

---

# Reproduction probabiliste de variables climatiques extrêmes par des réseaux antagonistes génératifs.

---

## Résumé :

*Cette étude s'intéresse à la reproduction probabiliste de variables climatiques extrêmes. La nature extrême de ces données est due ici à l'opérateur  $\max$  avec un support de 365 jours donnant ainsi des maxima annuels. La reproduction de ces variables sera effectuée à l'aide de réseaux antagonistes génératifs. Cette méthode provenant du domaine de l'apprentissage profond s'inscrit dans la classe des statistiques non-paramétriques. Ce cadre flexible permet alors d'éviter de choisir et donc de restreindre une famille de distributions particulière pour la modélisation. La reproduction de ces variables s'appuiera sur des données issues de simulations de plusieurs modèles climatiques, dans différents cadres expérimentaux. La méthode non-paramétrique choisie ici sera comparée aux distributions extrêmes généralisées, logiquement introduites par la nature des données. Les comparaisons de performances de reproduction seront basées sur l'apprentissage de distributions mais aussi sur la reproduction de séries temporelles et spatiales des observables climatiques. L'objectif global de ces modèles génératifs doit être de combler le manque de données inhérent au domaine des extrêmes. Les résultats sur les performances des réseaux antagonistes génératifs sont assez mitigés mais certaines pistes de recherches semblent prometteuses.*

**Mots clefs :** Réseaux antagonistes génératifs - Simulations probabilistes - Modèles climatiques - Maxima annuels - Théorie des valeurs extrêmes

Stage numérique encadré par : **Philippe Naveau**

[philippe.naveau@lsce.ipsl.fr](mailto:philippe.naveau@lsce.ipsl.fr) / tél. (+33) 1 69 08 41 58

Laboratoire des Sciences du Climat et de l'Environnement

Orme des Merisiers, Bat 714

91190 Saint-Aubin

France

<https://www.lsce.ipsl.fr/Pisp/philippe.naveau/>

## Remerciements

Je tiens à remercier toutes les personnes qui ont contribué au succès de mon stage et qui m'ont aidé lors de la rédaction de ce rapport.

Tout d'abord, j'adresse mes remerciements à Freddy Bouchet, directeur de recherche au CNRS, avec qui j'avais planifié de faire mon stage à l'origine. Il m'a permis de trouver ce stage en me donnant le contact de Philippe Naveau, chercheur au LSCE.

Je tiens donc à remercier vivement mon maître de stage, Philippe Naveau, pour son accueil, pour son temps et son expertise précieuse. Grâce à sa confiance j'ai pu prendre des initiatives et donner le meilleur de moi-même. Il a été d'une aide précieuse dans certains moments de doute et d'échec. J'ai pu grâce à lui expérimenter le domaine de la recherche.

Ensuite, je tiens à remercier le personnel du LSCE pour leur accueil. En particulier, je tiens à remercier Florence Gerry pour son aide dans la partie administrative qui a facilité mes démarches. De plus, je tiens à remercier le Centre Blaise Pascal de l'ENS de Lyon pour la mise à disposition de matériel informatique rendant possible l'exécution de certaines de mes simulations. Je remercie donc aussi Emmanuel Quemener pour sa patience et ses conseils vers des utilisations propres de ces ressources.

Enfin, je tiens à remercier toutes les personnes ayant relu mon rapport. Je pense donc naturellement à mon maître de stage, Philippe Naveau, pour son aide dans l'agencement global et dans les subtilités de ce rapport. Mais je tiens à remercier plus particulièrement ma compagne, Clara Druais, pour son aide et ses conseils dans la clarté et la justesse de ce rapport.

# Table des matières

<b>1</b>	<b>Contexte de l'étude</b>	<b>1</b>
1.1	Motivations . . . . .	1
1.2	Les modèles climatiques . . . . .	1
1.3	La théorie des valeurs extrêmes . . . . .	2
1.3.1	Valeurs extrêmes . . . . .	2
1.3.2	Lois limites . . . . .	2
1.4	Génération de données par des réseaux antagonistes génératifs . . . . .	3
1.4.1	Principe théorique . . . . .	3
1.4.2	État de l'art . . . . .	4
1.5	La reproduction de variables climatiques extrêmes . . . . .	5
<b>2</b>	<b>Méthodes numériques</b>	<b>6</b>
2.1	Les données climatiques . . . . .	6
2.2	Simulation et analyse des données . . . . .	8
2.2.1	Simulations de valeurs extrêmes . . . . .	8
2.2.2	Analyse statistique avec un diagramme quantile-quantile . . . . .	9
2.3	Les réseaux antagonistes génératifs . . . . .	9
2.3.1	Structure adoptée . . . . .	9
2.3.2	Précautions pour l'entraînement . . . . .	10
<b>3</b>	<b>Résultats</b>	<b>11</b>
3.1	Apprentissage de distributions . . . . .	11
3.2	Apprentissage de séries temporelles . . . . .	12
3.3	Apprentissage de séries temporelles et spatiales . . . . .	14
<b>4</b>	<b>Discussion et conclusion</b>	<b>16</b>
4.1	Résultats . . . . .	16
4.2	Limites . . . . .	17
4.3	Applications et perspectives . . . . .	18
	<b>Références</b>	<b>20</b>
	<b>Annexes</b>	<b>21</b>
<b>A</b>	<b>Algorithmes d'entraînement des réseaux antagonistes génératifs</b>	<b>21</b>
<b>B</b>	<b>Structures des réseaux</b>	<b>21</b>

# 1 Contexte de l'étude

Cette partie vise à présenter le cadre théorique de l'étude ainsi que ses motivations. Ces informations seront comparées à un bref état de l'art du domaine de recherche dans lequel s'inscrit cette étude.

## 1.1 Motivations

La climatologie est une discipline ancrée dans la société depuis la création de l'agriculture. Bien que les méthodes se soient largement améliorées, les enjeux sont toujours autant cruciaux. Prédire le climat peut permettre d'améliorer l'agriculture, le tourisme ou même la sécurité via une meilleure gestion des ressources et des populations. En particulier, l'anticipation d'événements extrêmes possède de grands enjeux sociétaux, notamment dans la prévention et dans l'assurance. Ces derniers, comme des pics de chaleurs ou de précipitation, peuvent s'avérer dévastateurs humainement et matériellement. L'étude de ces événements extrêmes devient d'autant plus importante que le climat se réchauffe. En effet, l'augmentation du nombre de sécheresses, de tempêtes ou l'augmentation de la valeur des maxima de température ne tendent qu'à augmenter sous le forçage anthropique actuel, forçage dû à l'activité humaine comme les émissions de CO<sub>2</sub>.

Pour prédire le climat futur il est nécessaire d'avoir un grand nombre de données sur le climat actuel mais aussi sur celui du passé. Les données climatiques peuvent permettre d'extrapoler des tendances, de mettre en lumière des cycles climatiques ou d'estimer des périodes de retour d'événements extrêmes. En ce sens, la modélisation de variables climatiques pourrait permettre de mieux prédire ces informations. Une modélisation spécifique pour chaque variable peut faire gagner en précision. Ces reproductions, plus que de simples prédictions, pourraient aussi donner une meilleure compréhension du climat.

Cependant, le manque de données est un problème récurrent qui survient dans le domaine de la climatologie, et plus particulièrement avec les phénomènes extrêmes. En effet, les données antérieures à l'ère pré-industrielle sont souvent manquantes ou imprécises, cela est d'autant plus vrai pour les événements extrêmes. Cela peut provenir de la fréquence d'occurrence de l'événement ou alors d'un échantillonnage comme pour les extrema annuels.

Ces manques peuvent être en partie comblés par les modèles climatiques, les données manquantes peuvent être extrapolées ou remplacées par des simulations. Cependant il apparaît peu pertinent de faire tourner des modèles climatiques globaux pour au final ne récupérer que la température et ensuite ne prendre que les maxima annuels. Il devient alors nécessaire d'explorer d'autres voies de simulation de variables, de phénomènes ou d'opérateurs extrêmes.

## 1.2 Les modèles climatiques

Les données physiques qui seront étudiées ici sont issues de simulations de modèles climatiques. Pour mieux comprendre ces données il est important de connaître le fonctionnement global d'un modèle climatique. De plus, il est important de comprendre les liens entre les différents modèles climatiques.

L'un des premiers modèles climatiques, au sens d'une modélisation globale du climat atmosphérique, date de 1965, [Manabe et al. \(1965\)](#). Ce modèle est notable car c'est le premier à inclure un cycle hydrologique, couplant ainsi l'atmosphère à l'océan. À l'époque, les simulations de modèles climatiques consistaient en la résolution d'équations de l'hydrodynamique, basées sur celles de Navier-Stokes avec les termes nécessaires de forçage, d'interactions et de bords. Ces équations étaient résolues par des méthodes de différences finies qui ont ensuite été remplacées par des méthodes spectrales, dont un des premiers modèles fut présenté dans [Manabe et al. \(1979\)](#). Ces méthodes spectrales sont toujours au cœur des modèles climatiques actuels, cependant on peut noter de grandes différences entre le modèle [Manabe et al. \(1979\)](#) et par exemple le dernier modèle de l'IPSL [Boucher et al. \(2020\)](#). La majeure amélioration des modèles n'a pas été apportée par la physique mais par l'innovation technologique autour des ordinateurs, ayant grandement amélioré la puissance des centres de calculs, permettant ainsi une discrétisation spectrale plus fine. Avec une meilleure discrétisation il a alors été possible de coupler plus précisément l'atmosphère à l'océan ainsi que de prendre en compte des phénomènes plus subtiles comme des interactions chimiques ou des gradients d'humidité. Les autres apports notables

de ces nouveaux modèles se trouvent aussi dans les termes de forçage anthropique comme l'interaction des gaz avec l'ozone conduisant à l'effet de serre ou la croissance économique permettant de prévoir le taux d'émissions de CO<sub>2</sub>. Ces implémentations précises des termes de forçage anthropique sont décrites dans [Lurton et al. \(2020\)](#) pour le modèle de l'IPSL.

Pour éviter que trop de modèles différents soient développés de façon indépendante plusieurs projets globaux ont vu le jour. Les données étudiées font partie du projet CMIP6 (Coupled Model Intercomparison Project phase 6), présenté par [Eyring et al. \(2016\)](#). Il s'agit d'un regroupement de modèles climatiques lancés conjointement pour obtenir une meilleure compréhension du climat. Un avantage supplémentaire réside dans l'uniformisation des conditions initiales et de forçage. Les comportements des différents modèles peuvent alors être comparés de façon objective. De plus, l'aspect de forçage anthropique est au centre du projet CMIP6, dont l'objectif final est la lutte contre le réchauffement climatique. Il s'agit alors de comprendre l'influence de certains types de forçage sur le climat. Ainsi on pourra distinguer deux types de simulations, les simulations avec forçage anthropique et celles sans, uniquement avec le forçage naturel. Ces deux types de simulations produisent deux types de données, et donc d'extrêmes, qui seront donc traitées en conséquence, notamment par l'étude de la théorie des valeurs extrêmes.

### 1.3 La théorie des valeurs extrêmes

#### 1.3.1 Valeurs extrêmes

Du point de vue des extrêmes l'étude sera centrée sur l'opérateur **max**. Bien que le cadre applicatif choisi ici soit la climatologie, la théorie des valeurs extrême reste d'une grande importance. Elle permet notamment d'obtenir des résultats puissants qui permettent une modélisation quantitative, introduits dans le chapitre 3 de [Coles \(2001\)](#) et dont certains résultats sont résumés dans la section 1.3.2. Dans un premier temps on peut de regarder des variables iid (indépendantes et identiquement distribuées). Bien que ce cadre soit restreint il permet de poser les bases de la théorie. Si on note  $n$  le support de l'opérateur **max**,  $X$  et  $(X_i)_{i \in [1, n]}$  les variables d'entrées iid,  $Z_n$  la variable issue du **max**, on a alors l'équation (1a),

$$Z_n = \max(X_1, \dots, X_n). \quad (1a)$$

La fonction de répartition de la variable  $Z$ ,  $G_n$  est alors donnée par l'équation (1b),

$$\forall z \in \mathbb{R}, \quad G_n(z) = F(z)^n, \quad (1b)$$

avec  $F$  la fonction de répartition de la variable  $X$ . Ces équations introduisent naturellement la question de l'étude de  $G_n$  en fonction de  $F$  et  $n$ . Un premier cas notable est celui des variables max-stables, propriété définie par l'équation (2),

$$\forall n \in \mathbb{N}, \exists (c_n, d_n) \in \mathbb{R}^2, \quad \max(X_1, \dots, X_n) \stackrel{d}{=} c_n X + d_n. \quad (2)$$

Dans ce cas la forme de la fonction  $G_n$  est la même que celle de la fonction  $F$ .

Même si le cadre des variables max-stables semble restreint il en fait étroitement lié au second cadre d'étude notable, l'étude asymptotique des extrêmes. Ce cadre, résumé dans la section 1.3.2, permet de généraliser la distribution des variables extrêmes ainsi que celle des variables max-stables.

#### 1.3.2 Lois limites

Reprenons la variable  $Z_n$  issue de l'opérateur **max** appliqué à  $n$  variables iid. L'étude asymptotique de cette variable montre que sa distribution limite ne peut être qu'une distribution GEV (generalized extreme value). Autrement dit, si l'équation (3) est vérifiée alors la distribution limite appartient au type de distribution définie par l'équation (4). Ce résultat est connu sous le nom de théorème des valeurs extrêmes, il a été officiellement formalisé dans [Gnedenko \(1943\)](#) après de nombreuses études asymptotiques sur cet opérateur.

$$\exists (a_n, b_n)_{n \in \mathbb{N}} \in \mathbb{R}_+^* \times \mathbb{R}^{\mathbb{N}}, \quad \forall z \in \mathbb{R}, \quad \mathbb{P}((Z_n - b_n)/a_n \leq z) \xrightarrow{n \rightarrow +\infty} G(z) \quad (3)$$

Cette propriété est tout à fait comparable au théorème central limite pour l'opérateur **somme**. La distribution limite n'est plus alors de type GEV, distribution donnée par l'équation (4),

$$G(z; \mu, \sigma, \xi) = \begin{cases} \exp\left(-\exp\left(\frac{z-\mu}{\sigma}\right)\right) & \xi = 0 \\ \exp\left(-\left[1 + \xi \frac{z-\mu}{\sigma}\right]^{1/\xi}\right) & \xi \neq 0 \text{ et } \xi \frac{z-\mu}{\sigma} > -1 \\ 0 & \xi > 0 \text{ et } \xi \frac{z-\mu}{\sigma} \leq -1 \\ 1 & \xi < 0 \text{ et } \xi \frac{z-\mu}{\sigma} \leq -1 \end{cases}, \quad (4)$$

mais de type gaussienne. Une description plus détaillée de cette loi limite est donnée dans le chapitre 3 de [Coles \(2001\)](#). Il y est notamment montré que cette famille de distributions est en fait la famille des distributions max-stables, équation (2).

Le cadre des variables iid peut être étendu à certains cas plus complexes. En effet, en laissant les paramètres de la distribution être dépendants d'un paramètre comme le temps, il est alors possible de modéliser des séries temporelles non-stationnaires. Dans ce rapport seules celles données par l'équation (5) seront étudiées, un approfondissement est fait dans le chapitre 6 de [Coles \(2001\)](#). Bien que toujours réduit, ce cadre constitue une extension non négligeable. En effet, dû au forçage anthropique certaines variables possèdent de fortes dépendances temporelles.

$$Z(t) = \mu(t) + \sigma(t)\tilde{Z}, \quad \text{avec } \tilde{Z} \sim G(z; 0, 1, \xi). \quad (5)$$

En pratique  $n$  est fini et il faut alors se poser alors la question de la pertinence de l'approximation par une distribution de type GEV. Ce questionnement est d'autant plus important en climatologie où la saisonnalité possède souvent une forte influence. En effet, si le maximum de température annuel survient tout le temps pendant la période estivale, le support du maximum n'est plus 365 jours mais de 90 jours. La convergence vers une GEV n'est alors pas forcément établie. Pour une loi de Student, distribution utilisée dans la partie 3, l'ajustement par une GEV est très correcte même pour de faibles tailles de bloc. Ce phénomène s'explique par la forme variable des GEV qui s'ajustent donc facilement à beaucoup de type de distributions.

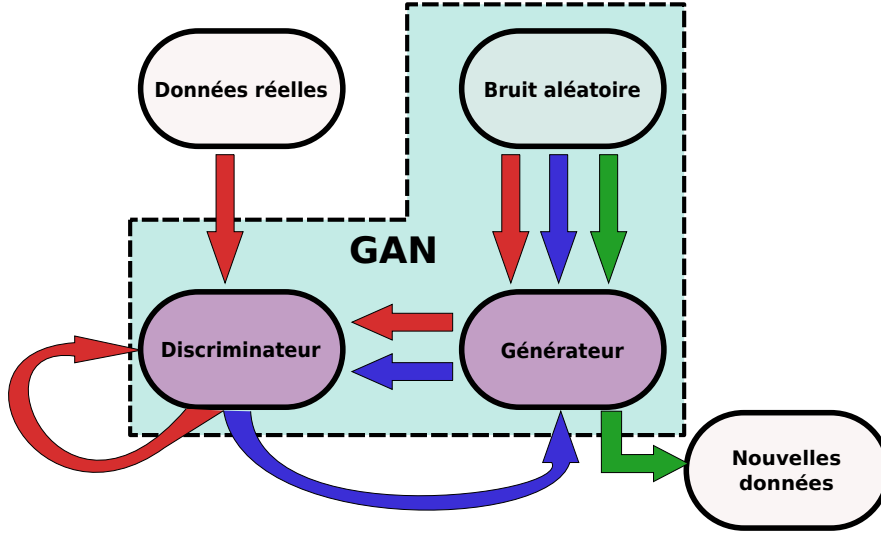
Cette approximation peut néanmoins rester assez mauvaise dans certains cas, notamment pour des distributions à queue très lourde avec peu d'échantillons. Ce genre de limitations ainsi que les problèmes liés aux dépendances dans les données justifient l'exploration de nouvelles méthodes.

## 1.4 Génération de données par des réseaux antagonistes génératifs

### 1.4.1 Principe théorique

Les GANs (Generative Adversarial Networks), aussi appelés réseaux antagonistes génératifs, présentés dans [Goodfellow et al. \(2014\)](#), appartiennent au champ de l'apprentissage profond. Ce domaine fournit un large panel de techniques d'inférence en statistique non paramétrique. Il est alors possible de créer des modélisations de données diverses sans devoir expliciter les règles sous-jacentes de ces modèles. Les GANs sont l'objet de cette étude car ils permettent de créer de nouvelles données proches des données originelles sans les répliquer. Ils sont formés de deux réseaux de neurones, le discriminateur et le générateur. Le principe est relativement intuitif, le discriminateur doit différencier les données générées des vraies données et le générateur doit tromper le discriminateur. Plus concrètement le générateur apprend une cartographie entre l'espace latent du bruit et l'espace réel des données. L'état final espéré, sous réserve de convergence, est constitué du générateur reproduisant la distribution des données d'entrée et du discriminateur qui donne une prédiction aléatoire. Une illustration du cycle d'apprentissage est donnée par la figure 1. Lors de ce cycle le discriminateur et le générateur sont entraînés tour à tour. Ce type d'optimisation n'étant pas convexe l'équilibre théorique n'est pas toujours atteint. Certaines méthodes doivent alors être mises en place pour améliorer la convergence du GAN.

Une partie des méthodes de [Salimans et al. \(2016\)](#) complétée par l'analyse statistique de [Chen et al. \(2020\)](#) sera présentée et discutée dans la section 2.3.2.



**Fig. 1** – Schéma de fonctionnement général d'un GAN. Le cycle rouge représente le cycle d'entraînement du discriminateur et le bleu celui du générateur. Le cycle vert est le cycle de production des nouvelles données après complétion de l'entraînement du GAN.

D'un point de vue plus formel, les deux réseaux sont entraînés dans le cadre de la théorie des jeux, ici un jeu à deux joueurs à somme nulle. Le discriminateur, de fonction  $D$  produit un niveau de confiance sur la nature de l'entrée  $x$ , réelle ou générée. Le générateur prend en entrée une variable aléatoire  $z$  et produit une donnée  $G(z)$ . L'équilibre mini-max recherché est alors donné par l'équation (6a),

$$\min_G \max_D V(G, D), \quad (6a)$$

où  $V(G, D)$  est donné par l'équation (6b),

$$V(G, D) = \mathbb{E}_{x \sim p_{\text{donnees}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (6b)$$

L'optimisation est faite sur la log-vraisemblance qui donne de meilleurs gradients ainsi qu'un apprentissage plus stable que la pure vraisemblance. L'équilibre partiel lors d'une étape d'entraînement pour le discriminateur est donné par l'équation (7),

$$D(x) = \frac{p_{\text{donnees}}(x)}{p_{\text{donnees}}(x) + p_g(x)}. \quad (7)$$

Ce discriminateur est en théorie sensé guider le générateur vers la distribution des données réelles. L'équilibre final est alors  $D(x) = 1/2$  et  $G(z) = x$  avec  $p_g(x) = p_{\text{donnees}}(x)$ .

En pratique l'équilibre est difficile à atteindre surtout à cause de l'instabilité de l'entraînement. Il y a notamment une asymétrie entre les deux boucles d'entraînement. En effet, le cadre de l'apprentissage du discriminateur est celui d'un problème de classification. Pour le générateur le cadre est relativement inhabituel, l'entraînement est fait de façon indirecte et commence sans réelle direction. Les données n'étant pas directement "vues" par le générateur, cela empêche le sur-apprentissage, mais peut alors défavoriser l'apprentissage pur.

#### 1.4.2 État de l'art

L'apparition des GANs a marqué un profond changement dans le domaine des modèles génératifs de l'apprentissage profond, mais pas seulement. Le cadre de l'apprentissage profond permet de ne pas



expliciter les règles du phénomène modélisé. Il permet donc une modélisation plus flexible et moins biaisée. Ensuite, les GANs possèdent l’avantage de ne pas devoir générer des exemples via une chaîne de Markov, à la différence des réseaux stochastiques génératifs ou de certains réseaux basés sur les machine de Boltzmann. Enfin, ils ne nécessitent aucune méthode d’inférence approximative présentes dans certains réseaux Bayésiens. C’est donc sans surprise que le domaine de la reproduction de données extrêmes par des GANs est actuellement en plein essor.

De récents travaux ont permis d’améliorer les performances des GANs dans la génération de phénomènes extrêmes. On peut d’abord noter l’apport de [Allouche et al. \(2021\)](#), qui présente une nouvelle paramétrisation du Générateur. Cette paramétrisation s’appuie sur la théorie des valeurs extrêmes en proposant une fonction d’activation s’appuyant sur l’indice de queue des distributions. Ce nouveau cadre démontre son intérêt dans l’approximation de distributions à queue lourde. L’application est faite sur des séries temporelles de données financières, plus particulièrement les indices de clôtures journaliers. Une autre publication en lien avec cette étude est [Bhatia et al. \(2021\)](#). Elle traite de la génération de données climatiques extrêmes en terme d’images. Leur apport principal se situe dans la technique de décalage de distribution. Les exemples générés sont sélectionnés pour progressivement décaler la distribution vers les extrêmes croissants. Ces deux nouvelles méthodes visent donc à combler les lacunes des GANs dans le domaine des distributions à queue lourde. En particulier, elles visent à améliorer l’ajustement en bout de queue, lieu des phénomènes les plus extrêmes.

Ces travaux présentent de nouveaux outils qui s’appuient sur la théorie des valeurs extrêmes. Ils seront donc comparés aux résultats de cette étude dans la partie 3. Ces résultats pourraient alors permettre de discuter de la pertinence de telles méthodes.

## 1.5 La reproduction de variables climatiques extrêmes

L’objectif de cette étude étant la reproduction de variables climatiques extrêmes, une étude de l’opérateur  $\max$  est nécessaire. L’apparente simplicité de cet opérateur cache le cadre de la théorie des valeurs extrêmes, section 1.3, qui apporte un premier support non trivial. Le reste de la complexité réside ensuite dans le choix du contexte et des méthodes d’études. En effet, l’objectif initial étant relativement large, un contexte ainsi que des motivations plus appliquées ont été choisies. En ce sens la climatologie apporte sa particularisation de l’opérateur  $\max$  appliquée à la physique. Cette particularisation s’accompagne de nombreuses motivations en parties décrites dans la section 1.1. Il sera donc question ici de reproduire différentes variables climatiques extrêmes a posteriori.

Les modèles climatiques, présentés en section 1.2, fournissent différents types de données. Ces différents types de données seront utilisés pour fournir une étude avec diverses enjeux et aspects. Par exemple, les maxima de températures issus de simulations avec forçage anthropique sont fortement corrélés, comme indiqué dans la section 2.1. Ils seront donc appropriés dans le cadre d’un apprentissage de tendance temporelle. Leurs fortes corrélations spatiales seront aussi un enjeu dans le pré-traitement et l’apprentissage. Au contraire les données de précipitations issues de simulation sans forçage anthropique sont très peu corrélées et pourront donc quasiment être considérées indépendantes. Les distributions théoriques de ces variables climatiques étant inconnues, des expériences de contrôle seront étudiées. Il s’agira donc de reproduire des maxima avec différentes structures selon la variable étudiée en s’appuyant sur la théorie des extrêmes présentée dans la section 1.3. Ces expériences sont décrites en détail dans la partie 3.

La nature variable et probabiliste du climat dirige le choix des outils d’étude vers des méthodes non-déterministes. Les outils choisis ici pour explorer cette question sont donc les GANs, présentés dans la section 1.4.1. Bien que le domaine d’étude des GANs soit en plein essor, les études sur les séries temporelles, en particuliers les séries extrêmes, sont assez récentes, section 1.4.2. Ce domaine, encore peu balisé, possède de nombreux avantages et ambitions. Les promesses de l’apprentissage profond résident dans la flexibilité des modèles et l’absence d’explicitation de règles. Les GANs quand à eux bénéficient du cadre de l’apprentissage profond et leur nature probabiliste augmente leur flexibilité. En pratique ces modèles génératifs extrêmes doivent répondre au manque de données intrinsèque au domaine des extrêmes. De plus, ils doivent ici saisir les différentes structures (temporelles et spatiales)



propres aux données climatiques.

Cette étude vise donc à explorer les capacités des GANs à résoudre ces différents problèmes de modélisations climatiques. Ces outils seront appliqués à reproduire des variables climatiques avec une structure spatiale et temporelle complexe.

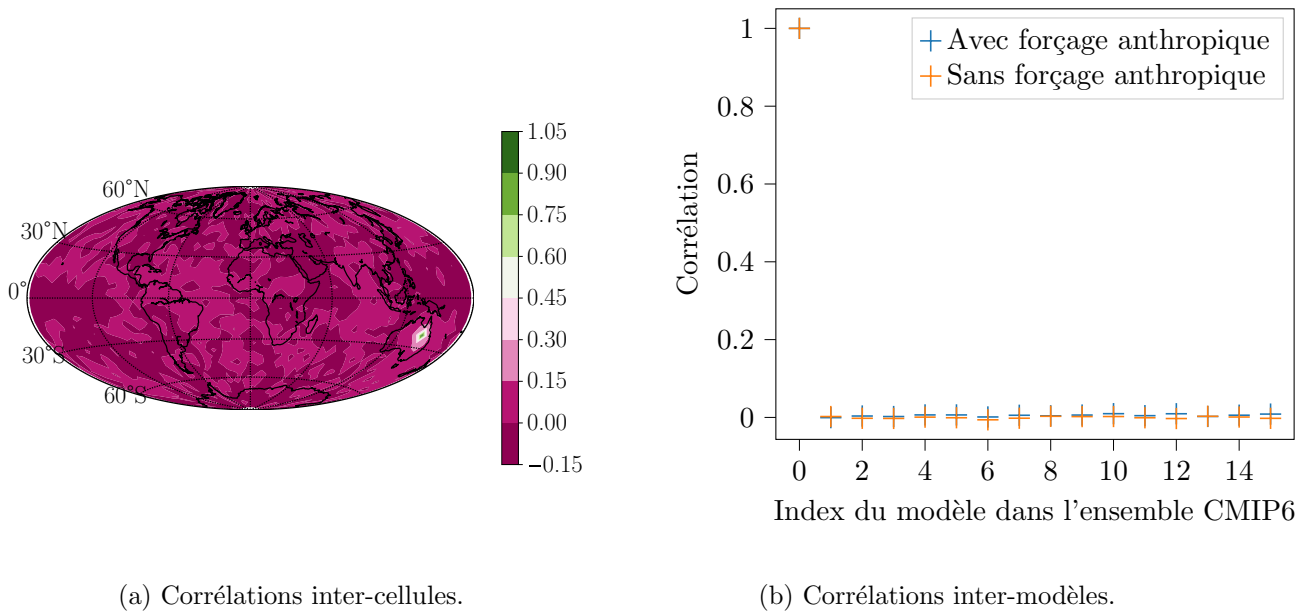
## 2 Méthodes numériques

Cette section présente les méthodes numériques utilisées pour la reproduction de variables climatiques extrêmes.

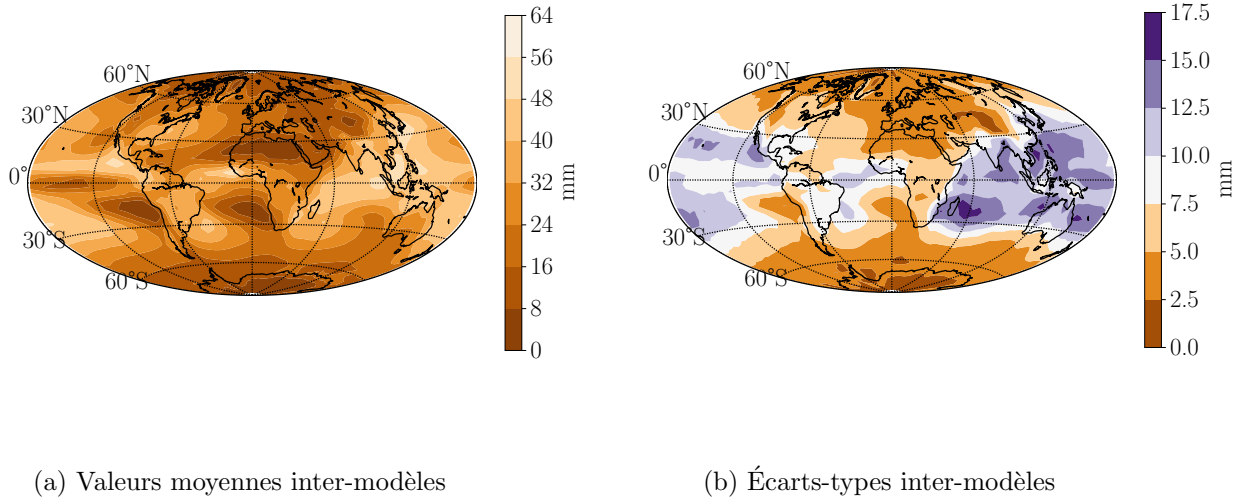
### 2.1 Les données climatiques

Dans cette étude deux variables climatiques extrêmes seront étudiées, les maxima annuels de température et de précipitation. Les résultats seront présentés dans la partie 3. Cependant il faut d'abord analyser la validité de certaines hypothèses ou degrés d'approximation. En effet, les données de température ou pluviométrique avec ou sans forçage anthropique sont de nature différentes et doivent être traitées en conséquence. Par exemple le forçage anthropique induira une tendance commune temporelle dans les signaux de maxima de température. Ces signaux pourront difficilement être traités comme indépendants temporellement.

Pour l'apprentissage d'une distribution pure, section 3.1, les exemples fournis aux modèles doivent être iid. La première approximation sera de remplacer l'indépendance par la non-corrélation. Le meilleur choix de données pour valider cette hypothèse sont les maxima de précipitation obtenus par des simulations sans le forçage anthropique. Les corrélations inter-cellules et inter-modèles sont présentés dans la figure 2. L'hypothèse de distribution identique peut-être questionnée en observant la structure spatiale dans la figure 3. Pour résoudre ce problème deux approximations sont possibles : considérer la distribution dans sa globalité (avec la structure spatiale) ou appliquer une transformation pour gommer cette structure spatiale. Du point de vue de l'apprentissage automatique aucun de ces deux cadres ne présente un réel avantage, cela est différent en ce qui concerne la théorie des valeurs extrêmes. En effet, le caractère extrême des données sera préservé après une transformation min-max. Cela permettra de comparer l'ajustement d'une GEV à l'apprentissage d'un GAN.

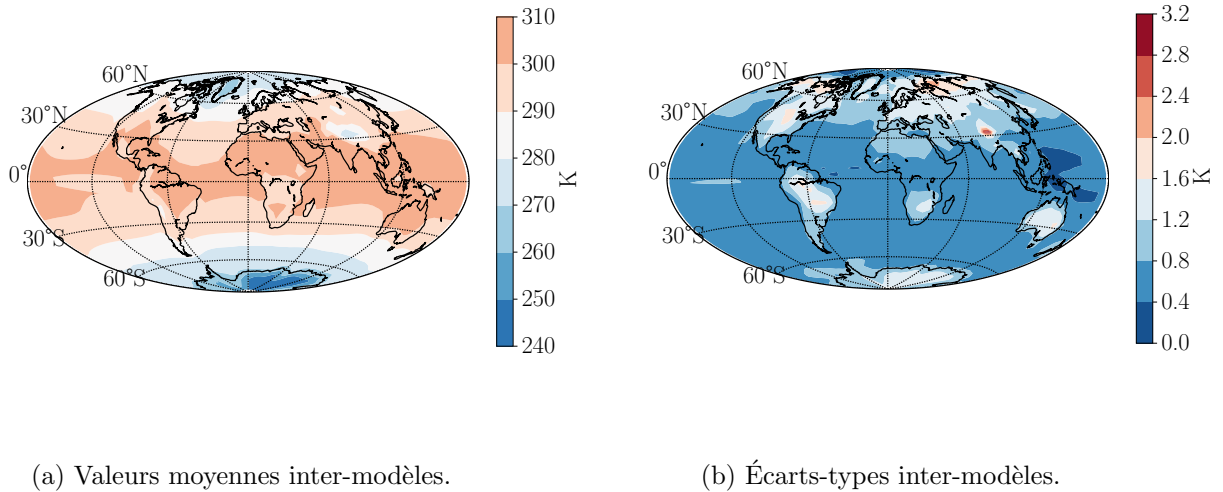


**Fig. 2** – (a) Corrélations moyennes inter-cellules pour une cellule de référence donnée. (b) Corrélations moyennes inter-modèles pour un modèle de référence donné. (a-b) Le coefficient de corrélation calculé est celui de Pearson, rendant compte de la corrélation linéaire. Les graphes de corrélations sont semblables pour différents modèles ou cellules de référence.



**Fig. 3** – (a) Valeurs moyennes du maxima de précipitation au cours des simulations et des modèles. (b) Écart-type du maxima de précipitation au cours des simulations et des modèles. (a-b) Ces deux cartes rendent compte de la structure spatiale des données.

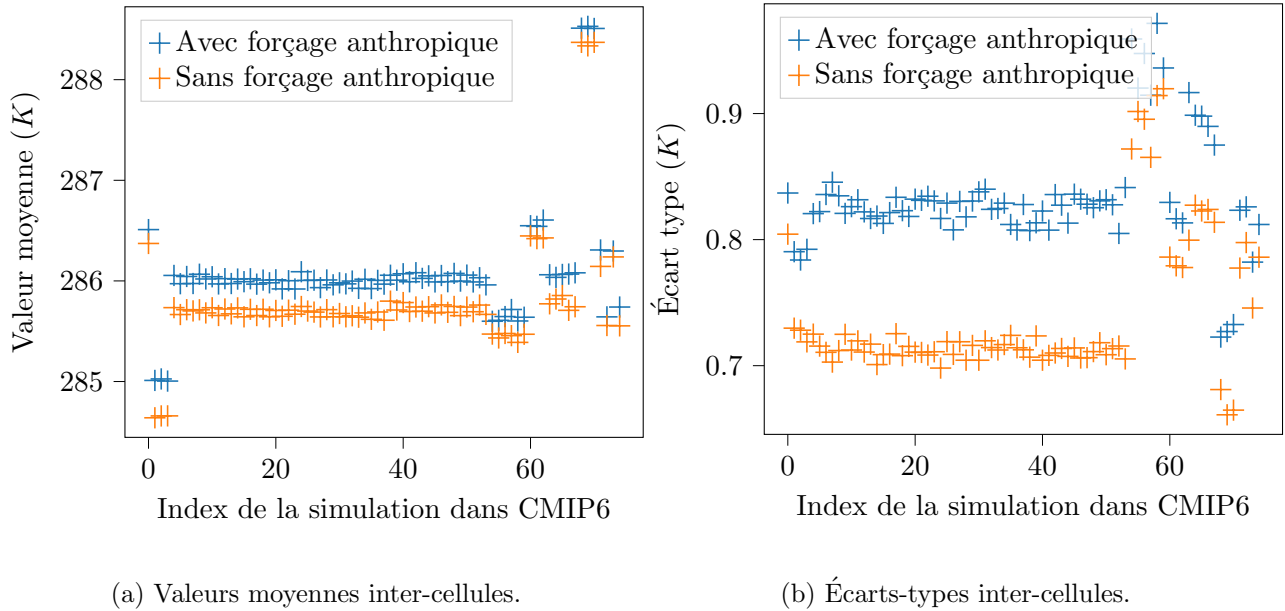
Pour l'apprentissage d'une tendance temporelle, section 3.2, il est nécessaire que les exemples possèdent une structure temporelle commune. Le choix le plus adapté dans les données physiques est clairement celui des maxima de température obtenus par des simulations avec forçage anthropique. La tendance à apprendre sera donc celle du réchauffement climatique global bruité et particularisé par les comportements locaux. Dans ces données deux structures apparaissent, une structure spatiale, figure 4, et une structure dans les modèles, figure 5.



**Fig. 4** – (a) Valeurs moyennes du maxima de température au cours des simulations et des modèles. (b) Écart-type du maxima de température au cours des simulations et des modèles. Ces deux cartes rendent compte de la structure spatiale des données.

Le problème déjà rencontré de la structure spatiale sera traité par une transformation spatiale. Le problème de la structure dans les modèles sera effacé en moyennant les signaux selon les différents modèles. Les avantages sont la réduction du bruit et l'accent qui est mis sur la tendance globale, l'apprentissage devrait alors être plus direct. L'inconvénient majeur et celui de la perte du caractère

extrême des données. Cependant, prendre le maximum entre les modèles n'aurait fait que sélectionner toujours les mêmes modèles et aurait donc privé les données de l'information contenue par les autres. Ce cadre devrait mettre en défaut l'ajustement par une GEV, ce qui justifierait la méthode générale d'apprentissage par un GAN.



**Fig. 5** – (a) Valeurs moyennes du maxima de précipitation au cours des simulations et des modèles. (b) Écart-type du maxima de précipitation au cours des simulations et des modèles. Ces deux cartes rendent compte de la structure spatiale des données.

Finalement, pour l'apprentissage d'une structure spatiale et temporelle, section 3.3, peu de traitement est réellement nécessaire. Il pourrait être intéressant de centrer et réduire les cellules pour ne pas avoir à apprendre des informations du premier ordre. L'accent serait alors mis sur l'apprentissage des différents comportements et distributions. Cependant il s'agit ici de guider l'apprentissage vers les caractéristiques de premier ordre, la structure spatiale sera donc conservée.

## 2.2 Simulation et analyse des données

### 2.2.1 Simulations de valeurs extrêmes

L'apprentissage sur les jeux de données climatiques possède un inconvénient majeur qui est celui du contrôle de l'apprentissage. La distribution exacte des variables étudiées étant inconnue le contrôle de l'apprentissage ne peut être exact. Il est toujours possible de comparer l'apprentissage du modèle avec les données originelles, seulement un sur-ajustement des données constitue un frein à la généralisation du modèle. Ainsi, chacune des expériences introduites dans la section 2.1 sera accompagnée d'une expérience de contrôle semblable. Les résultats ainsi que les descriptions plus détaillées de ces expériences sont donnés dans la partie 3.

Un autre avantage de générer des données de contrôle se situe dans l'entraînement du GAN. Il est plus simple d'ajuster certains hyper-paramètres sur un jeu de données particulier restreint et basique. De plus, il est alors possible de tester facilement les limites des GANs. Par exemple, l'apprentissage de distributions à queue lourde ou multi-modale se révèle bien plus compliqué que celui des distributions classiques.

D'un point de vue pratique les données seront générées de façon classique avec une distribution parente et avec l'opérateur `max`. La question du choix des distributions parentes pour les expériences de contrôle dépend fortement de la variable climatique d'étude. Le choix s'est porté sur la distribution de Student qui possède une grande flexibilité via son paramètre de nombre de degrés de liberté. Ce choix arbitraire n'est pas ici pénalisant car il ne s'agit pas de reproduire les variables climatiques, via cette

distribution, mais de simplement de fournir une expérience semblable. Pour le niveau de précipitation un faible nombre de degrés de liberté, ici 5 est choisi, modélise adéquatement une distribution à queue lourde. Pour la température le choix s’est porté sur une loi de Student à 50 degrés de liberté, s’approchant donc d’une loi gaussienne. L’effet de seuil dans le niveau de précipitation a été reproduit avec la désymétrisation de la loi de Student par la fonction valeur absolue. Le support de l’opérateur **max**, variable dans les phases d’ajustement d’hyper-paramètres, sera fixé à 90 dans les expériences de contrôle pour approximer le phénomène de saisonnalité.

### 2.2.2 Analyse statistique avec un diagramme quantile-quantile

Un des objets choisis pour contrôler l’apprentissage est le diagramme Q-Q (quantile-quantile). Ce diagramme permet de comparer deux distributions entre elles. Il sera donc possible de comparer l’ajustement par une GEV et l’apprentissage par un GAN à la distribution des données. Ce diagramme s’appuie sur la statistique d’ordre fournie par les quantiles des données. Le quantile théorique  $z_q$  associé au fractile  $q$  avec la fonction de répartition  $G$  vérifie l’équation (8),

$$z_q = G^{-1}(q). \quad (8)$$

Cette équation revient à dire que la probabilité que la variable  $Z$ , associé à  $G$ , soit inférieur à  $z_q$  est de  $q$ . Cependant, pour un ensemble de données,  $G$  n’est pas forcément connue. Une approximation du quantile théorique,  $z_q$ , est alors possible grâce à l’équation (9),

$$z_q = Z_{(\lfloor qN \rfloor)}(\lfloor qN \rfloor + 1 - qN) + Z_{(\lfloor qN \rfloor + 1)}(qN - \lfloor qN \rfloor), \quad (9)$$

avec  $N$  le nombre de données et  $(Z_{(i)})_{i \in [1, N]}$  les données ordonnées. L’équation (9) est une interpolation linéaire entre les deux données les plus proches du quantile. En pratique on choisira  $q$  tel que  $qN$  soit un entier.

Le principe du diagramme Q-Q est alors de tracer les quantiles d’une distribution en fonction de ceux d’une autre. Il devient alors possible d’interpréter la courbe obtenue pour comparer les deux distributions. Pour deux distributions identiques la courbe doit être la droite d’équation  $y = x$ . Pour deux distributions du même type mais avec des moyennes et écarts-types différents l’équation devient affine  $y = ax + b$ . De plus, la convexité de la courbe peut donner des indications sur les différences de symétrie des distributions. Une première inspection visuelle du graphe donnera donc une métrique pour vérifier l’apprentissage du type de distribution. Ensuite le coefficient de régression linéaire ainsi que la pente et l’ordonnée à l’origine renseigneront des comparaisons plus subtiles entre les distributions.

Avec une telle technique les données aberrantes ont moins d’influence. En effet, les comparaisons de quantiles sont moins sensibles aux données aberrantes que des comparaisons de moments. Ensuite, cette technique rend plus compte de la différence de forme des distributions comparé aux diagrammes probabilité-probabilité qui sont souvent trop linéaires. Enfin, les deux niveaux de lectures de ce type de diagrammes permettent de faire une analyse grossière ou précise selon la qualité de l’ajustement.

## 2.3 Les réseaux antagonistes génératifs

### 2.3.1 Structure adoptée

Une partie importante du développement d’un GAN réside dans le choix de la structure de ce dernier. Théoriquement il faut juste assez de capacité, i-e de neurones, pour que le GAN saisisse l’entière du jeu de données, [Goodfellow et al. \(2014\)](#). Cependant certaines instabilités d’entraînements ou certains problèmes de cartographie peuvent survenir, empêchant la convergence vers l’équilibre minimax. Si les deux réseaux sont déséquilibrés dans leur structure un réseau peu prendre le dessus sur l’autre. Ensuite, dans certains cas de sous-apprentissage le générateur associe trop d’exemples de l’espace latent au même exemple de l’espace réel, aussi appelé ”mode collapse”. Il sera donc l’objet ici d’adopter certaines techniques et lignes directrices quand à la structure du GAN.

La structure principale choisie pour les GANs de cette étude s’appuie sur le GAN proposé par [Radford et al. \(2016\)](#). Bien que ce GAN ait été développé pour des images la structure sera adaptée

pour traiter des séries temporelles. Les objets au centre de cette architecture sont les couches de convolution. Le système de partage de paramètre permet de traiter certaines symétries de la série temporelle comme l’invariance temporelle du résidu autour de la tendance temporelle. De plus, il sera alors plus aisé pour le générateur de produire des signaux temporellement corrélés sans sur-apprendre. Le choix de la structure proposé dans Radford et al. (2016), symétrique pour le générateur et le discriminateur, permet au générateur, respectivement au discriminateur, d’apprendre les liens de sur-échantillonnage, respectivement de sous-échantillonnage.

Un autre choix permettant de stabiliser l’apprentissage se situe dans celui des fonctions d’activation des réseaux. La première qui est ”imposée” par les données est la dernière couche du générateur. Pour les images l’usage est d’utiliser des fonctions bornées comme  $\tanh$ . Cependant ici les données n’étant pas théoriquement bornées le choix s’est porté sur la fonction linéaire,  $x \mapsto x$ . Ensuite, pour le générateur le choix s’est porté sur la fonction de rectification linéaire,  $x \mapsto \max(0, x)$ . Ce choix est tiré de la structure de Radford et al. (2016), utilisée dans Allouche et al. (2021), et dont les propriétés statistiques ont été étudiées dans Chen et al. (2020). Ce choix peut se justifier sur la stabilité apportée par cette fonction ainsi que l’efficacité d’apprentissage due à sa plage d’activation infinie. Pour le discriminateur la fonction d’activation finale choisie est la fonction logistique, donnée par l’équation (10),

$$\sigma : x \mapsto \frac{1}{1 + e^{-x}}. \quad (10)$$

Cette fonction est très adaptée pour le cadre de classification binaire dans lequel se trouve le discriminateur. Enfin, la fonction d’activation choisie pour les couches du discriminateur est une modification la fonction de rectification linéaire :  $x \mapsto \max(\epsilon x, x)$ . Bien qu’elle introduise un nouvel hyper-paramètre, son avantage principale réside dans la non-nullité de son gradient pour les entrées négatives, permettant une meilleure rétro-propagation.

Finalement, dans la section 3.3 une structure additionnelle de conditionnalité est ajoutée à la structure principale. Ce type de méthodologie a été introduit par Mirza & Osindero (2014). L’objectif de l’apprentissage pour le générateur devient la distribution conditionnelle  $p_{\text{donnees}}(x|c)$  où  $c$  peut représenter une étiquette, une classe ou un attribut. Cela permet notamment d’introduire une structure plus ou moins complexe des données. La structure générale décrite précédemment sera conservée, le GAN devra seulement traiter deux flux de données au lieu d’un seul. Une couche intermédiaire sera alors introduite entre les données de conditionnalité et le traitement par le GAN. Il s’agira d’une couche dense permettant d’augmenter la dimension des entrées. Cela est nécessaire pour ne pas à avoir à modifier la structure du GAN, mais cela présente aussi l’avantage laisser de la flexibilité au réseau pour apprendre la distribution conditionnelle.

La structure plus détaillée des réseaux utilisés pour la partie 3 est donnée dans l’annexe B. L’organisation globale, certaines caractéristiques ainsi que l’implémentation de l’aspect conditionnel des données sont approfondies.

### 2.3.2 Précautions pour l’entraînement

Après avoir présenté la structure générale des GANs utilisés il s’agit ici de présenter les lignes directrices adoptées pour l’entraînement. Il s’agira donc de traiter de la stabilité de l’entraînement mais aussi du sur-apprentissage et du sous-apprentissage.

Tout d’abord pour lutter contre le sur-apprentissage, des couches de ”Dropout” seront ajoutées au discriminateur. Ces couches permettent d’abandonner certains poids pour éviter l’encodage pur des données. Le générateur est moins sensible à ces subtilités car il bénéficie d’une rétro propagation indirecte par le discriminateur, il ne ”voit” donc jamais les données. Cette structure a pu être trouvée dans certains réseaux proposés dans Salimans et al. (2016). Enfin, la technique de bruitage des classes, Salimans et al. (2016), a permis en partie de lutter contre l’excès de confiance du discriminateur. Ce dernier pouvait aussi introduire de trop fort gradients.

L’augmentation de la stabilité de l’entraînement passe justement par la gestion de ces gradients. Pour éviter la divergence des gradients, leurs composantes ont été bornées, une technique similaire de coupure de gradient a été utilisée dans Bhatia et al. (2021) et inspirée de la coupure des poids

introduite par Arjovsky et al. (2017). Pour stabiliser plus avant l'apprentissage du GAN, des couches de normalisation de flux ont été ajoutées au discriminateur et au générateur comme conseillé par Radford et al. (2016). Enfin, pour lutter contre le déséquilibre d'apprentissage entre le générateur et le discriminateur, des taux d'apprentissage différents ont été attribués aux deux réseaux. Une explication théorique est la différence des cadres d'optimisation du générateur et du discriminateur qui peut introduire une différence de vitesse de convergence. En pratique cela réduit le sur-apprentissage du générateur, son taux d'apprentissage est choisi plus faible, cela permet donc de lutter contre le "mode collapse".

La méthode d'entraînement, brièvement décrite dans la section 1.4.1, est donnée dans l'annexe A de façon plus approfondie. Cette description et celle de l'annexe B contiennent la mise en place de certaines des méthodes décrites ici.

### 3 Résultats

Les résultats de reproduction de variables climatiques extrêmes seront présentés dans cette section. Plusieurs cadres d'expérience propres à différentes variables climatiques seront étudiés.

#### 3.1 Apprentissage de distributions

L'apprentissage probabiliste le plus simple et le plus direct est clairement celui de données iid. Dans ce cadre d'apprentissage, le but est de recouvrer une distribution inconnue à partir d'un nombre fini d'observations régies par cette distribution. Ce problème peut parfois se résoudre de façon directe lorsque l'on a une idée de la solution. Par exemple si le type de la distribution théorique est connu, la solution est obtenue par un simple ajustement d'hyper-paramètres par des méthodes d'optimisation.

Pour ce premier cadre d'apprentissage on se propose d'apprendre la distribution du maximum annuel de précipitation pour des simulations sans forçage anthropique. Le choix de ce type de simulation a été justifié dans la section 2.1. En effet, ce jeu de données possède très peu de corrélations temporelles ainsi qu'une distribution de type extrême une fois la structure spatiale gommée. Cette expérience sera accompagnée d'une expérience de contrôle similaire. Le choix de la distribution pour les données générées sera celui d'une loi de Student repliée entre  $[0, +\infty[$  avec 5 degrés de liberté. Peu de degrés de liberté ainsi que le fait d'asymétriser la distribution permet de générer des données semblables aux données de précipitation. Le support de l'opérateur `max` sera choisi comme une saison, soit 90 jours. Bien que les données climatiques soient des maxima annuels, il apparaît biaisé de prendre un support de 365 jours dû à la forte influence de la saisonnalité sur les variables étudiées.

Pour l'expérience de contrôle les quantiles théoriques sont donc donnés par l'équation (11),

$$z_q = F^{-1} \left( q^{1/2n} \right), \quad (11)$$

où  $F$  est la fonction de répartition de la loi de Student à 5 degrés de liberté,  $q$  le fractile,  $z_q$  le quantile associé et  $n$  le support de l'opérateur `max`.

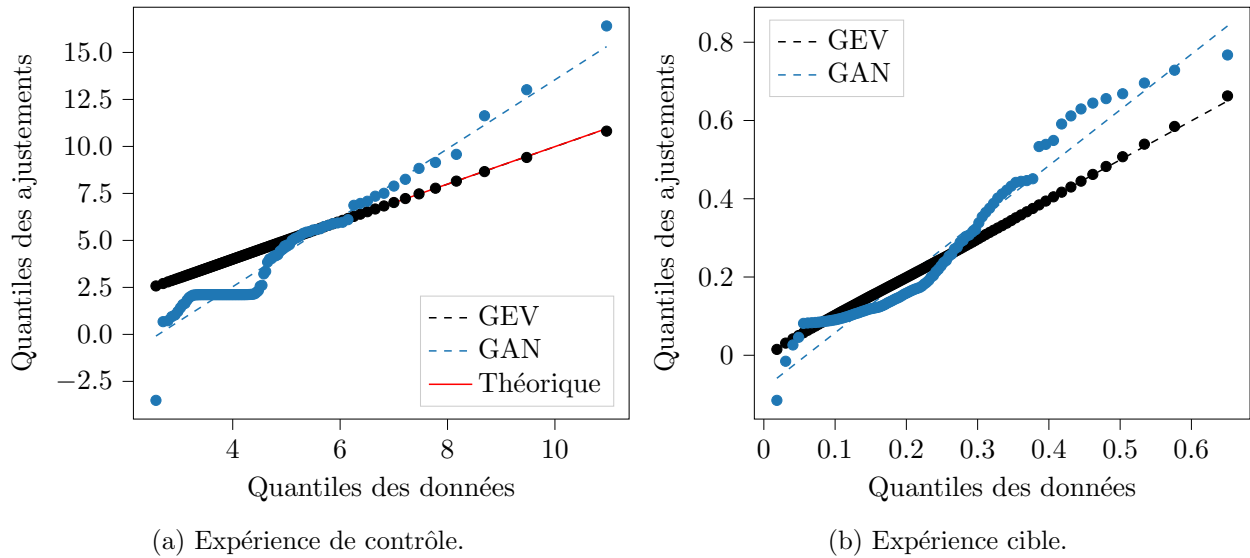
La méthode de comparaison choisie sera celle de l'alignement sur un diagramme Q-Q, présentée dans section 2.2.2. Cette méthode s'avère suffisante pour comparer les résultats présentés en figure 6. Les GANs des deux expériences ont été entraînés sur le même nombre d'exemples, environ  $1.5 \times 10^5$ . Ensuite, la GEV a été ajustée sur 1000 exemples par la méthode de maximisation de vraisemblance. Enfin, la distribution du GAN a été extrapolée sur 1000 exemples générés.

Sans surprise, dans l'expérience de contrôle la droite des quantiles théoriques se confond avec l'ajustement par une distribution GEV. Ce résultat était attendu connaissant la nature des données et la flexibilité d'une GEV. De plus, ce résultat semble se répéter pour les données climatiques. Bien que la distribution théorique soit inconnue, l'alignement des quantiles de la GEV avec ceux des données se fait sur la droite d'équation  $y = x$ . En effet, le coefficient  $R^2$  est très proche de 1, donc la courbe est bien une droite. Cela signifie que les distributions sont identiques. Ensuite, la dispersion des quantiles est identique car  $a \approx 1$  et les quantiles ne sont pas décalés car  $b \approx 0$ . On peut donc prendre l'ajustement par une GEV comme référence pour la figure 6(b).



Les résultats de la figure 6 montrent donc que l'apprentissage par un GAN n'est pas à la hauteur de l'ajustement par une GEV. Les données générées par le GAN sont un peu décalées et un peu plus dispersées, il n'a donc pas totalement appris ces informations de premier niveau. Dans l'expérience de contrôle figure 6(a) on observe un plateau. Ce dernier signifie que le GAN produit une distribution bi-modale. De plus, sur la figure 6(a), le premier quantile des données générées par le GAN est négatif, alors qu'aucune donnée du jeu d'apprentissage n'est négative.

Une explication assez pragmatique est que le réseau n'est sûrement pas assez entraîné. Cependant dans une recherche de performance d'entraînement un tel argument se révèle peu pertinent. Ce résultat n'est en fait pas très surprenant, il s'avère que dans les expériences précédentes l'atout majeur de la structure du GAN choisie, section 2.3.1, n'a pas été mise à profit. En effet, les couches de convolutions possèdent l'avantage de partager certains paramètres entre neurones, ce qui s'avère être un atout pour générer des données corrélées, mais pas ici. Ce résultat amène à l'exploration de l'apprentissage de séries temporelles non-stationnaires. Les corrélations temporelles pourraient ainsi compliquer l'ajustement par une GEV et donc justifier l'utilisation d'un GAN.



**Fig. 6** – (a) Diagramme Q-Q avec les données à apprendre en référence pour l'expérience de contrôle. Ajustement pour la GEV :  $R^2 = 0.9998$ ,  $a = 0.996 \pm 0.001$  et  $b = 0.020 \pm 0.006$ . Pour le GAN :  $R^2 = 0.947$ ,  $a = 1.83 \pm 0.04$  et  $b = -4.8 \pm 0.2$ . (b) Diagramme Q-Q avec les données à apprendre en référence pour l'expérience cible. Ajustement pour la GEV :  $R^2 = 1.000$ ,  $a = 0.999 \pm 0.002$  et  $b = 0.000 \pm 0.001$ . Pour le GAN :  $R^2 = 0.952$ ,  $a = 1.42 \pm 0.03$  et  $b = -0.085 \pm 0.008$ .

### 3.2 Apprentissage de séries temporelles

Dans cette section l'hypothèse d'indépendance temporelle des données ne sera plus nécessaire car l'objectif sera l'apprentissage de séries temporelles, supposées stationnaires du point de vue de la distribution. Ce point de vue semble plus pertinent dans le cadre de l'apprentissage de variables climatiques sur lesquelles la temporalité peut avoir un effet notable. Pour insister sur l'apprentissage d'une série temporelle, les données choisies seront celles possédant les plus fortes corrélations temporelles. Ce seront donc les maxima de température des simulations avec forçage anthropique. Cependant, comme précisé dans la section 2.1, la structure spatiale et de modèle devra être gommée. Pour insister sur la tendance temporelle, la moyenne inter-modèles centrée-réduite sera étudiée. Une autre possibilité aurait été de prendre le maximum inter-modèles puis d'appliquer une transformation de type min-max. L'information de la tendance du réchauffement climatique est portée par tous les modèles et ainsi la moyenne a plus de sens, permettant d'atténuer le biais de certains modèles. De plus, en appliquant une telle transformation, le caractère extrême des données est atténué, complexifiant l'ajustement par une GEV. Aucune distribution entre une GEV, une Student ou une gaussienne ne peut être réellement privilégiée. La méthode non paramétrique de l'ajustement par un GAN semble donc théoriquement la



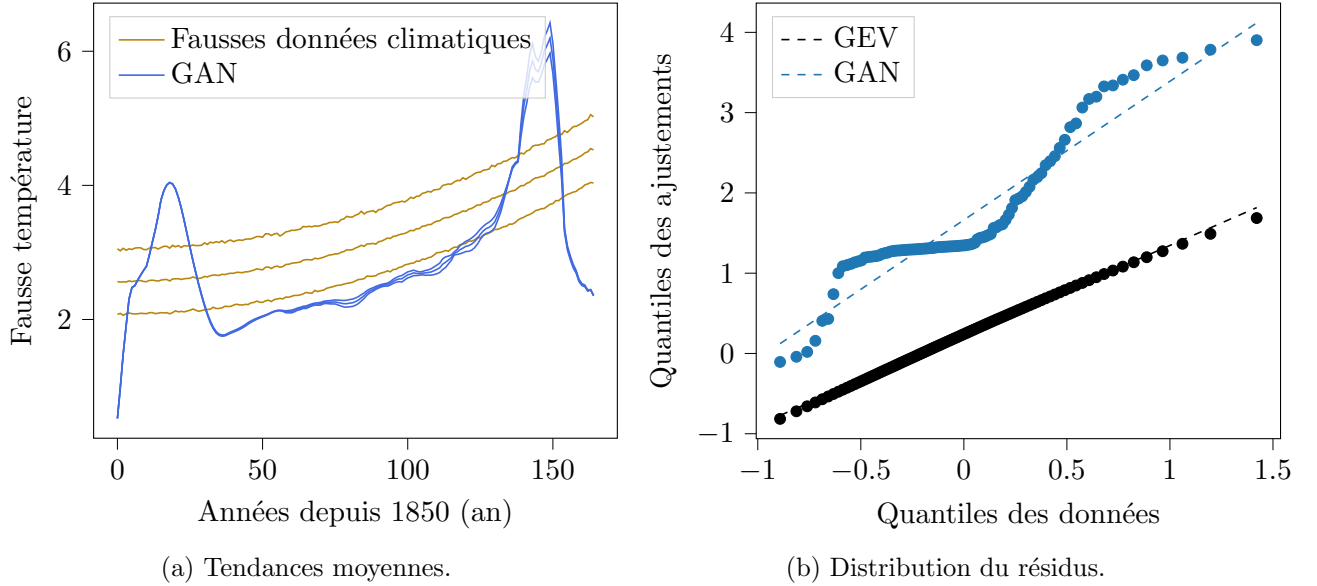
meilleure approche.

Pour l'expérience de contrôle, les données seront axées sur une tendance temporelle. Pour simuler un comportement similaire à la température, la distribution choisie sera une Student à 50 degrés de liberté standardisée,  $\mathcal{T}(0, 1, 50)$ . Le support de l'opérateur  $\max$ ,  $n$ , sera de 90 jours. Une série temporelle générée,  $Z(t)$ , sera donc décrite par l'équation (12),

$$Z(t) = \mu(t) + \max(X_1(t), \dots, X_n(t)) \quad \text{avec } \forall i \in [1, n], \forall t \in \mathbb{R}, X_i(t) \sim \mathcal{T}(0, 1, 50), \quad (12)$$

avec  $\mu(t)$  une fonction quadratique ici. L'apparente discontinuité introduite par cette définition est lissée par la discrétisation temporelle choisie pour être la même que celle des données climatiques.

Les résultats de l'expérience de contrôle sont présentés sur la figure 7 et ceux de l'expérience cible sur la figure 8. Ces résultats contiennent chacun la tendance temporelle moyenne des jeux de données ainsi que l'ajustement du résidu sur un diagramme Q-Q. Le résidu est obtenu en enlevant la tendance moyenne des données,  $\mu(t)$ , à toutes les séries temporelles. Enfin, les GANs ont été entraînés sur 2592 exemples, un par cellule, et les distributions ont été extrapolées à partir de 1000 exemples comme dans la section 3.1.

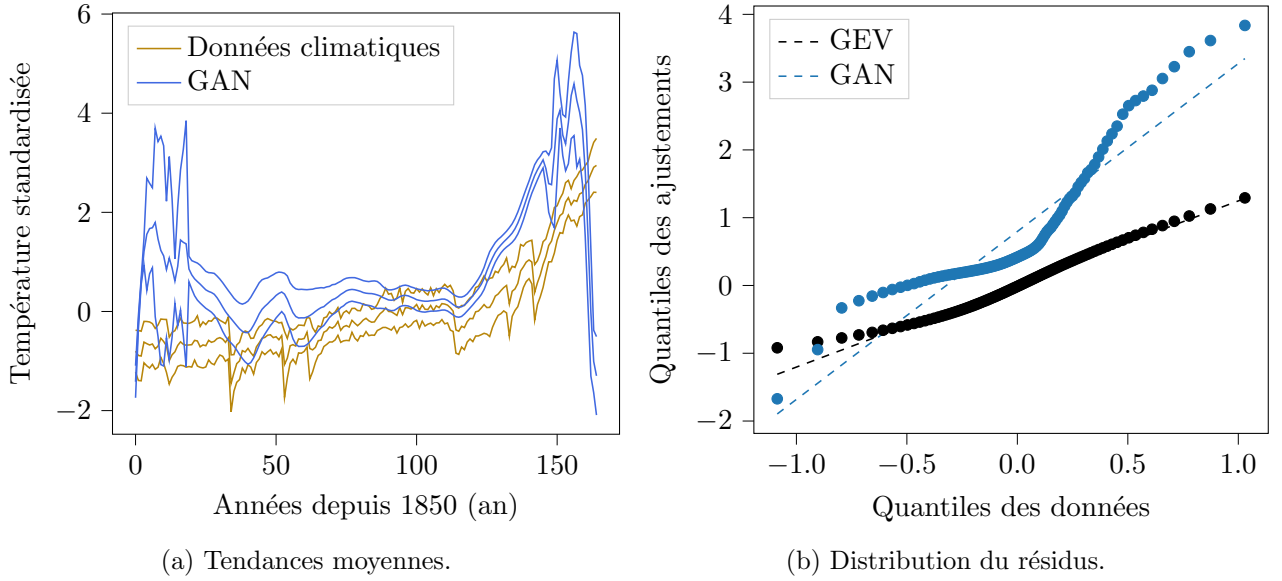


**Fig. 7** – (a) Tendances moyennes  $\mu(t)$ ,  $\mu(t) + \sigma(t)$  et  $\mu(t) - \sigma(t)$  pour l'expérience de contrôle. (b) Diagramme Q-Q pour le résidu de l'expérience de contrôle. Ajustement pour la GEV :  $R^2=0.998$ ,  $a = 1.119 \pm 0.005$  et  $b = 0.226 \pm 0.002$ . Pour le GAN :  $R^2=0.896$ ,  $a = 1.73 \pm 0.06$  et  $b = 1.66 \pm 0.03$ .

La tendance temporelle est en partie apprise par le GAN dans l'expérience cible, figure 8(a). On observe un léger décalage mais la plupart des variations climatiques sont reproduites par le GAN. Les extrémités des séries temporelles sont mal reproduites, cela peut rendre compte d'un manque d'entraînement ou de capacité du GAN. Dans l'expérience, figure 7(a), la faible variabilité des tendances indique la présence de "mode collapse". De plus, les extrémités étant mal ajustées on peut penser que le GAN manque d'entraînement.

Les erreurs d'apprentissage de la tendance se répercutent aussi sur le diagramme Q-Q. Cela est particulièrement vrai pour l'expérience de contrôle, figure 7. Les grandes différences aux extrémités des séries temporelles induit un décalage et un étalement des quantiles. De plus, la distribution du résidu produit par le GAN est alors bi-modale.

Dans la globalité, l'ajustement par une GEV, avec le paramètre de localisation non-stationnaire  $\mu(t)$ , s'avère plus performant. Cela était attendu pour l'expérience de contrôle connaissant la nature des données décrites par l'équation (12). On observe tout de même une concavité de la courbe obtenue en figure 8(b) ce qui traduit une légère asymétrie entre la distribution des données et la GEV.



**Fig. 8** – (a) Tendances moyennes  $\mu(t)$ ,  $\mu(t) + \sigma(t)$  et  $\mu(t) - \sigma(t)$  pour l’expérience cible. (b) Diagramme Q-Q pour le résidu de l’expérience cible. Ajustement pour la GEV :  $R^2=0.976$ ,  $a = 1.23 \pm 0.02$  et  $b = 0.022 \pm 0.007$ . Pour le GAN :  $R^2=0.865$ ,  $a = 2.5 \pm 0.1$  et  $b = 0.80 \pm 0.04$ .

Ces résultats sont donc mitigés car la tendance des séries temporelles étudiées ici n’est pas totalement reproduite alors qu’il s’agit de l’information principale des données. Gommée ici, la structure spatiale des données pourrait permettre d’amoinrir le phénomène de ”mode collapse”. En effet, la structure spatiale rajoute de la variabilité dans le jeu de données. On peut donc envisager le cadre des distributions conditionnelles pour introduire cette structure.

### 3.3 Apprentissage de séries temporelles et spatiales

Pour pousser encore plus loin l’apprentissage il est possible de garder la structure spatiale et de se servir de cette information pour obtenir une reproduction de variables climatiques plus proche et plus conforme à la réalité. Ainsi dans cette partie, l’objectif sera toujours l’apprentissage de séries temporelles mais en gardant leur structure spatiale. Cette information traduite par les coordonnées spatiales permettra au réseau de comprendre les liens entre les différentes séries temporelles. La localisation des séries temporelles sera passée en argument du générateur ainsi que du discriminateur. Il est alors nécessaire d’utiliser un réseau avec une structure conditionnelle, présenté dans la section 2.3.1. La distribution à apprendre pour le générateur est donc  $p_{\text{donnees}}(x|c)$  où  $c$  représente les coordonnées spatiales pour les données climatiques.

Les données choisies seront les maxima de température des simulations avec forçage anthropique, possédant différentes tendances selon la localisation. Le cadre de l’expérience précédente, section 3.2, sera conservé. La différence pour les données climatiques sera l’absence d’agrégation de modèles et de standardisation spatiale. Pour l’expérience de contrôle la conditionnalité portera sur la pente de la tendance temporelle. La série temporelle  $Z_c(t) \sim p_{\text{donnees}}(x|c)$  générée pour l’expérience de contrôle est décrite par l’équation (13),

$$Z_c(t) = c\mu(t) + \max(X_1(t), \dots, X_n(t)) \quad \text{avec } \forall i \in [1, n], \forall t \in \mathbb{R}, X_i(t) \sim \mathcal{T}(0, 1, 50). \quad (13)$$

Dans l’équation (13),  $c$  décrit la valeur de la pente choisie pour être dans l’intervalle  $[1/2, 10]$ . La notation  $Z_c$  sera conservée pour les séries temporelles climatiques, où  $c$  représentera les coordonnées d’une cellule donnée.

Pour contrôler l’ajustement de l’apprentissage il faut regarder l’apprentissage de la tendance temporelle mais aussi de la distribution conditionnelle. Les résultats de l’expérience de contrôle, figure 9, et de l’expérience cible, figure 10, contiennent deux graphes. Le premier est un tracé de la tendance

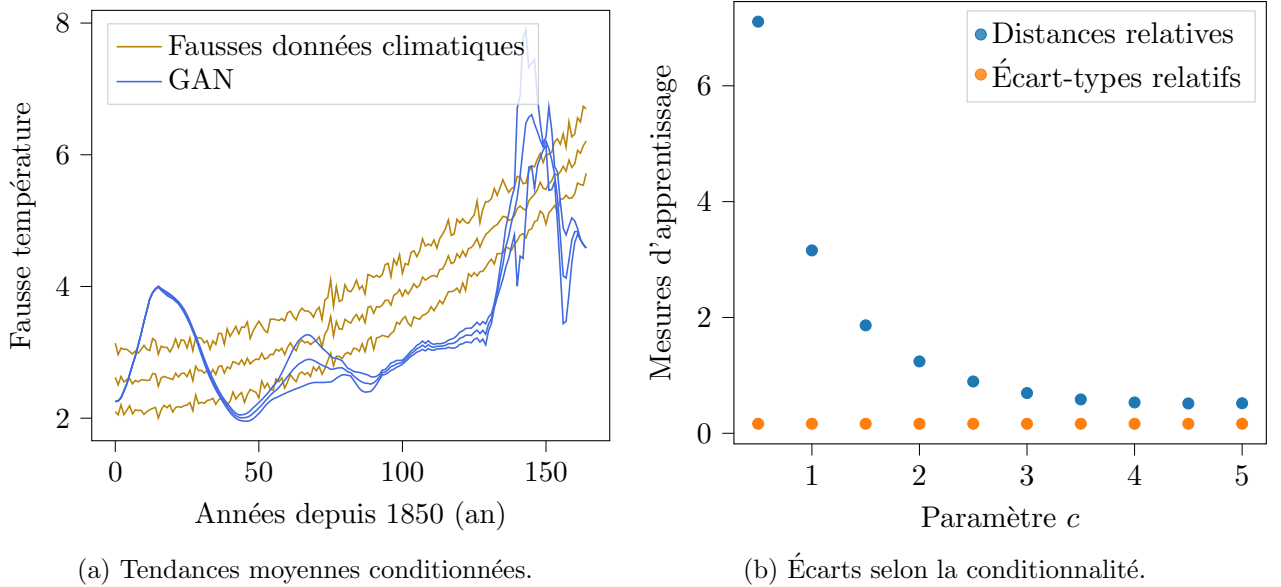
temporelle moyenne pour une valeur de  $c$  donnée, respectivement pour une cellule donnée. Il permettra une inspection visuelle de l'apprentissage de la tendance temporelle. Le deuxième est un graphe contenant deux types d'informations. Il contiendra d'abord la distance euclidienne moyenne normalisée entre les séries générées par le GAN et les séries des données le tout en fonction du paramètre  $c$ . Cette distance,  $d_c$ , dépendant du paramètre  $c$ , correspond à l'équation (14),

$$d_c = \frac{\|Z_{c,\text{donnees}}(t) - Z_{c,\text{GAN}}(t)\|_2}{\|Z_{c,\text{donnees}}(t)\|_2}, \quad (14)$$

où la norme 2 est prise sur les échantillons des séries temporelles. Ce deuxième graphe contiendra aussi la moyenne du rapport entre les écarts-types des séries temporelles,  $r_c$ , toujours en fonction du paramètre  $c$ . Le rapport  $r_c$  est donné par l'équation (15),

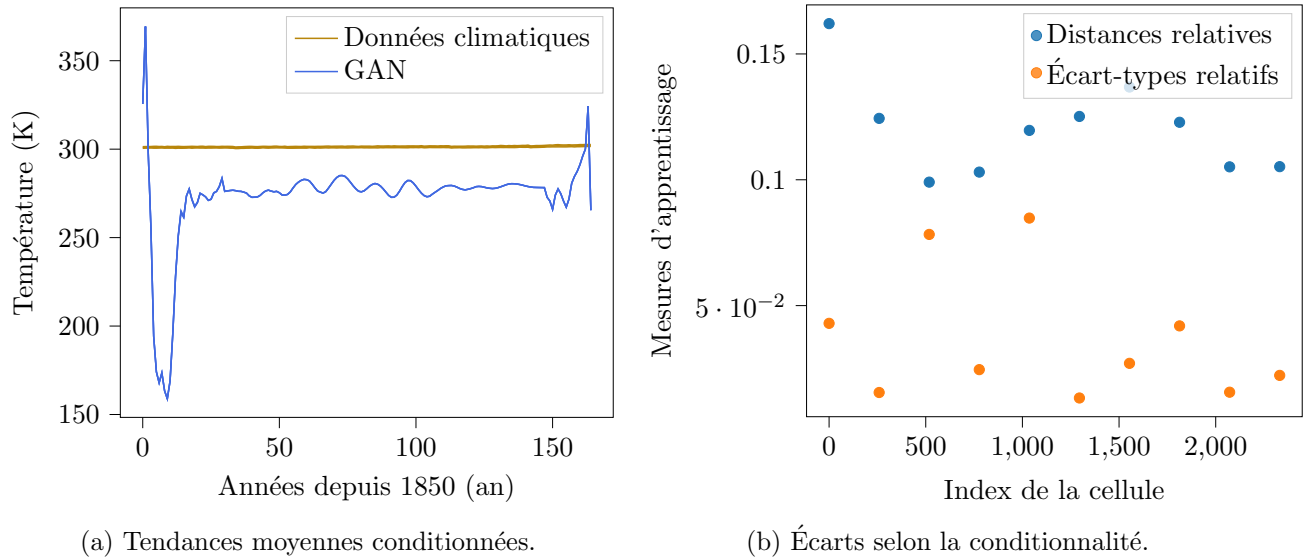
$$r_c = \frac{\sigma(Z_{c,\text{GAN}}(t))}{\sigma(Z_{c,\text{donnees}}(t))}. \quad (15)$$

La distance mesurera donc l'apprentissage de la tendance temporelle et le rapport des écarts-types mesurera la différence de dispersion des résidu des séries. En somme, le rapport des écarts-types donnera une mesure de la diversité des exemples créés par le GAN.



**Fig. 9** – (a) Tendances moyennes  $\mu(t)$ ,  $\mu(t) + \sigma(t)$  et  $\mu(t) - \sigma(t)$  pour une valeur de  $c \approx 3.4$  pour l'expérience de contrôle. (b) Rapport moyen entre les écarts-types des séries, et moyenne des distances euclidiennes normalisées entre les séries en fonction de la condition de pente  $c$ .

Pour l'expérience cible on observe un décalage global des tendances temporelles sur la figure 10(a). En regardant ces mêmes tendances pour des localisations différentes, on observe que la tendance moyenne des données générées par le GAN varie peu. L'explication se trouve en partie dans la figure 10(b). Le GAN n'a pas appris la structure conditionnelle de la distribution mais a simplement appris une distribution moyenne. Aux vues des figure 9(a) et figure 10(a) le problème semble aussi résider dans l'apprentissage des GANs. La convergence de l'apprentissage semble plus lente avec cette distribution conditionnelle. La première étape d'apprentissage, celle de l'apprentissage du comportement global des données, n'est pas totalement complète. Les écarts-types relatifs moyens sont bas dans la figure 9(b) et la figure 10(b). Cette faible variabilité peut aussi faire penser à la présence de "mode collapse", faible variabilité des exemples produits par le GAN. Cependant ces écarts-types relatifs ne sont pas totalement interprétables car les distances sont trop élevées. L'apparente proximité des séries temporelles dans la figure 10(b) est due à la relativité de la distance. L'information de premier niveau, la tendance temporelle, n'étant pas encore bien reproduite, la variabilité ne peut l'être.



**Fig. 10** – (a) Tendances moyennes  $\mu(t)$ ,  $\mu(t) + \sigma(t)$  et  $\mu(t) - \sigma(t)$  pour une cellule donnée,  $c = (67.5^\circ N, 17.5^\circ E)$  correspondant à l'index 1561, de l'expérience cible. (b) Rapport moyen entre les écarts-types des séries, et moyenne des distances euclidiennes normalisées entre les séries en fonction de la cellule.

Bien que peu convaincants, ces résultats ouvrent la voie vers de nombreuses applications dans le domaine de la génération de données. Certains problèmes récurrents comme celui des "mode collapse" nécessitent un approfondissement ainsi qu'une étude plus fine. La globalité de ces résultats sera discutée dans la section 4.1, certaines limitations, comme la variabilité des exemples, ainsi que de possibles améliorations seront elles discutées dans la section 4.2.

## 4 Discussion et conclusion

### 4.1 Résultats

Pour recontextualiser les résultats de la partie 3 avec les informations de la section 1, la figure 11 présente une carte mentale de l'étude. Ce schéma récapitule les outils et théories utilisés ainsi que les liens faits entre ces différents domaines.

Les résultats de l'apprentissage de distribution, section 3.1, montre que l'ajustement par un GAN d'une distribution, ici de type extrême, est moins performante que l'ajustement par une GEV. Ce résultat était en partie attendu. Connaissant la nature extrême des données, les distributions GEV fournies par la théorie des valeurs extrêmes s'avèrent plus pertinentes qu'une méthode non-paramétrique générale. Cependant cette méthode pourrait permettre de résoudre des problèmes avec des structures plus complexes comme dans les sections 3.2 et 3.3.

Les résultats de l'apprentissage de tendances temporelles, section 3.2, sont encourageants. En effet, pour l'expérience cible, avec des données climatiques, les données générées par le GAN reproduisent certaines variations visibles sur les données originelles, alors que l'on observe une légère baisse de précision de l'ajustement par une GEV. Cela est en partie dû au traitement des données, section 2.1 et 3.2. Ce cadre théorique plus complexe, avec la composition d'opérateurs, justifie l'utilisation de méthodes plus générales comme reproduction par un GAN.

Enfin, pour l'apprentissage de distributions conditionnelles, section 3.3, les résultats montrent un échec de l'apprentissage. La structure conditionnelle n'a pas du tout été apprise. Cela impose donc une remise en question de la méthodologie utilisée ainsi qu'une exploration des possibles causes et solutions.

Dans tous les résultats présentés en partie 3 plusieurs problèmes s'avèrent récurrents. Tout d'abord, le phénomène de "mode collapse" : lorsque la variabilité des exemples générés par le GAN est trop faible. Ensuite, la génération de données très éloignées et jamais vues par le GAN, pouvant être le

signe d'une divergence. Enfin, certains problèmes plus techniques ont pu empêcher la reproduction de la distribution des données. Par exemple, l'arrêt de l'apprentissage trop précoce et le manque de capacité des réseaux du GAN. Ces deux problèmes sont difficilement différentiables. Les solutions pour distinguer ces problèmes sont d'augmenter le nombre d'étapes d'entraînement ou d'augmenter la capacité des réseaux. Ces deux solutions demandent donc des temps de simulation et des ressources beaucoup plus importants, ce qui peut être compliqué à mettre en place.

La section 4.2 discute donc des problèmes de "mode collapse", de divergence et de certains aspects techniques.

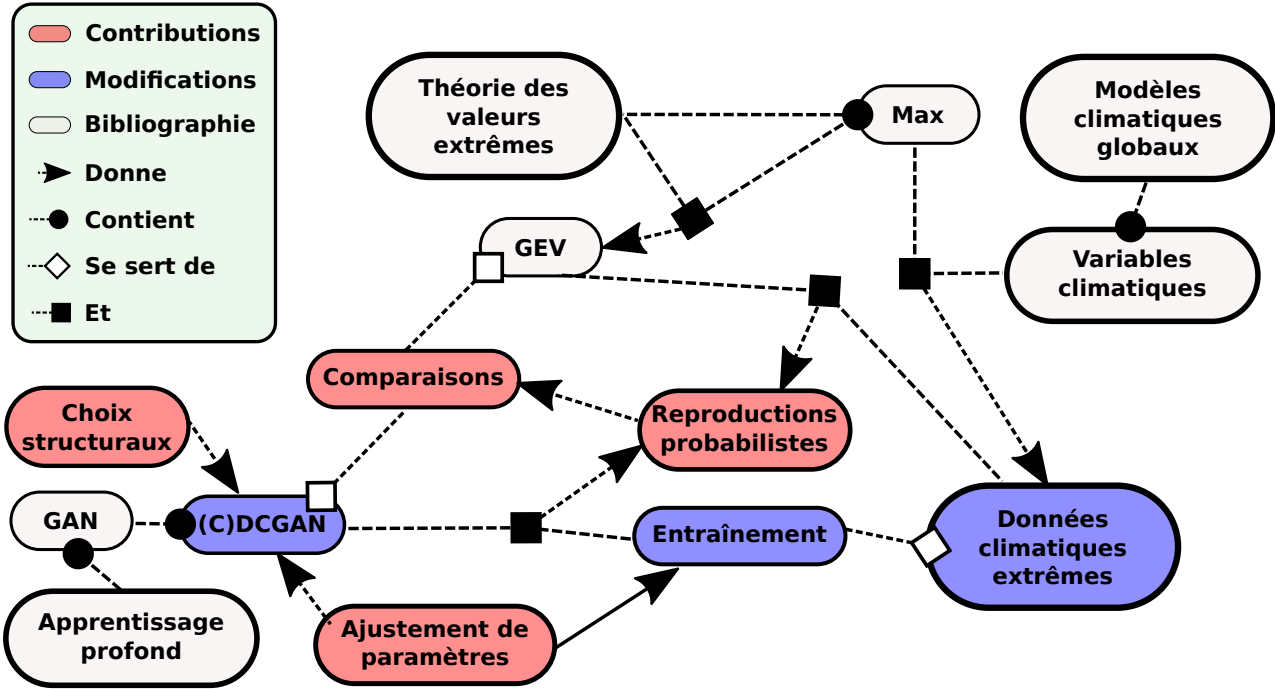


Fig. 11 – Schéma résumant les outils et théories utilisés ainsi que la plupart des contributions ou modifications apportées.

## 4.2 Limites

La première limite rencontrée dans cette étude réside dans la non-convergence de certains GANs utilisés. Son entraînement étant complexe et peu balisé, certaines divergences du générateur peuvent survenir. Plusieurs pistes peuvent être envisagées pour répondre à cette problématique en commençant par un ajustement plus fin des paramètres et de la structure. L'ajustement du taux d'apprentissage ainsi que de la hauteur des filtres de convolution a donné de bons résultats. Cependant pour être plus fin on pourrait utiliser des hauteurs variables ou des fonctions d'agréations de filtre.

L'apprentissage a aussi pu être asymétrique ou partiel. L'asymétrie provient d'un déséquilibre de structure ou de capacité entre les deux réseaux. La structure a donc été choisie symétrique et donnant les mêmes capacités aux deux réseaux. Cependant le cadre d'apprentissage étant différent, on peut penser qu'ils souffrent d'un déséquilibre naturel. L'étude de réseaux dissymétriques pourrait être une piste à envisager. Enfin, l'apprentissage partiel peut aussi provenir d'un arrêt trop précoce. Il serait donc intéressant de pousser le nombre d'étapes d'entraînement à l'extrême. Il faudrait alors conserver l'historique de l'entraînement du GAN pour ne pas obtenir un réseau ayant sur-appris.

Enfin, pour le phénomène de "mode collapse" la solution n'est pas forcément directe. Dans ce problème, le générateur est dans un cas de sous-apprentissage global. Cependant, il est localement dans un cas de sur-apprentissage. En effet, la plupart du temps le générateur produit le "mode", l'exemple le plus vraisemblable pour le discriminateur. Le générateur sur-ajuste donc le discriminateur. Pour amoindrir ce phénomène il faut réduire le taux d'apprentissage du générateur comparé à celui du discriminateur. Cette piste issue de Heusel et al. (2018) semble prometteuse.

Le domaine des GANs est majoritairement intéressé dans la production d'images. En ce sens, l'application des GANs aux séries temporelles possède moins de recul sur les limites et techniques de résolutions. La première possibilité de résolution serait d'utiliser des algorithmes et des métriques propres aux séries temporelles. La seconde serait de traiter la structure spatiale comme celle d'une image. Une alternative aux GANs conditionnels pourrait être envisagée comme celle présentée dans [Bhatia et al. \(2021\)](#). La conditionnalité est alors sous-jacente à la structure de l'image, et donc à la structure spatiale. La notion de carte est portée par la structure des pixels de l'image. Une application directe des techniques propres aux GANs développés pour les images serait alors possible.

### 4.3 Applications et perspectives

Une application possible des GANs réside dans l'augmentation de données, grande motivation dans le domaine des valeurs extrêmes, présenté dans la section 1.1. Au final, plus de données pourraient permettre une meilleure description des variables climatiques et ainsi conduire à une amélioration des prédictions climatiques. Notamment les périodes de retours de ces événements extrêmes.

De nombreuses améliorations des GANs sont possibles, notamment avec des techniques propres aux extrêmes comme présenté dans [Bhatia et al. \(2021\)](#) et [Allouche et al. \(2021\)](#). Une utilisation plus directe de la théorie des valeurs extrêmes aurait été aussi possible avec l'implémentation d'un résidu régi par une distribution de type GEV. Cependant la comparaison entre l'ajustement par une GEV et un GAN aurait pu être biaisée.

Actuellement, dans le domaine des modèles génératifs profonds, une technologie alternative au GANs est utilisée et développée. Il s'agit des VAE (Variational Auto Encoder) introduit par [Kingma & Welling \(2014\)](#). Ces derniers apprennent une cartographie entre l'espace latent et l'espace réel via l'apprentissage du codage et du décodage des données. Les VAE connaissent aussi de nombreuses améliorations récentes notamment avec des projets comme VAE-GAN [Makhzani et al. \(2016\)](#).

Les résultats obtenus dans la partie 3 ne sont donc pas définitifs, certaines améliorations et techniques pourraient permettre de surmonter les limites rencontrées lors de cette étude. Ainsi, une amélioration des GANs appellerait des outils de comparaison plus précis et subtils pour rendre compte de l'apprentissage. Des méthodes plus pertinentes d'analyse de données extrêmes comme celles présentées dans [Ribereau et al. \(2008\)](#) et [Naveau et al. \(2014\)](#) devront alors être utilisées pour comparer les performances des GANs dans ce cadre des extrêmes.

## Références

- ALLOUCHE, M., GIRARD, S. & GOBET, E. 2021 EV-GAN : Simulation of extreme events with ReLU neural networks. Working paper or preprint.
- ARJOVSKY, M., CHINTALA, S. & BOTTOU, L. 2017 Wasserstein GAN.
- BHATIA, S., JAIN, A. & HOOI, B. 2021 ExGAN : Adversarial Generation of Extreme Samples.
- BOUCHER, O., SERVONNAT, J., ALBRIGHT, A. L., AUMONT, O., BALKANSKI, Y., BASTRIKOV, V., BEKKI, S., BONNET, R., BONY, S., BOPP, L., BRACONNOT, P., BROCKMANN, P., CADULE, P., CAUBEL, A., CHERUY, F., CODRON, F., COZIC, A., CUGNET, D., D'ANDREA, F., DAVINI, P., DE LAVERGNE, C., DENVIL, S., DESHAYES, J., DEVILLIERS, M., DUCHARNE, A., DUFRESNE, J.-L., DUPONT, E., ÉTHÉ, C., FAIRHEAD, L., FALLETTI, L., FLAVONI, S., FOJOLS, M.-A., GARDOLL, S., GASTINEAU, G., GHATTAS, J., GRANDPEIX, J.-Y., GUENET, B., GUEZ, LIONEL, E., GUILYARDI, E., GUIMBERTEAU, M., HAUGLUSTAINE, D., HOURDIN, F., IDELKADI, A., JOUSSAUME, S., KAGEYAMA, M., KHODRI, M., KRINNER, G., LEBAS, N., LEVAVASSEUR, G., LÉVY, C., LI, L., LOTT, F., LURTON, T., LUYSSAERT, S., MADEC, G., MADELEINE, J.-B., MAIGNAN, F., MARCHAND, M., MARTI, O., MELLUL, L., MEURDESOF, Y., MIGNOT, J., MUSAT, I., OTTLÉ, C., PEYLIN, P., PLANTON, Y., POLCHER, J., RIO, C., ROCHETIN, N., ROUSSET, C., SEPULCHRE, P., SIMA, A., SWINGEDOUW, D., THIÉBLEMONT, R., TRAORE, A. K., VANCOPPENOLLE, M., VIAL, J., VIALARD, J., VIOVY, N. & VUICHARD, N. 2020 Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems* **12** (7), e2019MS002010, e2019MS002010 10.1029/2019MS002010.
- CHEN, M., LIAO, W., ZHA, H. & ZHAO, T. 2020 Statistical Guarantees of Generative Adversarial Networks for Distribution Estimation.
- COLES, S. 2001 *An introduction to statistical modeling of extreme values*. London : Springer.
- EYRING, V., BONY, S., MEEHL, G. A., SENIOR, C. A., STEVENS, B., STOUFFER, R. J. & TAYLOR, K. E. 2016 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9** (5), 1937–1958.
- GNEDENKO, B. 1943 Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics* **44**, 423.
- GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. & BENGIO, Y. 2014 Generative Adversarial Networks.
- HEUSEL, M., RAMSAUER, H., UNTERTHINER, T., NESSLER, B. & HOCHREITER, S. 2018 GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.
- KINGMA, D. P. & WELLING, M. 2014 Auto-Encoding Variational Bayes.
- LURTON, T., BALKANSKI, Y., BASTRIKOV, V., BEKKI, S., BOPP, L., BRACONNOT, P., BROCKMANN, P., CADULE, P., CONTOUX, C., COZIC, A., CUGNET, D., DUFRESNE, J. L., ETHE, C., FOJOLS, M. A., GHATTAS, J., HAUGLUSTAINE, D., HU, R. M., KAGEYAMA, M., KHODRI, M., LEBAS, N., LEVAVASSEUR, G., MARCHAND, M., OTTLÉ, C., PEYLIN, P., SIMA, A., SZOPA, S., THIÉBLEMONT, R., VUICHARD, N. & BOUCHER, O. 2020 Implementation of the CMIP6 forcing data in the IPSL-CM6A-LR model. *Journal of Advances in Modeling Earth Systems* **12**, e2019MS001940 [22 ].
- MAKHZANI, A., SHLENS, J., JAITLEY, N., GOODFELLOW, I. & FREY, B. 2016 Adversarial Autoencoders.
- MANABE, S., HAHN, D. & HOLLOWAY, J. 1979 Climate simulations with GFDL spectral models of the atmosphere : Effects of spectral truncation. *GARP Pub. Ser.* **22**, 41–94.



- MANABE, S., SMAGORINSKY, J. & STRICKLER, R. F. 1965 Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review* **93** (12), 769–798.
- MIRZA, M. & OSINDERO, S. 2014 Conditional Generative Adversarial Nets.
- NAVEAU, P., GUILLOU, A. & RIETSCH, T. 2014 A non-parametric entropy-based approach to detect changes in climate extremes. *Journal of the Royal Statistical Society : Series B* **76** (5).
- RADFORD, A., METZ, L. & CHINTALA, S. 2016 Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.
- RIBEREAU, P., GUILLOU, A. & NAVEAU, P. 2008 Estimating return levels from maxima of non-stationary random sequences using the Generalized PWM method. *Nonlinear Processes in Geophysics* **15** (6), 1033–1039.
- SALIMANS, T., GOODFELLOW, I., ZAREMBA, W., CHEUNG, V., RADFORD, A. & CHEN, X. 2016 Improved Techniques for Training GANs.

# Annexes

## A Algorithmes d'entraînement des réseaux antagonistes génératifs

La méthode d'entraînement des GANs a été brièvement décrite dans la section 1.4.1, son implémentation est un peu plus détaillée ici. Notamment l'utilisation de certaines techniques présentées dans la section 2.3.2. Une version simplifiée de l'entraînement utilisé dans cette étude est présentée dans l'algorithme 1. Certaines parties de l'algorithme de Goodfellow et al. (2014) ont été modifiées, comme l'ajout d'un bruit blanc pour éviter le sur-apprentissage du discriminateur. Dans l'algorithme 1, le discriminateur et le générateur sont entraînés un même nombre de fois, cependant, le discriminateur étant entraîné avec deux fois plus d'exemple, un léger déséquilibre se crée en sa faveur. La méthode de Goodfellow et al. (2014) où le discriminateur était entraîné plusieurs fois sans entraîner le générateur convergerait moins bien. Pour la remplacer, deux taux d'apprentissage différents ont été attribués aux différents réseaux, règle issue de Heusel et al. (2018). Cela permet de mieux gérer les asymétries d'apprentissage entre le générateur et le discriminateur. Enfin, une décroissance du taux d'apprentissage a été choisie. Pour conserver de la cohérence entre les expériences, le nombre de paquets d'entraînement par étape,  $k$ , a été fixé. Cette décroissance a été choisie telle que le taux d'apprentissage final soit 100 fois inférieur au taux originel.

---

**Algorithme 1 :** Descente de gradient stochastique pour un GAN. La taille de paquet  $m$  a été choisi pour que le nombre d'étape par boucle d'entraînement,  $k$ , soit de l'ordre de 100. ( $N = km$  avec  $N$  le nombre total d'exemples). Inspiré de l'algorithme d'entraînement de Goodfellow et al. (2014).

---

**pour** le nombre d'étapes d'entraînement **faire**

    Mélanger les exemples du jeu de données

**pour**  $k$  paquets de données de taille  $n$  **faire**

        Prendre  $m$  exemples  $\{z_{(1)}, \dots, z_{(m)}\}$  issues de la distribution du bruit  $p_z(z)$

        Bruiter les exemples  $\{z_{(1)}, \dots, z_{(m)}\}$  avec un faible bruit blanc

        Prendre  $m$  exemples  $\{x_{(1)}, \dots, x_{(m)}\}$  des données suivant  $p_{\text{donnees}}(x)$

        Modification du discriminateur par le gradient stochastique

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x_{(i)}) + \log (1 - D(G(z_{(i)})))]$$

        Prendre  $n$  nouveaux exemples  $\{z_{(1)}, \dots, z_{(m)}\}$  issues de la distribution du bruit  $p_z(z)$

        Modification du discriminateur par le gradient stochastique

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z_{(i)})))$$

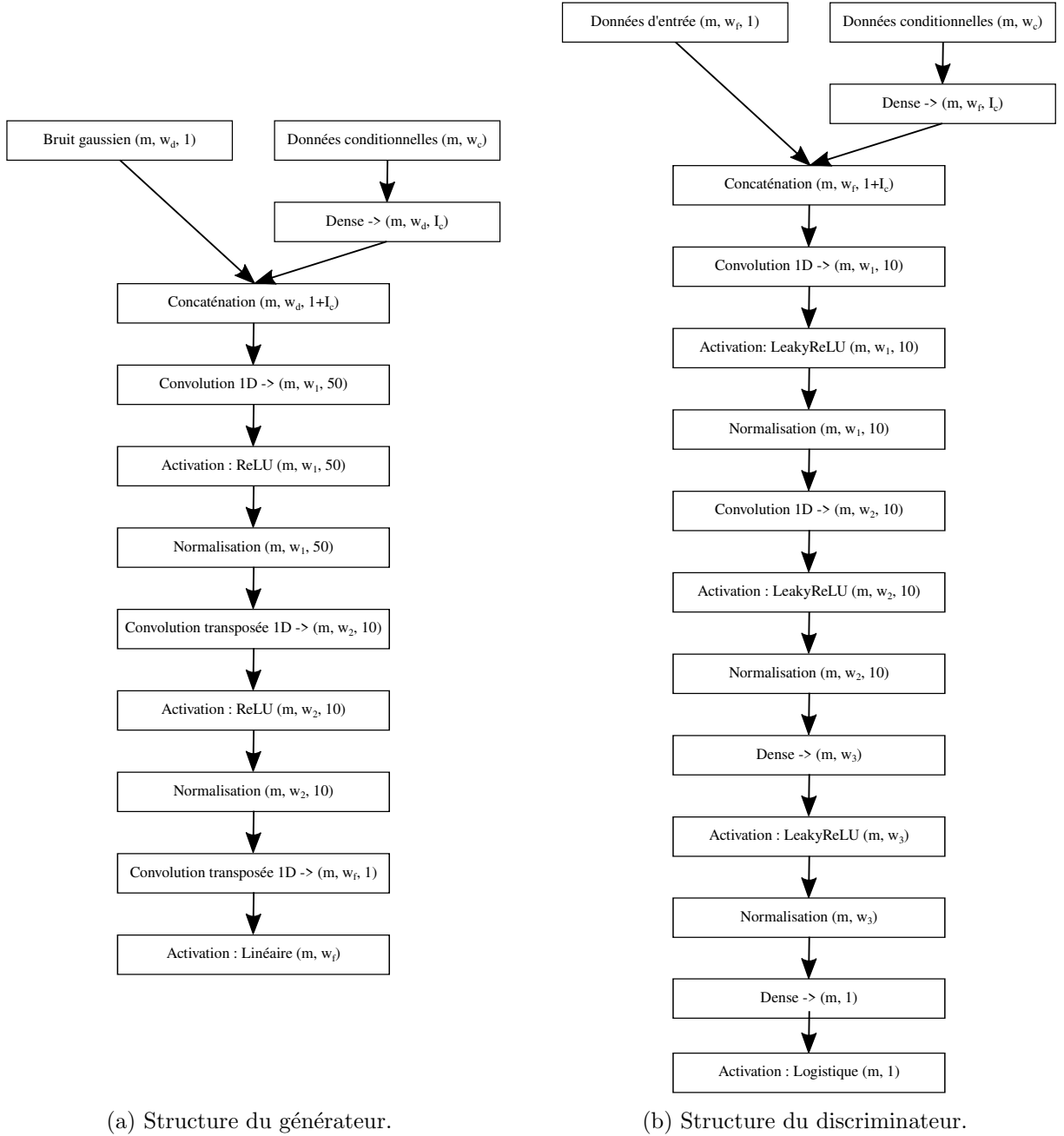
---

Ici  $\theta_d$ , respectivement  $\theta_g$ , représente l'ensemble des paramètres entraînables du discriminateur, respectivement du générateur. Cette apparente simplicité de notations cache une subtilité de la méthode d'entraînement. Le GAN est formé de la composition du générateur avec le discriminateur et l'apprentissage du générateur se fait par l'apprentissage du GAN complet. Dans cette structure, les paramètres du discriminateur,  $\theta_d$ , doivent donc être fixes pour que seul le générateur soit entraîné. Pour plus de détails sur la méthode d'entraînement des réseaux, le dépôt GitHub [Ext\\_c\\_dcgan](#) est disponible.

## B Structures des réseaux

Une ébauche de structure est présentée sur la figure 12. Elle présente l'agencement global des couches pour le générateur et le discriminateur. Certaines règles d'échelle pour les largeurs des couches

ont été adoptées. Tout d'abord, la largeur du bruit donnée par la couche de début  $w_d$  et celle des données d'entraînement la couche de fin  $w_f$  sont choisies égales pour permettre d'avoir une bijection entre l'espace latent et l'espace réel. Les différentes largeurs sont adaptées aux données d'entraînement. Ensuite, la hauteur des couches de convolution a été fixée pour les expériences présentées dans cette étude. Enfin, pour le discriminateur les couches de normalisation sont précédées de couches de "Dropout" pour éviter le sur-apprentissage.



**Fig. 12** – Schéma des structures des réseaux utilisés pour les résultats présentés.  $m$  représente la taille du paquet pour une étape d'entraînement,  $w_i$  représente les différentes largeurs des couches et  $I_c$  l'indicatrice de l'apprentissage conditionnel. La largeur de début  $w_d$  correspond à la dimension de l'espace latent du bruit. La largeur de fin  $w_f$  correspond à la dimension des données d'entraînement. La largeur  $w_c$  correspond au nombre de données conditionnelles. La fonction d'activation "ReLU" est la fonction  $x \mapsto \max(0, x)$ . La fonction d'activation "LeakyReLU" est la fonction  $x \mapsto \max(\epsilon x, x)$ , avec  $\epsilon = 0.2$  ici. La fonction d'activation logistique est la fonction  $x \mapsto \sigma(x)$  définie par l'équation (10).

Pour plus de détails sur la structure des réseaux, le dépôt GitHub [Ext\\_c\\_dcgan](#) est disponible.