



# **MODÈLES DE DISTRIBUTION D'ESPÈCES**

**ALVIN VEDEL | FLORIAN DUBOIS | MATIS BREILLAD**

# SOMMAIRE

**01**

## PRÉSENTATION

CONTEXTE ET PROBLÉMATIQUE

KAGGLE ET PROJET

DONNÉES À UTILISER

GESTION DU PROJET

**02**

## PREMIERS MODÈLES

MODÈLE NAIF ET KAGGLE

ALGORITHME DES K-VOISINS

RANDOM FOREST

**03**

## PROBLÈMES ET AMÉLIORATION

DONNÉES DE PRÉSENCES

AUTRES DONNÉES NON UTILISÉES

FUTURS MODÈLES

# ENVIRONMENTAL CONTEXT

- Le réchauffement climatique et autres
- Important de classifier les espèces



# PROJET KAGGLE

- Projet répond à la problématique et le contexte
- Compétition kaggle

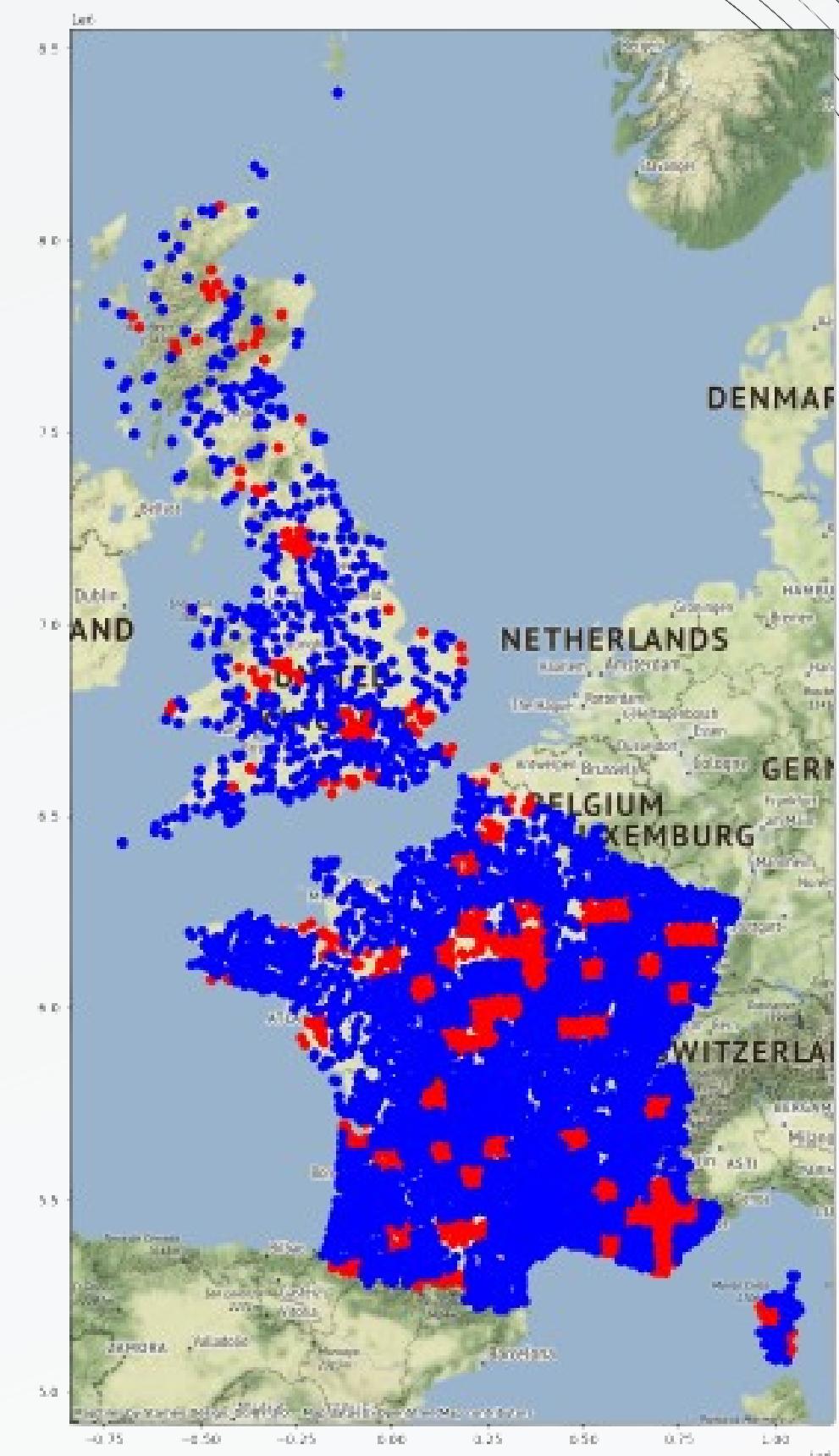


# NOS DONNÉES

- Des données de présences / absences, c'est à dire?

1 OBSERVATION = 1 SPECIESID + 1 PATCHID

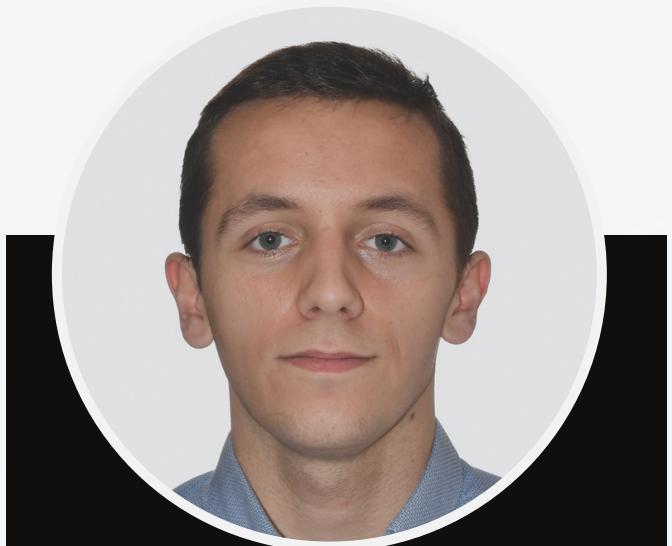
- Des données précises mais relativement peu nombreuses...



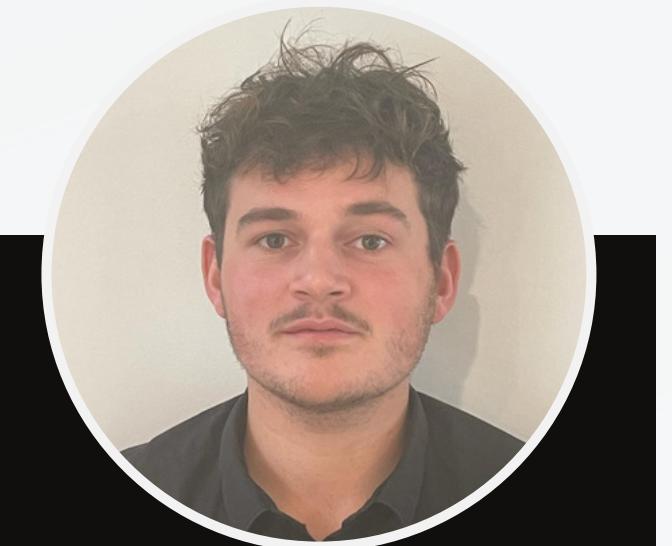
# GESTION DE PROJET



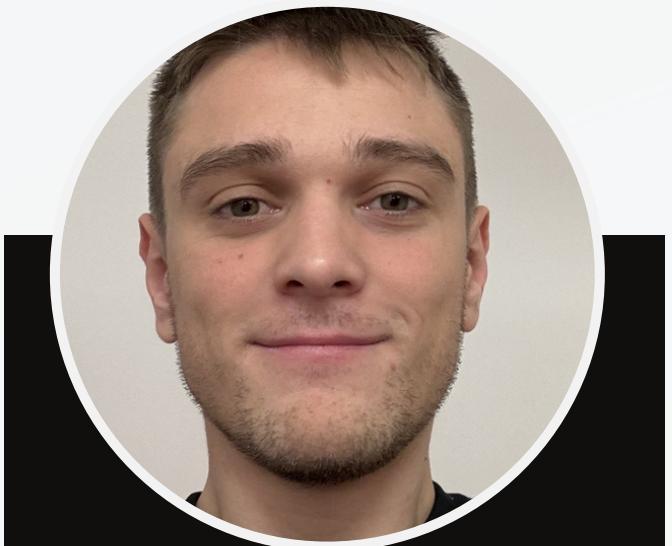
Maximilien  
Servajean  
Responsable



Alvin Vedel  
Chef du groupe



Matis Breillad  
Retardataire du  
groupe



Florian Dubois  
Il a fait les KNN

# MODELE NAIF ET RÉSULTATS

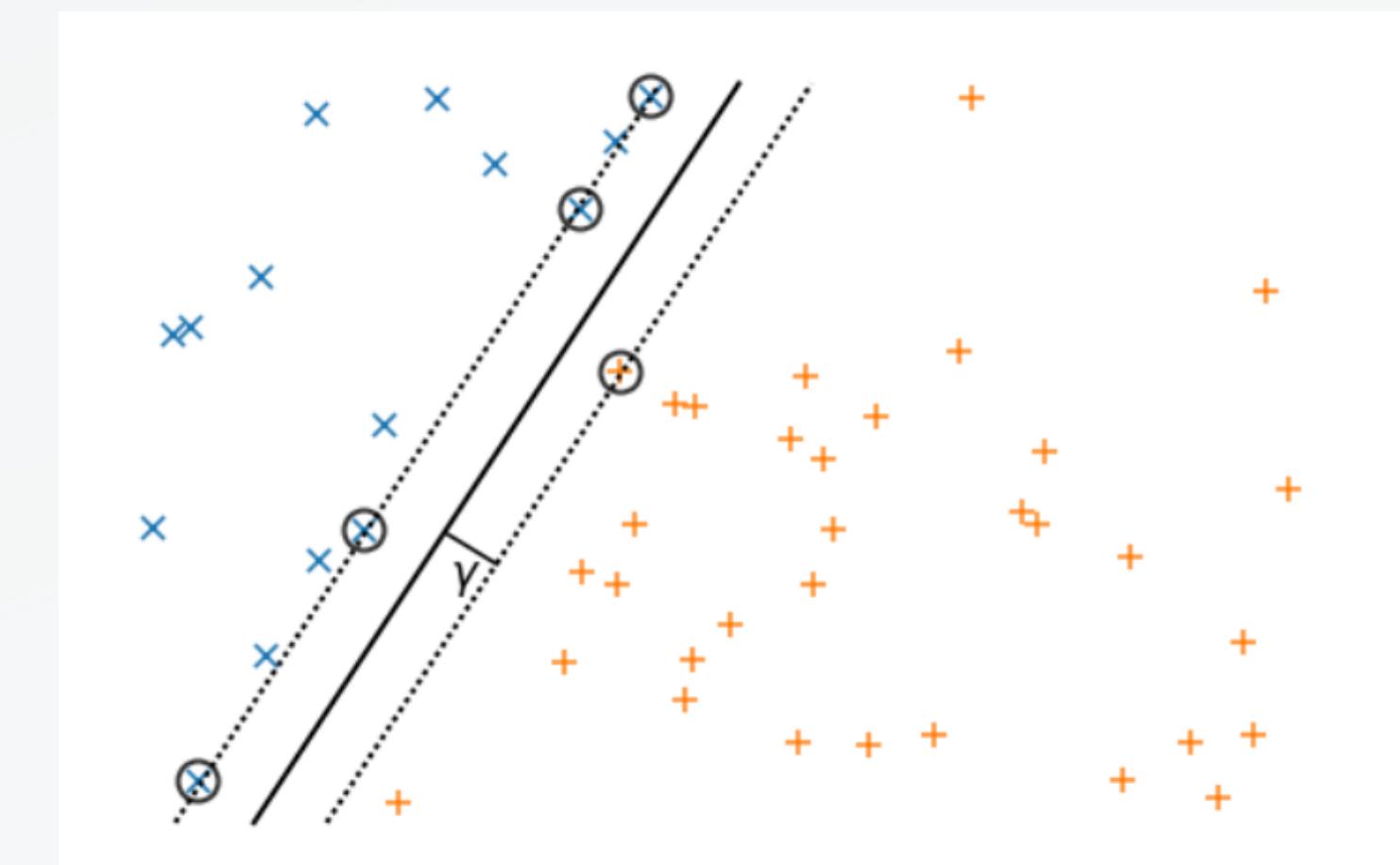
- 2 Types de modèles naïfs
  - Random
  - K plus présents

- Comment comparer les résultats ?

#	△	Team	Members	Score	Entries	Last	Solution
1	—	KDDI Research	  	0.27048	55	7mo	
		Maxent on PA		0.22445			
2	—	JiexunXu		0.22352	17	7mo	
3	—	Lucas Morin		0.20797	13	8mo	
4	—	Quantmetry	 	0.20646	14	7mo	
		Spatial Random Forest (PA)		0.19072			
		Environmental Random Forest (PA)		0.18847			
		Spatial KNN PA		0.16622			

# TENTATIVE D'UN MODÈLE SVM

- Pré-traitement des données
  - Le principe d'un SVM
- Résultats



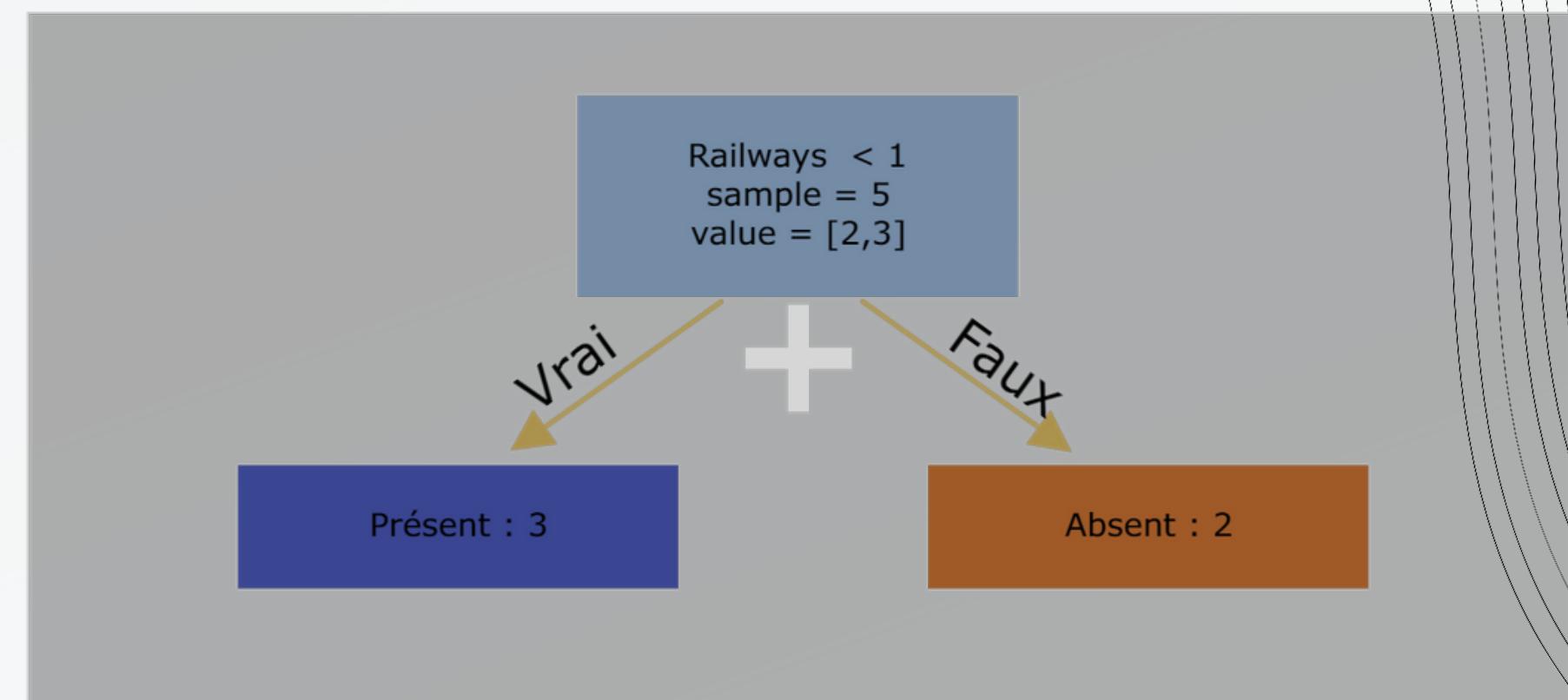
# ALGO DES K-VOISINS

- C'est quoi un K voisins
- Expliquer nos résultats

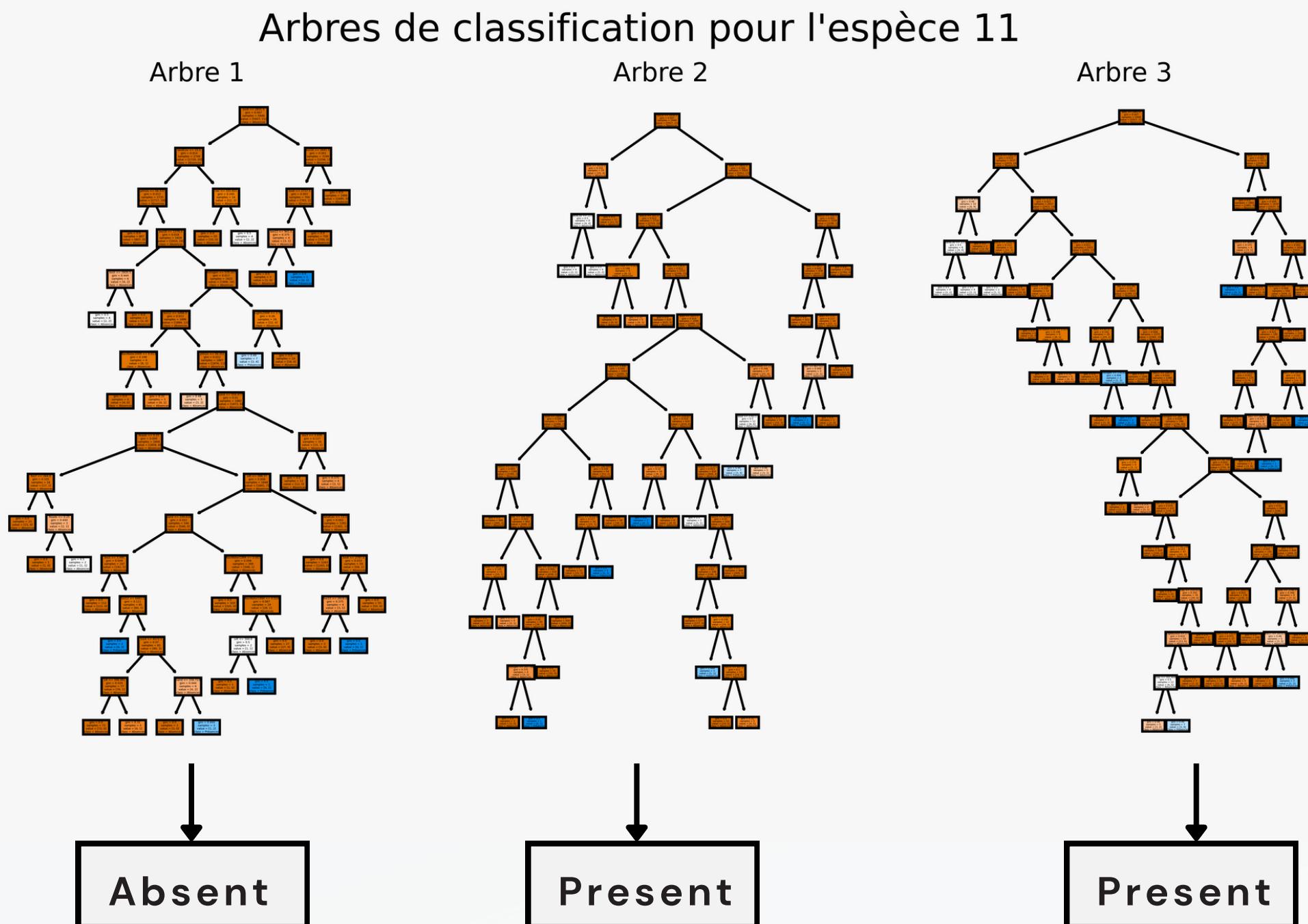
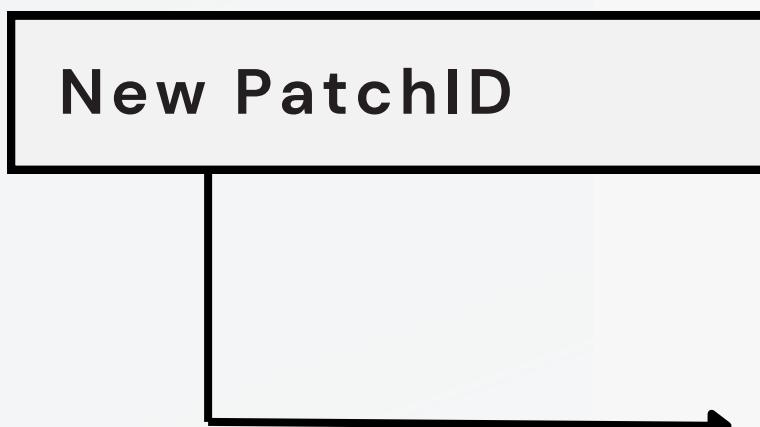
# RANDOM FOREST

patchID	bio1	bio2	Popdensity2010	Railways	11
3770	3010676	2839	50	7.000000	0.000000
3771	3010677	2839	50	7.000000	2.000000
3772	3010677	2839	50	7.000000	2.000000
3773	3010683	2839	50	7.000000	0.000000
3774	3010683	2839	50	7.000000	0.000000

Extrait de nos données



# RANDOM FOREST



Random Forest à 3 arbres

# RANDOM FOREST

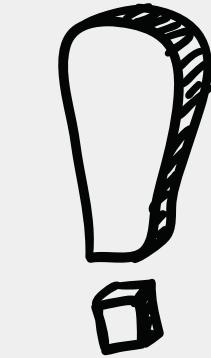
HitMap des scores obtenus

### 03 Problèmes

# DONNÉES DE PRÉSENCES UNIQUEMENT



Les données de présences uniquement :  
ressemblances et différences



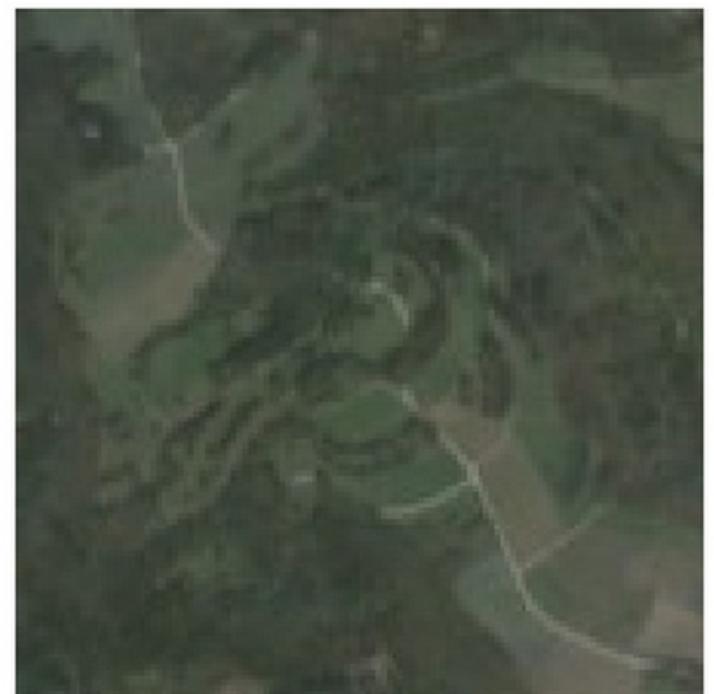
Des données basées à prendre en compte



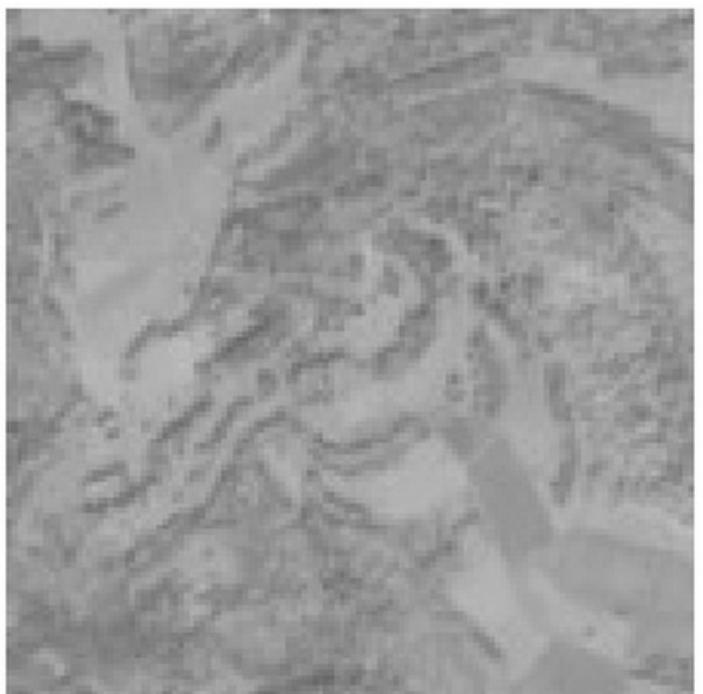
# AUTRES DONNÉES

- Images Satellites
- Série temporelles
- Images matricielles

**A. Satellite image patches**



**Red Green Blue (RGB)**

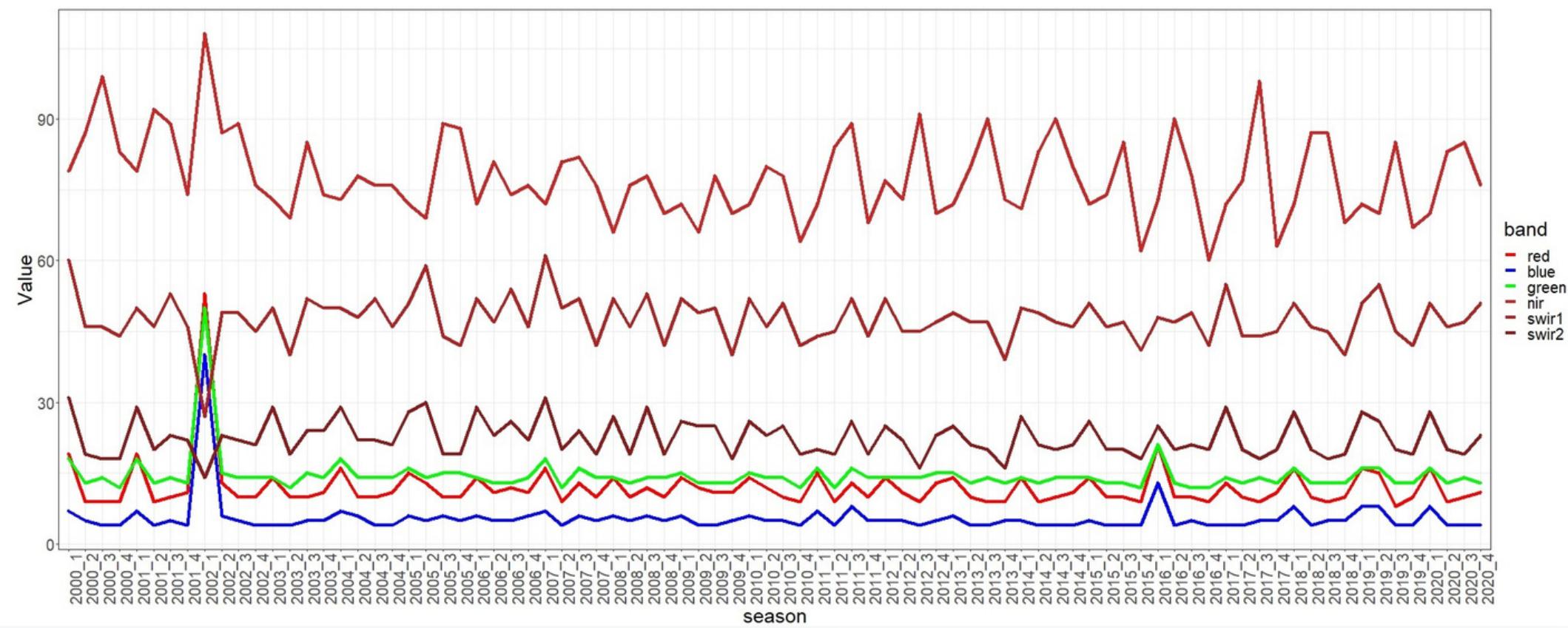


**Near Infra-Red (NIR)**

# AUTRES DONNÉES

- Images Satellites
- Série temporelles
- Images matricielles

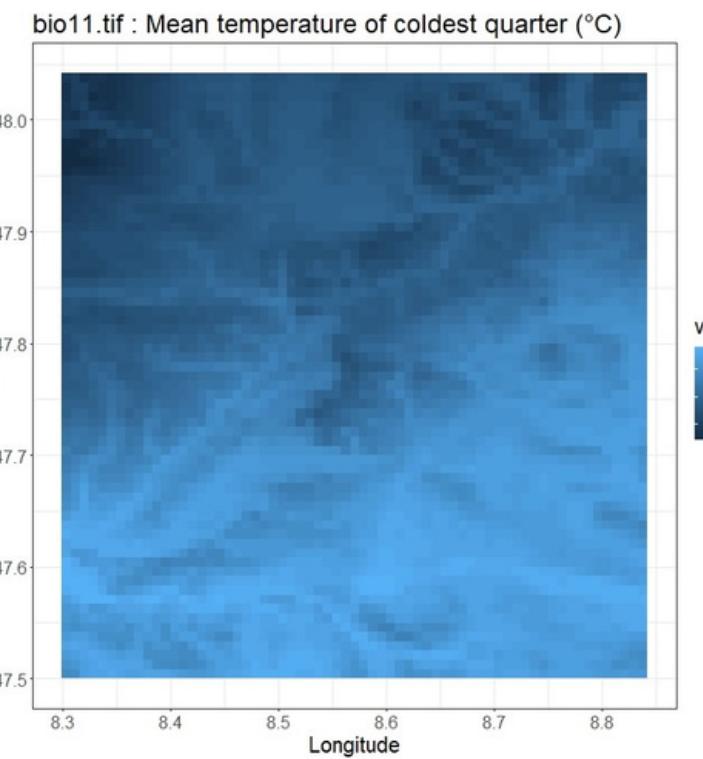
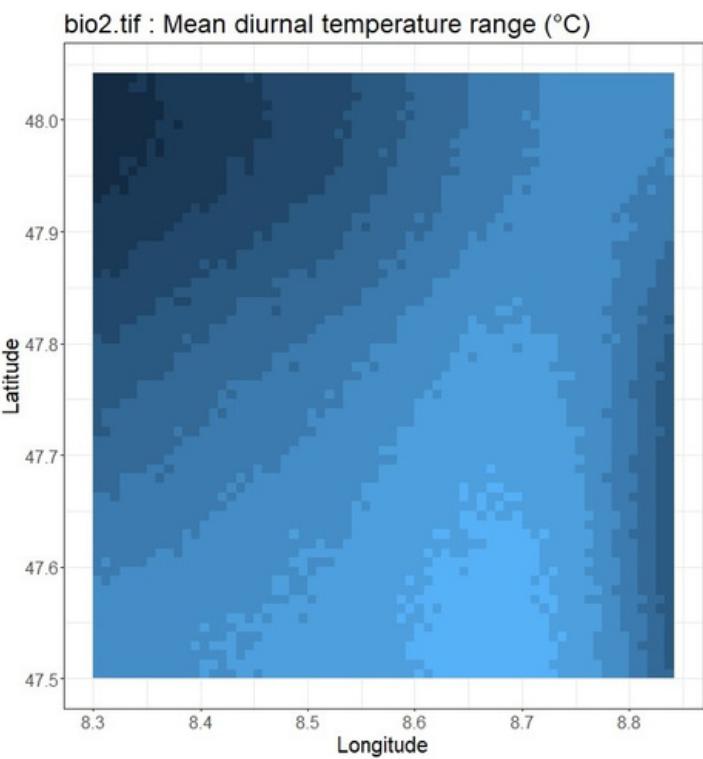
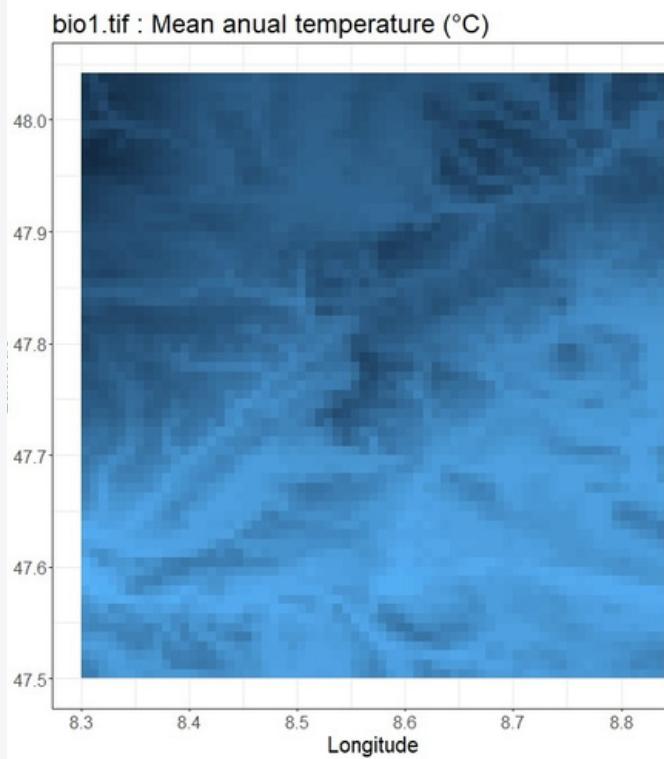
B. Satellite time series



# AUTRES DONNÉES

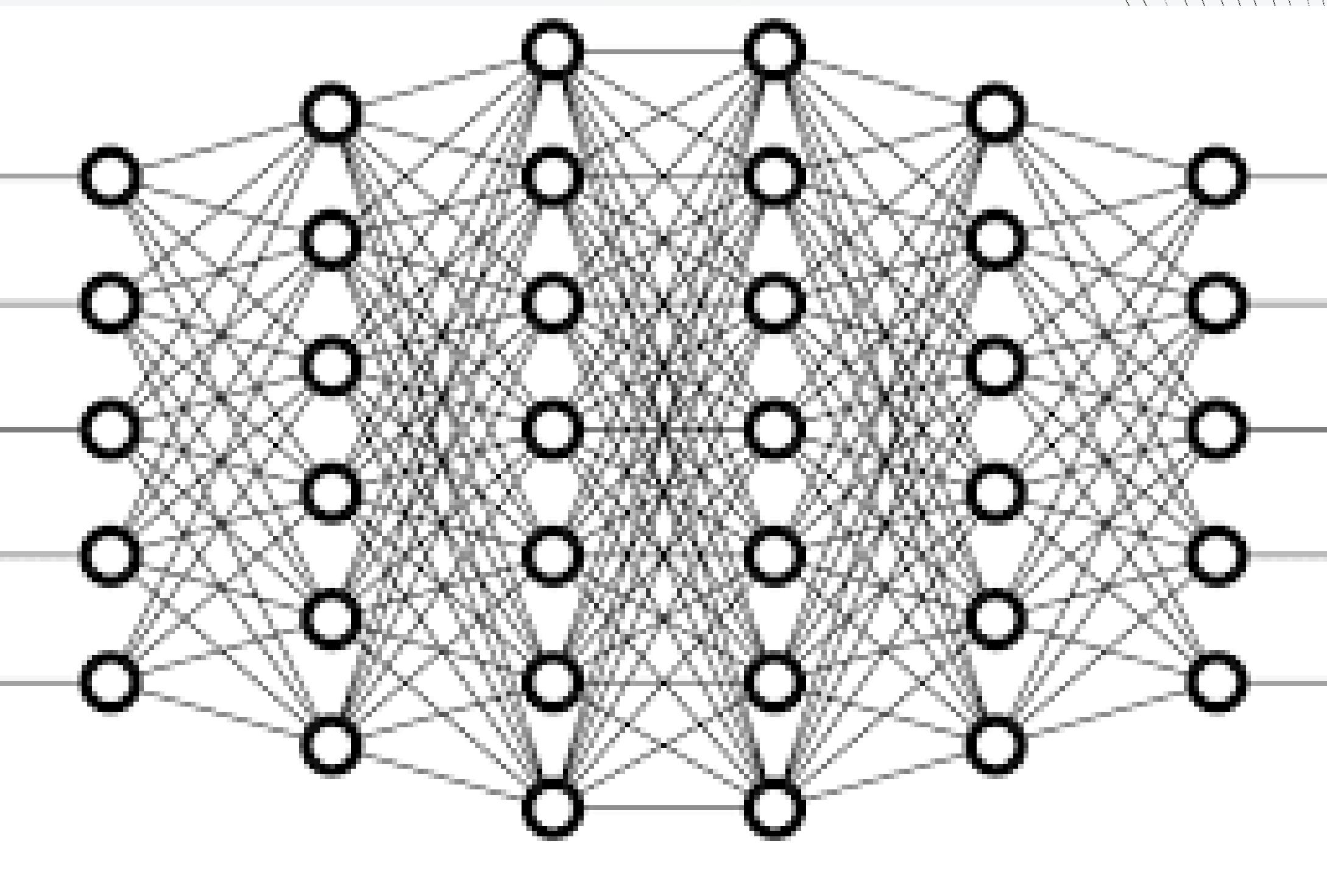
- Images Satellites
- Série temporelles
- Images matricielles

## C. Environmental rasters



# FUTURS MODÈLES

- Modèles simples avec nouvelles données
- Réseau de neurones
- FrameWork Loi de Poisson



# CONCLUSION



On a bien fait des trucs mais il faut d'autres



Avis personnelle sur le projet, satisfait, concrétisation.



# CONCLUSION



On a bien fait des trucss mais il faut d'autres



Avis personelle sur le projet, satisfait,  
concrétisation.

