Hackathon SDD 2024

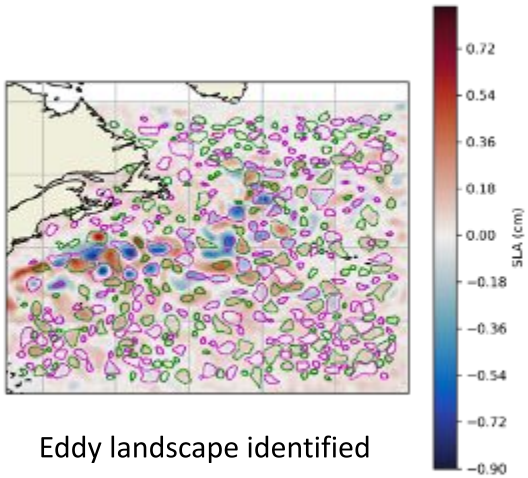
Wrap-up

Hack partners and subjects

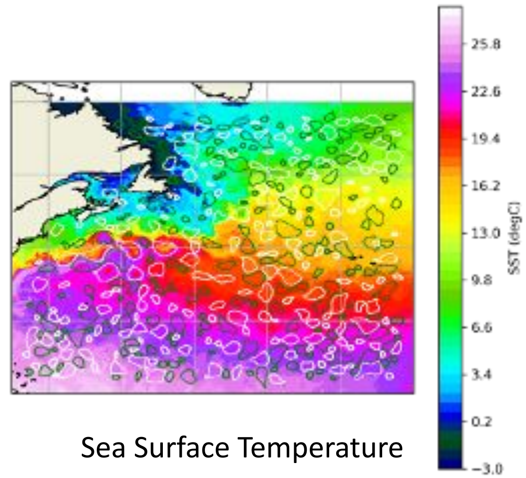
1. **Ocean Eddy Identification**
2. **Water Segmentation**
3. **Daily Rainfall Forecasting**
4. **Methane Detection**

Subject 1: Ocean Eddy Identification

Responsable : Daria Botvynko, Simon van Gennip



Eddy landscape identified
With Sea Level Anomaly



Sea Surface Temperature
(SST)

Oceanic Eddies are vortices of the order of ~10km in horizontal scale, whose signature is clearly visible in satellite products and well reproduced in ocean models. Sea Level Anomalies (SLA) are used for detecting such object by means of numerical techniques, yet such approach lacks accuracy in detection, namely because Sea Level Anomaly contains errors. Eddy signature is also visible in other variables such as Sea surface Temperature (SST) that do not suffer from such limitation.

The objective here is to **develop a Deep Learning approach to identify eddies using SST, SLA and the ocean velocity field.** For this students will have a dataset consisting of SLA, SST, and velocity components images which are slightly distorted relative to reality, together with labels (the eddy contours, see figure) obtained directly from reality.

MERCATOR OCEAN INTERNATIONAL


Merecastor

EDDY MITCHELL

HACKBAARRRR2BZZZZZZZ(H)
SCOTCH
Wicked

EDDY ET LES CAFARDS
OCEAN

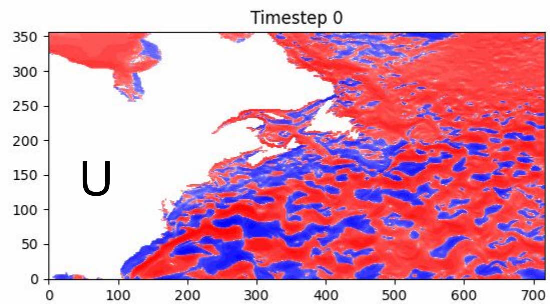
Présentation du problème

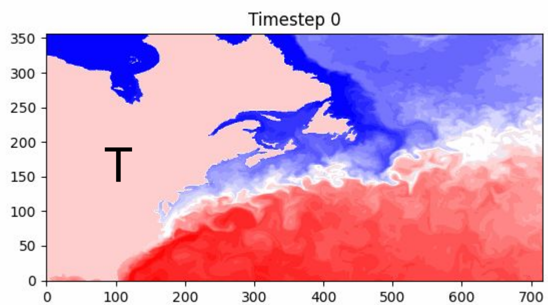
Problème :
- 10 jours de données physiques (u,v,H,T) -> forecast des vortex (eddies) sur les 7 jours suivants
- Données d'entrées sur 1 an: 8 mois entrainement, 4 mois evaluation (sur kaggle)

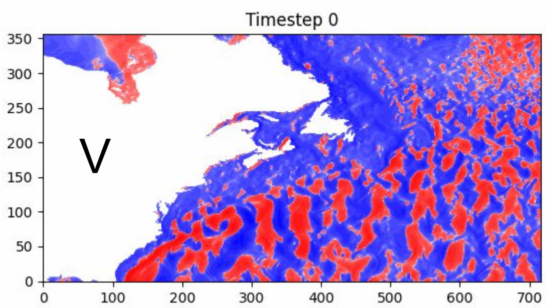
Stratégie : Décomposition en 2 sous-problèmes
- un problème de segmentation = prédire 1 jour de vortex map à partir de 1 jour de données physiques
- un problème de forecast = prédire le 11e jour de vortex map à partir des 10 jours précédents

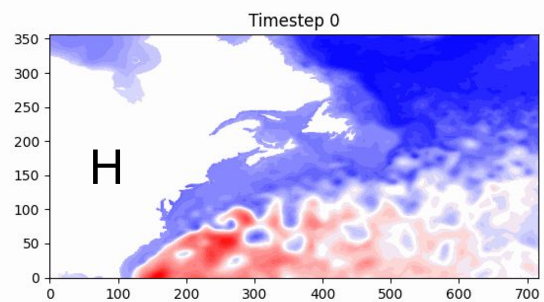
Modèles utilisés :
- UNet (~70 epoch) [segmentation]
- Réseau CNN classique [forecast]

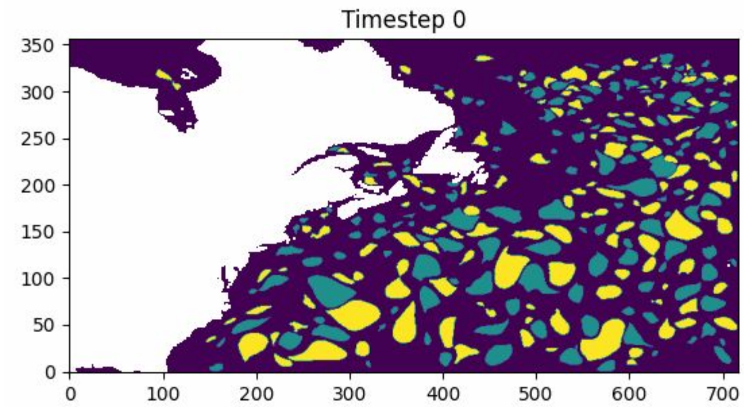
Data exploration and Feature selection







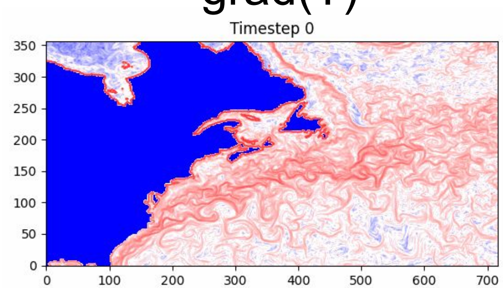




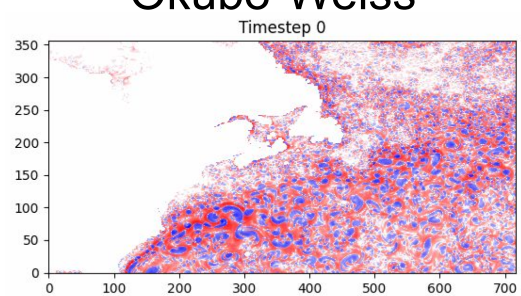
Label (Eddies)

+

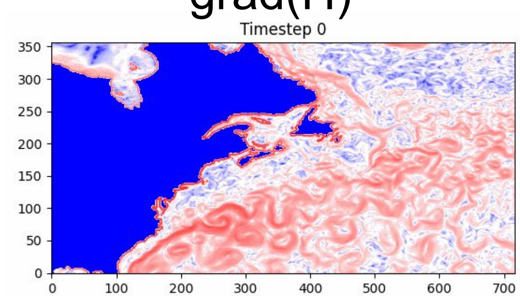
grad(T)



Okubo-Weiss

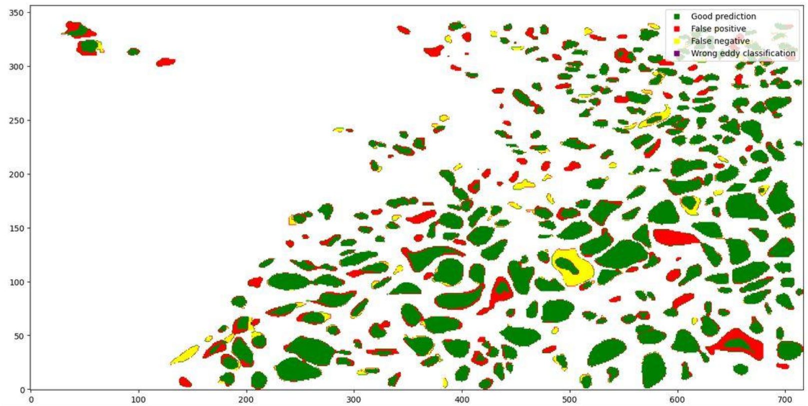


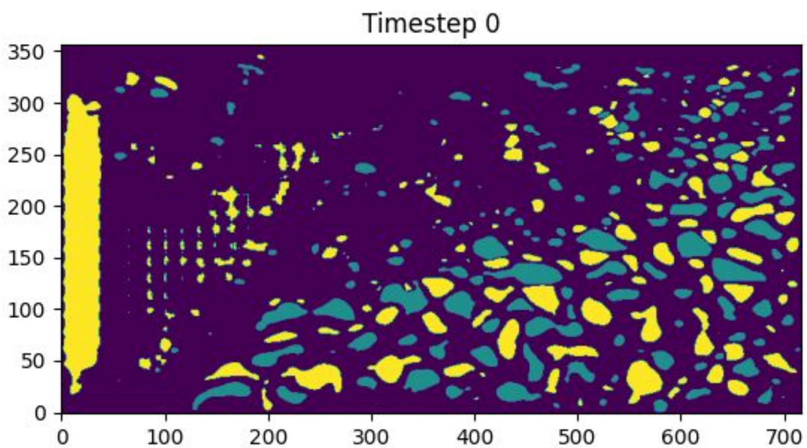
grad(H)



Résultats

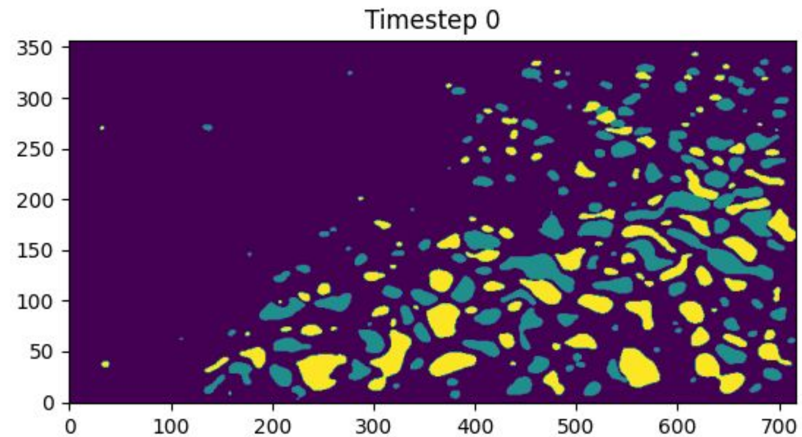
Prédictions





Accuracy 0.83

Forecast



Accuracy finale: 0.65

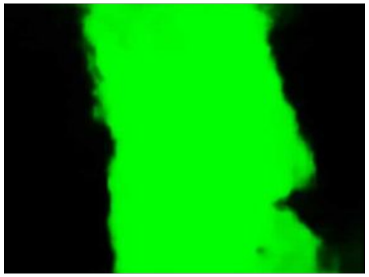
Subject 2: Water Segmentation

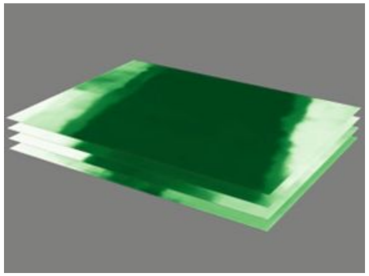
Responsable : Julianna Devillers

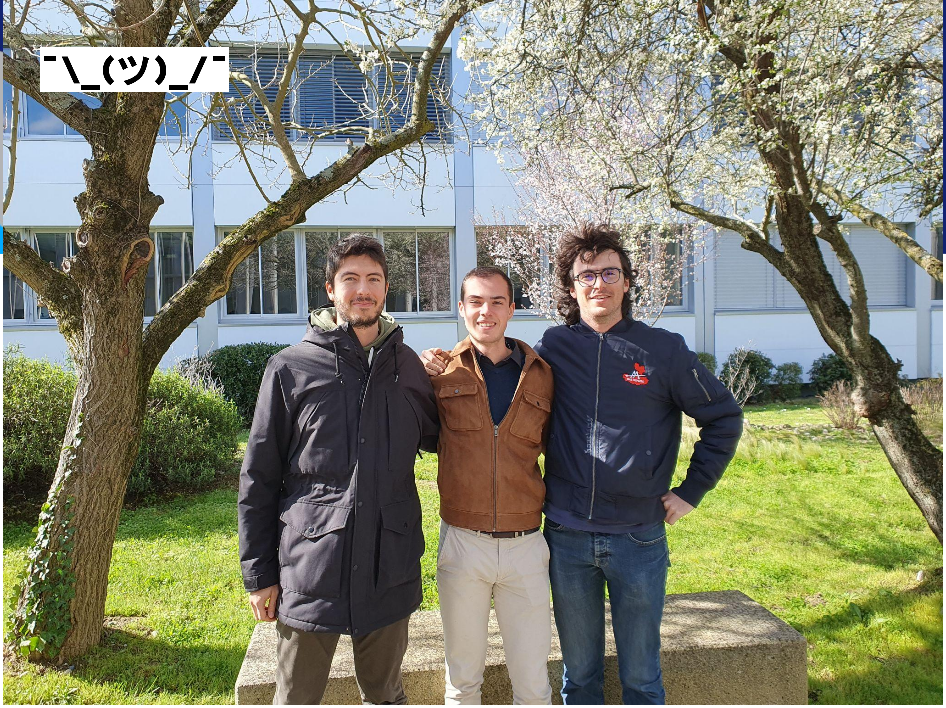
The knowledge of a watercourse's flow rate is important for various purposes, including **flood characterization** (or low-flow, drought), management of hydraulic structures, agriculture (through irrigation), survival of aquatic species, and many other subjects. This knowledge is even more crucial in addressing broader challenges such as **climate change**, water sharing, and biodiversity.

Vortex-io micro-stations already provide two essential real-time data points: water height and surface flow velocity. The remaining element to determine is thus the **geometry of the flow section**, crucial information for quantifying the volume of water in transit and, consequently, for generating an estimate of the flow rate.

The goal here is to provide an algorithm to identify **water-occupied areas in the micro-stations captured images**. Those "water masks" will allow for calculating the planar geometry of the watercourse at that specific height. By combining water height measurements with the complete history of captures at different levels, reconstructing the precise geometry occupied by the water surface in overlapping horizontal "layers" would become possible.






¯_(ツ)_/¯


waterbenders


OBJECTIF TOP 4


LOULOU Team

Slides d’équipe gagnante

Image segmentation : Loulou Team

Modèles essayés :

- U-net : différentes architectures
- DeeplabV3
- K-means (OpenCV)
- Threshold segmentation

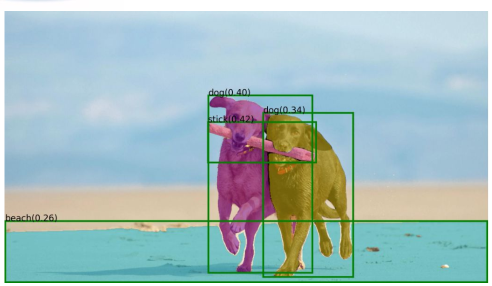
Modèle retenu :

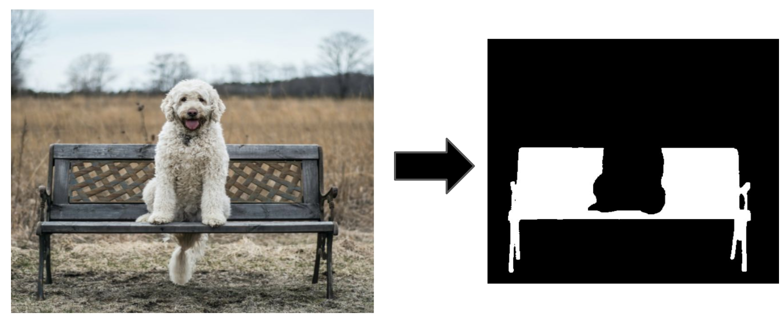
Grounded-SAM : pré-entraîné

Pre-processing et données :

- Data Augmentation
- Images jour/nuit
- Weight classes

Grounded-SAM





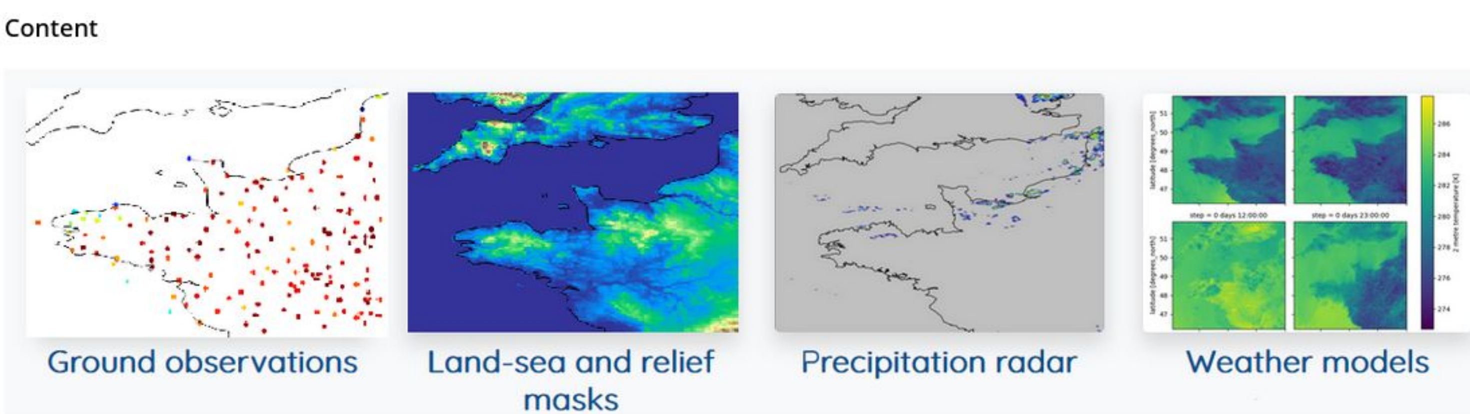
Subject 3: Daily Rainfall Forecasting

Responsable : Léa Berthomier

Forecasting the daily rainfall can prevent you from coming back from work soaked from head to toe, but more importantly, it can help anticipating extreme events such as floods or hurricanes.

Your mission, should you choose to accept it, will be to forecast the rainfall over the next 24h using data from the previous day : ground observations (temperature, wind, pressure, rainfall…) and numerical weather forecasts from METEO-FRANCE. (These forecasts are based on the equations of the physics of the atmosphere).
You will have to learn from the past errors of the numerical weather models to forecast the rainfall on several ground stations.

The data comes from MeteoNet, an open weather dataset for AI : https://www.kaggle.com/katerpillar/meteonet
It spans the North-West quarter of France from 2016 to 2019.




Totanalist Spies


Evelyne Dhéliack
SUPAERO

Construction du dataset d'entraînement et pre-processing

AROME + ARPEGE

Données de prévision de précipitation heure par heure

→ **Application de métriques pour un <u>résumé journalier</u> :**

- moyenne
- médiane
- min
- max

\+

Observations des précipitations à la journée à J-1

→ **Compromis : Remplacement des valeurs manquantes (NaN) par la médiane**

→ Dataset de 120 colonnes

Variables

Prévisions des modèles numériques AROME et ARPEGE

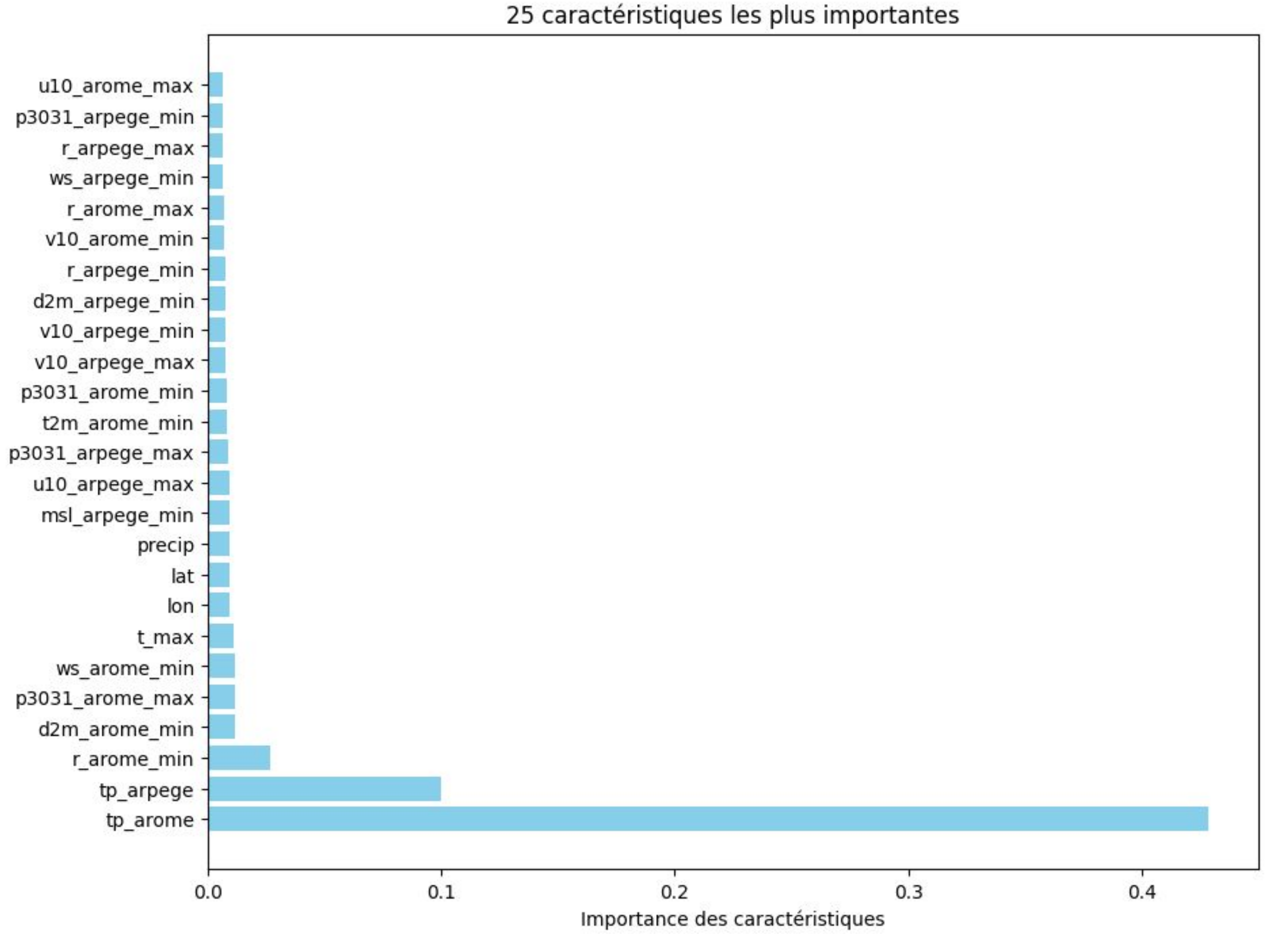
- number_sta : n° de la station
- time : jour de la prévision
- valid_time : heure de validité de l'échéance
- t2m : température à 2m (K)
- d2m : Point de rosée à 2m (K)
- r : relative humidity (%)
- msl : mean sea level pressure (Pa)
- ws : wind speed (m/s)
- p3031 : wind direction (°)
- u10, v10 : composantes vecteur vent à 10m (m/s)
- tp : total precipitation (kg/m^2)

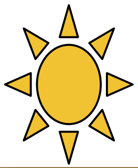
+

Mesures d'observations

- date de la mesure
- n° de la station
- Id : number_sta + index day + hour → need Xi to predict Yi
- ff : Wind speed, m.s-1
- t : Temperature, Kelvin (K)
- td : Dew point, Kelvin (K)
- hu : Humidity, percentage (%)
- dd : Wind direction, degrees (°)
- precip : Precipitation during the reporting period (kg.m2 = mm)

Sélection de variables : RandomForest

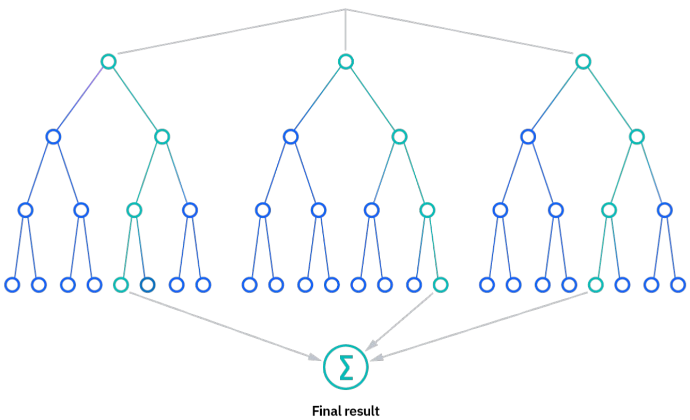




→ Mieux !

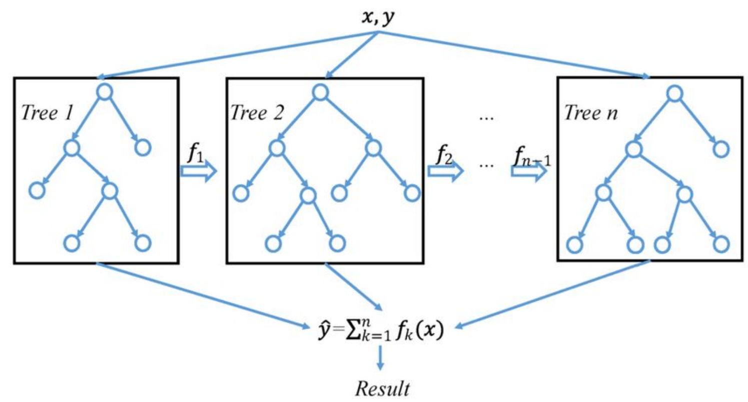
Sélection du modèle

RANDOM FOREST CLASSIFIER



Classifieur pour prédire s'il y a de la pluie ou non

LGBM REGRESSOR

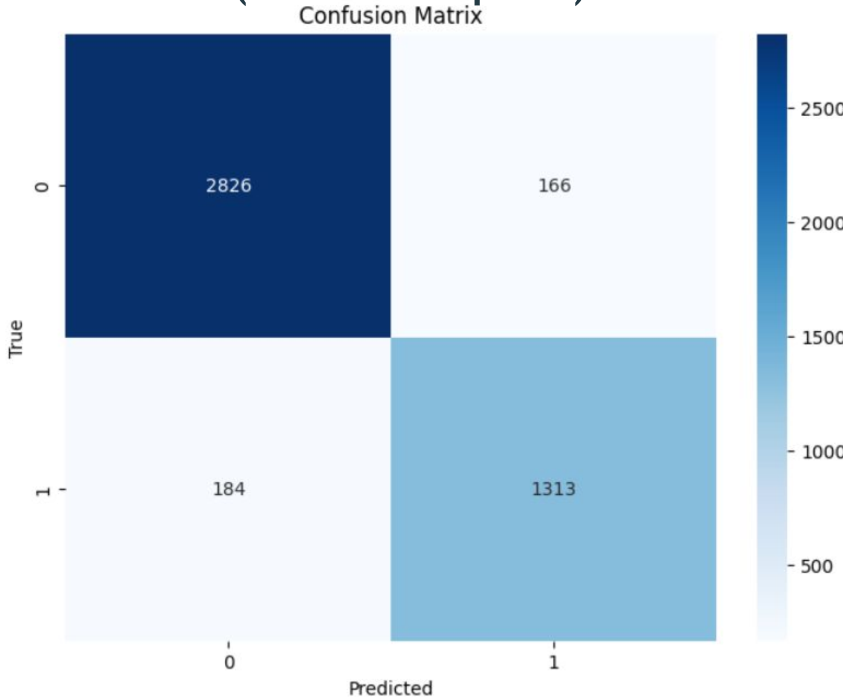


Régression pour estimer la quantité de pluie sur la journée

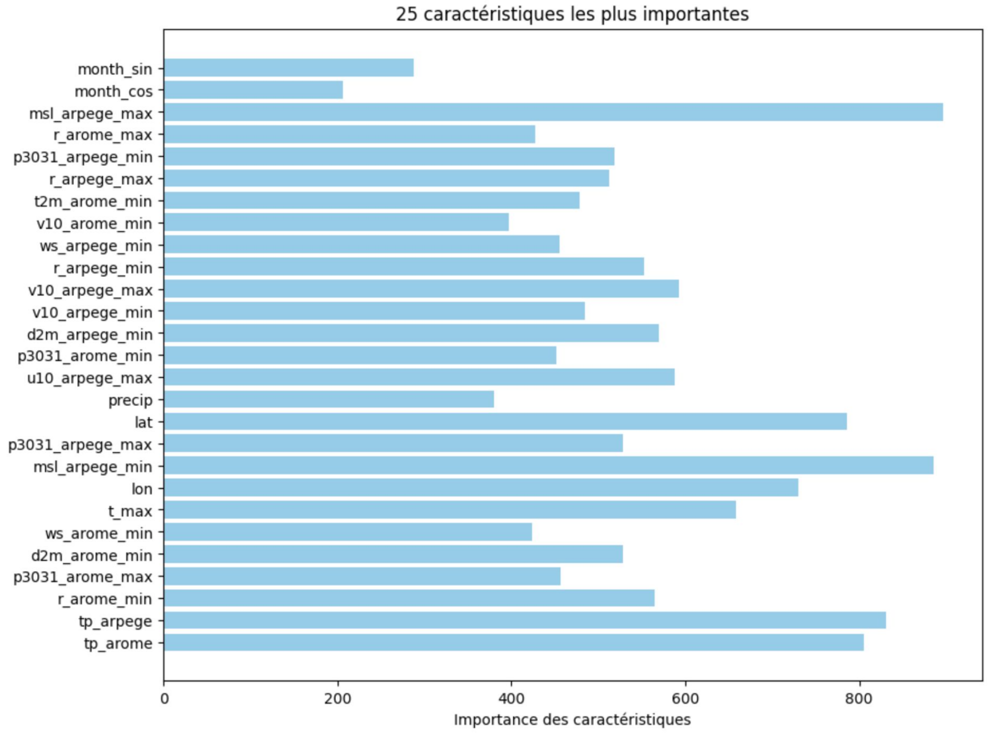
Utilisation de la loss L1 et de l'erreur absolue pour optimiser notre métrique (MAPE)

Sélection du modèle

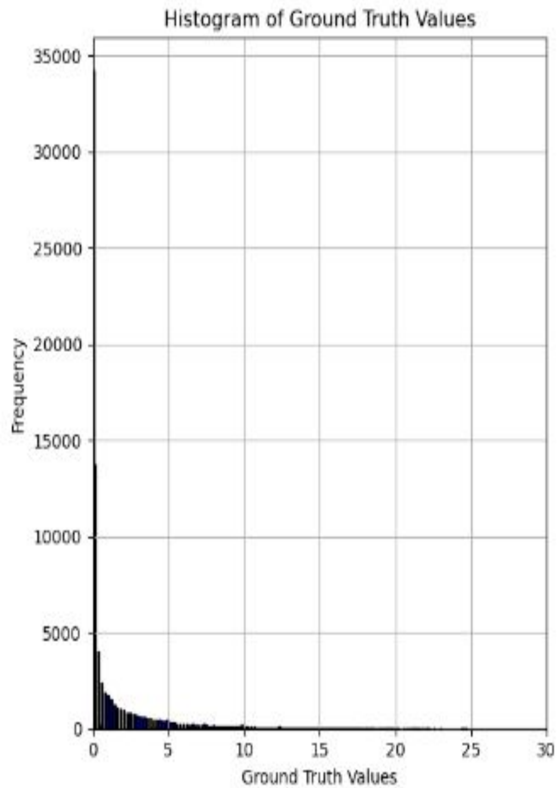
Matrice de confusion de la classification (oui ou non pluie)

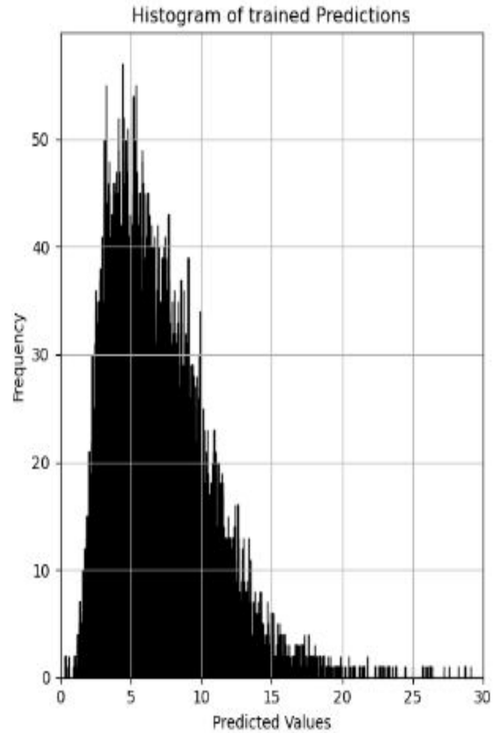


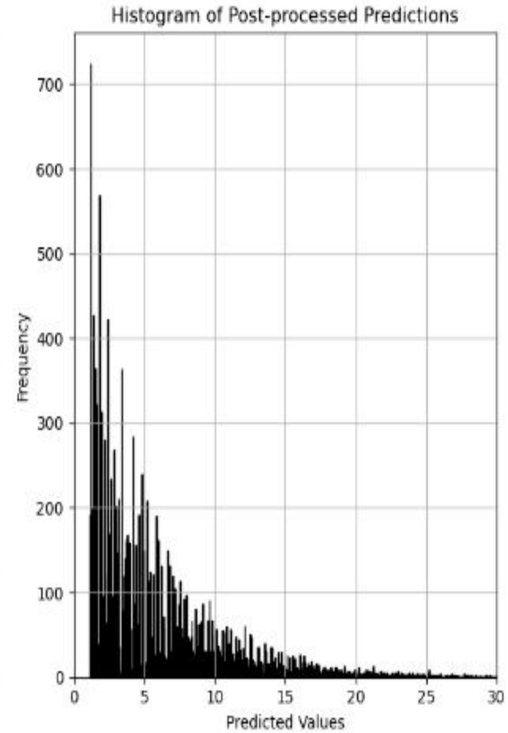
Importances des caractéristiques selon LGBM



Postprocessing: égalisation d’histogramme

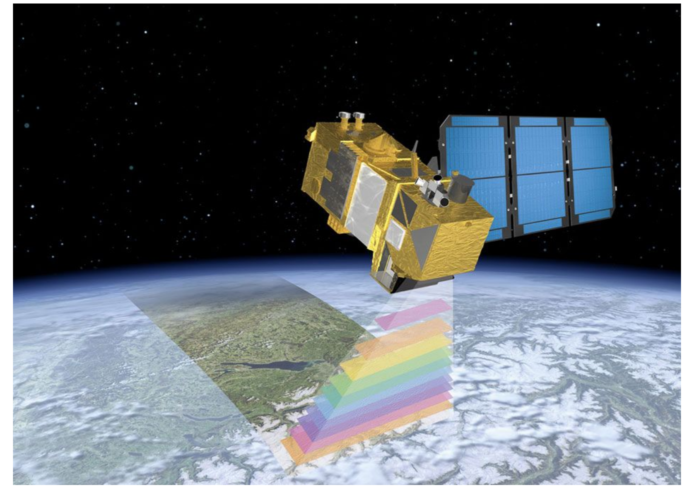






Motivation: Modifier la distribution des prévisions sur les ground truth de l'entraînement pour mieux coller aux précipitations réelles.

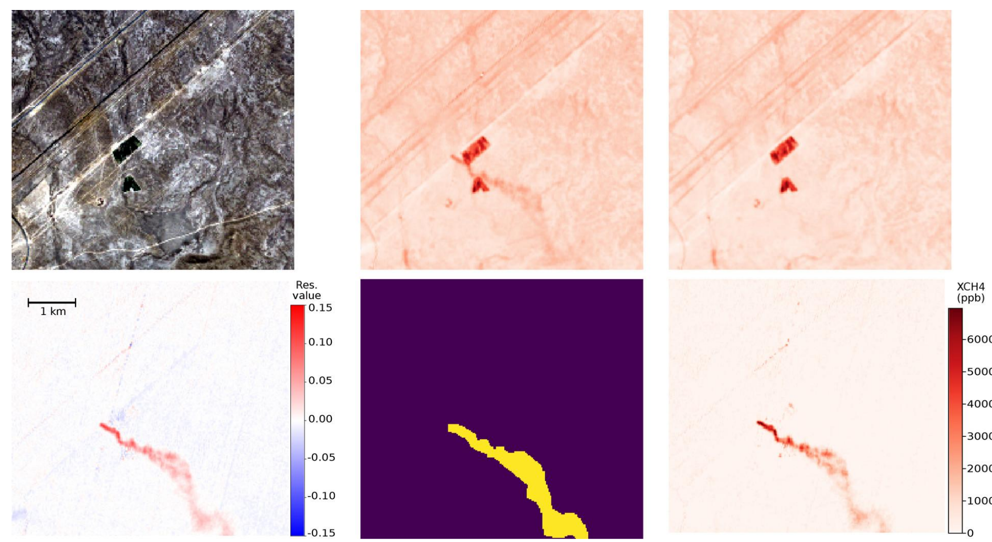
Subject 4: Methane Detection



Responsable : Dennis Wilson

Methane, a potent greenhouse gas with over 80 times the warming potential of carbon dioxide over a 20-year timeframe, plays a significant role in climate change. Understanding and mitigating global methane emissions is crucial for mitigating its impact on the planet.

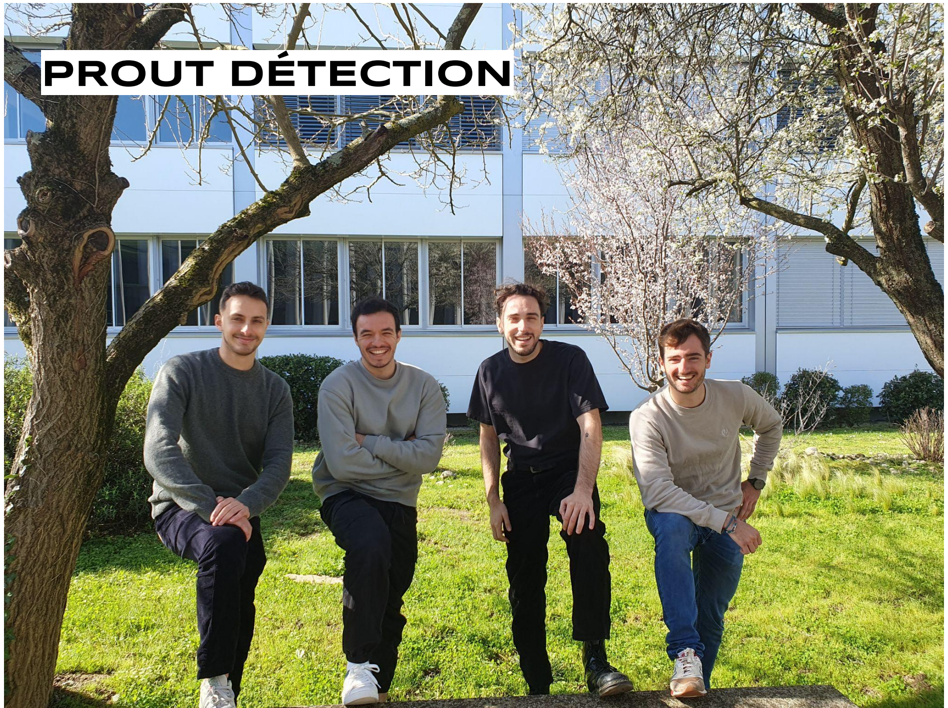
Goal: While satellite data holds immense potential for methane detection, leveraging it effectively requires sophisticated analysis techniques. Deep learning has emerged as a promising approach for extracting meaningful insights from large datasets like satellite imagery. This competition aims to harness the power of deep learning for methane detection from Sentinel-2 satellite imagery.

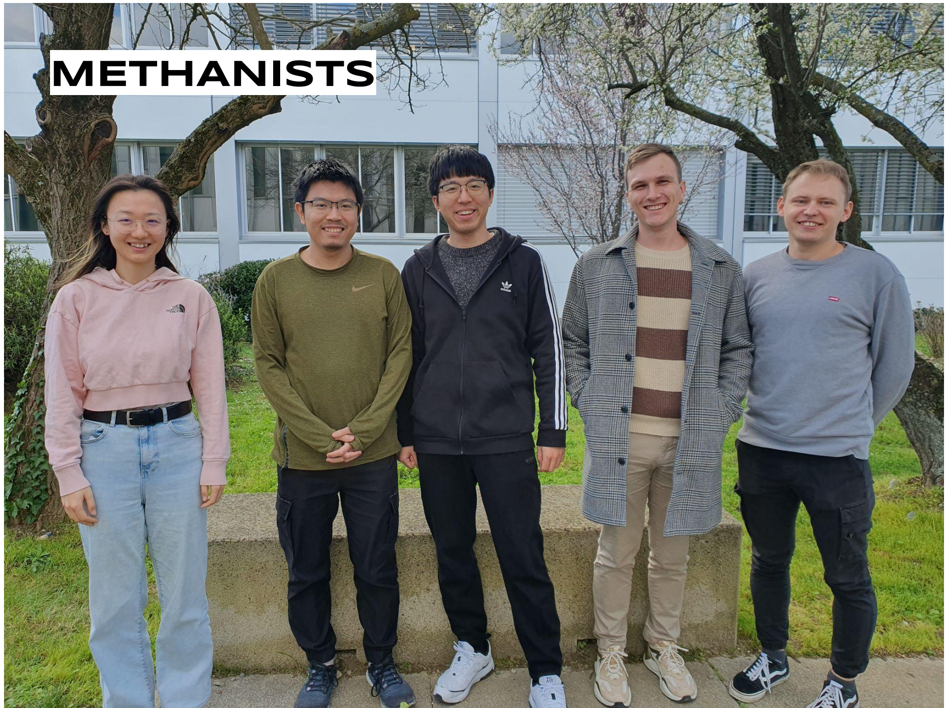


Institut Supérieur de l'Aéronautique et de l'Espace
SUPAERO

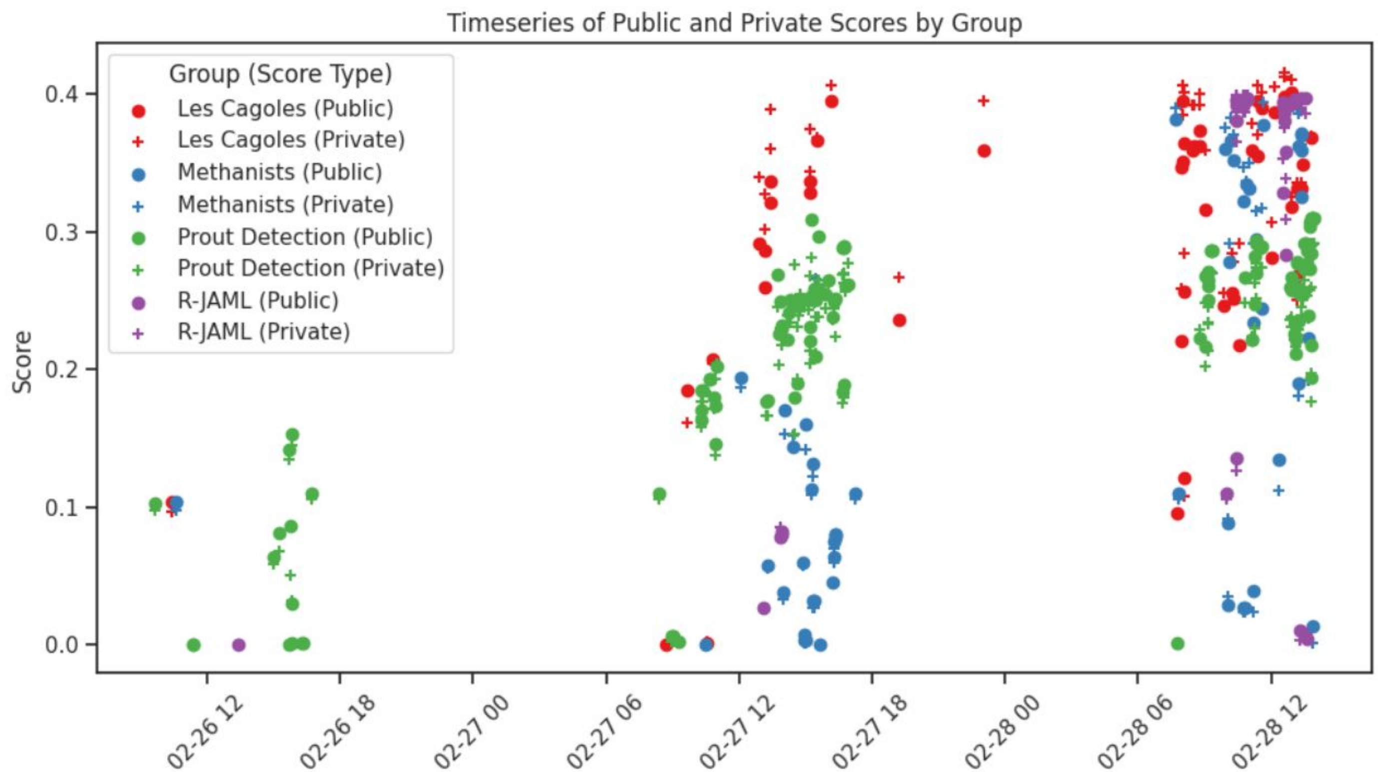

R-JAML


LES CAGOLES

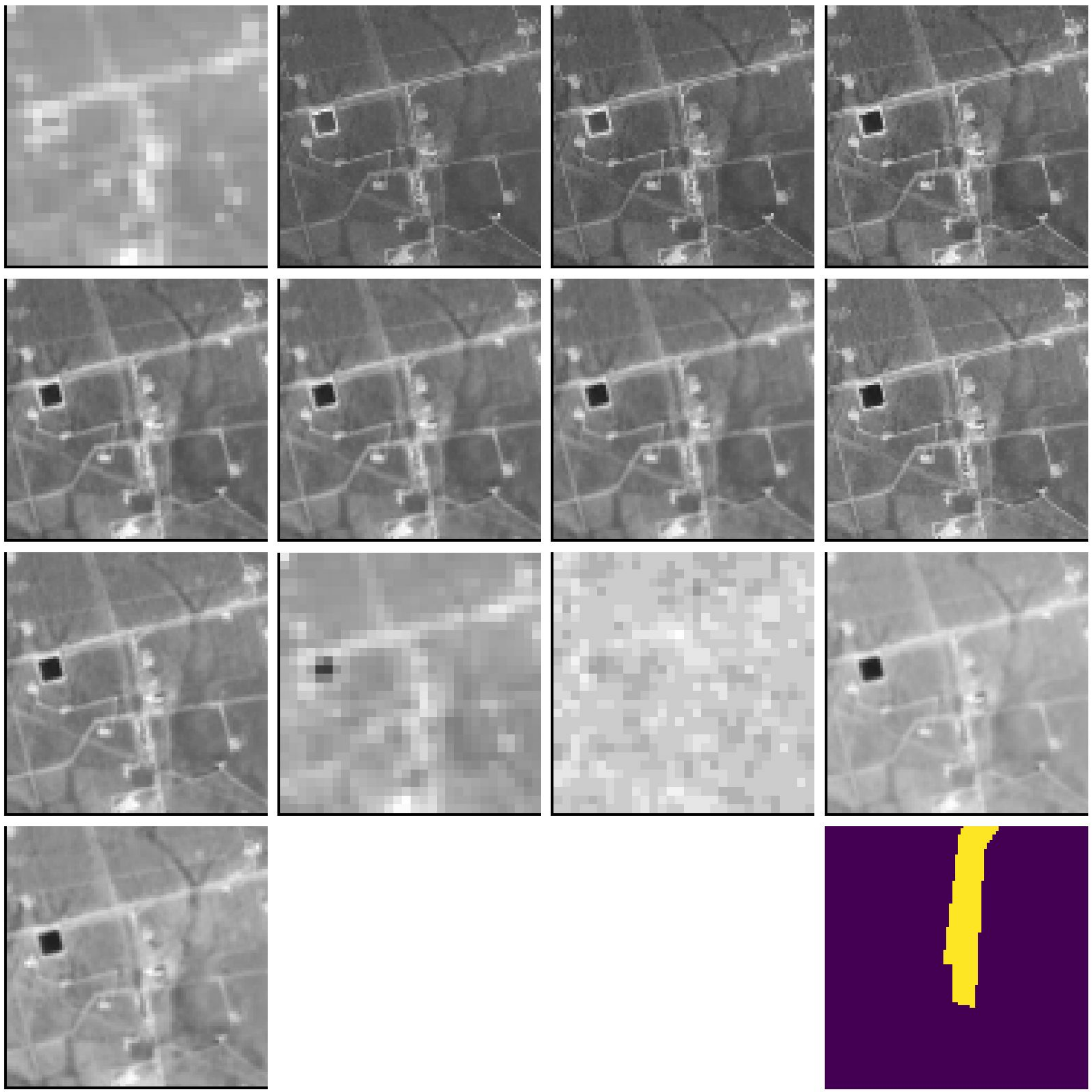

PROUT DÉTECTION


METHANISTS

#	△	Team	Members	Score	Entries	Last	Solution
1	—	Les Cagoles		0.41593	48	39m	
2	—	R-JAML		0.39932	36	1h	
3	—	Methanists		0.39348	52	32m	
4	—	Prout Detection		0.29155	124	30m	
		submission_data.csv		0.08879			



Input: 13 channels / Output: 1 binary mask



Methodology

Modèle utilisé : **UNet (~80 epochs)**

Data engineering = normalisation par canal sur toutes les images
Pas de Data augmentation

Modification de la fonction loss:

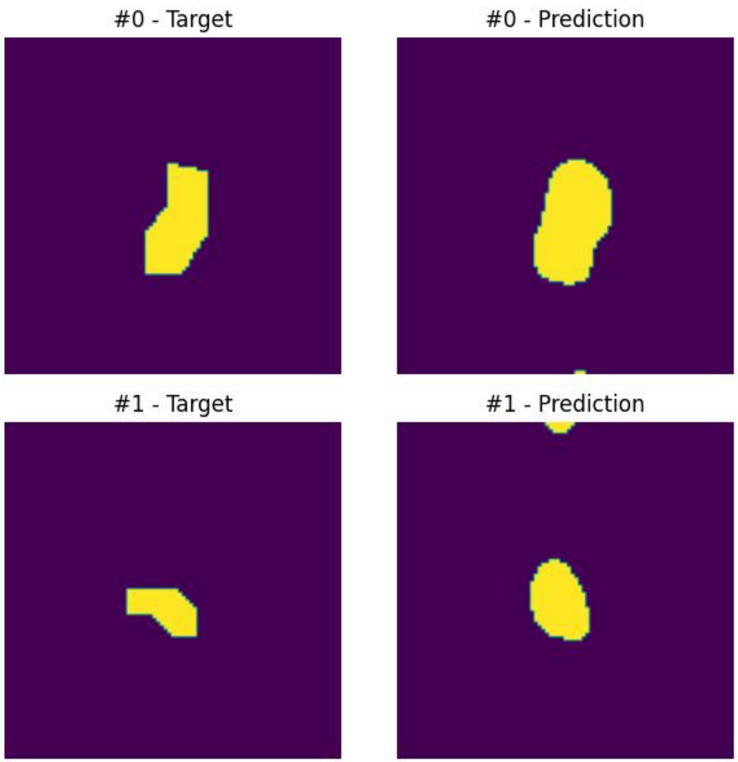
$$\textbf{\textit{BCE}} + \alpha * \textbf{\textit{DiceLoss}} - \beta * \textbf{\textit{PixelMean}}$$

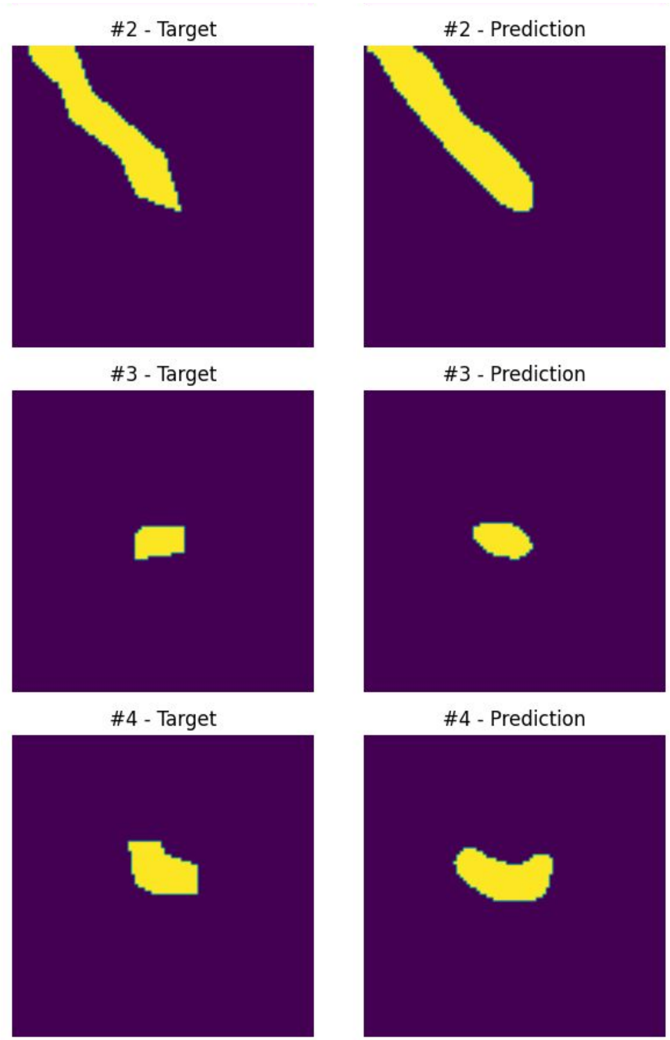
- BCE = Binary Cross Entropy
- DiceLoss = 1 - Dice coefficient
- PixelMean = Mean value of the predicted pixels

Modèles testés : 3 layers Convolution, UNet, FCN_ResNet50 (0.20), DeepLab_V3 (0.23), LARSPP + dropout (0.30), UNet avec 13 canaux + prédiction d'un UNet entraîné (0.36)

Results

Résultats – Prédictions





Hackathon SDD 2024

Wrap-up