

Rapport de Stage de fin d'études

*Prédiction de précipitations
à partir d'images infrarouges satellite
via Deep Learning Génératif*

Camille Francois-Martin

IMT ATLANTIQUE
Stage réalisé au LATMOS
Équipe Météo-Climat
Avril - Août 2025



Table des matières

Remerciements	2
Résumé	3
1 Introduction	4
1.1 Le LATMOS et l'équipe Météo-Climat	4
1.2 Contexte	5
1.3 Problématique	6
2 État de l'art	8
2.1 Point de départ	8
2.2 Diffusion	9
2.3 Conditionnement	11
3 Dataset et environnement de travail	12
3.1 Sources de données	12
3.1.1 Données satellites	12
3.1.2 Données radar	13
3.2 Construction du jeu de données	15
3.3 Ressources de calcul et environnement de développement	17
4 Modèle et principaux hyperparamètres	19
4.1 Prétraitement des données	19
4.2 Description de l'architecture du modèle	20
4.3 Algorithme de Diffusion	21
4.3.1 Scheduler	21
4.3.2 Algorithme	23
5 Résultats	26
5.1 Choix des métriques	26
5.2 Résultats obtenus	26
5.3 Comparaison des résultats	30
6 Analyse Critique et Réflexions Personnelles	34
6.1 Portée, limites et impacts du projet	34
6.2 Réflexion personnelle sur l'expérience	36
Conclusion	37
Bibliographie	38

Remerciements

Je souhaite tout d'abord remercier mes encadrants de stage, Cécile Mallet et Nicolas Viltard. Leur accompagnement bienveillant et notamment le temps précieux et l'attention qu'ils m'ont accordés ont permis de rendre ce stage particulièrement enrichissant et agréable. J'ai énormément appris au cours de ces cinq mois et ai découvert le monde de la recherche de la meilleure des manières grâce à vous.

Merci également à tous les membres de l'équipe précipitations notamment Laurent Barthès et Aymeric Chazottes avec qui j'ai pu discuter et qui m'ont aidé à maintes reprises sur les problèmes que je pouvais rencontrer.

Je souhaite aussi remercier Olivia, Nicolas, Nathan, Noah, Naël, Matthias et Mohammed pour les parties de ping-pong et d'échecs et tout ce qu'on a pu échanger lors des pauses déjeuner.

Enfin, un grand merci à Matthieu Meignin, qui m'a accompagné tout au long du stage et avec qui j'ai beaucoup échangé. Sa disponibilité, ses conseils et sa bonne humeur ont rendu cette expérience à la fois stimulante et plaisante. Travailler avec lui m'a vraiment aidé à progresser, à prendre du recul et à mieux structurer ma réflexion. J'espère t'avoir appris au moins autant au ping-pong.

Résumé

Il s'agit de mon rapport de stage de fin d'études concluant ma formation d'ingénieur à IMT Atlantique. J'ai débuté ce stage le 1er avril 2025 au sein du LATMOS (Laboratoire Atmosphère et Observation Spatiale). Il a duré jusqu'au 31 août 2025.

J'ai choisi ce cadre afin de m'initier au monde de la recherche tout en travaillant sur un sujet présentant des enjeux environnementaux. Pendant ce stage, j'ai eu l'opportunité de travailler sur des modèles de diffusion (generative deep learning) permettant de reconstituer des images de pluie à partir d'images satellites. On s'intéresse à l'impact du changement climatique sur les précipitations. Afin d'étudier celui-ci, il est nécessaire de disposer de données de précipitations sur l'entièreté de la planète. L'équipe cherche donc à développer un algorithme de restitution de la pluie à partir d'observations infrarouges obtenues par satellites géostationnaires qui couvrent la quasi-totalité de la surface du globe avec une fréquence élevée. Cependant, cette tâche est non triviale, car la relation entre les données infrarouges et les précipitations est indirecte, ce qui rend l'inversion délicate.

Ce travail m'a permis de découvrir et de mobiliser des approches modernes en intelligence artificielle, tout en contribuant à une problématique concrète ayant un impact sociétal réel. J'ai donc pu construire des modèles de diffusion, de type DDPM (Denoising Diffusion Probabilistic Model) en étudiant de manière rigoureuse les aspects théoriques me permettant d'effectuer mes choix de structure et des hyperparamètres. Ce travail a notamment permis de comparer les performances de ces modèles génératifs avec celles de réseaux convolutifs plus classiques, déjà utilisés par le laboratoire. Ce rapport présente le cadre scientifique du stage, les méthodes développées, les résultats obtenus, ainsi qu'une analyse critique de leur portée et de leurs limites.

Chapitre 1

Introduction

1.1 Le LATMOS et l'équipe Météo-Climat

Le LATMOS (Laboratoire Atmosphères et Observations Spatiales) est une unité mixte de recherche (UMR 8190) spécialisée dans l'étude des atmosphères terrestres et planétaires, en lien avec le CNRS¹, l'UVSQ², Sorbonne Université et le CNES³. Il développe des instruments scientifiques déployés sur Terre ou dans l'espace, ainsi que des modèles numériques pour analyser les observations. Créé en 2009, il regroupe environ 230 personnes sur les sites de Guyancourt et Paris, et fait partie de l'Institut Pierre Simon Laplace (IPSL).

J'ai intégré l'équipe "Météorologie et Climat" plus communément appelée "MétéoClimat" située sur le site de Guyancourt qui étudie les processus météorologiques à différentes échelles, en particulier dans les milieux polaires et urbains, pour mieux comprendre la formation des nuages, brouillards et précipitations, ainsi que leurs extrêmes. Elle développe et déploie des instruments de pointe (lidars, radars) dans le cadre de missions satellitaires comme EarthCARE (ESA-JAXA), C2OMODO (CNES) et WIVERN (ESA), trois missions dédiées à l'étude des nuages, des précipitations et de leur rôle dans le climat, combinant radar, lidar et radiomètres micro-ondes pour observer la structure, la dynamique et les échanges radiatifs de l'atmosphère.

L'équipe explore également l'usage de l'intelligence artificielle pour analyser la variabilité climatique et les précipitations en collaborant avec des projets internationaux comme CORDEX⁴ et INRIA ARCHES⁵. J'ai donc rejoint mes encadrants, Cécile Mallet et Nicolas Viltard dans la sous branche "précipitations" de l'équipe MétéoClimat, afin de travailler sur l'analyse des précipitations à l'aide de Deep Learning.

-
1. Centre National de la Recherche Scientifique
 2. Université Saint Quentin en Yvelines
 3. Centre national d'études spatiales.

4. CORDEX (Coordinated Regional Downscaling Experiment) est une initiative internationale qui coordonne la modélisation du climat régional afin de produire des projections détaillées pour la recherche, l'adaptation et l'évaluation des impacts climatiques.

5. ARCHES (AI Research for Climate Change and Environmental Sustainability) est une équipe-projet d'INRIA dédiée à la recherche en intelligence artificielle pour le climat et la durabilité environnementale, en collaboration avec Sorbonne Université, UVSQ et le CNRS via le LATMOS.

1.2 Contexte

La prévision précise des précipitations est un enjeu majeur pour la prévention des catastrophes naturelles telles que les inondations ou les sécheresses, qui affectent chaque année des millions de personnes. Bien que les radars météorologiques au sol soient une source fiable de mesure, leur couverture reste limitée (Voir Figure 1.1), laissant de vastes régions non observées, comme les océans, certaines zones des pays en développement dépourvues de radars fiables, ou encore les zones montagneuses où le relief complique la couverture radar. Observer les précipitations à l'échelle du globe est également un enjeu scientifique important pour améliorer la compréhension et la quantification des liens entre intensification du cycle de l'eau et changement climatique.

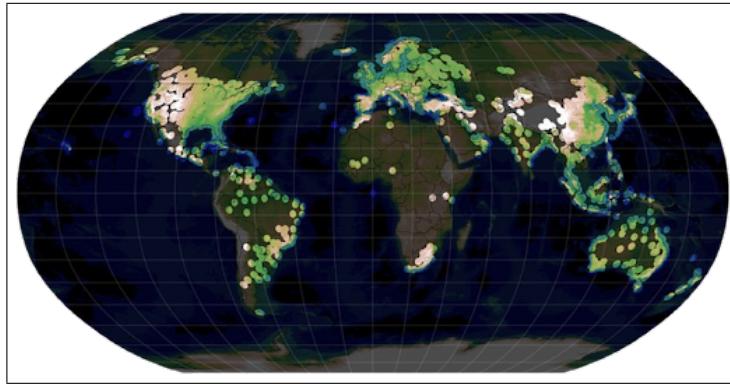


FIGURE 1.1 –
Carte de la couverture mondiale des radars météorologiques terrestres.
From Saltikoff et al. (2019) [1]

Afin de bénéficier d'une couverture spatiale étendue, le LATMOS s'est tourné vers l'exploitation des données issues de satellites météorologiques. Dans ce cadre, l'équipe "précipitations" a conçu plusieurs stratégies d'estimation des pluies, notamment à partir d'observations fournies par des satellites en orbite basse. Le projet DRAIN[2] s'inscrit dans cette démarche en s'appuyant sur les températures de brillance observées par les radiomètres micro-ondes passifs, combinées à des techniques d'apprentissage profond. Le modèle retenu, un réseau U-Net, est entraîné à l'aide de données conjointes issues du radiomètre GMI⁶ et du radar DPR⁷, tous deux à bord du satellite GPM⁸. En traitant les images globalement plutôt que de manière ponctuelle comme cela était fait dans les approches antérieures, le modèle atteint des performances comparables, voire supérieures, à celles d'un algorithme de référence GPROF, notamment sur les surfaces continentales et littorales.

La proximité des satellites utilisés par le projet Drain leur permet d'utiliser des micro-

6. GMI (GPM Microwave Imager) est un radiomètre micro-ondes passif qui mesure la température de brillance pour estimer les précipitations.

7. DPR (Dual-frequency Precipitation Radar) est un radar à double fréquence qui fournit des mesures actives de la structure verticale des précipitations.

8. Satellite GPM (Global Precipitation Measurement), une mission internationale dédiée à la mesure globale des précipitations.

ondes pour détecter la présence d'eau, donc de précipitations, avec une bonne précision spatiale et temporelle. Cependant, en raison de leur orbite, ces satellites ne peuvent observer un même point que brièvement une ou deux fois par jour, ce qui limite leur fréquence d'acquisition. De plus, leur faible altitude réduit la largeur de leur fauchée, et donc l'étendue couverte pour chaque image acquise. Les précipitations, qui varient beaucoup dans le temps, ne peuvent pas être quantifiées de manière fiable à partir d'observations aussi peu nombreuses. En effet, la pluie présente une forte intermittence temporelle : il pleut très rarement et, la plupart du temps, aucune précipitation n'est observée. Les épisodes pluvieux sont ponctuels et séparés par de longues périodes sèches, ce qui donne l'impression qu'« il ne pleut jamais » sur une période donnée. Cette rareté et cette discontinuité font que la pluie n'apparaît pas comme un phénomène régulier, mais plutôt comme une succession de courts événements isolés dans le temps.

1.3 Problématique

Le projet de ce stage vise donc à exploiter les satellites géostationnaires, qui permettent une observation continue de larges zones palliant ainsi aux principaux problèmes de couverture spatiale et temporelle des satellites à orbite basse. Cependant, les satellites géostationnaires ne disposent pas de capteurs capables de mesurer directement les précipitations, contrairement aux satellites qui utilisent des micro-ondes. En revanche, ils enregistrent des données infrarouges liées aux températures des sommets nuageux, une donnée d'autant plus corrélée à la présence de pluie que ces dernières sont intenses.

En effet, cette corrélation s'explique par la physique de la convection atmosphérique. Lorsqu'une masse d'air chaud et humide s'élève, elle se refroidit en altitude jusqu'à atteindre le point de condensation, formant des nuages. Si l'ascendance est suffisamment intense, le nuage continue de croître verticalement, formant des nuages convectifs profonds tels que les cumulonimbus, dont les sommets peuvent atteindre de hautes altitudes, où les températures chutent fortement (jusqu'à -60 °C). Ce refroidissement est un indicateur de la hauteur du nuage, et donc indirectement de la vigueur du mouvement ascendant et du potentiel de précipitations. Dans ces nuages, les gouttelettes d'eau et les cristaux de glace s'accumulent, fusionnent ou gèlent, ce qui favorise la formation de précipitations. Ainsi, une température très basse au sommet d'un nuage est souvent le signe d'un développement vertical important et d'un environnement propice à la pluie, voire à des orages. Voir Figure 1.2

Les températures des sommets des nuages constituent une source unique pour estimer la pluie, mais elles restent difficiles à interpréter en raison de la forte variabilité spatio-temporelle et de la complexité des processus physiques qui la régissent. En effet, la relation entre la température des sommets nuageux et la pluie est indirecte (voir Figure 1.3), influencée par de nombreux facteurs atmosphériques, ce qui rend difficile l'exploitation directe des données infrarouges. Pour en extraire le maximum d'information, il est nécessaire de recourir à des méthodes d'inversion capables de modéliser ces relations complexes.

Dans ce contexte, le deep learning s'est récemment imposé comme un outil particulièrement performant [3]. Grâce à sa capacité à apprendre des représentations non linéaires à partir de grandes quantités de données, il permet de modéliser les liens entre les images satellites infrarouges et les phénomènes météorologiques sous-jacents. Bien que l'informa-

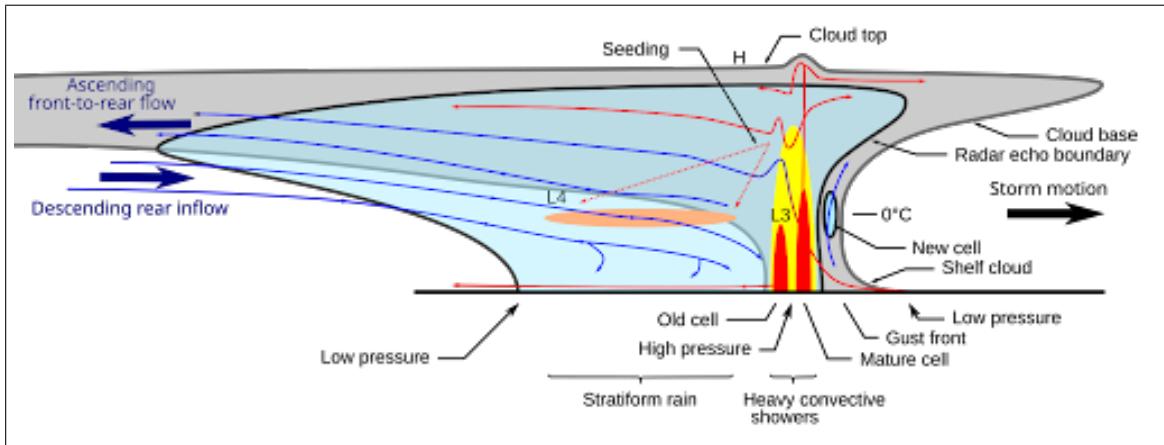


FIGURE 1.2 –
Modèle conceptuel d'un système convectif from Wikipedia : Rear-inflow Jet

tion issue de l'IR soit indirecte, elle contient des indices précieux sur la structure spatiale et l'évolution des nuages (forme, extension, texture, organisation convective) qui sont étroitement liés aux précipitations. De nombreuses architectures, notamment les réseaux convolutionnels (CNN) et les transformateurs (Transformers), exploitent efficacement ces caractéristiques pour estimer et prévoir les précipitations à partir d'observations satellites.

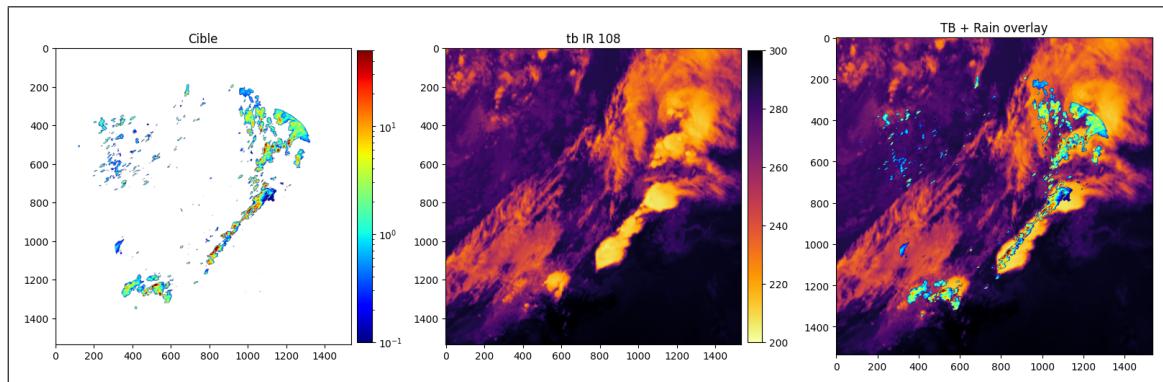


FIGURE 1.3 – Corrélation Température de Brillance et Taux de Pluie
Superposition d'un champ de pluie sur une Température de brillance

Le but de ce stage est de développer des modèles génératifs permettant de reconstituer la pluie à partir d'images de radiomètres IR embarquées à bord de satellites géostationnaires, puis de comparer leurs performances avec celles des modèles existants ou en développement. Pour cela, nous explorons notamment l'usage des modèles de diffusion (DDPM)[4], une classe récente de modèles d'apprentissage profond génératif qui s'est révélée très performante pour la génération d'images complexes, et dont le potentiel pour la modélisation des précipitations reste encore peu exploré.

Chapitre 2

État de l'art

2.1 Point de départ

En parallèle de ce stage, l'équipe Météo-Climat développe une approche basée sur des réseaux de neurones convolutionnels (CNN), tels que le U-Net, pour reconstituer les précipitations à partir d'images infrarouges.

Un réseau de neurones est un modèle d'intelligence artificielle inspiré du fonctionnement du cerveau humain. Il est constitué de couches de neurones artificiels interconnectés, capables d'apprendre à partir des données en ajustant des poids internes. L'objectif est de modéliser la relation entre les entrées et les sorties à travers des combinaisons linéaires des poids et des entrées, transformées ensuite par des fonctions d'activation.

Les réseaux de neurones convolutionnels utilisés ici (CNN) sont particulièrement bien adaptés au traitement d'image. Ils utilisent des filtres de convolution pour extraire automatiquement des caractéristiques locales (bords, textures, formes), et des couches de pooling pour réduire la dimensionnalité tout en conservant l'essentiel de l'information. Le U-Net est une architecture CNN symétrique composée d'un encodeur (qui compresse l'information et extrait le contexte global) et d'un décodeur (qui reconstruit l'image à la résolution d'origine). Sa particularité réside dans les connexions de type "skip", qui relient les couches de l'encodeur et du décodeur de même niveau (voir Figure 4.1). Cela permet d'extraire les caractéristiques de l'image à différentes échelles, ce qui est très utile pour la reconstitution de pluie composée de cellules et d'événements de taille variable.

Le U-Net prend en entrée une ou plusieurs images infrarouges représentant la température de brillance (TB) et, à l'aide de sa fonction de transfert (une combinaison linéaire de ses paramètres suivie d'une activation), produit en sortie une image estimée de la pluie. Lors de la phase d'entraînement, la sortie du réseau est comparée à la vérité terrain (précipitations réellement mesurées par les radars de Météo-France pour la même image TB). La différence entre la prédiction et la cible, mesurée par une fonction de coût (loss), permet d'ajuster les poids du réseau. Ce processus est répété sur un grand nombre d'epochs afin de réduire progressivement l'erreur et d'améliorer la qualité des prédictions.

Cette approche cherche à minimiser l'écart entre la pluie prédite et la pluie réelle à l'aide d'une fonction de perte, ici la Mean Squared Error (MSE), qui correspond à la moyenne des carrés des écarts entre chaque pixel. Cependant, les précipitations suivent une distribution

dite “à queue lourde” : la majorité des pixels d’une image contiennent peu ou pas de pluie. Le réseau, pour minimiser la MSE, a donc tendance à prédire de très faibles intensités de pluie. En moyenne, cela conduit les modèles classiques comme le U-Net à sous-prédirer les quantités de précipitations. De plus, la corrélation entre les températures de brillance (TB) et les précipitations est relativement faible, ce qui complique l’apprentissage de la localisation précise des pluies. Par prudence, et pour minimiser la perte, le réseau produit souvent des champs de pluie très lissés, qui manquent de réalisme et ne capturent pas bien la variabilité spatiale des événements pluvieux.

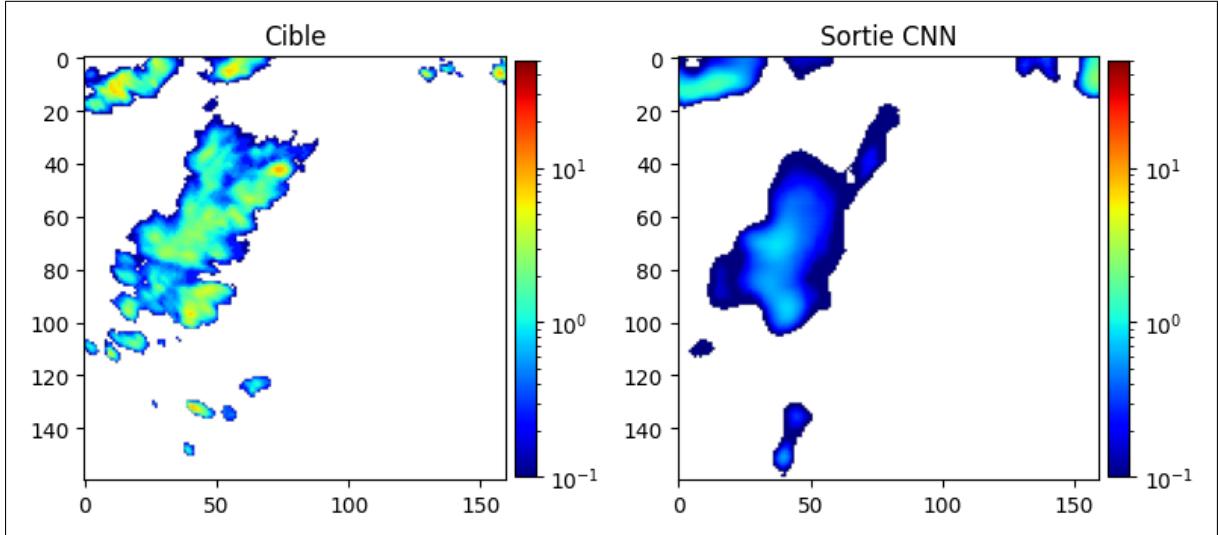


FIGURE 2.1 – Sortie du CNN
CNN par Matthieu Meignin

2.2 Diffusion

Pour surmonter les problèmes de sous-prédition et de lissage souvent rencontrés avec les modèles convolutionnels classiques, ce stage propose d’explorer l’utilisation de modèles génératifs. Ces derniers permettent de reconstruire des images similaires à celles du jeu d’apprentissage tout en conservant fidèlement la distribution de pluie présente dans les données d’entraînement. Dans ce cadre, nous nous intéressons particulièrement aux modèles de diffusion, et plus précisément aux Denoising Diffusion Probabilistic Models (DDPM).

Les DDPMs sont une classe de modèles génératifs probabilistes récemment popularisés pour leur capacité à produire des images de haute qualité. Les DDPMs ont été appliqués avec succès dans divers domaines : génération d’images réalistes [5], restauration d’images, super-résolution [6], et même modélisation de séquences ou synthèse vocale. Leur fonctionnement repose sur un processus à double sens : une phase directe, appelée diffusion, durant laquelle on corrompt progressivement les données en leur ajoutant du bruit gaussien, et une phase inverse, ou génération, où le modèle apprend à reconstituer les données originales en supprimant ce bruit ajouté lors de la phase directe de manière itérative (voir Figure 2.2).

Durant la phase de diffusion, on ajoute progressivement du bruit gaussien à une image issue du jeu de données d'entraînement x_0 , sur un grand nombre d'étapes T (dans notre cas 500), jusqu'à ce que l'image soit totalement bruitée et ne contienne plus d'information visuelle discernable. Ainsi ce processus à l'étape t est modélisé par une distribution $q(x_t | x_{t-1})$, une marche aléatoire gaussienne définie à chaque étape t .

La phase inverse, ou génération, consiste à estimer le bruit ajouté lors de la phase de diffusion à chaque étape à l'aide d'un réseau de neurones (ici encore un U-Net). En soustrayant ce bruit de manière non triviale à l'image d'entrée x_t on peut retrouver l'image précédente x_{t-1} . Ainsi, en partant d'un échantillon de bruit pur $x_T \sim \mathcal{N}(0, I)$, le modèle génère une image réaliste issue de $p_\theta(x_0)$, étape par étape, en l'échantillonnant à partir des distributions apprises. En combinant cette estimation du bruit avec une modélisation probabiliste, on échantillonne une image x_{t-1} à partir de la distribution $p_\theta(x_{t-1} | x_t)$ avec θ les paramètres du réseau de neurones.

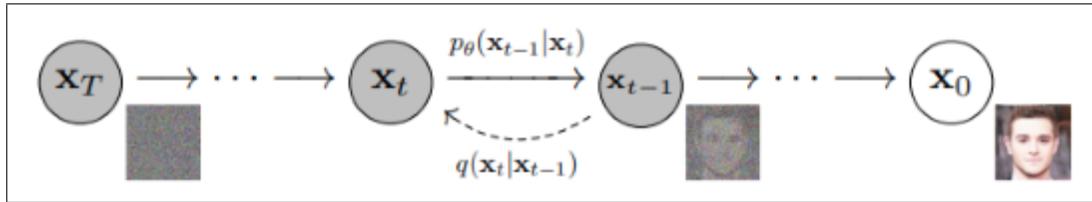


FIGURE 2.2 – Processus de Génération
from Denoising Diffusion Probabilistic Model [4]

Ici, la fonction de perte utilisée est l'écart quadratique moyen (MSE) entre le bruit prédict par le réseau ϵ_0 et le bruit réellement ajouté à l'image ϵ lors de la phase de diffusion (voir équation (2.1)).

$$\mathcal{L}_{\text{MSE}} = \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2] \quad (2.1)$$

Cette fonction de perte dépend donc de t car le réseau apprend différemment à prévoir le bruit ajouté en fonction du pas de temps (timestep) du processus de diffusion.

Ce processus d'échantillonnage permet de produire des images aléatoires suivant une distribution apprise $p_\theta(x_0)$, censée se rapprocher de la véritable distribution des données $p_{\text{data}}(x_0)$. En générant plusieurs exemples $x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(n)} \sim p_\theta(x_0)$, on peut estimer de manière pratique comment se répartissent les précipitations, et calculer des indicateurs comme la moyenne, la variabilité ou la probabilité d'un événement extrême.

Les DDPMs offrent des performances compétitives, voire supérieures, à d'autres modèles génératifs comme les GANs (Generative Adversarial Networks)[9], tout en étant plus stables à l'entraînement et exempts des problèmes de mode collapse¹.

En résumé, le caractère stochastique des DDPMs leur permet de ne pas se contenter d'une régression vers la moyenne, mais de reconstruire la diversité structurelle des champs de précipitations, en s'alignant mieux avec la distribution réelle observée dans les données.

1. problème où le générateur produit un ensemble limité de sorties très similaires

2.3 Conditionnement

Dans la continuité de l'approche précédemment décrite, nous nous intéressons à la diffusion conditionnée, une extension naturelle des DDPMs qui permet de guider la génération d'images à partir d'informations additionnelles.

Par exemple, dans un jeu de données contenant des images d'animaux, il est possible de conditionner la génération sur une classe (comme « chien » ou « chat »), ce qui permet au modèle d'apprendre, durant l'entraînement, à reconstruire des images en tenant compte de cette information. Ainsi, lors de l'inférence, on peut demander au modèle de générer une image correspondant à une classe spécifique.

Dans notre cas, la condition ne prend pas la forme d'une étiquette catégorielle, mais d'une image TB. L'objectif est alors de générer un champ de pluie réaliste correspondant à cette image de TB, en exploitant les relations statistiques entre les structures observées dans les données infrarouges et les précipitations associées. Le processus reste similaire à celui des DDPMs classiques : l'image cible (le champ de pluie) est progressivement bruitée au cours de la phase de diffusion, puis le modèle apprend à supprimer ce bruit de manière itérative pendant la phase de génération. La différence clé réside dans le fait que, durant chaque étape de cette phase inverse, le modèle reçoit en entrée à la fois l'image bruitée et l'image de TB conditionnelle. Cette information contextuelle est injectée dans le réseau de débruitage via une concaténation avec le champ de pluie bruité, permettant au modèle de moduler dynamiquement ses prédictions en fonction du contexte fourni.

Ainsi, au lieu d'estimer une distribution globale $p_\theta(x_0)$, le modèle apprend à approximer une distribution conditionnelle $p_\theta(x_0|y)$, où y désigne l'image de TB. Cette approche permet de produire des échantillons réalisistes compatibles avec l'observation fournie, tout en préservant la diversité stochastique propre aux modèles de diffusion. Appliquée à la prévision de la pluie, la diffusion conditionnée favorise la cohérence physique et spatiale des champs générés par rapport aux images de TB d'entrée, mais cette cohérence dépend de la capacité du modèle à capturer les relations entre TB et pluie.

Chapitre 3

Dataset et environnement de travail

3.1 Sources de données

Dans cette étude, deux sources de données complémentaires sont exploitées afin de construire une base cohérente pour la régression des précipitations : d'une part, les observations infrarouges (IR) issues de satellites géostationnaires, et d'autre part, les mesures radar fournies par le réseau opérationnel de Météo-France.

3.1.1 Données satellites

Les données satellites proviennent de capteurs embarqués sur les satellites géostationnaires *Meteosat*, qui offrent une couverture temporelle régulière avec une image toutes les 15 minutes d'une même région terrestre. Les observations utilisées dans cette étude couvrent cinq canaux infrarouges thermiques :

- IR087 : 8.7 µm
- IR097 : 9.7 µm
- IR108 : 10.8 µm
- IR120 : 12.0 µm
- IR134 : 13.4 µm

À ceux-ci s'ajoutent deux canaux dans l'infrarouge moyen (*Mid-Wave InfraRed, MWIR*), sensibles à la vapeur d'eau :

- MW062 : 6.2 µm
- MW073 : 7.3 µm

Ces différents canaux mesurent la température de brillance, qui correspond à une estimation de la température au sommet des nuages. En général, plus cette température est basse, plus le sommet nuageux est froid, ce qui est une caractéristique typique des nuages convectifs, souvent associés à des épisodes de précipitations.

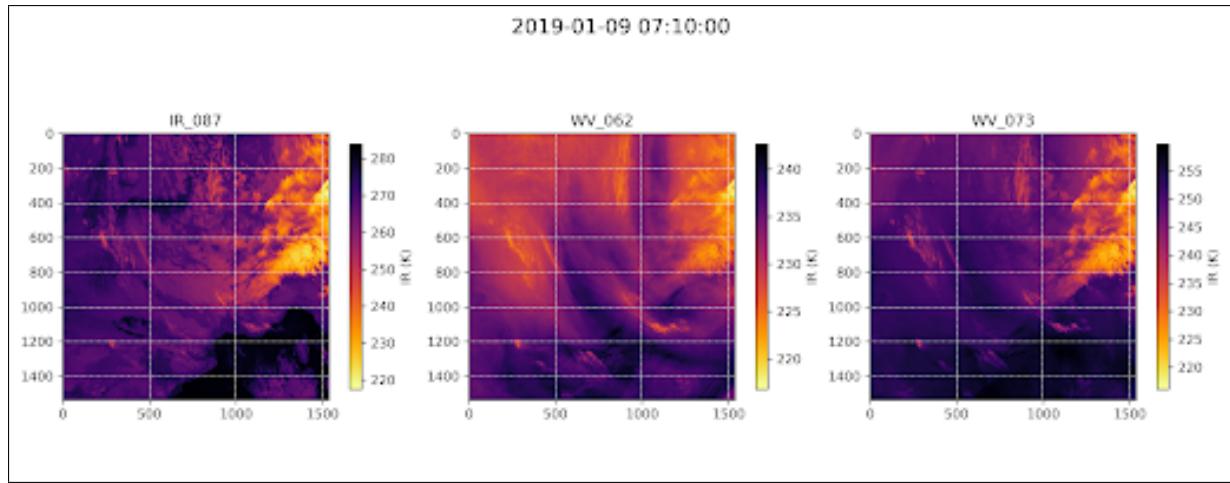


FIGURE 3.1 – Différents canaux d’images de TBs

Ainsi, les images IR fournissent une information indirecte sur le potentiel précipitant des structures nuageuses. Les canaux IR sont globalement corrélés entre eux mais présentent de légères différences dues à leur sensibilité spectrale spécifique, ce qui permet d’enrichir l’analyse atmosphérique.

3.1.2 Données radar

Les données radar proviennent du réseau de radars au sol de Météo-France, qui couvre l’ensemble du territoire national via une mosaïque de radars. Chaque radar fournit des mesures à haute résolution temporelle (toutes les 5 minutes) :

- Taux de précipitation (en mm/h)
- Indice de qualité de la mesure (en %)

La résolution spatiale des radars est de 1 km (1 pixel = 1 km), résolution avec laquelle nous travaillerons. Les images satellites ont quant à elles une résolution native de 3 km. Nous effectuons un suréchantillonnage (upsampling) des images satellites de 3 km à 1 km afin d’assurer la compatibilité avec la résolution des données radar.

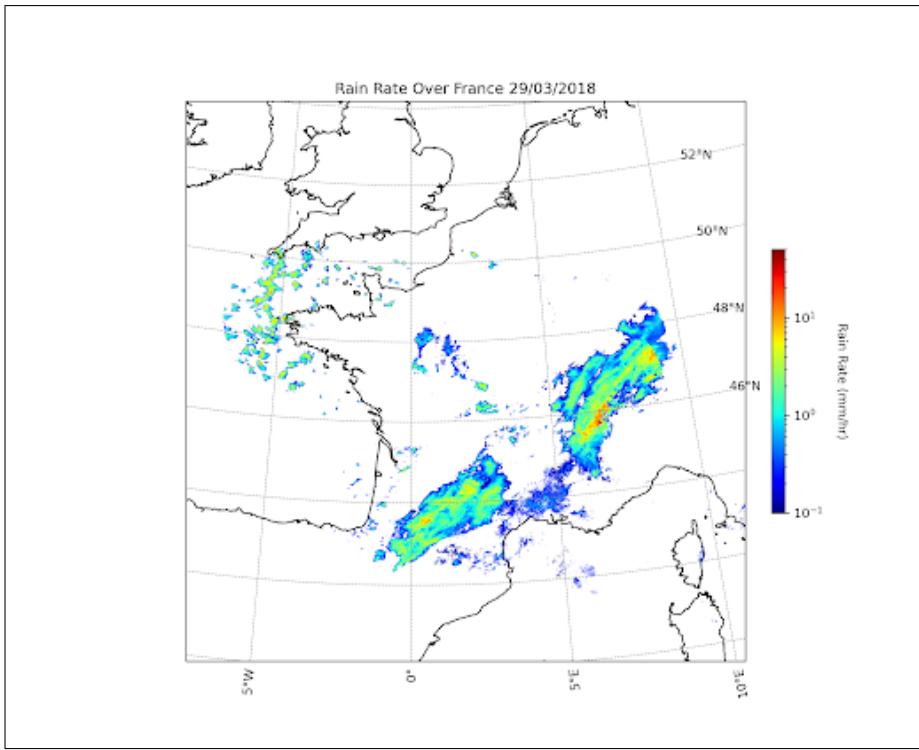


FIGURE 3.2 – Taux de pluie (mm/h) mesurée par les radars météo France

Nous avons également extrait l’indice de qualité radar sur une échelle de 0 à 100, permettant d’avoir une idée de la fiabilité des mesures au sol. La précision de ces mesures dépend fortement de la position géographique : la qualité diminue à mesure que l’on s’éloigne du centre du radar, et peut également être perturbée par des obstacles naturels comme les montagnes. Des cartes de la médiane de qualité illustrent clairement cette dégradation spatiale du signal radar.

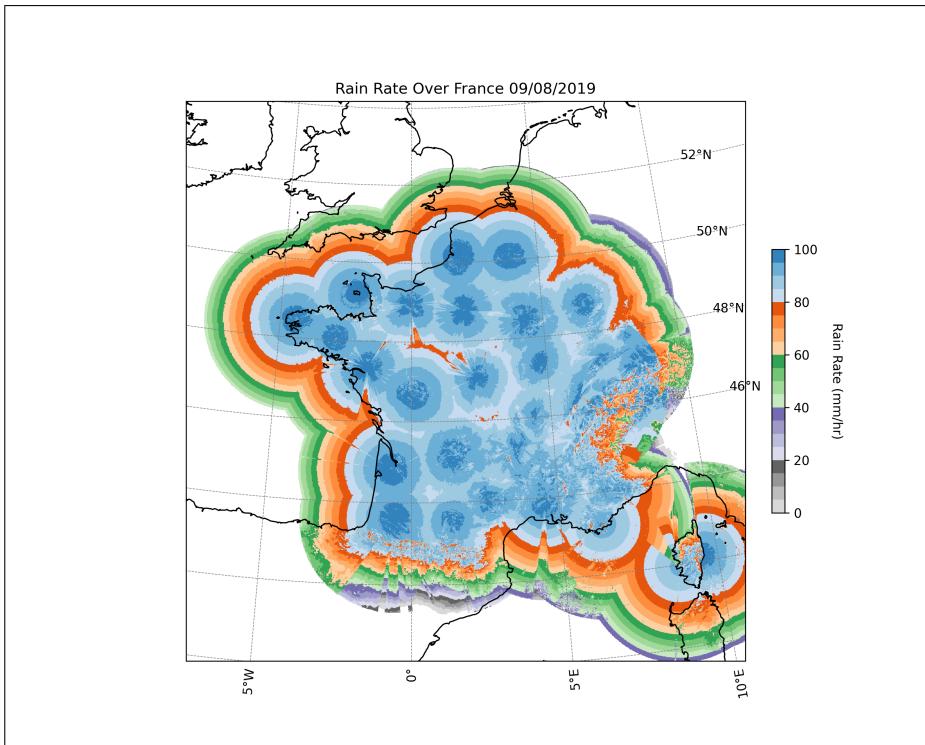


FIGURE 3.3 – Qualité des radars météo France

3.2 Construction du jeu de données

Lors de la constitution du jeu de données, chaque image de températures de brillance (TBs) a été associée à un champ de pluie provenant de la mosaïque radar de Météo-France. Les observations infrarouges sont espacées de 15 minutes, une contrainte imposée par la fréquence de passage des satellites, tandis que les radars de Météo-France disposent d'une fréquence d'acquisition plus élevée (5 minutes).

Les scènes initiales couvrent une surface d'environ $1500 \text{ km} \times 1500 \text{ km}$. Toutefois, ces images sont à la fois trop volumineuses pour être traitées directement par un réseau de neurones et majoritairement composées de zones sans pluie. Ces zones peu informatives limitent la capacité d'apprentissage du réseau, qui apprend rapidement qu'une absence de nuages correspond généralement à une absence de précipitations.

Pour pallier ce problème, des sous-images de taille 128×128 pixels ont été extraites. Plusieurs méthodes de sélection ont été testées, dont la méthode de clustering DBSCAN. L'approche retenue repose sur un découpage en grille associé à un décalage aléatoire, avec la contrainte qu'au moins 10 % des pixels de chaque extrait soient pluvieux. Cette méthode s'est révélée la plus efficace pour obtenir un nombre suffisant de sous-images pertinentes.

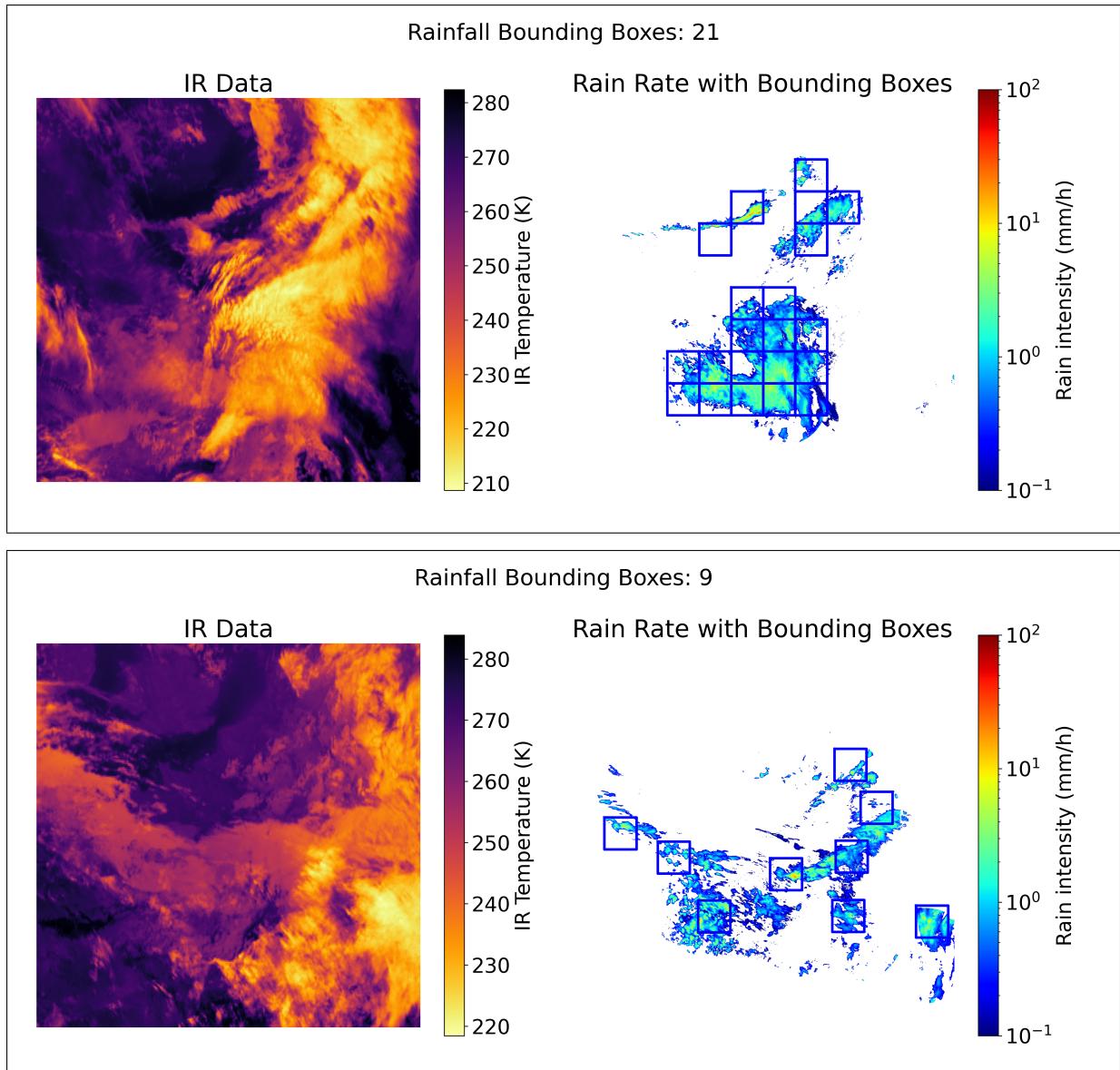


FIGURE 3.4 – Extraction Grille vs DBscan
extraction d’imagettes 128x128 par grille (haut) et DBscan (bas)

Par souci de qualité, tous les pixels dont la qualité radar est inférieure à 65 ont été supprimés afin d’éliminer les valeurs aberrantes. Toutefois, pour certaines applications (notamment la diffusion), il est nécessaire de disposer de cartes complètes. Les pixels manquants (non valides) ont alors été interpolés par la méthode des plus proches voisins, permettant d’obtenir des champs de pluie sans lacunes tout en maintenant une cohérence spatiale. Finalement, les cartes contenant plus de 5 % de pixels interpolés (qualité < 65) ont été écartées du jeu de données.

Au total, cette procédure a permis de constituer une base de données de 106 000 images, réparties comme suit :

- **Jeu d’entraînement** : 83 117 images, correspondant à l’ensemble des données restantes après la sélection des jeux de validation et de test.
- **Jeu de validation** : 20 105 images, correspondant aux images extraites des 2, 7,

- 14, 18, 25 et 30 de chaque mois.
- **Jeu de test** : 2 711 images, comprenant l'ensemble des scènes de l'année 2019, afin de couvrir toutes les saisons.

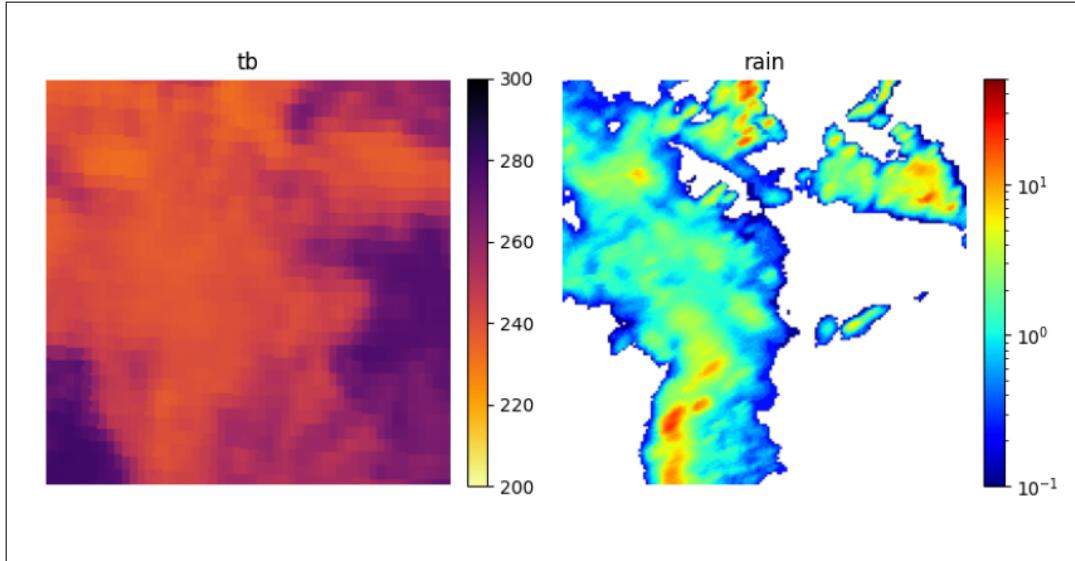


FIGURE 3.5 – Element du dataset
image 128x128 résolution 1pixel = 1km

Le seuil de détection des précipitations a été fixé à 0,1 mm/h, conformément au seuil de détection des radars de Météo-France. Ainsi, toutes les intensités inférieures ou égales à cette valeur ont été considérées comme non pluvieuses. Dans le jeu de données, les précipitations sont fortement déséquilibrées, la majorité des images ne contenant pas de pluie. Pour y remédier, lors de notre méthode de découpage en grille, nous avons filtré les images selon des seuils de présence et d'intensité de pluie.

3.3 Ressources de calcul et environnement de développement

L’Institut Pierre-Simon Laplace (IPSL) est un groupement de laboratoires de recherche franciliens spécialisés dans l’étude du climat, de l’environnement et du système Terre. Il rassemble des expertises complémentaires en modélisation, observation et analyse des processus climatiques à différentes échelles, du local au global. Le LATMOS (Laboratoire Atmosphères, Milieux, Observations Spatiales) est l’un des laboratoires affiliés à ce groupement.

Les phases d’entraînement des modèles ainsi que les calculs intensifs, en particulier ceux liés à l’inférence avec des modèles de diffusion, reconnus pour leur coût computationnel élevé, ont été réalisés sur le cluster de calcul **HAL**, mis à disposition par l’IPSL. Ce cluster est équipé d’une dizaine de GPU, couvrant une large gamme de générations :

- **NVIDIA RTX 2080 Ti** (2019), offrant une puissance de calcul de 14,2 TFLOPS, avec les caractéristiques suivantes :
 - 4352 coeurs CUDA,

- fréquence de base de 1350 MHz,
- 11 Go de mémoire GDDR6,
- bande passante mémoire de 616 Go/s.
- **NVIDIA RTX 5090** (2025), parmi les GPU les plus récents et performants disponibles, avec une puissance de calcul avoisinant les 90 TFLOPS, incluant :
 - 21 760 cœurs CUDA,
 - fréquence de base plus élevée (non spécifiée),
 - entre 24 et 32 Go de mémoire GDDR7 ultra-rapide,
 - bande passante mémoire supérieure à 1700 Go/s.

L'accès à ces ressources matérielles a permis d'accélérer de manière significative les phases d'apprentissage et d'inférence, en particulier pour les modèles avancés de type diffusion, particulièrement exigeants en termes de calcul. Le développement du code a été réalisé sous l'environnement **Visual Studio Code (VSCode)** en utilisant le langage **Python**, largement adopté dans le domaine du deep learning. En particulier, la bibliothèque **PyTorch** a été utilisée pour la définition, l'entraînement et l'inférence des modèles neuronaux.

Chapitre 4

Modèle et principaux hyperparamètres

4.1 Prétraitement des données

Dans notre étude, les données de pluie présentent une distribution à queue lourde, caractérisée par une très forte proportion de faibles intensités et quelques valeurs extrêmes. Le choix de la normalisation a donc une importance cruciale, car cette dissymétrie statistique complique l'apprentissage : sans normalisation adaptée, nous avons observé que le modèle tend à ignorer ou à sous-estimer les épisodes de forte pluie. Les modèles de diffusion sont en effet sensibles à la dynamique des écarts et des gradients dans les données. Une mauvaise calibration de l'échelle peut conduire à une génération trop lissée ou à une incapacité à reproduire correctement les événements rares.

La normalisation poursuit ici un double objectif :

- Réduire la dynamique entre les pixels afin d'équilibrer la contribution des différents régimes de pluie (faible, modérée, forte) durant l'apprentissage.
- Stabiliser l'échelle des entrées et des sorties.

Plusieurs méthodes ont été évaluées :

- **Min-Max** : écartée car elle écrasait excessivement les faibles valeurs, nombreuses mais peu informatives.
- **Logarithmique** ($\log(1 + p)$) : abandonnée en raison d'instabilités numériques lors de la transformation inverse (exponentielle), produisant parfois des valeurs irréalistes.
- **standardisation** (centrage-réduction). Elle ramène la majorité des valeurs dans l'intervalle $[-1, 1]$ tout en conservant la forme globale de la distribution.

Finalement, c'est cette dernière qui a été retenue car elle permet également d'aligner les données sur la distribution du bruit injecté lors du processus de diffusion, supposé suivre une loi $\mathcal{N}(0, I)$, facilitant ainsi l'apprentissage.

Les images de température de brillance ont été standardisées canal par canal selon la même logique, afin de garantir des valeurs compatibles avec celles des champs de pluie et une concaténation efficace des deux sources d'informations.

4.2 Description de l'architecture du modèle

Structure générale

Le U-Net est structuré de manière symétrique :

- **Encodeur (downsampling)** : réduit la taille spatiale tout en augmentant la profondeur via des convolutions successives.
- **Bottleneck** : capture les motifs globaux de l'image à l'aide de blocs ResNet et de modules d'attention.
- **Decodeur (upsampling)** : reconstruit progressivement l'image cible, en réinjectant les détails via des *skip connections*.

Chaque étage est enrichi de blocs ResNet intégrant :

- Des **temporal position embeddings** (sinusoïdaux), injectant l'information temporelle liée au niveau de bruit,
- Un **module d'attention**, améliorant la modélisation des dépendances spatiales à longue portée, crucial pour reconstruire des structures pluvieuses étendues.

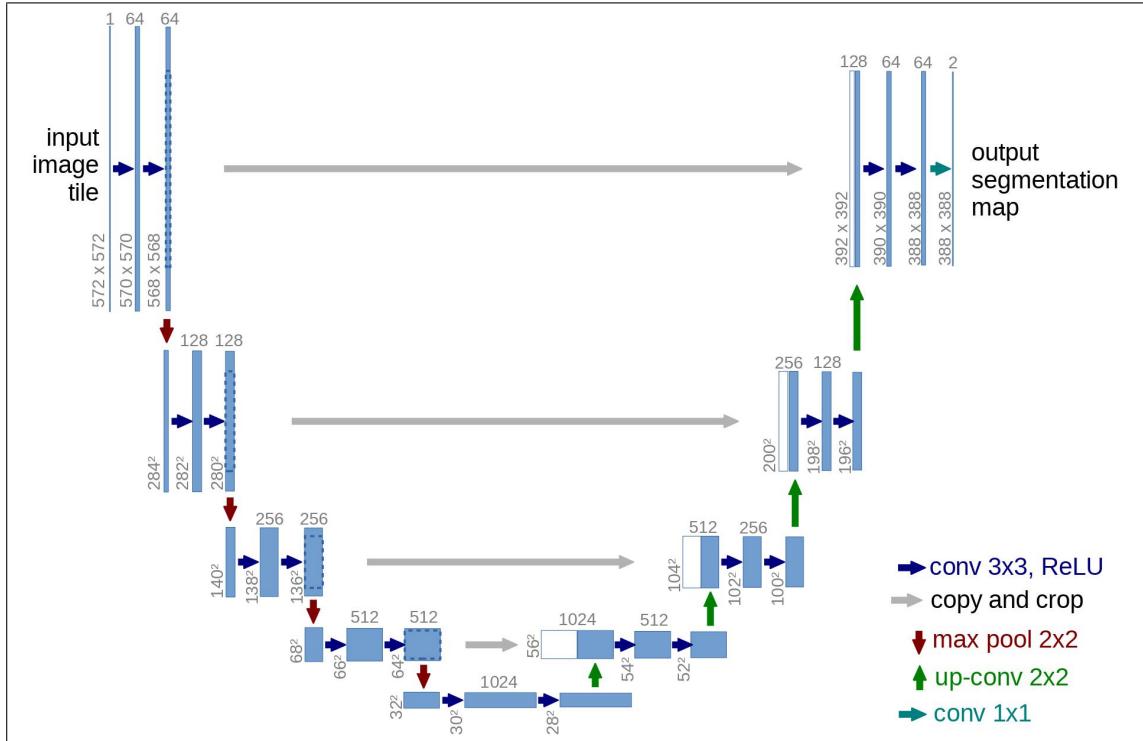


FIGURE 4.1 – Architecture d'un Unet
Ronneberger et al. 2015

Fonctionnement du réseau

L'encodeur applique des convolutions avec stride pour capturer des motifs de plus en plus abstraits. À chaque étage, les blocs ResNet sont suivis d'un module d'attention, puis d'une réduction de taille spatiale. Le goulot central agit comme une couche de synthèse. Il contient deux blocs ResNet et un module d'attention, permettant de modéliser des structures globales et complexes. Le décodeur effectue l'opération inverse : un upsampling suivi de blocs ResNet et d'un module d'attention. Les skip connections permettent de préserver

et de réinjecter des informations spatiales fines perdues à l'encodage, essentielles pour la précision des structures de pluie. Une dernière convolution terminale produit une estimation du bruit à soustraire à l'image bruitée. Le modèle prédit donc une version débruitée à chaque étape du processus de génération.

Le réseau totalise environ **9,9 millions de paramètres** pour une taille mémoire estimée à **39,6 Mo**. Sa structure détaillée est la suivante :

- Couche de convolution initiale : **640** paramètres,
- MLP temporel (embeddings) : **82,4k**,
- Blocs d'encodage (Downs) : **1,6M**,
- Goulot central :
 - 2 blocs ResNet : **1,3M** chacun,
 - 1 bloc d'attention résiduel : **131k**,
- Blocs de décodage (Ups) : **5,3M**,
- Bloc ResNet final : **152k**,
- Convolution de sortie : **65** paramètres.

Cette architecture reste relativement compacte tout en étant suffisamment expressive pour modéliser la complexité des champs de pluie. Un modèle plus large aurait pu être envisagé, notamment pour traiter des images de plus grande taille, mais la mémoire des GPUs disponibles limite cette possibilité.

Avantages pour notre tâche

Le U-Net conditionnel utilisé dans notre approche s'est révélé particulièrement adapté à la génération de champs de pluie :

- Il capture efficacement les structures fines et localisées (noyaux convectifs),
- Il reconstruit également les structures étendues (zones frontales, stratiformes),
- Il peut être guidé par des observations auxiliaires, comme les images IR, grâce à son conditionnement explicite.

4.3 Algorithme de Diffusion

4.3.1 Scheduler

Un des éléments fondamentaux dans le fonctionnement d'un modèle de diffusion est la manière dont l'image est bruitée au cours du processus direct de diffusion [7]. À chaque étape $t \in [0, T]$ (avec $T = 500$ dans notre cas), un bruit gaussien $\epsilon \sim \mathcal{N}(0, I)$ est ajouté à l'image x_t de la manière suivante, pour les modèles de type DDPM :

$$x_t = \sqrt{\beta_t} \cdot x_{t-1} + \sqrt{\alpha_t} \cdot \epsilon, \quad (4.1)$$

où $\alpha_t = 1 - \beta_t$.

Par simplification, et afin de faciliter les calculs lors de l'entraînement, on utilise l'approximation suivante :

$$x_t = \sqrt{\bar{\beta}_t} \cdot x_0 + \sqrt{\bar{\alpha}_t} \cdot \epsilon, \quad (4.2)$$

où $\bar{\beta}_t$ et $\bar{\alpha}_t$ représentent les produits cumulés des coefficients β_t et α_t respectivement jusqu'à l'instant t :

$$\bar{\beta}_t = \prod_{s=1}^t \beta_s, \quad \bar{\alpha}_t = \prod_{s=1}^t \alpha_s.$$

Le choix de la séquence des β_t (le *scheduler*) a une influence directe sur la dynamique du bruit injecté au fil du temps, et donc sur la performance du modèle. Différents types de schedulers existent :

- **Scheduler linéaire** :

$$\beta_t = \beta_{\text{start}} + \frac{t}{T-1} \cdot (\beta_{\text{end}} - \beta_{\text{start}})$$

C'est le scheduler le plus couramment utilisé dans la littérature, apprécié pour sa simplicité, sa stabilité, et sa capacité à produire des résultats de qualité sans réglages fins.

- **Scheduler quadratique** :

$$\beta_t = \left(\frac{t}{T} \right)^2 \cdot (\beta_{\text{max}} - \beta_{\text{min}}) + \beta_{\text{min}}$$

Ce scheduler augmente le bruit plus rapidement dans les étapes avancées, ce qui peut être utile pour certaines tâches de génération synthétique.

- **Scheduler sigmoïde** :

$$\beta_t = \text{sigmoid} \left(\gamma \cdot \left(\frac{t}{T} - 0.5 \right) \right) \cdot (\beta_{\text{max}} - \beta_{\text{min}}) + \beta_{\text{min}}$$

Il applique une croissance douce du bruit en début et fin de processus, avec une transition plus marquée au centre, contrôlée par le facteur γ .

- **Scheduler cosinus** (*cosine scheduler*) :

$$\bar{\alpha}_t = \frac{\cos^2 \left(\frac{1+s \cdot \pi}{2} \right)}{\cos^2 \left(\frac{t/T+s}{1+s} \cdot \frac{\pi}{2} \right)} \quad \text{pour } t \in [0, T],$$

avec s un paramètre de régularisation, souvent fixé à 0.008 dans la pratique.

Dans notre étude, nous avons opté pour le cosine [8]. Ce choix se justifie par plusieurs avantages spécifiques au contexte de génération de champs de pluie :

- Il permet de préserver davantage l'information structurelle dans les premières étapes du débruitage, ce qui est crucial pour maintenir les détails spatiaux fins, comme les cellules convectives.
- En ralentissant la décroissance de $\bar{\alpha}_t$ au début du processus, il limite la perte d'information essentielle et assure une meilleure stabilité de génération.

- Contrairement au scheduler linéaire, il introduit le bruit de manière plus progressive, ce qui permet au réseau d'apprendre des reconstructions plus cohérentes, même sur des structures complexes comme les zones frontales ou les noyaux de pluie.

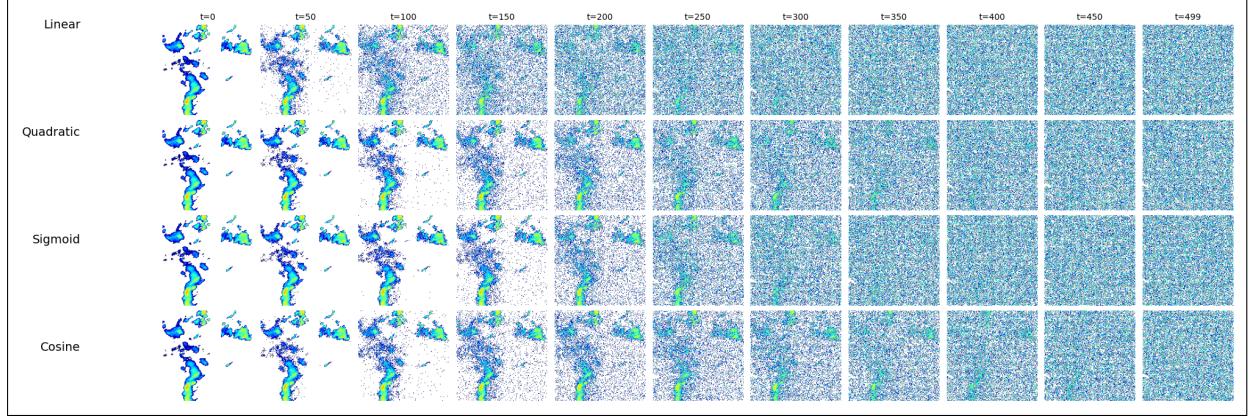


FIGURE 4.2 – Différents schedulers

4.3.2 Algorithme

Soit $\mathbf{x}_0 \sim p_{\text{data}}(\mathbf{x}_0)$ une image issue de la distribution réelle des données (ici x_0 représente un champ de pluie). Et la suite de coefficients $\{\beta_t\}_{t=1}^T$ le scheduler.. En utilisant la formule fermée, on peut échantillonner \mathbf{x}_t directement à partir de \mathbf{x}_0 :

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

L'objectif du réseau de prédiction $\boldsymbol{\varepsilon}_\theta$ est d'estimer le bruit $\boldsymbol{\varepsilon}$ à partir de \mathbf{x}_t et t . Pour cela, on minimise une perte au sens des moindres carrés :

$$\mathcal{L}_{\text{MSE}}(\theta) = \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\varepsilon}} [\|\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t)\|_2^2].$$

Dans le cas d'un modèle conditionnel, on ajoute une variable \mathbf{y} (ici nos Températures de Brillance) et la perte devient :

$$\mathcal{L}_{\text{MSE cond}}(\theta) = \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\varepsilon}} [\|\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t \mid \mathbf{y})\|_2^2].$$

Algorithme d'entraînement :

1. Échantillonner une image réelle $\mathbf{x}_0 \sim p_{\text{data}}(\mathbf{x}_0)$.
2. Tirer un pas de temps $t \sim \mathcal{U}(\{1, \dots, T\})$.
3. Échantillonner un bruit gaussien $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
4. Construire l'image bruitée :

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\varepsilon}.$$

5. Prédire le bruit avec le réseau :

$$\boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t \mid \mathbf{y})$$

6. Calculer la perte MSE :

$$\mathcal{L} = \|\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}\|_2^2$$

7. Mettre à jour les paramètres θ par descente de gradient (Backpropagation).

Ce schéma d'entraînement est conçu pour apprendre au modèle un processus de débruitage progressif à partir d'échantillons corrompus. Lors de l'inférence, la génération de nouvelles données s'effectue en échantillonnant un bruit initial $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ et en appliquant itérativement l'inverse du processus de diffusion.

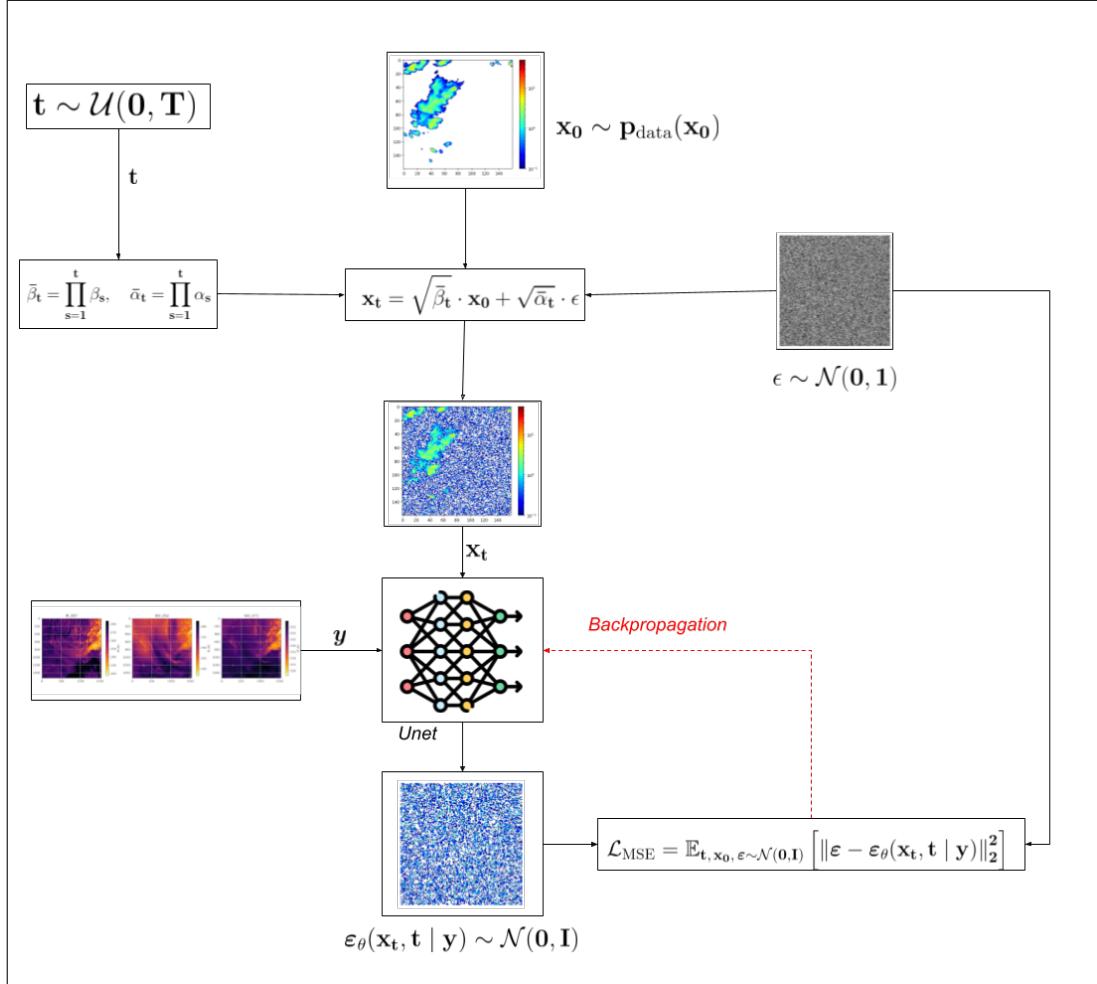


FIGURE 4.3 – Algorithme d’Entrainement

Algorithme d’inférence : À l’inférence, le modèle génère de nouvelles images à partir d’un échantillon de bruit Gaussien.

$$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

puis applique successivement l’inverse du processus de diffusion pour reconstruire progressivement une image propre. Pour chaque pas de temps décroissant $t = T, T-1, \dots, 1$, on calcule :

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\varepsilon}_\theta(\mathbf{x}_t, t | \mathbf{y}) \right) + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}', \quad \boldsymbol{\epsilon}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

où $\varepsilon_\theta(\mathbf{x}_t, t \mid \mathbf{y})$ est le bruit prédict par le réseau conditionné sur \mathbf{y} , et $\sqrt{1 - \alpha_t}$ est le niveau de bruit ajouté à l'étape t . L'ajout de bruit gaussien $\sqrt{1 - \alpha_t}\epsilon'$ à chaque étape permet de préserver la diversité des échantillons générés et d'assurer une exploration correcte de la distribution conditionnelle, évitant ainsi que le processus ne devienne déterministe trop tôt.

La variable aléatoire ϵ' est nulle pour $t = 1$, garantissant une sortie déterministe et nette. Ce processus permet de passer d'un bruit initial à une image synthétique cohérente et conditionnée sur \mathbf{y} .

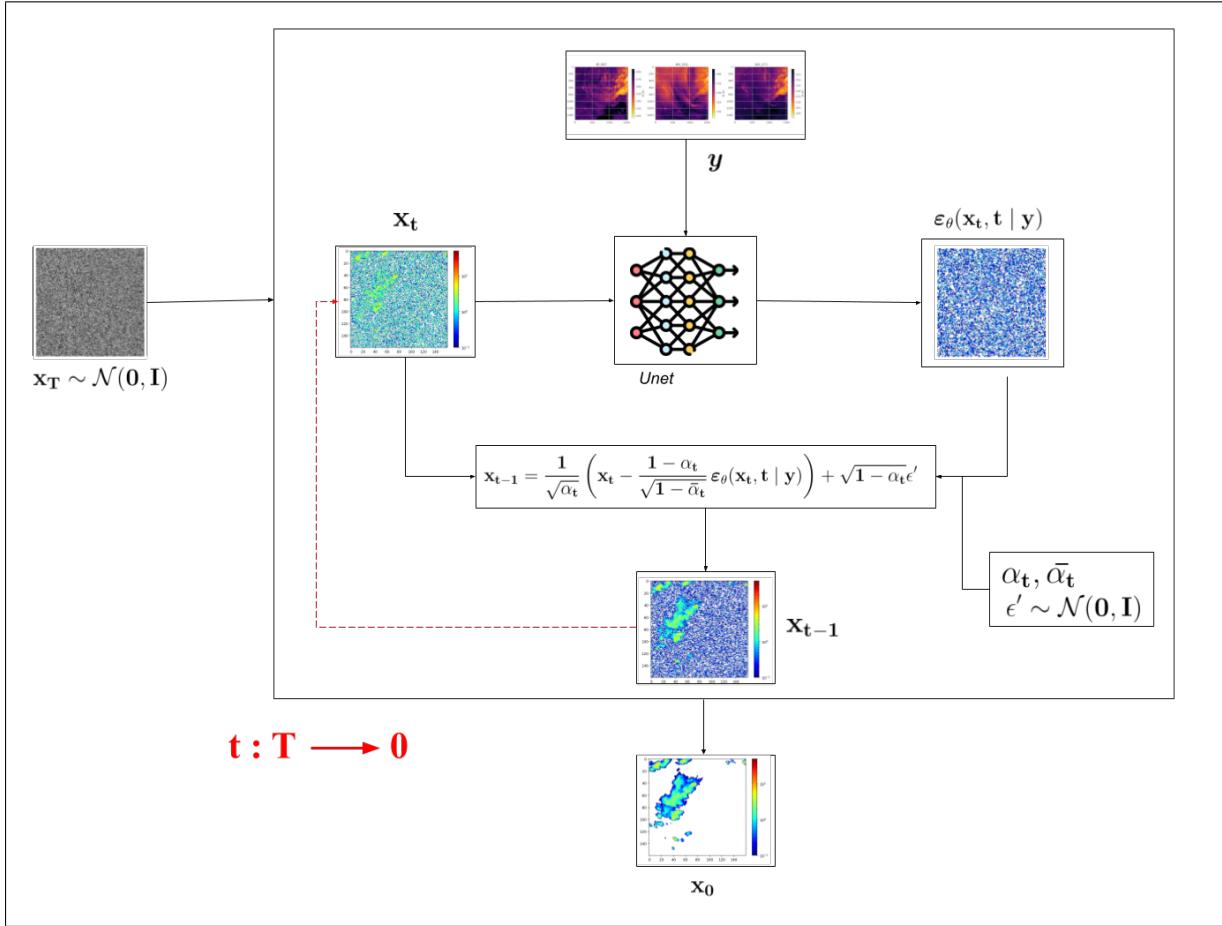


FIGURE 4.4 – Algorithme d’Inférence

Chapitre 5

Résultats

5.1 Choix des métriques

Pour évaluer la qualité des champs de pluie restitués, plusieurs métriques ont été mobilisées afin de couvrir à la fois l'aspect quantitatif et spatial des précipitations. La RMSE et la MAE mesurent l'écart pixel par pixel entre champs prédits et observés : la RMSE, plus sensible aux fortes erreurs, est adaptée aux pluies intenses, tandis que la MAE complète l'analyse sur les pluies faibles à modérées. Pour évaluer la détection des zones pluvieuses ($> 0,1 \text{ mm/h}$), des métriques binaires sont utilisées : le CSI pour les événements rares, le F1-score pour le compromis précision/rappel, le recall pour mesurer la couverture des pluies réelles, la précision pour limiter les fausses alertes, et l'accuracy pour une vision globale malgré le déséquilibre des classes.

En complément de ces scores calculés sur l'image entière, nous allons analyser les corrélations tranche de pluie par tranche de pluie (non-pluie, faible, moyenne, forte, violente). Cela permet d'identifier les forces et faiblesses du modèle selon l'intensité de la précipitation. Par exemple, que le modèle surestime les faibles pluies mais échoue à capturer les fortes intensités. Enfin, l'analyse de la distribution des intensités de pluie générées par le modèle comparée à celle des données cibles permet de vérifier si le modèle respecte les propriétés statistiques du champ de pluie, notamment la fréquence, l'étalement et la proportion des différents régimes. Cette évaluation statistique est cruciale pour s'assurer que le modèle ne présente pas de biais systématique (par exemple, en générant trop de pluie faible ou en sous-estimant les extrêmes).

5.2 Résultats obtenus

Pour l'entraînement du réseau de neurones, plusieurs paramètres ont été définis afin d'optimiser les performances et de stabiliser la convergence.

Hyperparamètre	Valeur
Nombre d'epochs total	460
Temps d'entraînement	4 jours
Batch size	16
Learning rate	1×10^{-4} (ajusté à 5×10^{-5} à 250 epochs)
Optimiseur	Adam
Fonction de perte	MSE
Métrique d'arrêt	RMSE \times (1 - CSI) (minimisation)
Epoch d'arrêt	399
Canaux	IR087, IR108, IR120, WV062, WV073, IR134, IR097, CNN output
Taille de l'image	128x128
Data augmentation	Random Cropping
Scheduler	Cosine

TABLE 5.1 – Résumé des hyperparamètres du réseau de neurones

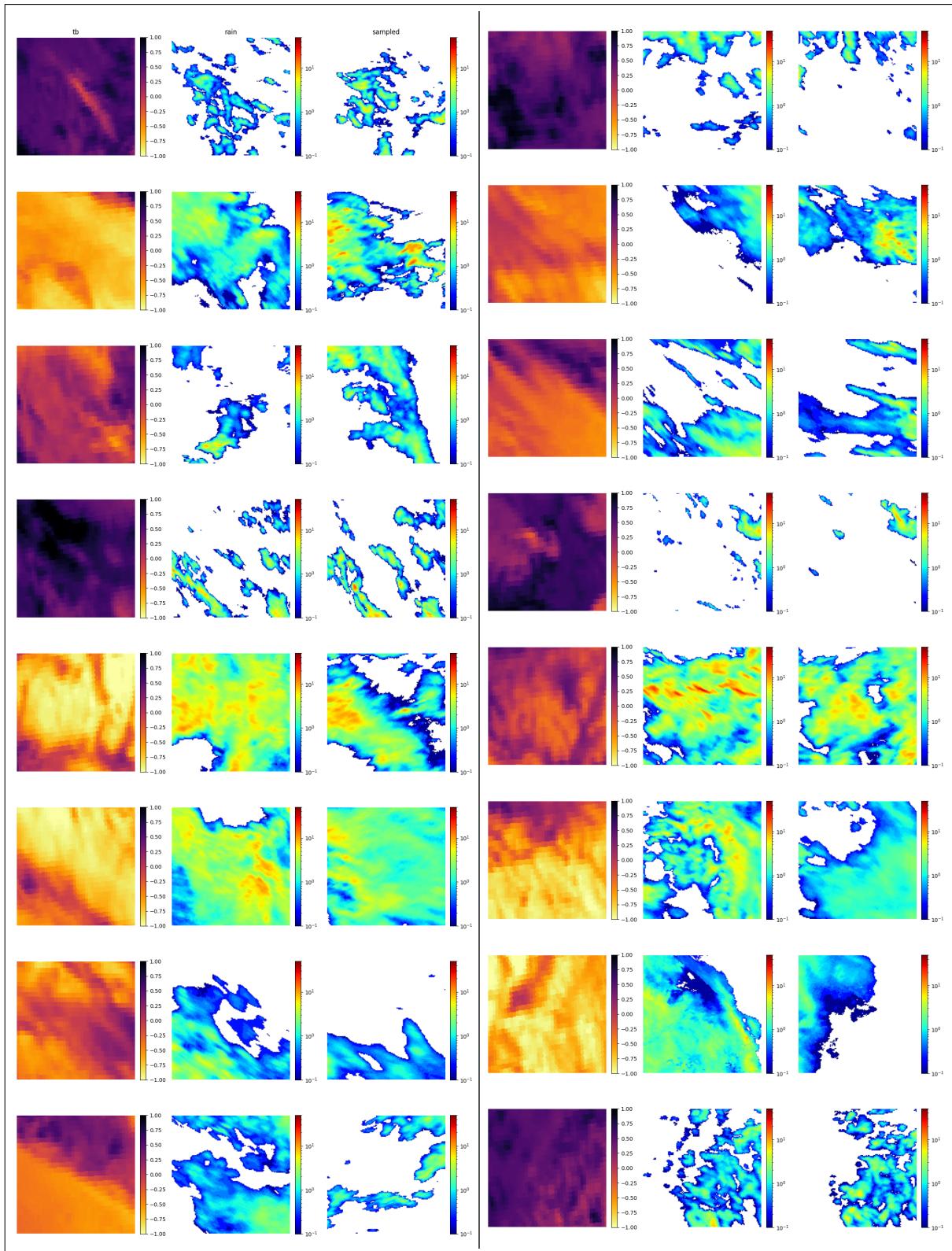


FIGURE 5.1 – Images générées tirées aléatoirement
De gauche à droite : TB, cible, sortie générée

Nos résultats de la Figure 5.1 nous montrent des images de pluie réaliste, il est difficile de discerner à l'œil nu la différence entre les images obtenues et les images de vérité terrain. Des valeurs de forte pluie sont retrouvées et la distribution globale est respectée comme

vue dans la Figure 5.2.

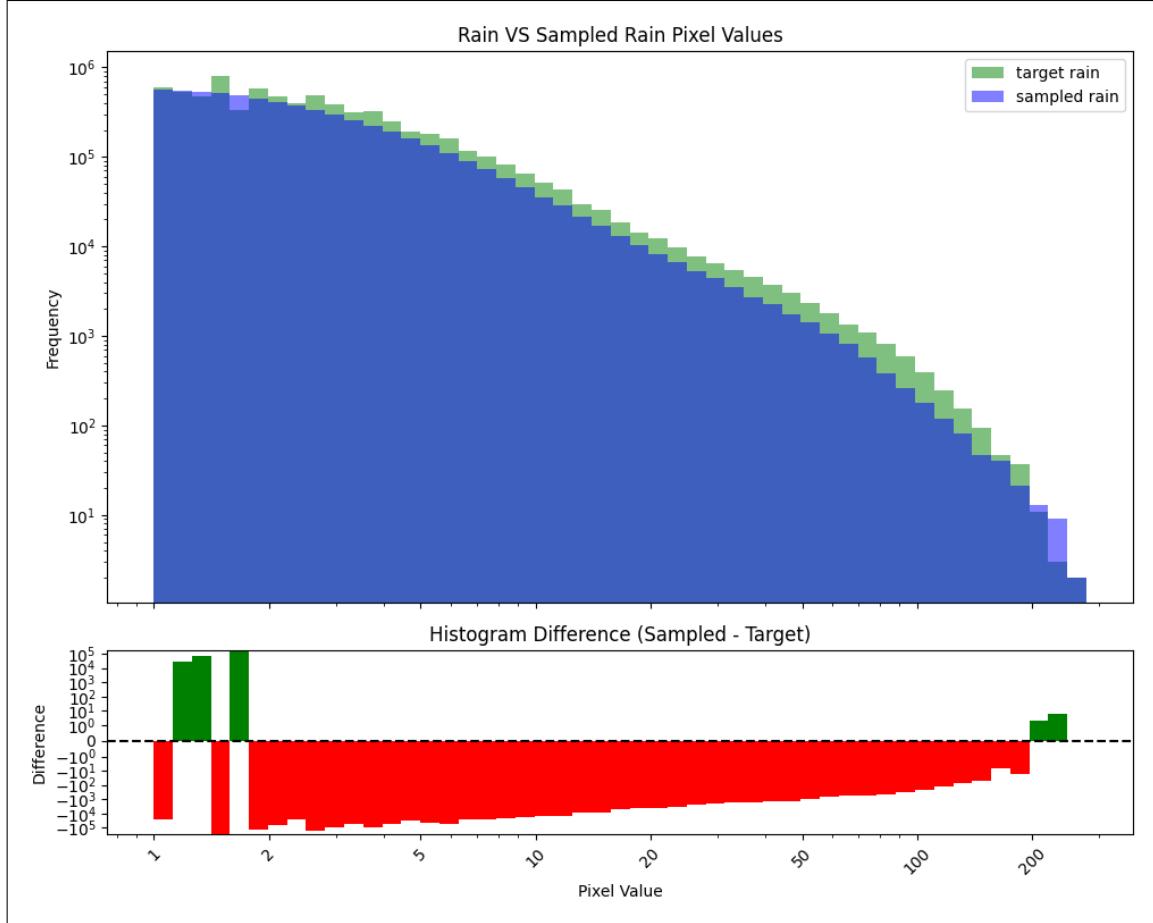


FIGURE 5.2 – Distribution de pluie générée vs vérité terrain
En bleu : pluie générée, en vert : cible

Les métriques suivantes (Table 5.2) sont obtenues sur l'ensemble du set de test (année 2019).

Métrique	Valeur
RMSE	1.6760
MAE	0.6815
Précision	0.6434
Rappel	0.6034
F1-score	0.5979
Accuracy	0.7577
CSI	0.4497
Critère d'arrêt : $RMSE \times (1 - CSI)$	0.8584

TABLE 5.2 – Performances du modèle sur l'ensemble de test

Le critère d'arrêt d'entraînement a été défini comme le produit $RMSE \times (1 - CSI)$. Ce choix permet de combiner deux aspects essentiels de l'évaluation : la précision globale des valeurs prédites (capturée par le RMSE) et la capacité du modèle à détecter correctement les zones pluvieuses (mesurée par le CSI). En multipliant le RMSE par $(1 - CSI)$, on favorise les prédictions qui sont à la fois précises et bien positionnées.

CSI), on pénalise davantage les modèles qui présentent à la fois une erreur élevée et une mauvaise détection spatiale des précipitations. On aurait également pu retenir la cohérence de distribution comme critère, mais celle-ci ne garantit pas la précision spatiale des précipitations prédites.

5.3 Comparaison des résultats

Nous comparons les performances de notre approche à celles obtenues avec la méthode CNN (utilisant seulement un U-net classique vu en Partie 2.1) développée en parallèle de ce projet, afin d'évaluer l'apport potentiel des modèles de diffusion dans la restitution des champs de pluie. Pour mener cette comparaison, nous avons retenu comme étude de cas l'orage du 9 août 2019, un événement particulièrement intense qui a fortement marqué la France et qui figure dans notre base d'entraînement. Ce choix permet de tester les modèles sur une situation météorologique extrême, mettant en évidence leur capacité à reproduire fidèlement à la fois l'intensité et la structure spatiale des précipitations.

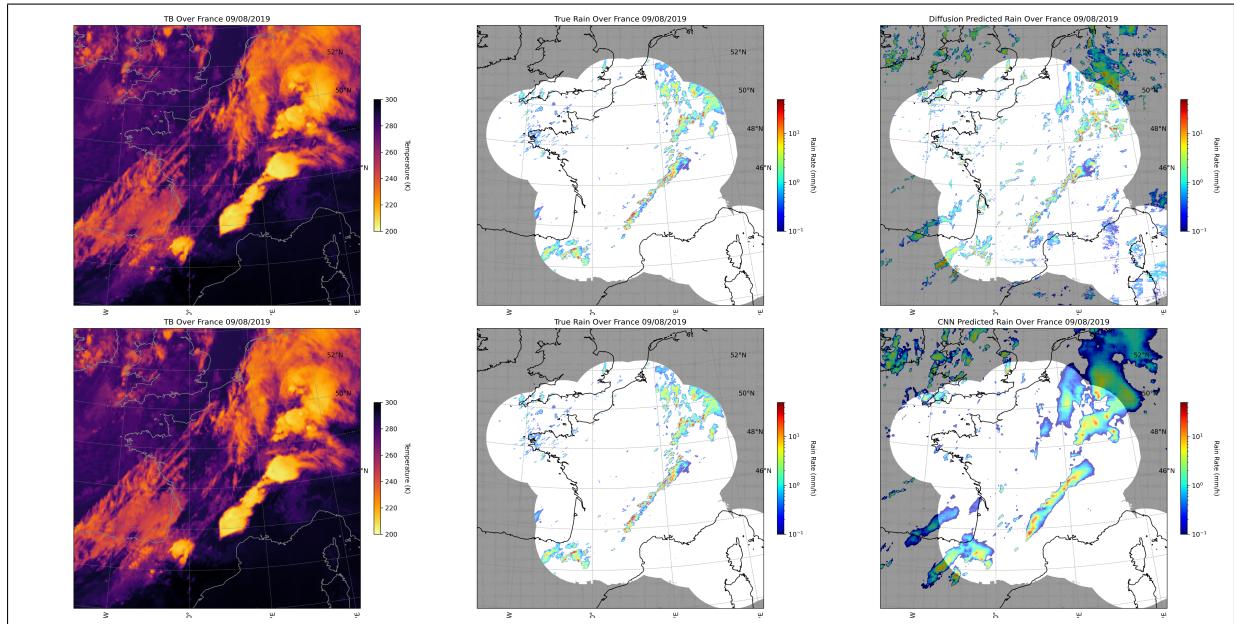


FIGURE 5.3 – Sortie Diffusion vs CNN
Résultats obtenus à partir de diffusion (haut) vs a partir du CNN (bas),
de gauche à droite : TBs, Cible, Output

La figure de diffusion a été obtenue en découplant l'image originale de la France, de taille 1500×1500 pixels, en une grille d'images plus petites de 128×128 pixels (le réseau ne pouvant pas traiter des images de tailles différentes à cause des couches d'attention qu'il possède). Chaque « imagette » ainsi extraite est ensuite traitée individuellement par le modèle de diffusion. En revanche, la structure du CNN utilisée dans ce projet permet d'entrer directement l'image complète, sans découpage préalable. Cette différence peut entraîner un léger impact sur la qualité et la cohérence spatiale des prédictions, en particulier au niveau des bords des imagettes, où des discontinuités peuvent apparaître.

Nous pouvons observer que, comme espéré, le modèle de diffusion permet de reconstruire des images de pluie visuellement plus réalistes et détaillées, évitant ainsi l'effet de lissage

souvent rencontré avec les CNN classiques. En revanche, le modèle de diffusion présente une tendance à générer des cellules de pluie artificielles, c'est-à-dire des zones pluvieuses inexistantes dans la vérité terrain, même en l'absence de conditions nuageuses (en bas à droite de l'image 5.3). Ce phénomène aberrant (il est physiquement impossible d'avoir de la pluie sans nuage) pourrait être expliqué par l'écart entre la vérité terrain et le dataset d'entraînement (comportant une quantité plus élevée de pluie).

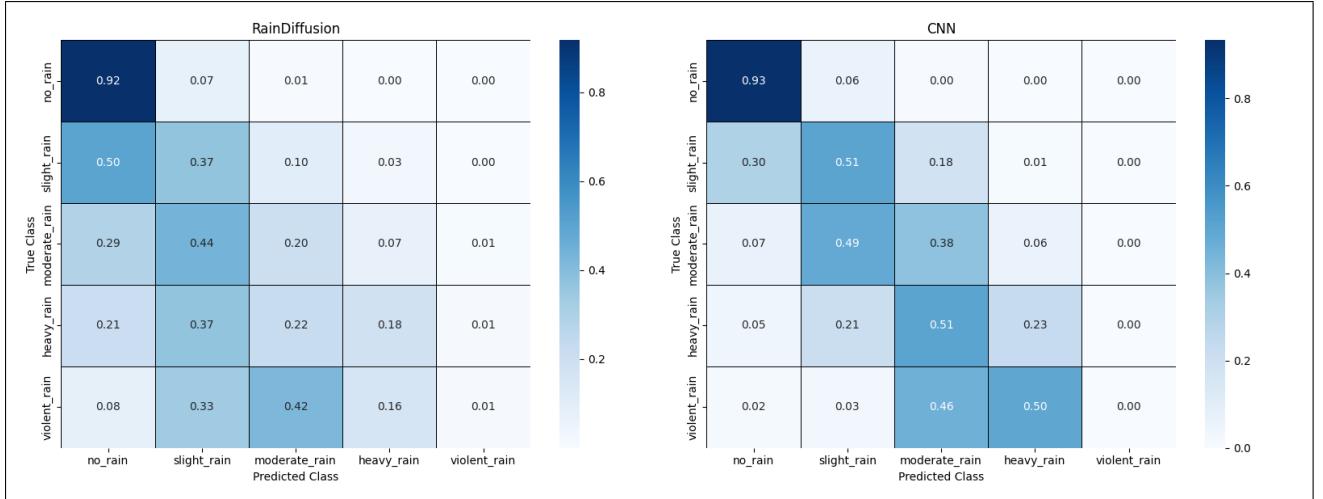


FIGURE 5.4 – Corrélation des tranches de pluie
Matrice de corrélation entre la vérité terrain et les sorties de réseau de neurones par tranche de pluie

Nous pouvons observer sur la Figure 5.4 que la diffusion a tendance à moins bien placer les tranches de pluie¹ que le CNN, ce qui peut être expliqué par un décalage spatial entre les cellules de pluie prédites et la vérité terrain. La diffusion prédit bien de la pluie, mais pas au bon endroit précisément.

Le Tableau 5.3 montre clairement que le CNN affiche des performances supérieures selon l'ensemble des métriques considérées. On note également une dégradation des résultats sur l'étude de cas par rapport à l'ensemble du jeu de test, à l'exception de la MAE qui est plus faible. Cette dernière s'explique par la forte proportion de zones sans pluie, où les valeurs proches de zéro réduisent naturellement l'écart moyen absolu.

Cependant, la Figure 5.5 illustre que le modèle de diffusion restitue beaucoup mieux la distribution globale des valeurs de précipitations. Bien que ce dernier tende à légèrement surestimer les faibles intensités de pluie, probablement en raison de la présence de cellules de pluie artificielles générées par le modèle, il capture néanmoins la variabilité et les extrêmes de manière plus fidèle que le CNN. Celui-ci manifeste une tendance marquée à sous-estimer systématiquement les intensités, ce qui conduit à une représentation moins réaliste de la répartition des précipitations. Cette différence souligne l'intérêt des modèles de diffusion pour reproduire des distributions complexes. Le modèle de diffusion reproduit des structures de pluie réalistes, mais la faible corrélation avec les données infrarouges

1. no rain : < 0.1 mm/h | slight rain : 0.1-2.5 mm/h | moderate rain : 2.5-10 mm/h | heavy rain : 10-50 mm/h | violent rain : > 50 mm/h

Métrique	Diffusion	CNN
RMSE	3.030	2.167
MAE	0.395	0.273
Précision	0.353	0.479
Rappel	0.562	0.763
F1-score	0.434	0.589
Accuracy	0.891	0.921
CSI	0.277	0.417
Critère d'arrêt : $(\text{RMSE} \times (1 - \text{CSI}))$	2.191	1.263

TABLE 5.3 – Métriques pour l'étude de cas du 9 août 2019, Diffusion vs CNN

limite la précision de la localisation des intensités. Une analyse à différentes résolutions pourrait préciser l'échelle à laquelle le modèle demeure pertinent.

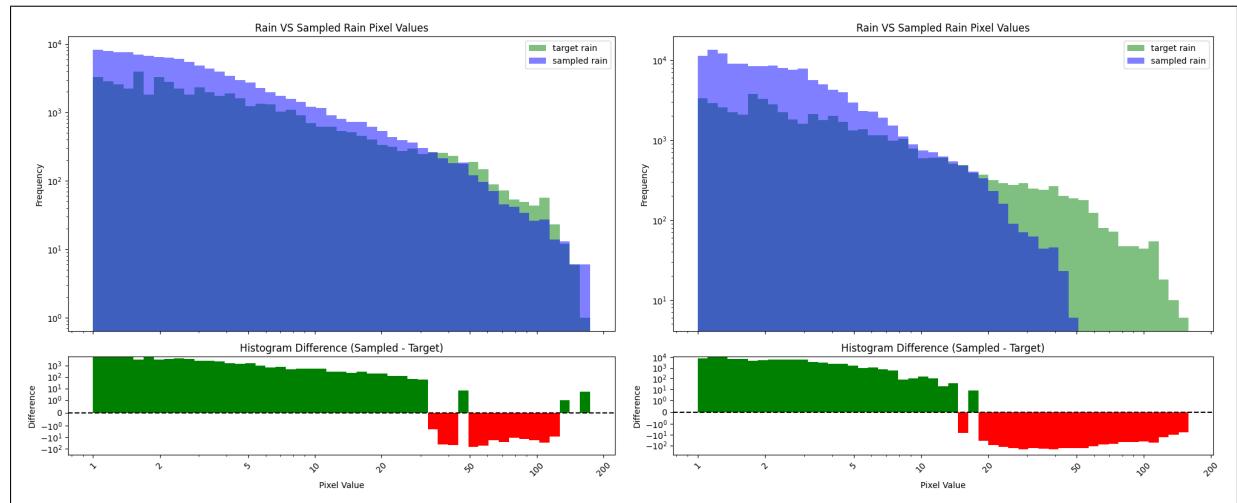


FIGURE 5.5 – Histogramme Diffusion vs CNN
Distribution des valeurs de pluie pour la diffusion et le CNN

Une des observations majeures de ce projet est que le caractère stochastique inhérent aux modèles de diffusion induit une forte variabilité lors de l'inférence, conduisant à des sorties différentes à partir d'une même image de température de brillance (TB). En calculant la moyenne sur plusieurs prédictions de diffusion, on obtient des métriques proches de celles produites par le CNN, avec des effets de lissage et de sous-prédition. Cela indique que le réseau parvient à modéliser la relation moyenne entre TB et pluie. Cependant, il peine à localiser précisément les fortes valeurs : il identifie leur existence mais pas leur position exacte. Comme on peut le voir sur la Figure 5.6, l'écart entre la diffusion et la vérité terrain est beaucoup plus hétérogène que celui observé avec le CNN, présentant des écarts plus forts et plus variés aux alentours des pluies intenses, tant en excès qu'en déficit.

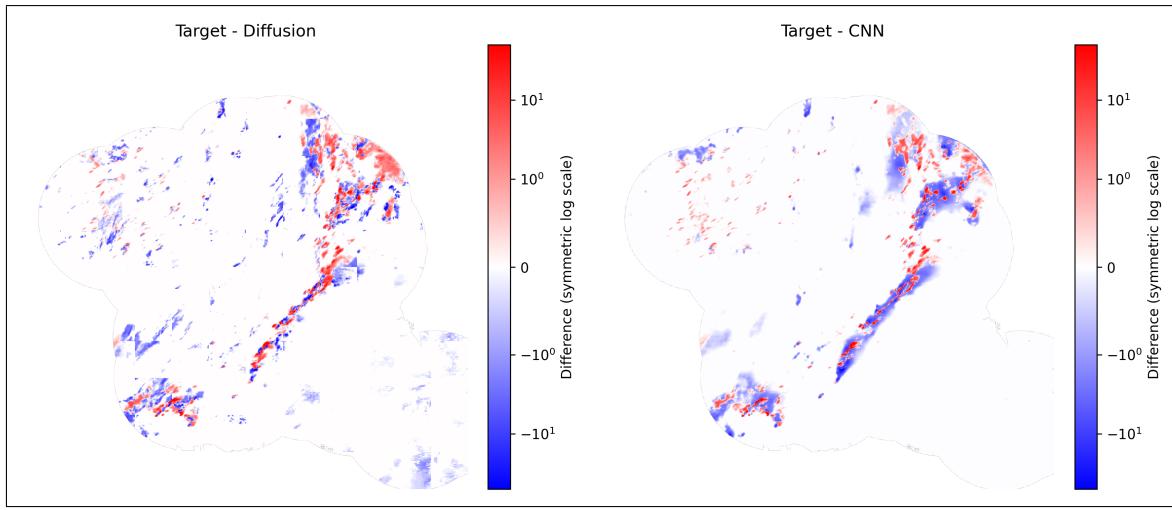


FIGURE 5.6 – Différence entre la pluie réelle et la pluie prédictée
Soustraction des valeurs de pluie des champs de pluie du 09/09/2019 pour la diffusion et le CNN

Cette limitation soulève la question de la résolution spatiale pertinente, et de savoir si une précision de 1 km n'est pas excessive, en particulier avec des TBs de résolution plus faible. La variabilité introduite par le processus stochastique permet de restituer la distribution globale et les valeurs extrêmes de la vérité terrain, ce qui est un avantage par rapport aux CNN. Cependant, cette variabilité, généralement recherchée dans les applications classiques de diffusion, où l'objectif est de générer plusieurs réalisations plausibles à partir d'une même condition, peut devenir problématique, car l'objectif est d'obtenir une prédiction stable et précise plutôt que des échantillons diversifiés. On peut émettre l'hypothèse que cette difficulté provient du faible degré de corrélation entre TBs et pluie, et que l'information nécessaire à la localisation fine des fortes précipitations est absente des TBs.

Chapitre 6

Analyse Critique et Réflexions Personnelles

6.1 Portée, limites et impacts du projet

Comme nous l'avons vu précédemment, ce projet propose une approche alternative aux réseaux de neurones convolutifs pour la prédiction des précipitations à partir de données satellitaires infrarouges. Bien que les modèles de diffusion restent moins performants que les CNN selon les métriques classiques, ils montrent un potentiel prometteur dans certaines situations. En particulier, ils reproduisent mieux la variabilité et les extrêmes des précipitations, ce qui est essentiel pour certaines applications météorologiques, malgré des erreurs persistantes de localisation. Leur capacité à restituer fidèlement la distribution des intensités de pluie ouvre des perspectives pour l'analyse et la simulation de conditions météorologiques rares ou extrêmes.

L'enjeu principal consiste à évaluer la capacité du modèle à exploiter la corrélation existante entre les températures de brillance et les précipitations, tout en considérant les limites imposées par les informations disponibles en entrée. L'observation de persistances d'erreurs de localisation, même en optimisant l'exploitation de cette corrélation, suggère que l'intégration de données complémentaires ou de modalités additionnelles pourrait être nécessaire pour améliorer la précision spatiale des prédictions.

Une autre approche innovante pourrait consister à combiner la puissance des CNN et des modèles de diffusion en divisant la tâche. Par exemple, le CNN pourrait être utilisé pour produire une première estimation globale de la pluie, tandis que la diffusion serait appliquée uniquement aux résidus, c'est-à-dire aux différences entre la vérité terrain et la prédiction CNN. Cette décomposition permettrait de simplifier la tâche de chaque modèle, d'améliorer la précision globale, et de tirer parti des avantages spécifiques de chacune des architectures (voir Figure 6.1).

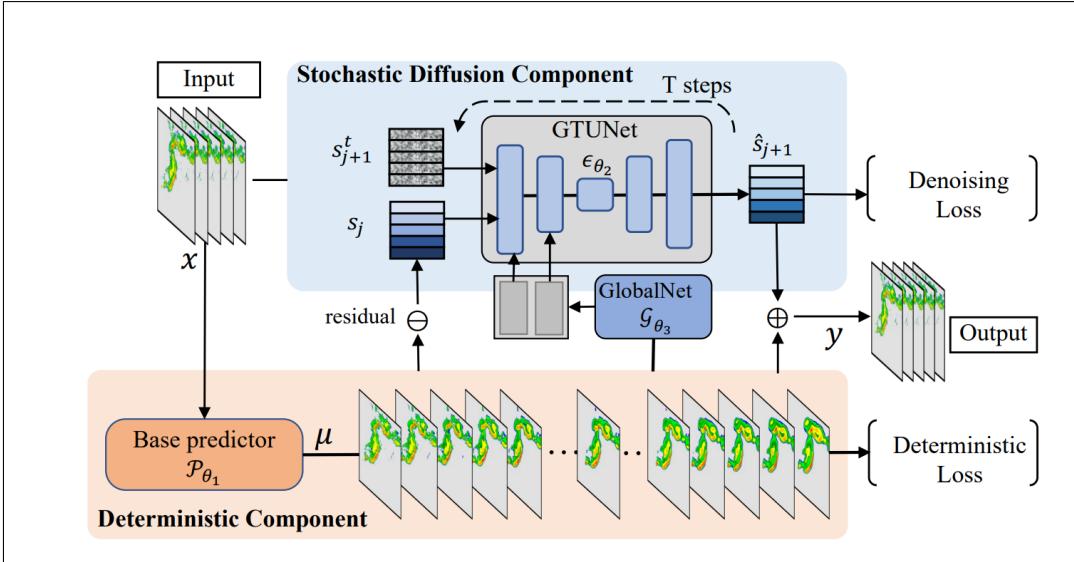


FIGURE 6.1 – Structure de modèle de diffusion appliquée à des résidus
Structure de DiffCast [10]

Une des principales limites rencontrées durant ce projet a été le temps d’entraînement ainsi que la puissance de calcul requise. En effet, dans les conditions matérielles et temporelles actuelles, il n’a pas été possible d’explorer plusieurs axes d’amélioration importants, comme l’augmentation de la taille du réseau ou l’utilisation d’images de résolution plus élevée. Ces deux leviers sont essentiels pour capturer des motifs pluviométriques à plus grande échelle, ce qui permettrait au modèle d’exploiter davantage d’informations contextuelles.

Par ailleurs, les modèles de diffusion sont connus pour leur temps de calcul important, aussi bien à l’entraînement qu’à l’inférence. Contrairement aux approches classiques fondées sur la comparaison directe des pixels prédictifs et observés, la fonction de perte employée dans les modèles de diffusion mesure la capacité du réseau à estimer correctement le bruit injecté à chaque étape du processus, orientant ainsi l’apprentissage vers un débruitage progressif plutôt qu’une correspondance pixel par pixel. Cette fonction de perte était donc plus difficile à analyser. Ce qui a mené au choix de procéder aux calculs de différentes métriques entre images générées et cibles sur notre set de validation, afin de mieux évaluer les dynamiques de nos entraînements. Cela a impliqué de réaliser l’intégralité du processus d’inférence sur un ensemble d’images conséquent (entre 200 et 500 par epoch dans notre cas).

Ce processus, d’une durée moyenne d’environ quatre jours, a considérablement limité le nombre d’expérimentations et d’itérations possibles, ralentissant ainsi la capacité à tester différentes architectures, hyperparamètres ou autres.

À terme, ce projet, actuellement limité à la France, devra être étendu à l’échelle mondiale. Une telle extension pose plusieurs défis, car les comportements nuageux et les régimes de précipitations observés en France ne peuvent pas être directement transposés à d’autres régions, compte tenu de la diversité climatique et géographique de la planète. Pour assurer une bonne capacité de généralisation, il sera nécessaire de réaliser un fine-tuning à partir de bases de données issues de radars mobiles couvrant une large variété de zones dans le

monde. Cette étape permettra d'adapter le modèle aux particularités locales et régionales des phénomènes météorologiques, et ainsi d'améliorer sa robustesse et sa précision dans différents contextes géographiques.

6.2 Réflexion personnelle sur l'expérience

Durant ce stage, j'ai pu acquérir de solides apports scientifiques et techniques, notamment en approfondissant mes connaissances sur les modèles génératifs de type diffusion et leur application à la prédiction météorologique. J'ai également développé des compétences pratiques en programmation et optimisation de réseaux de neurones. L'intégration au sein d'une équipe de recherche a été une expérience à la fois particulièrement agréable et enrichissante, me permettant de collaborer étroitement avec des experts du domaine, d'échanger sur des problématiques complexes.

Ce stage a été très formateur, tant sur le plan technique, en approfondissant ma compréhension des modèles complexes, que sur le plan méthodologique, en renforçant ma rigueur et ma capacité à analyser les résultats de manière critique. Il a également confirmé mon intérêt pour la recherche appliquée en intelligence artificielle, notamment dans le domaine environnemental, tout en me donnant une vision claire des compétences et des enjeux nécessaires pour évoluer dans ce secteur et m'encourageant à envisager, à terme, une thèse dans ce domaine.

J'ai aussi pu réfléchir à l'impact écologique du projet étant à la fois positifs et négatifs. Du côté négatif, l'entraînement des modèles de diffusion nécessite une importante puissance de calcul, entraînant une consommation énergétique élevée sur plusieurs jours. Cette consommation génère une empreinte carbone non négligeable. Ces coûts environnementaux soulignent la nécessité de réfléchir à des approches plus efficientes des recherches en intelligence artificielle. Cependant, le projet porte aussi un potentiel impact positif à plus long terme. Le but étant à terme d'étudier l'impact du réchauffement climatique sur les précipitations, les anciens modèles n'utilisant pas de machine learning étaient très peu efficaces, ainsi l'efficacité accrue des modèles pourrait participer à une meilleure adaptation au changement climatique, justifiant en partie les investissements énergétiques initiaux.

Conclusion

Ce stage au sein du LATMOS a été une expérience particulièrement enrichissante, tant sur le plan scientifique que personnel. Il m'a permis de m'immerger dans un environnement de recherche dynamique et pluridisciplinaire, et d'aborder des problématiques concrètes à fort enjeu environnemental, liées à l'amélioration de la compréhension et de la restitution des précipitations à l'échelle globale. Le travail réalisé sur les modèles génératifs de type diffusion a apporté un nouvel éclairage sur les possibilités offertes par l'intelligence artificielle appliquée à l'observation météorologique par satellite. Malgré des résultats encore inférieurs aux approches classiques basées sur les réseaux convolutifs, ces modèles montrent un potentiel intéressant, notamment pour restituer la distribution statistique des précipitations avec une plus grande fidélité. Les pistes d'amélioration envisagées ouvrent des perspectives prometteuses pour optimiser la précision et la robustesse de ces modèles.

Ce projet a aussi mis en lumière les défis liés à la complexité algorithmique et aux besoins importants en ressources de calcul, soulignant la nécessité d'un équilibre entre performance scientifique et impact environnemental. Cette prise de conscience s'inscrit pleinement dans les enjeux actuels de la recherche responsable.

Bibliographie

- [1] Elena Saltikoff, Katja Friedrich, Joshua Soderholm, Katharina Lengfeld, Brian Nelson, Andreas Becker, Rainer Hollmann, Bernard Urban, Maik Heistermann, Caterina Tassone, *An Overview of Using Weather Radar for Climatological Studies : Successes, Challenges, and Potential*, Bulletin of the American Meteorological Society, vol. 100, no. 9, pp. 1739–1764, 2019.
- [2] N. Viltard, V. Sambath, P. Lepetit, A. Martini, L. Barthès, and C. Mallet, *Evaluation of DRAIN, a Deep-Learning Approach to Rain Retrieval from GPM Passive Microwave Radiometer*, arXiv :2303.01220, 2023.
- [3] Junchao Gong, Lei Bai, Peng Ye, Wanghan Xu, Na Liu, Jianhua Dai, Xiaokang Yang, Wanli Ouyang, *CasCast : Skillful High-resolution Precipitation Nowcasting via Cascaded Modelling*, in *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 15809–15822, PMLR, 2024.
- [4] Jonathan Ho, Ajay Jain, Pieter Abbeel, *Denoising Diffusion Probabilistic Models*, Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851, 2020.
- [5] Prafulla Dhariwal, Alex Nichol, *Diffusion Models Beat GANs on Image Synthesis*, Proceedings of NeurIPS 2021 (Advances in Neural Information Processing Systems, vol. 34), pp. 8780–8794, 2021.
- [6] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, Mohammad Norouzi, *Image Super-Resolution via Iterative Refinement*, arXiv :2104.07636 [eess.IV], 2021 ; publié également dans IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 1338–, 2023.
- [7] Tero Karras, Miika Aittala, Timo Aila, Samuli Laine, *Elucidating the Design Space of Diffusion-Based Generative Models*, Advances in Neural Information Processing Systems, vol. 35, pp. 27138–27151, 2022.
- [8] Alexander Quinn Nichol, Prafulla Dhariwal, *Improved Denoising Diffusion Probabilistic Models*, in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, vol. 139, pp. 8162–8171, 2021.
- [9] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, *Generative Adversarial Nets*, Advances in Neural Information Processing Systems (NeurIPS) 27, pp. 2672–2680, Montréal, Canada, 2014.
- [10] Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, Xunlai Chen, *DiffCast : A Unified Framework via Residual Diffusion for Precipitation Nowcasting*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 27758–27767, 2024.