

# Prévision du risque orageux à l'aide de réseaux de neurones

**Mélanie BOSCH**

*Doctorante 2ème année*

*ONERA, DPHY/FPA et DTIS/SAPIA*

**Directeur de thèse** : Adrien CHAN-HON-TONG (DTIS/SAPIA)

**Encadrante ONERA** : Aurélie BOUCHARD (DPHY/FPA)

**Encadrant extérieur** : Dominique BEREZIAT (Sorbonne Université – LIP6)

**Financement** : ONERA et ALBATROS

# Contexte :



<https://www.albatros-horizon.eu/>

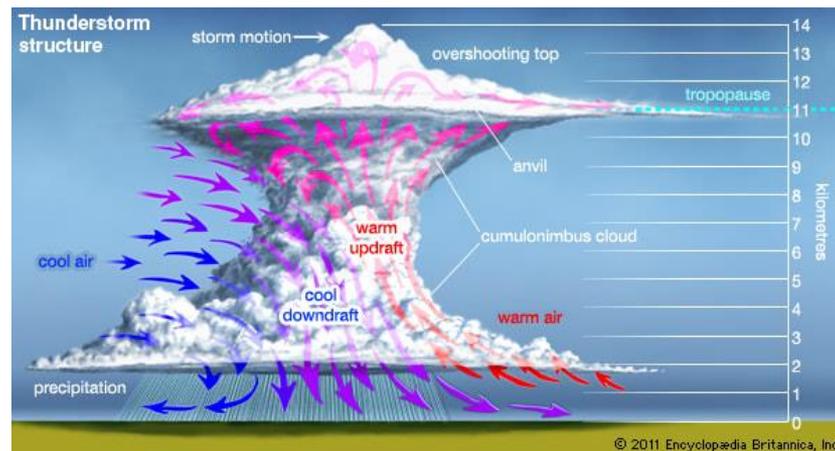
ALBATROS (Advanced systems and soLutions for Better practices Against hazaRDs in the aviatiOn System) → **Renforcer la sécurité aérienne.**

**Anticiper et prévenir l'émergence de dangers** dans l'aviation en développant des modèles de risque pour la sécurité.

Quels sont les dangers ?

**Cumulonimbus** : Nuage d'orage

Présence de **turbulences** dans le nuage, apparition de **fortes pluies** et de **grêle** → **dangers pour les avions.**



© 2011 Encyclopædia Britannica, Inc.

<https://www.britannica.com/science/thunderstorm>

# Objectifs :

Cumulonimbus : **principaux générateurs d'éclairs sur Terre.**

Chaque avion de ligne est en moyenne **foudroyé une fois par an.**

Une **opération de maintenance** est nécessaire après chaque foudroiement → **perte de temps et d'argent**

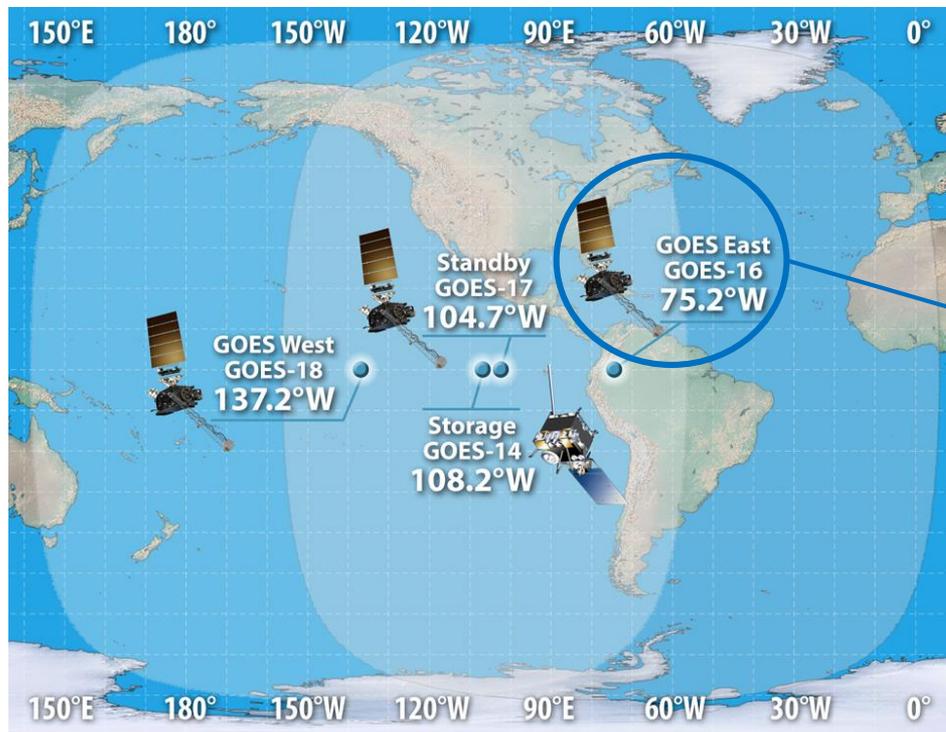
## Objectifs :

- Prévoir la position de l'**activité électrique**
- Sous forme de **cartes de risque de probabilités**
- A **très court terme** (<1h et toutes les 5 minutes)
- Avec des méthodes de **réseaux de neurones**
- Permettrait aux avions d'**éviter** les zones à fort risque de foudroiement
- Prévoir aussi **au dessus de la mer**, donc sans **données radar**



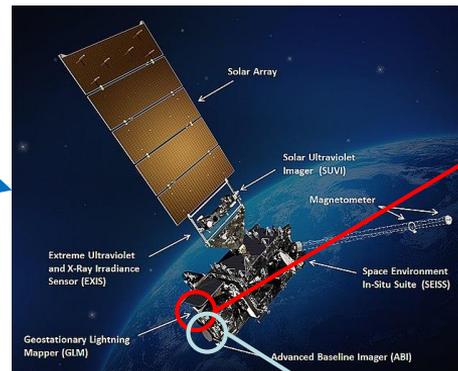
<https://makeagif.com/gif/amazing-airplane-struck-by-lightning-plane-crash-sskKzn>

# Le satellite GOES-R et ses instruments :



<https://www.goes-r.gov/mission/mission.html>

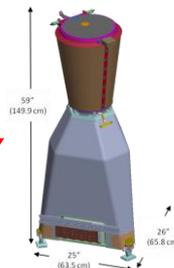
Champ de vue de la constellation GOES



[https://fr.wikipedia.org/wiki/Geostationary\\_Operational\\_Environmental\\_Satellite](https://fr.wikipedia.org/wiki/Geostationary_Operational_Environmental_Satellite)

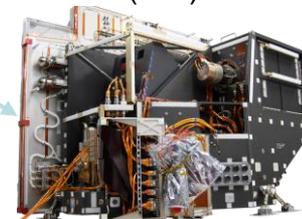
Geostationary Operational Environmental Satellite (GOES-R)

Geostationary Lightning Mapper (GLM)



<https://www.goes-r.gov/spacesegment/glm.html>

Advanced Baseline Imager (ABI)



<https://spaceflight101.com/goes-r/goes-r-instruments/>

# Les capteurs utilisés :

## Advanced Baseline Imager (ABI)

- Spectromètre/Radiomètre
- Résolution spatiale de 0,5 km dans le visible
- Résolution de 2 km pour l'IR
- 16 bandes de longueur d'onde
- Bande 13 : 10,3  $\mu\text{m}$ , plus sensible à la classification nuageuse
- Images toutes les 5 minutes

## Geostationary Lightning Mapper (GLM)

- Caméra
- Résolution spatiale de 8 km
- Observe l'activité électrique
- 70-90% de détection de la foudre
- Fonctionne jour et nuit
- Images toutes les 20 secondes

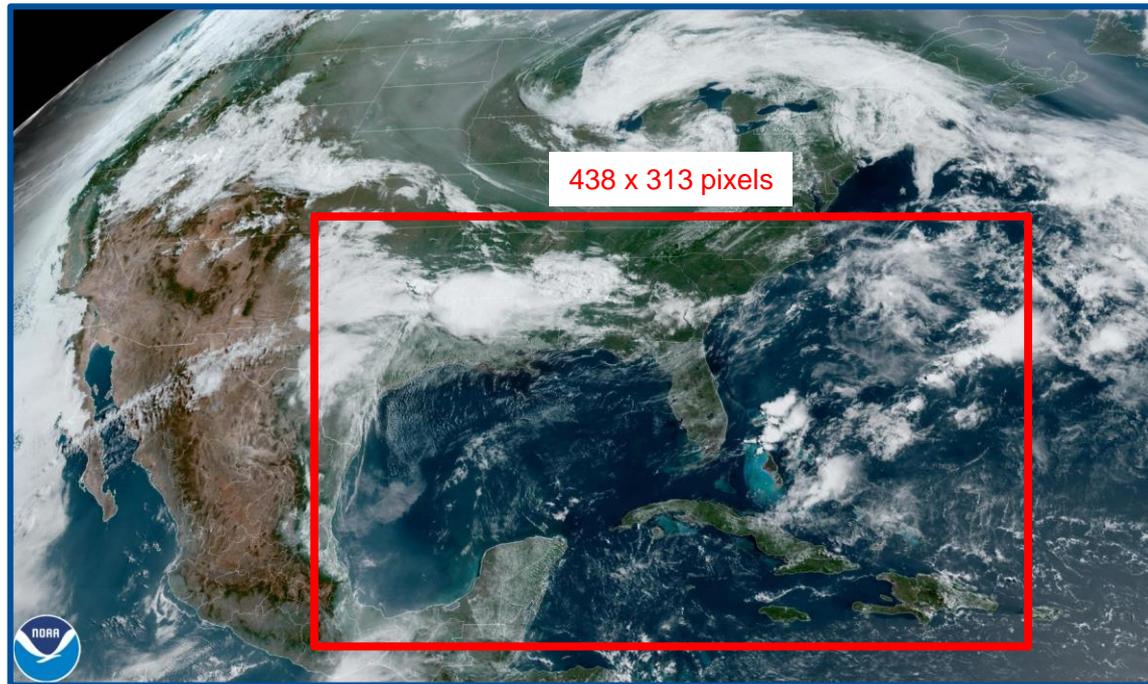
# La zone géographique étudiée :

## Pourquoi CONUS ?

- Capteurs avec bonne résolution spatiale et temporelle
- Gratuitement téléchargeables en ligne
- Proche de la zone de convergence intertropicale → climat chaud et humide qui favorise les orages et donc la foudre

## Pourquoi cette zone précise ?

- Permet une restriction spatiale



13 Jun 2023 14:21Z - NOAA/NESDIS/STAR GOES-East - GEOCOLOR Composite - Day(0.47 um - blue, 0.64 um - red, and 0.86 um - near IR)

<https://www.star.nesdis.noaa.gov/GOES/conus.php?sat=G16>

## Continental United States (CONUS)

# Données satellites :

## Données satellite GOES-R

### Données :

- Température de brillance (ABI)
- Groupes (Eclairs et leur étalement spatial, GLM)

### Prétraitement :

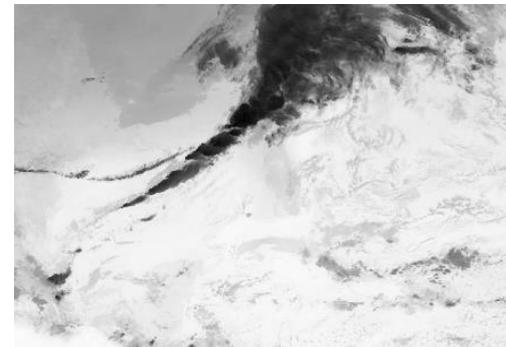
- Traitement spatial : maillage commun de  $0,08 \times 0,08^\circ$  (sous-échantillonnage)
- Traitement temporel : regroupement des éclairs sur 5 minutes

Résolution spatiale : 8,8 km

Résolution temporelle : 5 minutes

Images de TB : Pixels les plus sombres  $\rightarrow$  TB la plus basse donc sommet des nuages plus hauts  $\rightarrow$  Cumulonimbus

Images de la position de l'activité électrique : en blanc les pixels qui représentent les éclairs et en noir le background



Température de brillance



Groupes

# Données de modèle de prévision numérique du temps :

Modèle GFS  
Global Forecasting System

## Ajout de deux paramètres :

Lifted Index (LI) : Différence de température entre une parcelle d'air étudiée et son environnement

Plus  $LI < 0$  plus la zone est instable et donc possibilité de convection

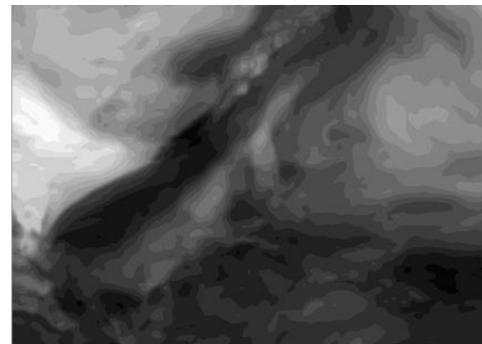
$$LI = T_{\text{environnement}} - T_{\text{parcelle}}$$

Humidité relative : Quantité de vapeur d'eau présente dans l'air ( $e$ ) par rapport à la quantité maximale que l'air peut contenir à une température donnée ( $e_s$ )

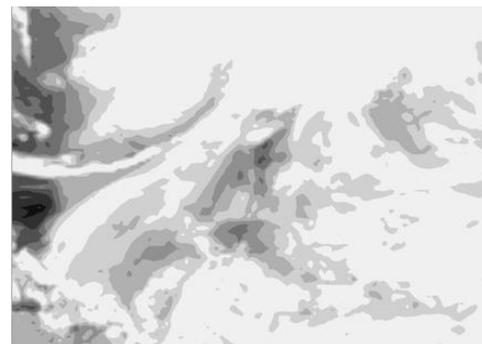
$$RH = \frac{e}{e_s} * 100$$

Lifted Index : Pixels plus sombres → LI négatif donc présence d'instabilité

Max humidité relative : Pixels les plus clairs → Taux d'humidité plus grand



Lifted Index



Max humidité relative

# Création du dataset :

---

## Données récupérées pour :

- Certains jours des mois d'hivers : Janvier, Février, Décembre
- De 00h à 05h du matin
- Pour les années allant de 2020 à 2023
- En tout 154 jours récupérés
- Environ 50% de jours avec sans éclairs et 50% de jours avec éclairs (*pour ne pas biaiser l'algorithme*)

## Données satellites :

154 jours, images toutes les 5 minutes, 30 images par jour [00h:05h] pour chaque type de données → **18 480 images**

## Données GFS :

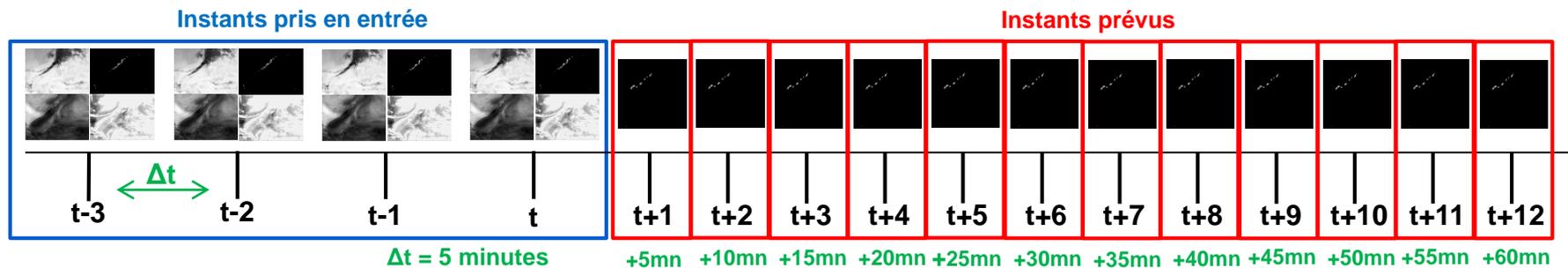
154 jours, 3 images par jour [00h:05h] pour chaque type de données → **924 images**

Une même image de 00h à 1h30, une autre de 1h30 à 04h30 et une autre de 4h30 à 05h → Redondance

**Répartition dataset :** 70% pour l'entraînement et 30% pour le test

# Méthodologie :

Prendre 4 instants en entrée (soit 20 minutes), prévoir toutes les 5 minutes jusqu'à 1 heure :



Séquences spatio-temporelles des quatre types de données.

On veut réaliser une tâche de **segmentation sémantique** :

- Générer un masque en sortie du modèle.
- Chaque pixel de ce masque obtient un score d'appartenir à la classe éclair ou à la classe background.

# Résultats attendus :

→ On veut obtenir des cartes avec des probabilités d'avoir de l'activité électrique dans une zone.

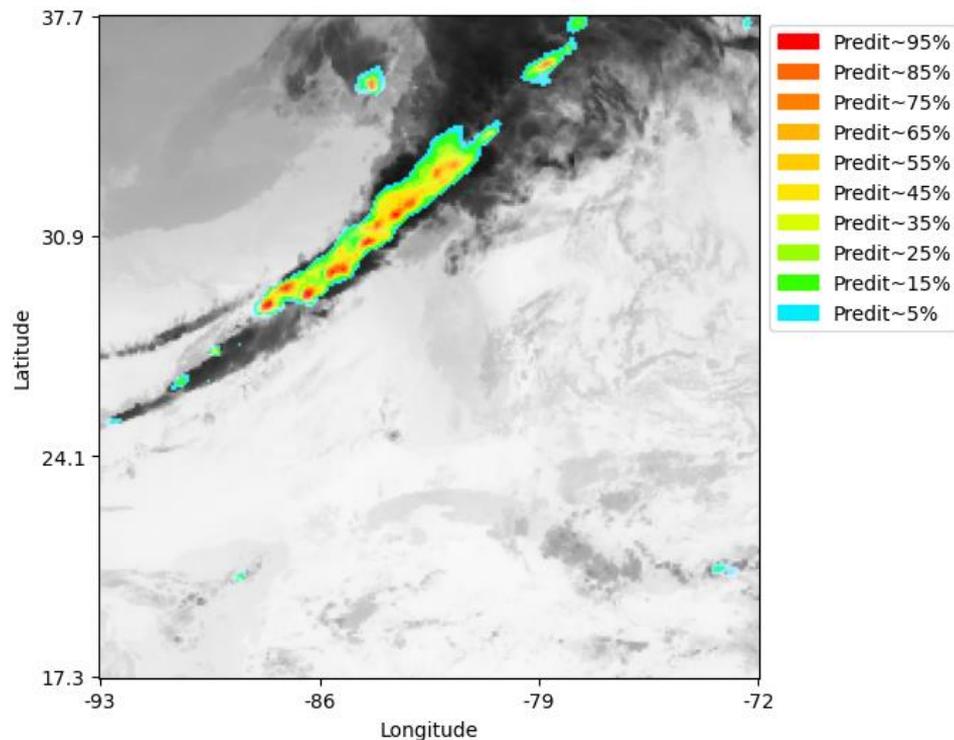
→ Pour obtenir ces résultats, on ne peut pas utiliser n'importe quel réseau de neurones car ils ne sont pas tous calibrés.

## **Zones en rouge :**

Très forte probabilité d'avoir de l'activité électrique

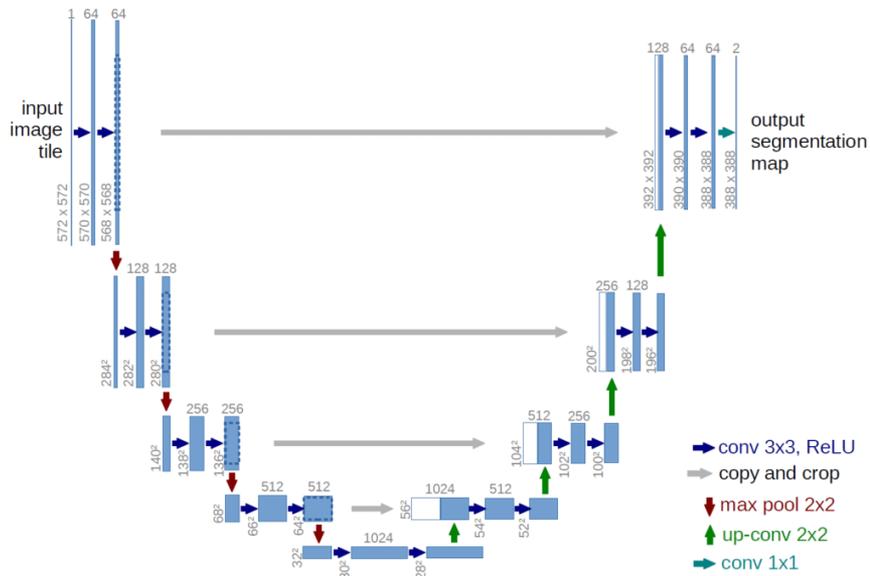
## **Zones en bleu :**

Faible probabilité d'avoir de l'activité électrique



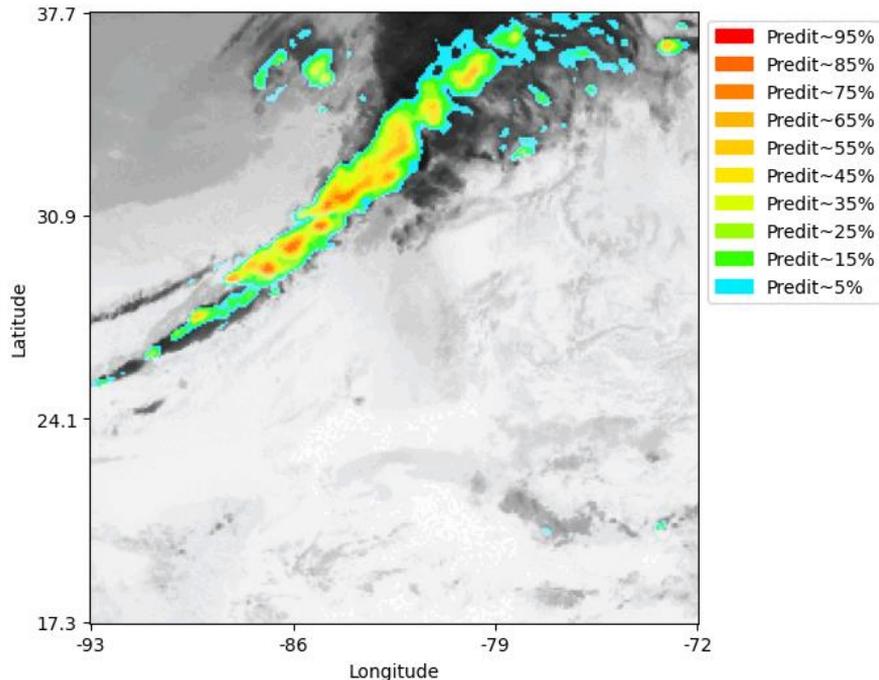
Carte représentant le pourcentages de risque d'avoir de l'activité électrique (13/01/23 à 01h11, prévisions à 5mn)

# Test avec U-Net :



Ronneberger, O et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.

Modèle classiquement utilisé pour faire de la segmentation sémantique



Carte représentant des zones de risque d'avoir de l'activité électrique (prévisions à 5mn)

# Calibration avec U-Net :

En sortie du réseau, chaque pixel obtient un score d'appartenance à la classe éclair (score de confiance) :

0,2	-10	0,6	-3	5	0,02	0	0,06	0	0,5
3	-24	2	0,4	0,1	0,3	0	0,2	0,04	0,01
-15	7	15	-20	1	0	0,7	0,9	0	0,1
0,3	-24	30	25	-6	0,03	0	0,98	0,95	0
2	-4	0,2	1,5	0,8	0,2	0	0,02	0,15	0,08

→ Softmax

0,02	0	0,06	0	0,5
0,3	0	0,2	0,04	0,01
0	0,7	0,9	0	0,1
0,03	0	0,98	0,95	0
0,2	0	0,02	0,15	0,08

Masque de prévision obtenu en sortie  
(Pas des probabilités)

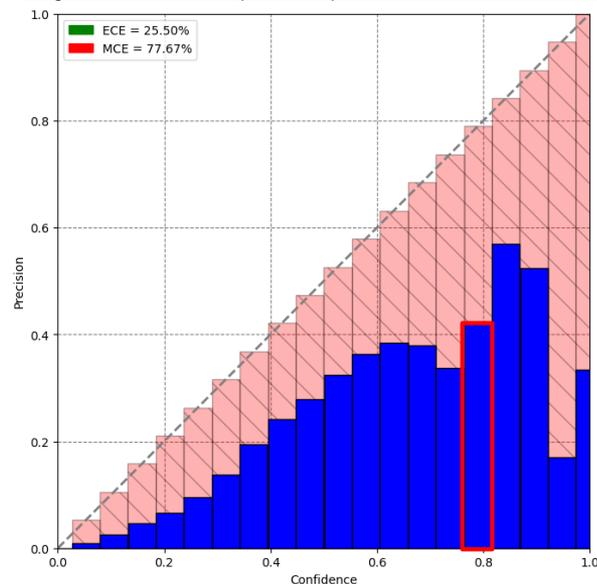
Scores de confiance entre 0 et 1  
(Toujours pas des probabilités)

Utiliser la fonction softmax en sortant du réseau n'assure pas la calibration

**Pour être sûr que ce sont des probabilités :**

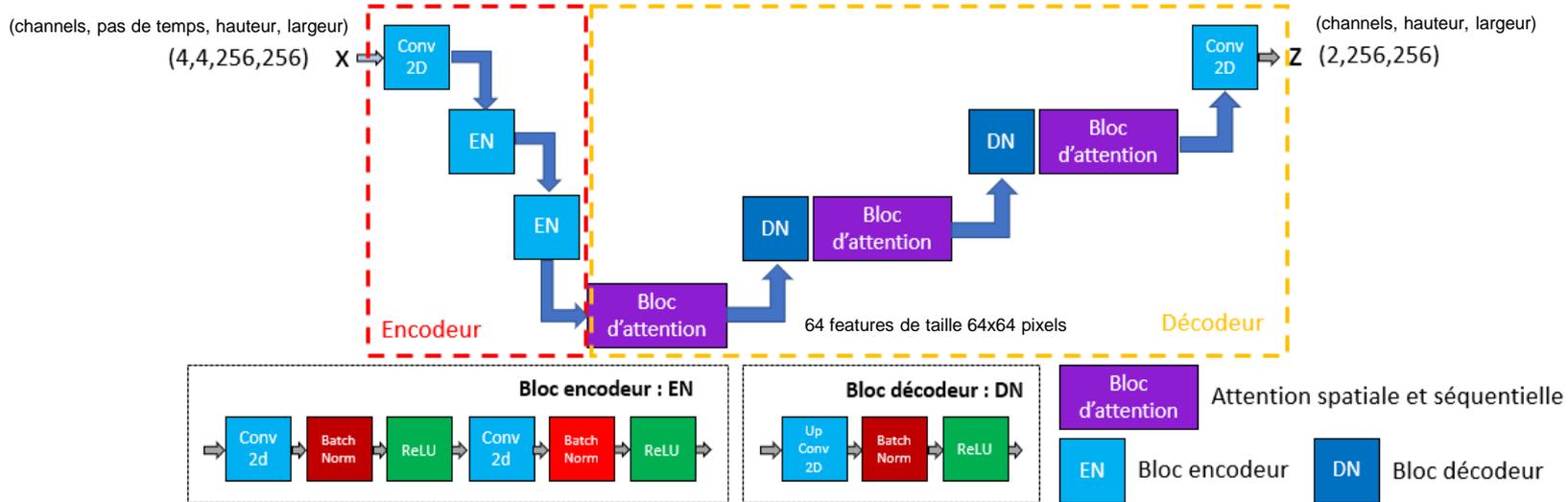
- Vérifier que les **probabilités prédites** reflètent la **fréquence des évènements réels**.
- Pour tous les pixels classés dans le bin [0.35, 0.45], il faut qu'on ait effectivement environ 40% de ces pixels qui soient des éclairs sur la vérité terrain.

Diagramme de fiabilité pour des prévisions à 5mn avec U-Net



Tous les pixels ayant un score de confiance d'environ 0,8 sont composés de 42% de vrais éclairs, donc le réseau surestime la présence d'éclairs

# Architecture du réseau ED-DRAP :



Che, H et al. (2022). ED-DRAP: Encoder–Decoder deep residual attention prediction network for radar echoes. *IEEE Geoscience and Remote Sensing Letters*.

Encodeur : Extrait les caractéristiques et dépendances spatiales et temporelles des séquences d'images.

Décodeur : Reconstitue le masque de prévision en sortie.

Utilisation d'attention spatiale et séquentielle.

# Fonction de perte :



Le dataset est **très déséquilibré** (environ 1% de pixels blancs sur les labels)  
→ le réseau va avoir tendance à prévoir seulement du **background** (la classe majoritaire) car les **métriques** resteront bonnes.

**Solution** : Modifier la fonction de perte.

**Ajout de la fonction dice dans la fonction de perte pour obtenir gérer le déséquilibre :**

$$loss = CrossEntropy(y, z) + \alpha * diceLoss(y, z) = -\left( \sum_{x \in classes} y(x) \log(z(x)) \right) + \alpha * \left( 1 - 2 * \frac{y \cap z}{y + z} \right)$$

Ajout de la **fonction dice** pour permettre au réseau de **détecter plus d'éclairs**.

Si  $\alpha$  augmente, la POD augmente mais on perd en calibration.

$x = Pixel$   
 $y = Vérité terrain$   
 $z = Prédiction$

**Objectif** : Trouver la bonne balance entre détecter assez d'éclairs et être assez calibré.

**Priorité** : Avoir un réseau calibré pour estimer des probabilités d'avoir de l'activité électrique.

# Calibration avec ED-DRAP :

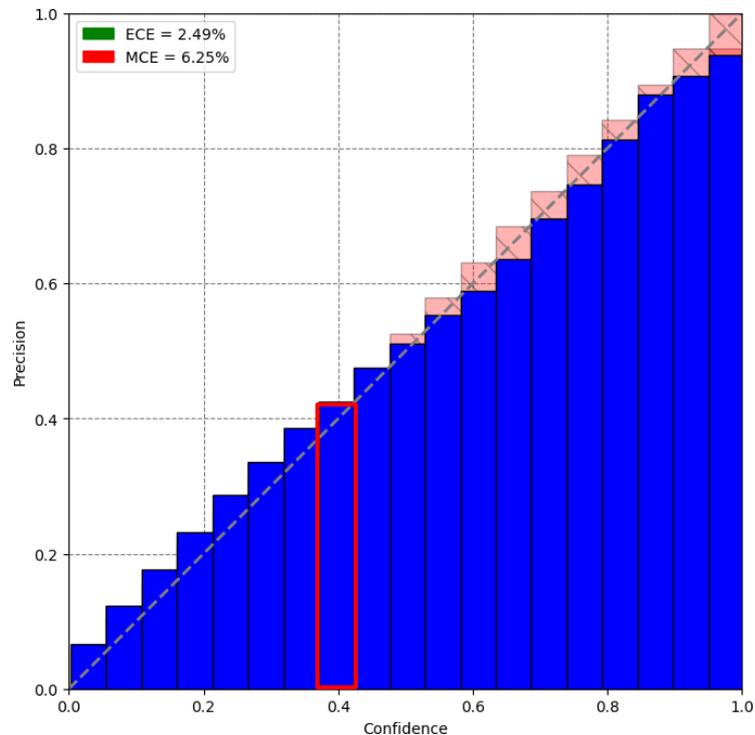
Résultats **calibrés** avec ED-DRAP pour des prévisions à toutes les échéances.

## Pourquoi on obtient des résultats calibrés ?

- Choix des données d'entrée
- Utilisation d'un réseau plus adapté (ED-DRAP)
- Choix d'une bonne fonction de perte (bon coefficient  $\alpha = 0.001$ )

## Exemple pour le bin autour de 0,4 :

- Tous les pixels dont le score de confiance était d'environ de 0,4 sont bien composés de 40% d'éclairs.
- Donc le réseau estime correctement les probabilités d'apparition de l'activité électrique



**Diagramme de fiabilité pour des prévisions à 30 minutes**

# Carte de prévision à 5 mn :

Carte représentant le pourcentages de risque d'avoir de l'activité électrique dans ces zones :

## Zones en rouge :

Très forte probabilité d'avoir de l'activité électrique

## Zones en bleu :

Faible probabilité d'avoir de l'activité électrique

## Prévision :

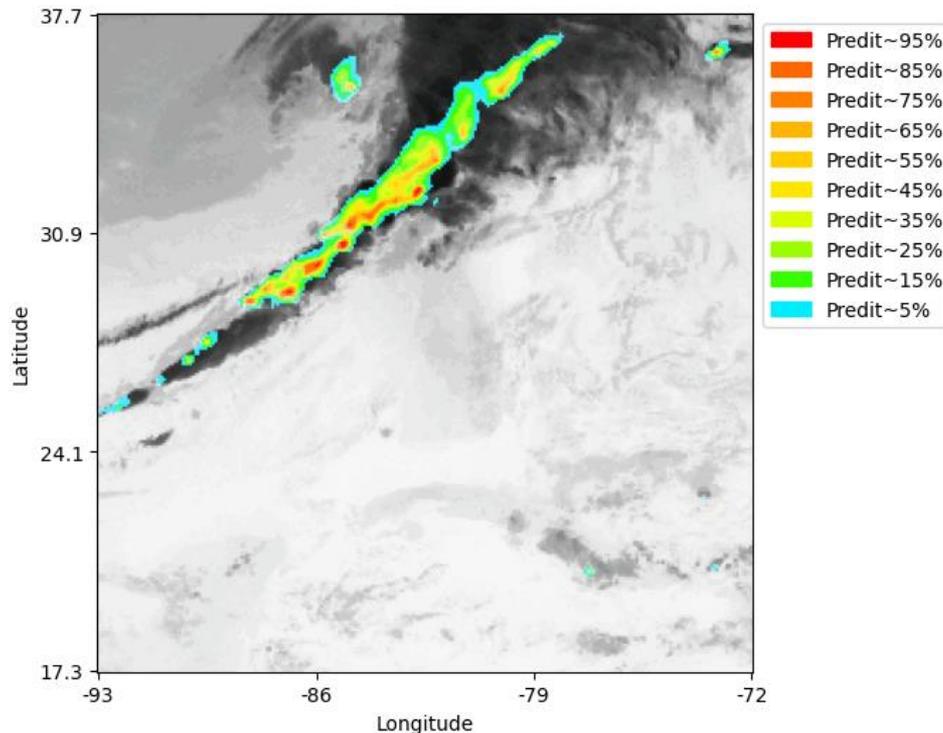
- Echéance : 5 minutes
- Séquence prise en entrée : 20 minutes

## Date :

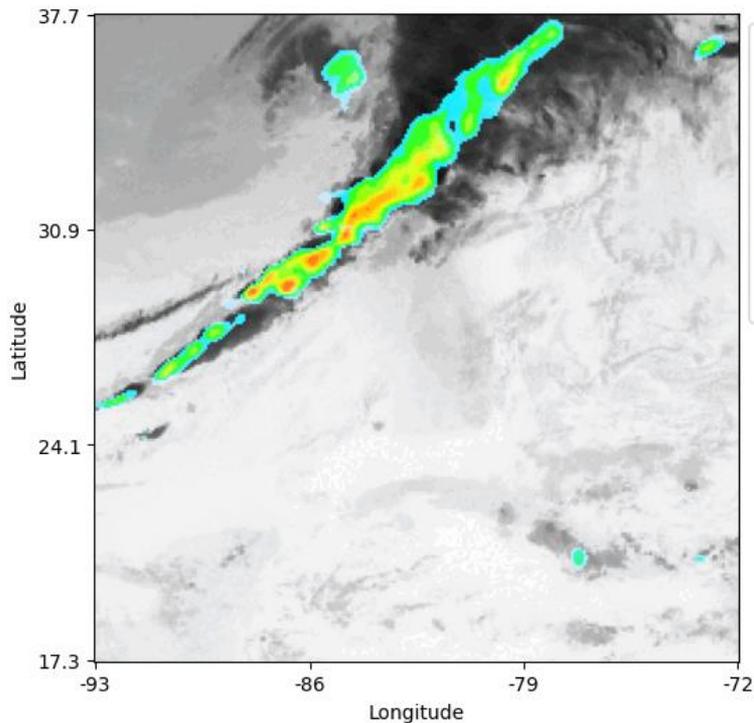
- 13 Janvier 2023
- De 00H30 à 5h

## Performances :

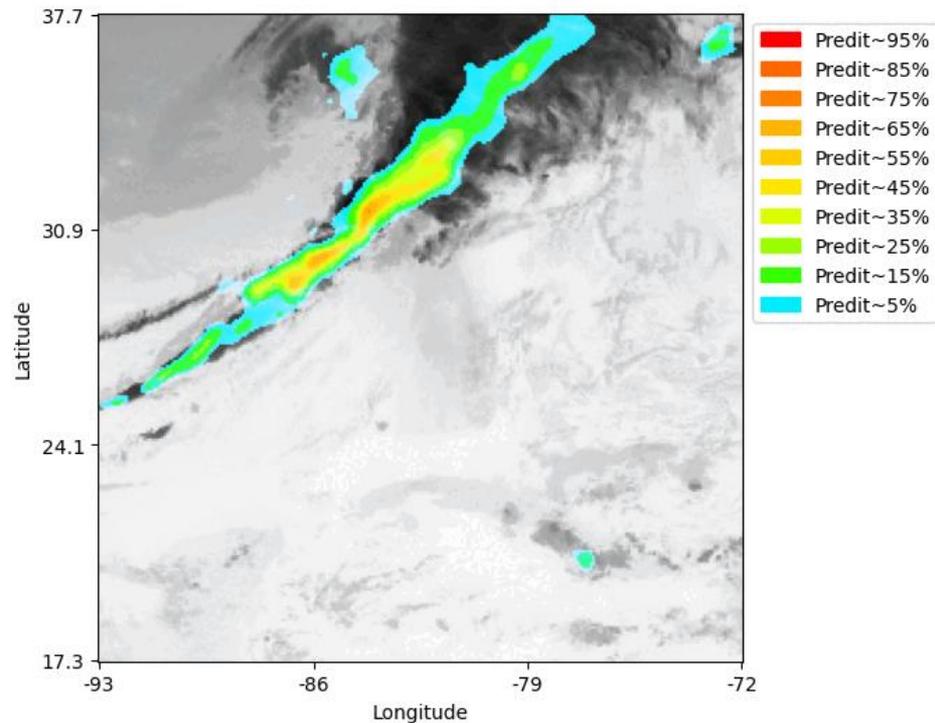
- Résultats calibrés, c'est-à-dire interprétable physiquement comme des probabilités



# Carte de prévision à 30mn et 55mn :



Prévisions à 30 minutes



Prévisions à 55 minutes

# Conclusion :

---

## Ce qui a été réalisé :

- ✓ Collecte et préparation de données météorologiques pour créer une base de données liée au risque orageux
- ✓ Utilisation de ED-DRAP, calibration et prévisions à 1 heure toutes les 5 minutes.
- ✓ Etablissement de cartes de risque de foudroiement adaptables à l'utilisateur.

## Perspectives :

- Ajouter d'autres données en entrée comme des données radar pour améliorer les performances
- Utiliser les données provenant du satellite MTG et de ses deux capteurs similaires pour faire des prévisions en Europe.
- Tester la robustesse de la méthode sur de nouvelles zones géographiques non vues par le réseau.

---

**Merci pour votre attention  
Avez-vous des questions ?**