même, les auteurs de [4] ont utilisé un réseau de neurones récurrents pour prédire les précipitations annuelles dans différentes parties de l'Inde, et ont obtenu des résultats qui concordent avec les observations historiques.

De plus, La technique de backward optimization est une méthode relativement nouvelle dans le domaine de la climatologie pour comprendre comment les différentes variables climatiques contribuent aux changements climatiques. Cette approche est basée sur backward optimization des entrées d'un modèle de réseau de neurones pour maximiser la contribution de chaque entrée à la prédiction de la variable de sortie. Plusieurs études ont montré l'efficacité de cette technique dans l'analyse des changements climatiques.

Par exemple, Wang et al. dans [5] ont utilisé la technique de backward optimization pour identifier les contributions des différentes variables climatiques aux changements de température à la surface de l'océan dans l'Atlantique Nord. Les auteurs ont appliqué une descente de gradient inversée sur les variables climatiques pour maximiser leur contribution à la prédiction de la température, et ont identifié les régions où chaque variable a la plus forte contribution. Les résultats ont montré que la fonte de la glace de mer dans l'océan Arctique était la principale cause des changements de température observés dans l'Atlantique Nord.

De même, Makkonen et al. [6] ont utilisé la technique de backward optimization pour identifier les contributions des différentes sources de pollution atmosphérique aux niveaux de particules fines dans l'air. Les auteurs ont appliqué une descente de gradient inversée sur les émissions de polluants pour maximiser leur contribution aux niveaux de particules fines observés, et ont identifié les sources de pollution les plus importantes. Les résultats ont montré que les émissions de véhicules étaient la principale source de particules fines dans l'air.

Enfin, Toms et al. [7] ont utilisé la technique de backward optimization pour identifier les contributions des différentes variables climatiques à la variabilité de la précipitation en Australie. Les auteurs ont appliqué une descente de gradient inversée sur les variables climatiques pour maximiser leur contribution à la variance de la précipitation, et ont identifié les variables les plus importantes dans différentes régions du pays. Les résultats ont montré que l'oscillation australe El Niño était la principale cause de la variabilité de la précipitation dans la plupart des régions d'Australie.

## 4 Méthodologie

### 4.1 Présentation des données

Les données utilisées dans cette étude comprennent des simulations de modèles climatiques et des observations. Les simulations de modèles climatiques ont été obtenues à partir des sorties des modèles de circulation générale océan-atmosphère (GCM) du projet CMIP6 (Coupled Model Intercomparison Project)[1]. Les simulations historiques (HIST) ont été effectuées pour la période 1850-2014 en utilisant 12 modèles différents, avec des variations de gaz à effet de serre, d'aérosols, de l'ozone, de l'activité solaire et de l'utilisation des terres. En outre, des simulations à forçage unique ont été utilisées pour des scénarios avec des gaz à effet de serre (GHG), des aérosols (AER) et des forçages naturels (NAT) provenant d'éruptions volcaniques et de variations solaires. Les données d'observation(OBS) proviennent de la température de l'air à 2 mètres sur les continents, provenant du HadCRUT4[8] combiné avec la température de surface de la mer provenant du HadISST4[9]. Dans ces simulations, les modèles climatiques sont utilisés avec en entrée un forçage qui varie les autres étant conservés à une valeur moyenne

Les chiffres dans le tableau indiquent le nombre de membres de simulation pour chaque forçage utilisé dans chaque modèle. Les chiffres dans les colonnes GHG, AER et NAT sont très déséquilibrés, certains modèles ayant un grand nombre de membres tandis que d'autres ont très peu. Par conséquent, des méthodes doivent être mises en place lors de l'étape d'apprentissage pour équilibrer le nombre de membres pour chaque modèle.

	GHG	AER	NAT	HIST	References
CESM2	3	3	2	11	Danabasoglu et al. 2020
IPSL-CM6A-LR	10	10	10	32	Boucher et al. 2020
ACCESS-ESM1-5	3	3	3	30	Ziehn et al. 2019
BCC-CSM2-MR	3	3	3	3	Wu et al. 2019
CanESM5	50	50	30	65	Swart et al. 2019
CNRM-CM6-1	9	10	10	10	Volodire et al. 2019
FGOALS-g3	3	3	3	6	Li et al. 2020
HadGEM3	4	4	4	5	Roberts et al. 2019
MIROC6	3	3	3	50	Tatebe et al. 2019
MRI-ESM2.0	5	5	5	7	Seiji et al. 2019
NorESM2-LM	3	3	3	3	Seland et al. 2020
GISS-E2-1-G	10	12	20	19	Kelley et al. 2020

FIGURE 1 : Modèle et simulations utilisés dans cette étude. Les chiffres dans les colonnes GHG, AER et NAT fournissent le nombre d'instances de simulation utilisées pour chaque modèle

Toutes les données mensuelles ont été converties en moyennes annuelles et ont été corrigées pour être comparables. Les valeurs manquantes ont été remplies par kriging. Les anomalies de température ont été estimées en utilisant le climat préindustriel comme référence pour la période 1900-2014. On fait l'hypothèse que avant 1990 (ère préindustrielle) on n'observe pas de réchauffement dus aux forçages. Pour chaque simulation, la moyenne de la température a été calculée pour la période 1850-1900 et soustraite de la température. Les séries temporelles ont ensuite été normalisées pour faciliter la comparaison entre les différents modèles.

Dans les fichiers, il y a une dimension "région" qui comprend un numéro de région parmi les 58 régions définies dans Iturbide et al. (2020)[10] pour The sixth IPCC assessment report (AR6). Les régions couvrent les terres et les océans (ar6.all) et sont également divisées en catégories terres (ar6.land) et océans (ar6.ocean). Les données sont des moyennes annuelles pour 165 années.

En outre, une analyse plus détaillée a été menée pour les régions de l'océan Pacifique équatorial(région 48 :EPO) et de l'Atlantique Nord (région 50 :NAO),qui sont connues pour leur grande

variabilité interne. Il est important de vérifier si le modèle de Détection Attribution utilisé à l'échelle mondiale est applicable aux études régionales, car le climat régional peut varier considérablement d'une région à l'autre du globe.

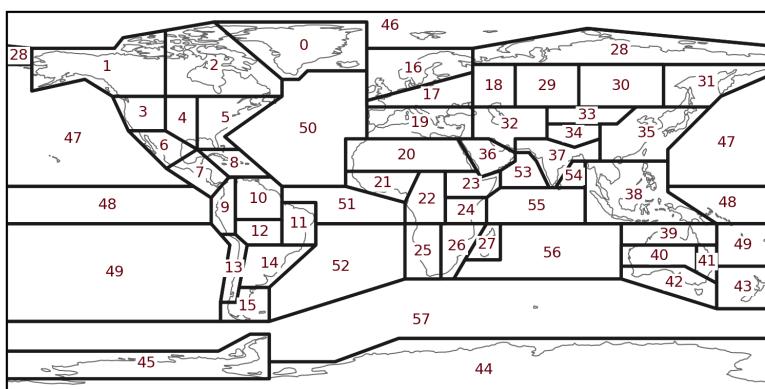


FIGURE 2 : La carte de zones définie pour The sixth IPCC assessment report (AR6)[10]

Pour le région 48, EPO signifie Equatorial Pacific-Ocean, qui fait référence à la région de l'océan Pacifique équatorial située près de l'équateur. Cette région est influencée par plusieurs facteurs, notamment les courants océaniques, les vents alizés et les variations de température de surface de l'océan, qui peuvent affecter le climat et les conditions météorologiques dans la région ainsi que dans les zones avoisinantes. C'est aussi la région des océans où le phénomène climatique de El Niño est le plus fortement observé apportant une forte variabilité aux températures de surface. Cette variabilité ayant une certaine periodicité.

Pour le région 50, NAO signifie North Atlantic Oscillation, qui est une variation de la circulation atmosphérique dans la région de l'Atlantique Nord. La NAO est souvent associée à des variations de température et de précipitations dans la région de l'Atlantique Nord, en particulier en Europe et en Amérique du Nord.

## 4.2 Procédures de traitement et d'analyse des données.

Les données sont dans des fichiers au format NetCDF, avec un fichier par forçage et par modèle, soit un total de 48 fichiers. A l'intérieur des NetCDF, les différents membres de chaque fichier apparaissent comme des séries temporelles avec des noms de variables différents. Chaque fichier contient une dimension "région" avec un numéro correspondant à l'une des 58 régions possibles, ainsi qu'une dimension "time" donnant l'année, car il s'agit de moyennes annuelles. Les données couvrent la période de 1850 à 2014 et représentent les anomalies de température par rapport à la période préindustrielle (avant 1850), sans lissage ni filtre appliqué. La période de simulation préindustrielle a été soustraite de la moyenne de chaque simulation.

### 4.2.1 Visualisation des données et suppression des valeurs aberrantes

Tout d'abord, nous avons analysé les données afin de mieux comprendre leurs caractéristiques et leur distribution. Nous avons effectué une visualisation des données et avons identifié des valeurs aberrantes dans tous les modèles. En d'autres termes, certaines données étaient nettement plus élevées que les autres. La moyenne des variations de température se situe généralement entre -1,5 et 1,5, mais nous avons trouvé des valeurs extrêmement supérieures à cette plage. Par conséquent, suite à une discussion avec nos encadrants, nous avons choisi de supprimer les colonnes contenant des valeurs supérieures à 10.

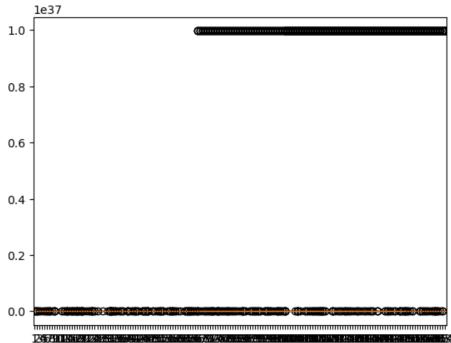


FIGURE 3 : Boxplot avant la suppression des valeurs aberrantes pour les données d'historique de modèle ‘MIROC6’

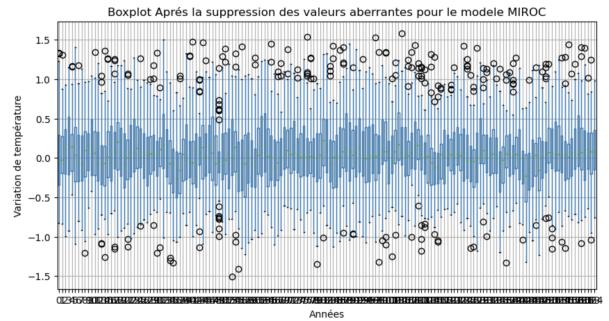


FIGURE 4 : Boxplot après la suppression des valeurs aberrantes pour les données d'historique de modèle ‘MIROC6’

Une analyse des données a été réalisée pour étudier les corrélations entre les membres de chaque année. En étudiant les variables associées, une certaine corrélation a été observée entre les membres au cours de la période de 1900 jusqu'à 2014. Parfois, une forte corrélation a été remarquée entre les années consécutives. (Un exemple de diagramme de dispersion et de matrice de corrélation est présenté en annexe).

#### 4.2.2 Génération des fichiers All-sim, models et forcings

A partir des données NetCDF mentionnées, nous allons préparer plusieurs fichiers de données pour chaque région étudiée :

- All-sim : Un fichier contenant un tableau de dimensions (nb-forcings, nb-models), indiquant le nombre de membres disponibles pour chaque modèle et chaque forçage. Il y a trois forçages et une variable de simulation, ce qui fait un total de 4 colonnes. Les modèles sont représentés par des colonnes et les forçages par des lignes.
- models : Un fichier contenant simplement la liste des modèles participant aux données, dans l'ordre des colonnes du tableau mentionné dans le fichier All-sim.
- forcings : Un fichier contenant les étiquettes des forçages dans l'ordre correspondant aux lignes du fichier All-sim, qui incluent le nombre de membres pour chaque modèle et chaque forçage.

#### 4.2.3 Génération de fichiers pour chaque forçage et filtrage des données pour tous les forçages

Comme mentionné précédemment, les données sont enregistrées en séparant par type de forçage et par modèle. Afin de pouvoir effectuer l’entraînement et l’apprentissage ultérieurs avec le CNN, nous devons donc générer 4 fichiers différents pour chaque type de forçage. Nous atteignons cet objectif en bouclant sur tous les fichiers NetCDF associés à chaque type de forçage, puis en les concaténant selon la dimension des lignes. Cela donne lieu à la création de 4 nouveaux fichiers différenciés par type de forçage pour chaque région, avec les noms et dimensions suivants :

Les données originales, représentées par la courbe orange sur la figure ci-dessus, montrent une variation très importante et comportent beaucoup de bruit en termes de variation de température. Pour éliminer les hautes fréquences de bruit, nous avons utilisé la méthode `signal.butter()` et `signal.filtfilt()` du module Python ‘scipy’ pour filtrer toutes les variables HIST, GHG, NAT et AER. Le résultat est une courbe bleue plus uniforme et plus lisse, qui représente les données filtrées. Les abscisses représentent les années, et l’ordonnée représente la variation de température.

TABLE 1 : Les noms et dimensions des fichiers générés

Forçage	dimension
GHG	(103,167)
AER	(89, 167)
NAT	(160, 167)
HIST	(256, 167)

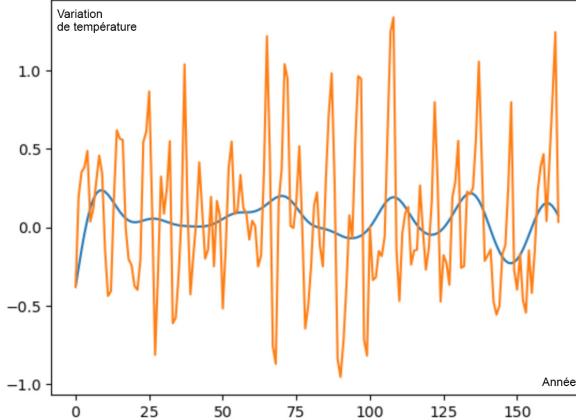


FIGURE 5 : Les abscisses représentent les années, et l'ordonnée représente la variation de température : Les données d'origine(orange) et les données filtrés(bleue)

#### 4.2.4 Génération des données Train/Test/Inversion

Les données qui seront utilisées pour l’entraînement et inversement des membres sont composées d’une entrée ‘X’ composée de trois forçages ‘GHG’, ‘AER’ et ‘NAT’ associées à une sortie ‘Y’ avec un profil ‘HIST’. La composition des ensembles de données pour l’entraînement est donc une affaire de combinaisons des forçages. Voici le tableau des modèles et les combinaisons des simulations :

n	MODEL	nombre de simulations G-A-N-H par forçage	nombre max de combi. G-A-N	nombre max de combi. G-A-N-H
0)	BCC-CSM2-MR	[3, 3, 3, 3]	27	81
1)	FGOALS-g3	[3, 3, 3, 6]	27	162
2)	CanESM5	[50, 30, 50, 65]	75000	4875000
3)	CNRM-CM6-1	[10, 10, 10, 30]	1000	30000
4)	ACCESS-ESM1-5	[3, 3, 3, 40]	27	1080
5)	<b>IPSL-CM6A-LR</b>	<b>[10, 10, 10, 33]</b>	<b>1000</b>	<b>33000</b>
6)	MIROC6	[3, 10, 44, 50]	1320	66000
7)	HadGEM3-GC31-LL	[4, 4, 4, 5]	64	320
8)	MRI-ESM2-0	[5, 5, 5, 11]	125	1375
9)	GISS-E2-1-G	[10, 15, 20, 35]	3000	105000
10)	CESM2	[3, 2, 3, 11]	18	198
11)	NorESM2-LM	[3, 3, 3, 3]	27	81
12)	GFDL-ESM4	[1, 1, 3, 3]	3	9

FIGURE 6 : Les modèles et les combinaisons des simulations

Comme le tableau le montre, il existe une grande différence entre le nombre de simulations par forçage pour les différents modèles, ce qui entraîne une différence considérable dans le nombre maximum de combinaisons. Le nombre minimum est seulement de 3, tandis que le nombre maximum atteint 75 000. Si nous utilisons directement de telles combinaisons pour l’entraînement, la précision de l’entraînement pour les modèles avec un nombre de combinaisons plus faible pourrait

être mise en doute. Par conséquent, nous avons décidé de définir un seuil de combinaison de 1000. Pour les modèles ayant un nombre de combinaisons inférieur à 1000, nous allons répéter les combinaisons existantes jusqu'à ce que leur nombre atteigne 1000. Pour les modèles ayant un nombre de combinaisons supérieur à 1000, nous allons choisir aléatoirement 1000 combinaisons parmi toutes les combinaisons disponibles pour équilibrer les données dans chaque modèle.

Ensuite, une fois les données préparées, nous commençons à diviser les ensembles de données en vue de l'entraînement, du test et de l'inversion. Tout d'abord, nous divisons l'ensemble de données en 85% pour l'ensemble d'entraînement et 15% pour l'ensemble de test pour un apprentissage direct. Ensuite, nous divisons l'ensemble d'entraînement en ensembles d'entraînement et de validation pour l'inversion.

### 4.3 Apprentissage direct

Dans cette étude, nous avons utilisé une approche d'apprentissage profond basée sur un réseau de neurones convolutifs (CNN) pour la détection et l'attribution des changements climatiques. Tout d'abord, nous avons déterminé la relation reliant les simulations historiques (HIST) aux simulations avec forçages de gaz à effet de serre (GHG), d'aérosols (AER) et de forçages naturels (NAT). Pour cela, nous avons entraîné le CNN en utilisant les séries temporelles de GHG, AER et NAT comme entrée du réseau de neurones, avec une taille de (3,115), et HIST comme cible, avec une taille de (1,115). Le CNN obtenu permet d'estimer l'anomalie de la température globale moyenne sur la période historique à partir des séries temporelles de GHG, AER et NAT.

Le CNN est constitué de trois couches de convolution unidimensionnelles, avec une taille de noyau de 7. La fonction tangente hyperbolique est utilisée comme fonction d'activation sur les couches cachées. La phase d'entraînement est divisée en trois étapes utilisant la norme de l'erreur quadratique moyenne (MSE) :

- (1) nous sélectionnons aléatoirement un modèle climatique.
- (2) nous sélectionnons aléatoirement une instance de chaque simulation (GHG, AER, NAT et HIST) du modèle.
- (3) nous entraînons le réseau en utilisant les séries temporelles correspondantes de GHG, AER, NAT et HIST comme entrée et cible (en utilisant une taille de batch de 100).

Dans notre étude, nous avons examiné 12 modèles climatiques en utilisant 85% de la base de données initiale pour l'apprentissage. L'architecture et les hyperparamètres, ont été fixés dès le début par le tuteur prenant en compte les travaux précédents, ces derniers ont été obtenus après plusieurs essais avec une technique de validation simple. Par la suite, nous avons entraîné un CNN en explorant 1000 combinaisons possibles des trois types de forçages existants (GHG, NAT et AER), afin d'assurer une utilisation équitable de chaque modèle climatique lors de l'entraînement du réseau. L'erreur de validation moyenne a été estimée pour chaque combinaison, et l'architecture retenue a été celle qui présentait l'erreur de validation la plus faible.

### 4.4 Régression linéaire

L'objectif de la régression multalinéaire est d'établir une relation linéaire entre une variable de sortie et plusieurs variables d'entrée. Dans notre cas, nous utilisons cette approche pour examiner l'hypothèse additive des forçages climatiques, qui suppose que les effets des différents forçages s'additionnent de manière linéaire. Cependant, cette hypothèse peut présenter certaines limites, notamment dans le cas où des interactions complexes entre les forçages sont à l'œuvre.

Pour tester l'hypothèse additive, nous avons utilisé les valeurs moyennes des forçages GHG, AER et NAT ainsi que les données d'observation OBS pour chaque modèle. Les trois types de forçages (GHG, AER et NAT) sont considérés comme variables d'entrée, tandis que la valeur d'observation (OBS) est utilisée comme variable de sortie. Nous avons séparé les données en un ensemble d'entraînement et un ensemble de test afin d'évaluer la performance du modèle de régression linéaire dans la simulation de notre problème.

En procédant ainsi, nous pouvons évaluer si l'hypothèse additive est suffisante pour expliquer les variations observées, ou si des méthodes plus complexes, telles que celles basées sur l'apprentissage profond, pourraient être nécessaires pour mieux comprendre et quantifier les contributions des différents forçages climatiques aux changements observés.

Pour la région 48 et région 50, nous avons obtenu les résultats suivants :

TABLE 2 : Les résultats obtenus par la régression linéaire

Region	Poids	RMSE
48	[0.37341347 -0.39172237 0.06687582]	0.5068
50	[1.12963748 0.72312223 1.29274067]	0.3831

Dans les figures ci-dessous, la couleur rouge représente les observations tandis que la couleur pourpre illustre les prévisions d'observation obtenues par le modèle linéaire. Il y a une différence significative entre les deux, et pour la région sélectionnée, les poids des variables indépendantes ne sont pas égaux à 1.

Cela démontre que les différents facteurs ont des impacts distincts sur le changement climatique et qu'ils ne sont pas purement additifs. Cela met en évidence les limites du modèle linéaire pour simuler la complexité du changement climatique. Cela suggère que des approches non linéaires et plus sophistiquées pourraient être mieux adaptées pour modéliser les relations complexes entre les différents forçages et les variations de température observées.

Dans l'ensemble, la régression linéaire permet de mettre en lumière les relations linéaires potentielles entre les forçages et les changements climatiques observés, tout en offrant un point de comparaison pour évaluer les méthodes plus avancées.

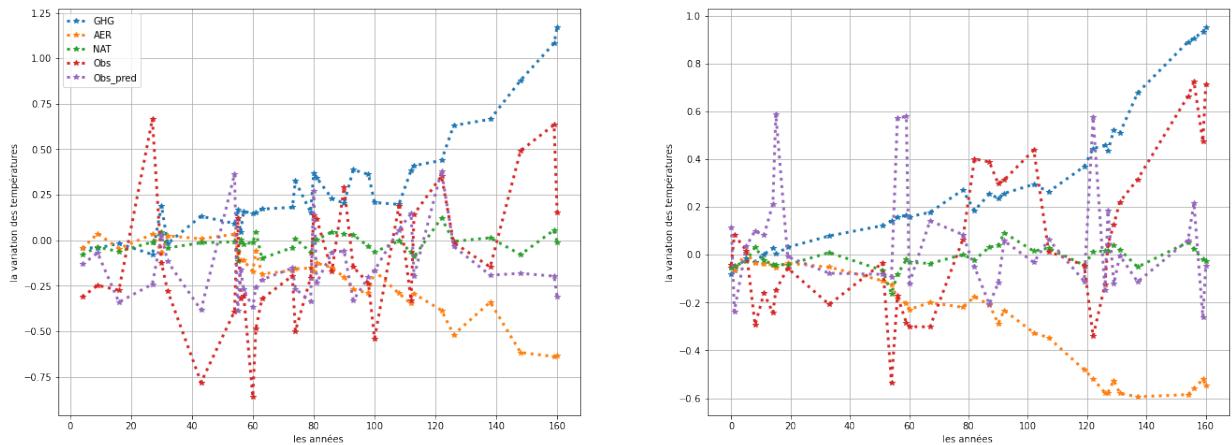


FIGURE 7 : Résultat de la régression linéaire pour les régions 48 (à gauche) et 50 (à droite) : la moyenne de l'ensemble pour les (Bleu) gaz à effet de serre, (Orange) aérosols anthropiques, (Vert) forçages naturels, (Rouge) observations et (Violet) prédictions des observations

## 4.5 Apprentissage par inversion (Backward Optimisation)

Dans le contexte de l'attribution et la détection des changements climatiques, la technique de backward optimization fait référence à une approche qui consiste à modifier les entrées (par exemple, dans notre cas c'est les forçages GHG, NAT, AER) plutôt que les poids du réseau de neurones. Cette technique permet d'identifier les contributions relatives de chaque variable d'entrée à la prédiction de la variable de sortie qui est la température en appliquant une descente de gradient inversée sur les entrées plutôt que sur les poids du réseau de neurones. Cette méthode est utile pour comprendre comment les différentes variables climatiques contribuent aux changements climatiques, ainsi que pour identifier les régions géographiques où ces contributions sont les plus importantes.

Le processus d'inversion implique la recherche de la combinaison optimale des forçages climatiques GHG, NAT et AER, qui correspond le mieux aux observations de température. Pour ce faire, la fonction de coût est formulée comme suit :

$$J(X) = \text{MSE}(\text{OBS}, \text{CNN}(X)) + B \times \text{MSE}(X, X_{ref}) \quad (1)$$

La première partie de cette équation,  $\text{MSE}(\text{OBS}, \text{CNN}(X))$ , mesure l'écart entre les observations réelles de température et les prédictions du modèle d'apprentissage profond (CNN) pour les entrées X. L'objectif de cette partie de la fonction de coût est de forcer l'inversion à se rapprocher autant que possible de l'observation réelle. La deuxième partie de l'équation,  $B \times \text{MSE}(X, X_{ref})$ , mesure l'écart entre les entrées inversées (X inversé) et les entrées de référence ( $X_{ref}$ ), qui sont les moyennes des forçages climatiques pour tous les modèles climatiques. Cette partie de la fonction de coût est utilisée pour s'assurer que les entrées estimées restent physiquement cohérentes et proches des moyennes des forçages climatiques pour tous les modèles climatiques. Nous avons conservé les hyperparamètres trouvés pour l'étude complète qui sert de modèle, y compris le facteur d'échelle B fixé à 0.02 pour doser l'influence de la deuxième partie dans la fonction de coût finale. Le processus d'inversion implique donc une recherche itérative des entrées optimales pour minimiser la fonction de coût  $J(X)$ , en utilisant des méthodes de gradient pour estimer les paramètres de contrôle. Ce processus peut être sensible à l'initialisation et nécessite souvent la répétition de l'optimisation à partir de différents points de départ pour obtenir une solution stable et cohérente.

Cette minimisation est répétée jusqu'à ce que  $\text{MSE}(\text{OBS}, X)$  soit inférieure à  $0.0025^{\circ}\text{C}/^{\circ}\text{C}$ . Le problème consiste donc à choisir l'initialisation  $X_{st}$  optimale, et pour ce faire, nous avons utilisé plusieurs initialisations physiquement cohérentes.

Pour chacun des douze modèles climatiques, nous sélectionnons aléatoirement un maximum de 100 triplets de membres de GHG, AER et NAT, en ne faisant pas des triplets répétés. Cela fait, pour les modèles climatiques ayant peut de membres, moins de 100 triplets. En tout pour les 12 modèles climatiques présents nous avons 790 triplets uniques pour l'initialisation, et générerons ainsi autant d'inversions variationnelles.

## 5 Résultats et discussions

Dans cette section, nous analysons les résultats obtenus à la suite des diverses expériences menées sur chacune des deux régions (48 et 50). Pour résumer, notre approche se déroule en deux étapes principales :

Dans un premier temps, nous avons effectué un apprentissage supervisé en utilisant un réseau de neurones convolutif (CNN) pour développer un modèle capable de recevoir en entrée les données relatives aux membres des trois types de forçages (GHG, NAT, AER) et de produire en sortie les historiques d'évolution de la température pour la période 1900-2014.

La seconde étape consiste en l'application de l'inversion, dont l'objectif est de prédire les entrées les plus adaptées à une observation donnée. Autrement dit, il s'agit d'identifier les facteurs les plus influents de ces observations, qui correspondent aux trois forçages (GHG, NAT, AER).

Dans les paragraphes suivants, nous examinerons en détail les résultats obtenus pour chacune de ces étapes, en discutant de leur pertinence et de leur impact sur notre compréhension des relations entre les forçages et l'évolution de la température.

## 5.1 Présentation et analyse des résultats obtenus.

### Apprentissage Direct

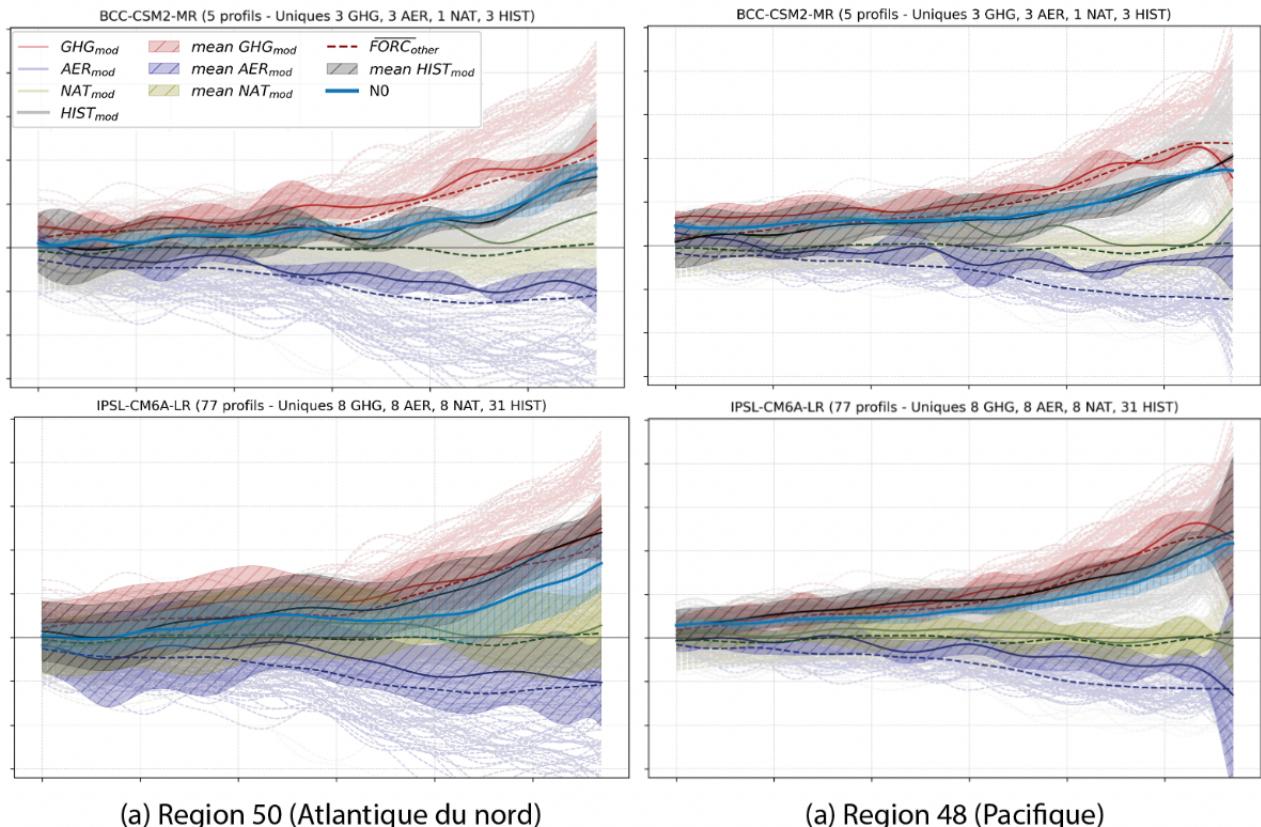


FIGURE 8 : Évaluation de la performance du CNN avec les modèles IPSL et BCC

Les courbes dans la figure 8 montrent en arrière plan (lignes pointillées en couleurs claires) tous les membres des forcages des 12 modèles climatiques : en rouge les GHG, en bleu les AER et en vert les NAT. Ce sont donc les mêmes courbes par région dans la figure. Les images montrent aussi en couleurs plus foncées, traits continues et barres d'erreur (zones hachurées) les moyenne et écart-types des données du modèle en question, BCC-CSM2MR en haut et IPSL-CM6A-LR en bas, et pour chaque région 50 et 48, gauche et droite. Les traits discontinus montrent les moyennes des données par forçage pour tous les modèles climatiques à l'exception du modèle en cours. Ils sont différents pour chaque région. Ces forçages moyens ce sont les forçages de référence intervenant dans l'inversion. Enfin, la figure montre les données historiques du modèle, leur moyenne et écart-type en noir et en gris hachuré surmontés de la sortie de la fonction directe en trait continu bleu clair et identifiée comme N0 dans la légende.

Dans la figure 1, on observe que les courbes des concentrations de gaz à effet de serre (GHG) suivent une tendance similaire pour les deux régions, avec une augmentation constante sur la période étudiée. Cela indique que le réchauffement dû aux GHG affecte de manière similaire les deux océans.

Les courbes des forçages naturels (NAT) montrent une certaine stabilité, avec quelques variations mineures dues aux éruptions volcaniques et aux variations solaires. La comparaison entre les deux régions révèle que les forçages naturels ont un impact similaire sur les températures des deux océans.

Enfin, les courbes des aérosols (AER) indiquent un effet de refroidissement sur les températures. Cependant, la magnitude et la fréquence des variations sont différentes entre les deux régions, ce qui suggère que l'impact des AER varie selon la région.

L'analyse des données et la comparaison entre les deux régions montre clairement des perturbations périodiques dans les températures dans la région 48 (Bien visible avec les autres modèles climatiques présentés en annexe) et cela est due au phénomène de El Niño qui a une influence sur la région de l'océan Pacifique équatorial. Les variations de température associées à El Niño sont visibles sur les courbes et semblent avoir un impact aussi sur les forçages climatiques, en particulier les GHG et les AER.

En comparant les courbes des températures observées (HIST) et celles estimées par le CNN pour les deux régions, on constate une grande proximité entre elles. Cette similitude indique que le modèle CNN est capable d'estimer avec précision les historiques des températures en utilisant les données des trois forçages climatiques. Cette performance est observée pour les 12 modèles climatiques testés (voir annexe).

### Conclusion

L'analyse des courbes des forçages climatiques et des températures pour l'océan Atlantique du Nord et l'océan Pacifique équatorial montre que les GHG réchauffent, les AER refroidissent et les NAT sont relativement stables pour les deux régions. Le phénomène El Niño influence les températures de l'océan Pacifique équatorial en sud, avec des répercussions sur les forçages climatiques. Le modèle CNN utilisé pour estimer les historiques des températures se révèle précis et performant, démontrant l'efficacité de cette approche pour modéliser les interactions entre les forçages climatiques et les températures régionales.

### Inversion

Dans cette partie, nous examinons la figure qui représente les résultats de l'étape d'inversion (Backward optimization) pour les deux régions étudiées, avec pour modèle climatique de référence IPSL. L'objectif de l'inversion est d'estimer les entrées (GHG, NAT, AER) à partir des sorties données qui sont les historiques des températures observées dans les deux régions.

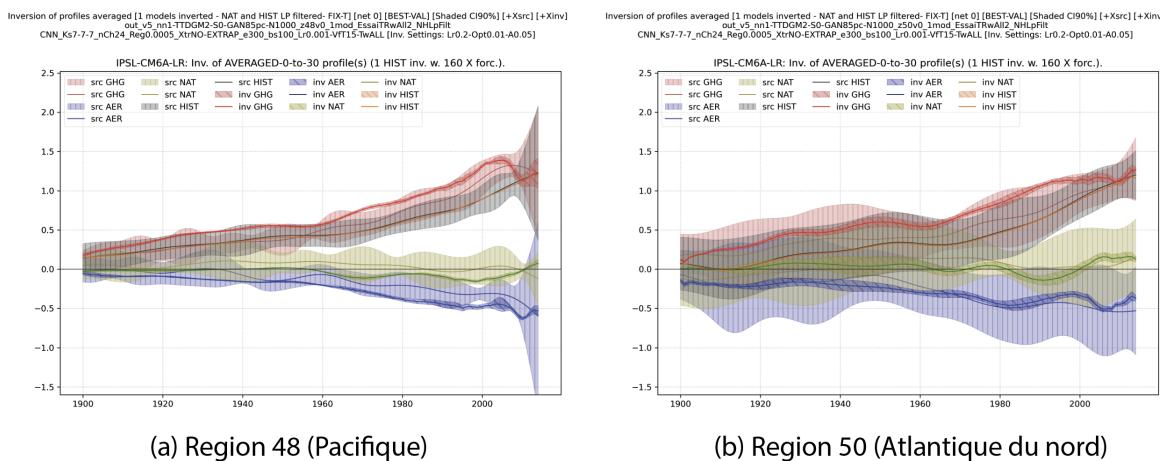


FIGURE 9 : Résultat d'inversion avec le modèle IPSL

La figure 9 montre les courbes résultant de l'inversion. Avant les résultats de l'inversion, la figure montre en arrière plan, distingué par un hachuré composé de trait verticaux, les écart-types des données des forçages pour le modèle de l'IPSL et leur moyenne en couleur foncé, en rouge les GHG, en bleu les AER, en vert les NAT en gris pour les HIST. Ces courbes sont identifiées comme « src XXX » dans la légende. Ce sont en quelque sorte les données source. La figure montre aussi les résultats de l'inversion d'abord sur les trio d'entrée, les X : GHG, AER et NAT, montrant la moyenne et écart-type des ensembles des profils inversés et utilisant un hachuré de forme anti-flash. Enfin il y a les courbes des Y de l'inversion : en noir c'est la courbe du profil historique moyen à inverser et en orange le profil après inversion. Il y a un hachuré qui correspond à l'écart type de différentes inversions du même profil, il y a ici 160 inversions effectuées pour le même profil, mais on ne aperçoit pas vraiment ces barres d'erreur, tellement les inversions sont proches.

Nous pouvons observer sur la figure que les courbes d'inversion sont situées à l'intérieur de l'intervalle de variabilité des données GHG, NAT, AER, ce qui indique une bonne correspondance entre les entrées prédites par le modèle et les données réelles. Nous pouvons également noter que les entrées prédites par l'inversion sont proches des moyennes connues qui sont à priori les entrées les plus adéquates pour le modèle IPSL. Ces résultats indiquent que notre modèle d'inversion est capable de retrouver des entrées cohérentes avec les séries des observations de température qu'on veut étudier, ce qui renforce sa fiabilité et sa pertinence pour estimer les entrées à partir des sorties. Il est aussi important de noter que les sorties Y issues de l'inversion sont également très proches des données historiques sources pour les deux régions étudiées.

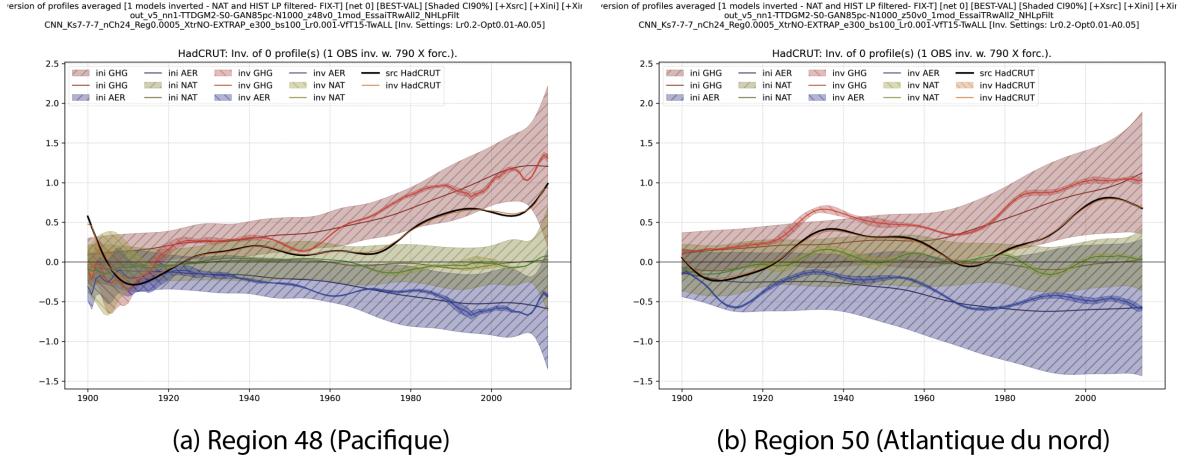


FIGURE 10 : Résultat d'inversion avec les observations HadCRUT

La figure 10 montre les résultats de l'étape d'inversion pour les deux régions étudiées, mais cette fois-ci avec uniquement les données d'observation HadCRUT (sorties connues, mais sans informations sur les entrées cohérentes).

Tout d'abord, nous pouvons remarquer que la courbe prédite par notre modèle d'inversion suit relativement bien la moyenne des données (GHG, NAT, AER) de tous les modèles tout en restant dans l'intervalle d'erreur. Cependant, nous pouvons également observer que la courbe prédite sort légèrement de l'intervalle durant les années 1930 jusqu'à 1940 après cela ça refroidit pour revenir à la normale, ce qui est probablement dû aux perturbations présentes dans la région mais aussi la qualité des données que l'on dispose.

Il est important de noter que les premières années des courbes prédites pour la région 48 présentent un certain niveau de bruit. Ce bruit est dû aux effets de bord, c'est-à-dire que notre modèle a des difficultés à capter les données aux bords. Cela peut être dû à la faible quantité de données disponibles aux bords, qui ne permettent pas une bonne estimation des entrées. Malgré cela, notre modèle reste capable de retrouver les tendances générales des courbes observées.

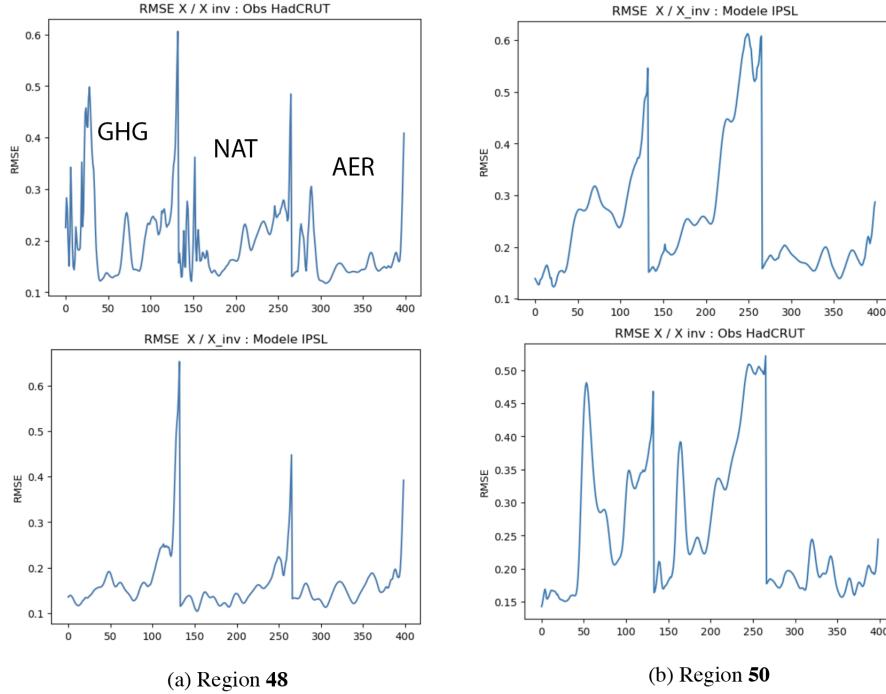


FIGURE 11 : Évaluation de la performance du modèle d'inversion en terme de valeur RMSE pour les deux régions étudiées

La figure ci-dessous montre les valeurs de RMSE (Root Mean Square Error) des prédictions faites avec notre modèle d'inversion pour les deux régions étudiées.

Tout d'abord, pour la région 50 (Atlantique du nord), nous pouvons remarquer que les valeurs de RMSE sont souvent relativement élevées par rapport à celles de la région Pacifique équatorial. Cela indique une plus grande difficulté à estimer les entrées pour cette région. Nous pouvons également remarquer que les valeurs de RMSE les plus élevées sont présentes aux bords des courbes. Comme mentionné précédemment, cela peut être expliqué par le fait que notre modèle a plus de difficultés à capter les informations des dernières années, car les données sont moins nombreuses et donc moins représentatives.

En revanche, pour la région 48 (Pacifique équatorial), nous pouvons observer que les valeurs de RMSE sont généralement faibles, avec des valeurs comprises entre 0,1 et 0,25. Cela indique une bonne capacité de notre modèle à estimer les entrées pour cette région, même en l'absence d'informations cohérentes sur les entrées.

## 5.2 Comparaison avec les résultats de la régression linéaire

Il est intéressant de comparer les résultats obtenus avec notre modèle d'inversion avec ceux obtenus à l'aide d'une autre méthode, comme la régression linéaire.

En effet, une comparaison de la RMSE obtenue avec la régression linéaire et celle obtenue avec notre modèle d'inversion montre que la RMSE est plus élevée avec la régression linéaire pour les deux régions étudiées. Cela suggère que notre modèle d'inversion est plus performant que la régression linéaire pour estimer les entrées à partir des sorties dans le contexte de la modélisation climatique.

En outre, il convient de souligner que la régression linéaire est une méthode qui ne prend pas en compte la complexité du système climatique et ne tient pas compte de l'influence de tous les facteurs qui peuvent avoir un impact sur les sorties. En revanche, notre modèle d'inversion utilise une approche basée sur un réseau de neurones convolutifs et une méthode de minimisation des coûts qui prend en compte les relations non linéaires entre les entrées et les sorties, ce qui permet une estimation plus précise des entrées à partir des sorties.

## 6 Conclusion

En conclusion de ce stage, nous avons mis en œuvre une méthode d'inversion basée sur un réseau de neurones convolutifs pour estimer les entrées (GHG, NAT, AER) à partir des sorties (températures) dans le contexte de la modélisation climatique.

Nous avons ainsi développé un modèle capable de prédire les entrées les plus adéquates à une observation donnée, en identifiant les facteurs les plus influents de ces observations. Nous avons réalisé des expériences sur deux régions océaniques distinctes (Atlantique du nord et Pacifique équatorial) en utilisant un ensemble de données d'observations de température et de forçages climatiques (GHG, NAT, AER).

Les résultats obtenus montrent que notre modèle est capable de retrouver les spécificités climatiques de chaque région, même dans des situations de forte variabilité. Nous avons également comparé notre modèle d'inversion avec une méthode de régression linéaire et montré que notre modèle est plus performant pour estimer les entrées à partir des sorties dans le contexte de la modélisation climatique.

Enfin, nous avons mis en évidence les limites de notre modèle, notamment en ce qui concerne les effets de bord et les perturbations dues à des phénomènes spécifiques tels que le phénomène El Niño. Nous avons proposé des pistes d'amélioration, telles que l'utilisation de méthodes de régularisation pour améliorer la robustesse de notre modèle.

## 7 Annexes

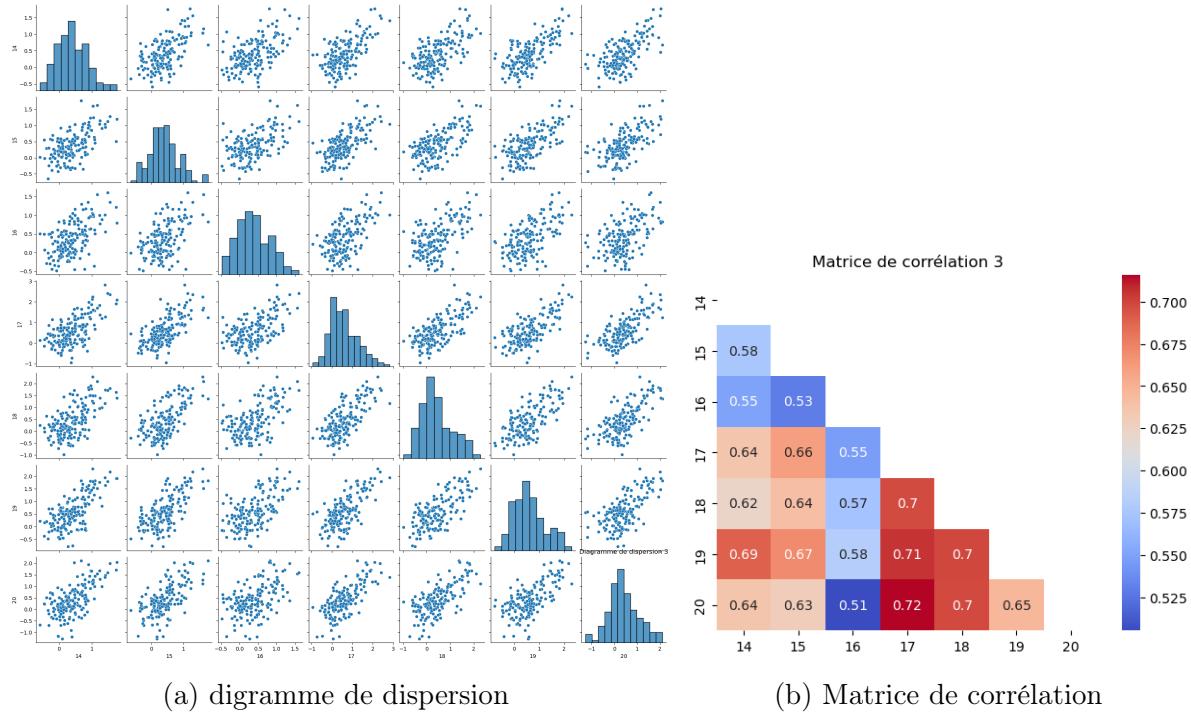


FIGURE 12 : Résumé graphique montrant les corrélations entre les membres de 7 années consécutives

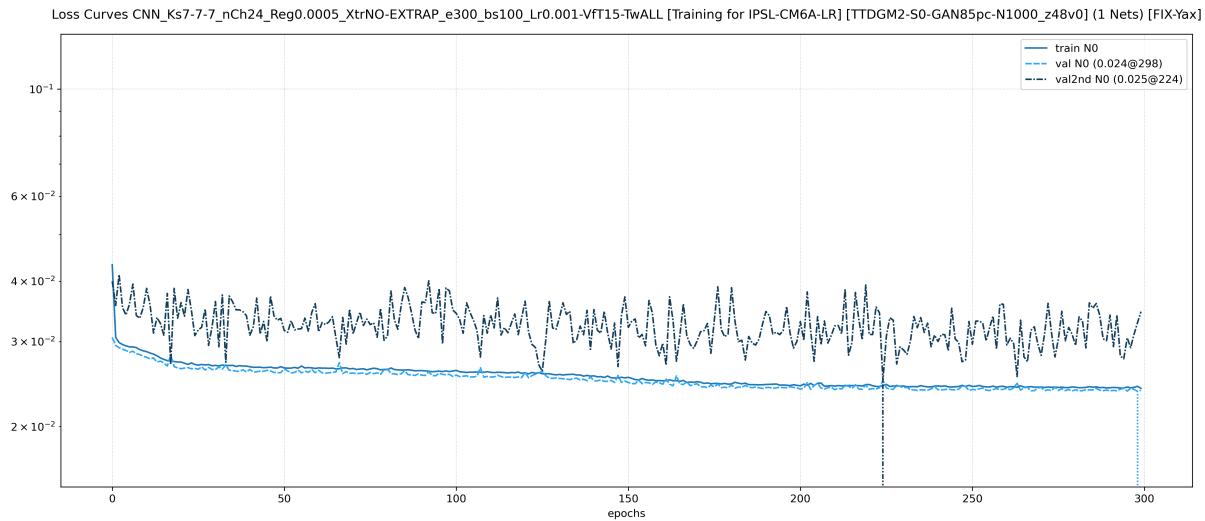


FIGURE 13 : Suivi de la variation de la fonction de coût pendant l'apprentissage (region 48)

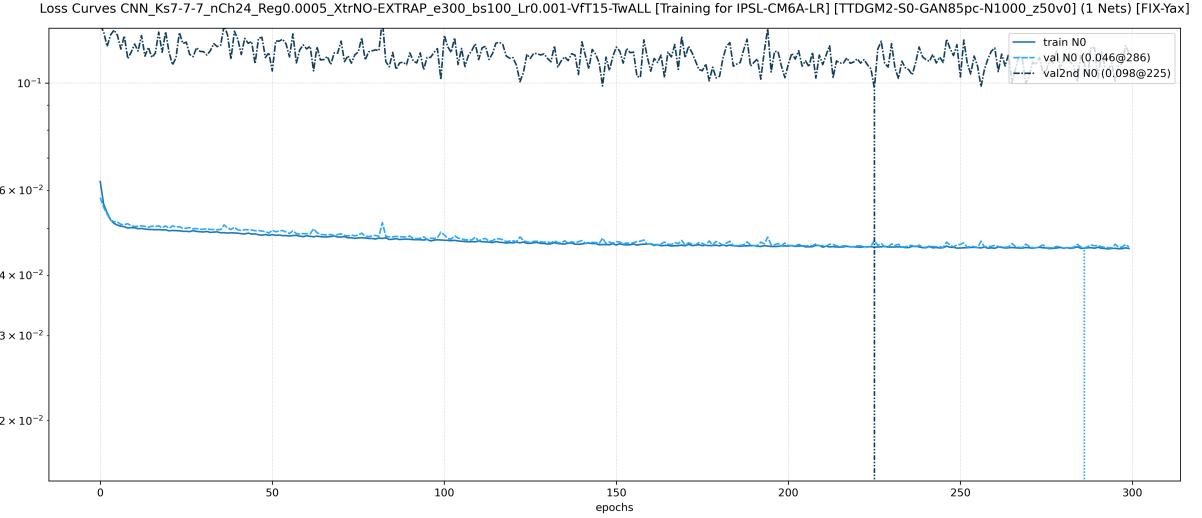


FIGURE 14 : Suivi de la variation de la fonction de coût pendant l'apprentissage (region 50)

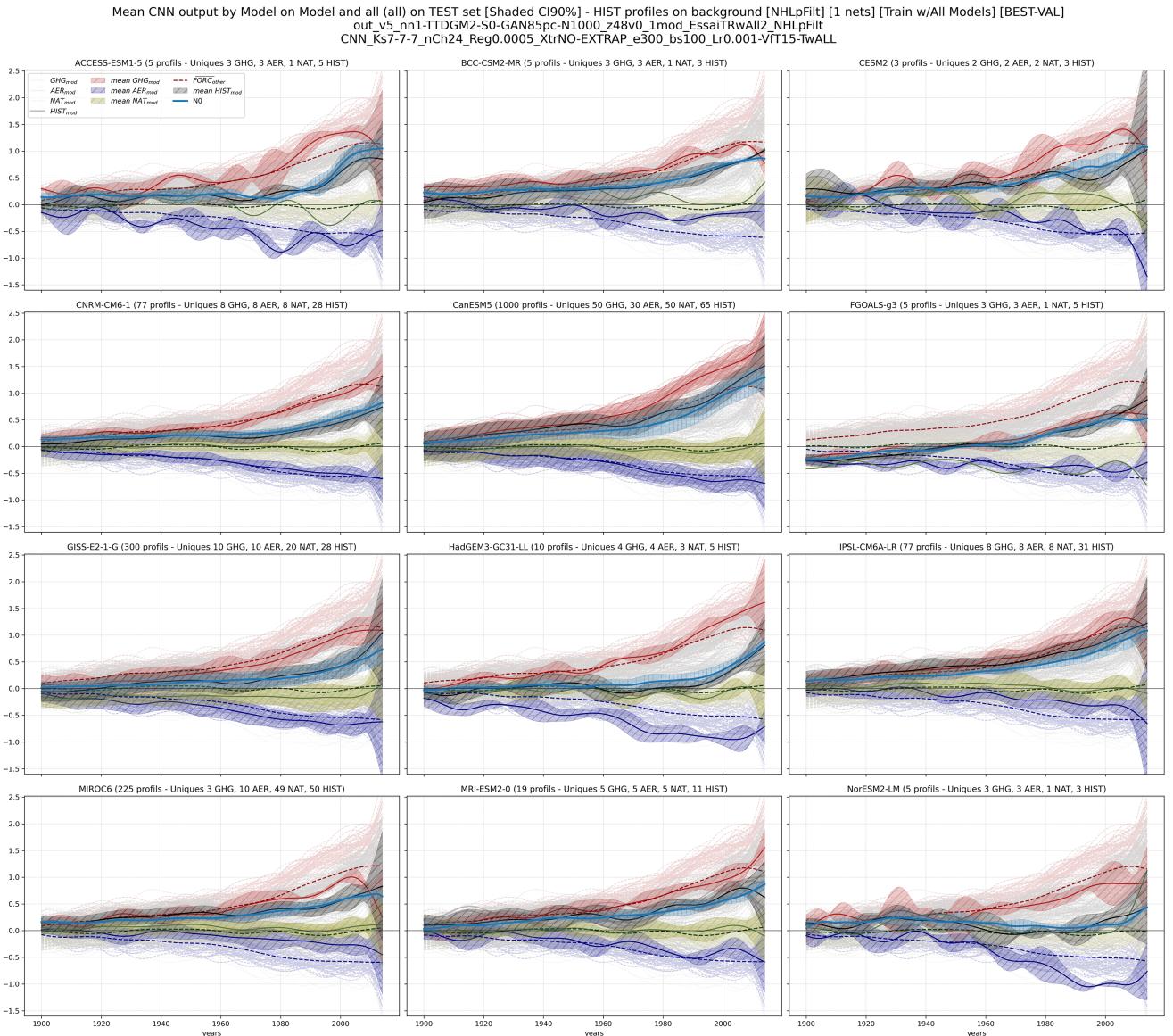


FIGURE 15 : Évaluation de la performance du CNN avec les 12 modèles climatiques (Region48)

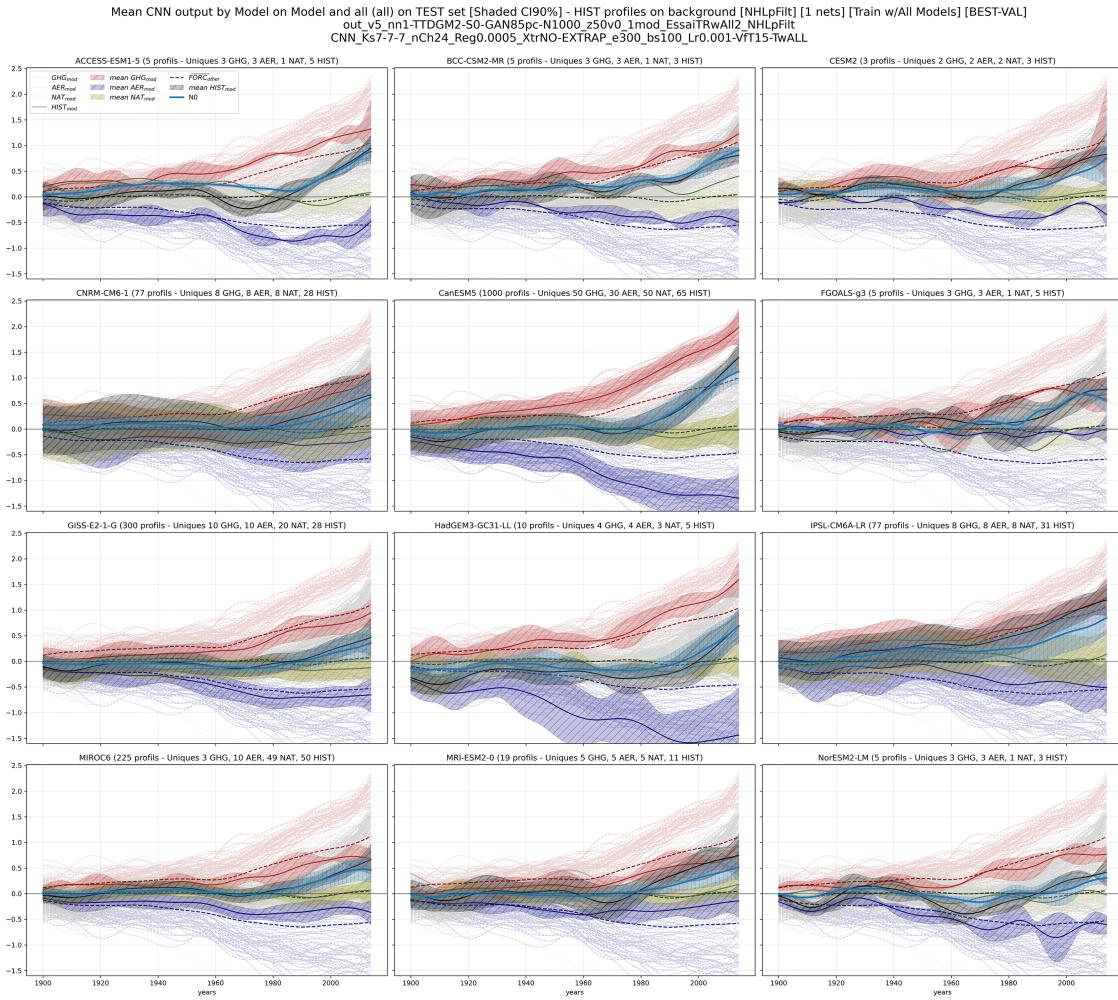


FIGURE 16 : Évaluation de la performance du CNN avec les 12 modèles climatiques (Région 50)

## Références

1. EYRING, V. ; BONY, S. ; MEEHL, G. A. ; SENIOR, C. A. ; STEVENS, B. ; STOUFFER, R. J. ; TAYLOR, K. E. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*. 2016, t. 9, pp. 1937-1958. Available from DOI : 10.5194/gmd-9-1937-2016.
2. LIU, Yunjie ; RACAH, Evan ; PRABHAT ; CORREA, Joaquin ; KHOSROSHAHI, Amir ; LAVERS, David ; KUNKEL, Kenneth ; WEHNER, Michael ; COLLINS, William. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. *arXiv preprint arXiv :1605.01156*. 2016. Available from DOI : 10.5194/essd-12-2959-2020.
3. LIU, Jia ; YANG, Chunxue ; GUO, Huan ; LI, Zhengguo. Deep learning for predicting annual mean temperature in China. *Journal of Geophysical Research : Atmospheres*. 2020, t. 125, n° 3, e2019JD031797.
4. KUMAR, Sanjay ; RATHORE, LS ; SINGH, Jitendra. Precipitation prediction using recurrent neural networks over different regions of India. *Theoretical and Applied Climatology*. 2020, t. 141, n° 1-2, pp. 185-199.
5. WANG, Q. ; ZHANG, R. ; ZHANG, S. ; YANG, S. ; HUANG, B. Attribution analysis of Atlantic Ocean surface temperature change with a deep learning algorithm. *Advances in Atmospheric Sciences*. 2020, t. 37, n° 5, pp. 513-525. Available from DOI : 10.1007/s00376-020-9223-6.
6. MAKKONEN, R. ; KUKKONEN, J. ; KARPPINEN, A. ; LÄHDE, T. Source apportionment of PM2.5 using backward optimization of a neural network model. *Atmospheric Environment*. 2021, t. 249, p. 118282. Available from DOI : 10.1016/j.atmosenv.2020.118282.
7. TOMS, R. B. ; UKKOLA, A. M. ; SHARMA, A. Identifying atmospheric drivers of precipitation variability using backward optimization. *Journal of Climate*. 2021, t. 34, n° 6, pp. 2307-2321. Available from DOI : 10.1175/JCLI-D-20-0385.1.
8. MORICE, C. P. ; KENNEDY, J. J. ; RAYNER, N. A. ; JONES, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates : The HadCRUT4 data set. *J. Geophys. Res.* 2012, t. 117, p. D08101. Available from DOI : 10.1029/2011JD017187.
9. RAYNER, N. A. ; PARKER, D. E. ; HORTON, E. B. ; FOLLAND, C. K. ; ALEXANDER, L. V. ; ROWELL, D. P. ; KENT, E. C. ; KAPLAN, A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* 2003, t. 108, p. 4407. Available from DOI : 10.1029/2002JD002670.
10. ITURBIDE, M. ; GUTIÉRREZ, J. M. ; ALVES, L. M. ; BEDIA, J. ; CEREZO-MOTA, R. ; CIMADEVILLA, E. ; COFIÑO, A. S. ; DI LUCA, A. ; FARIA, S. H. ; GORODETSKAYA, I. V. ; HAUSER, M. ; HERRERA, S. ; HENNESSY, K. ; HEWITT, H. T. ; JONES, R. G. ; KRAKOVSKA, S. ; MANZANAS, R. ; MARTÍNEZ-CASTRO, D. ; NARISMA, G. T. ; NU-RHATI, I. S. ; PINTO, I. ; SENEVIRATNE, S. I. ; HURK, B. van den ; VERA, C. S. An update of IPCC climate reference regions for subcontinental analysis of climate model data : definition and aggregated datasets. *Earth Syst. Sci. Data*. 2020, t. 12, pp. 2959-2970. Available from DOI : 10.5194/essd-12-2959-2020.