



20/10/2017

Christophe BOTELLA

Alexis JOLY

Pierre BONNET

Pascal MONESTIEZ

François MUNOZ

Apprentissage profond de la distribution d'espèces de plantes: Retour d'expérience.

Objectifs et dispositif

- **Objectif:** Evaluer le potentiel de modèles à réseaux de neurones profonds et convolutifs en comparant empiriquement les performances à celles de SDM de référence (MAXENT).

- **Données:**

- Observations flore de la FCBN en France métropolitaine agrégées à une maille de 10x10km.
- Descripteurs environnementaux variés (33).

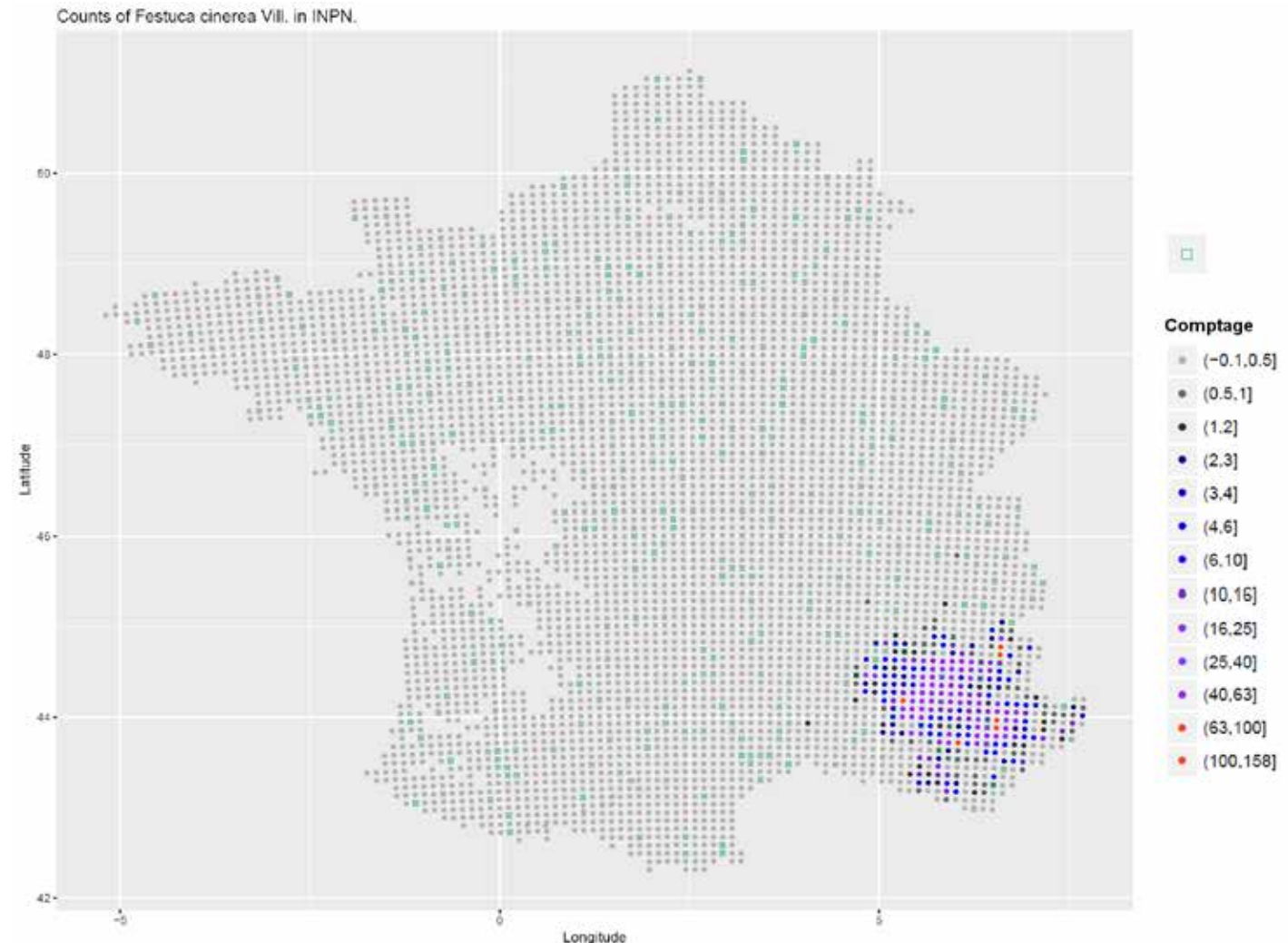
- **Modèle de comptage**

$$y \sim \mathcal{P}(\lambda_{m,\theta}(x))$$

- **4 architectures évaluées**

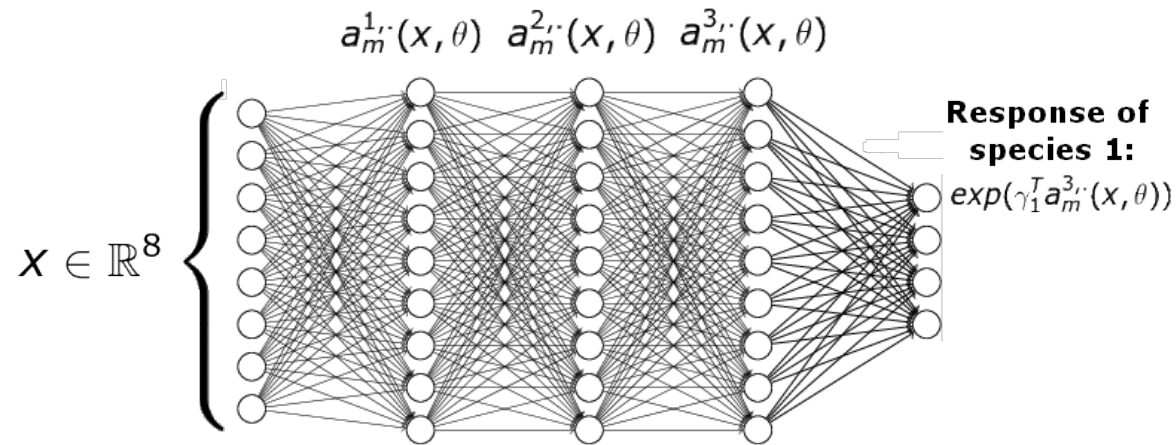
- Loglinear model (LGL) - **baseline**
- Single layer NN (SNN)
- Deep NN (DNN)
- Convolutional NN (CNN)

- **Versions mono et multi-réponses**
(Plusieurs espèces prédites en sortie)



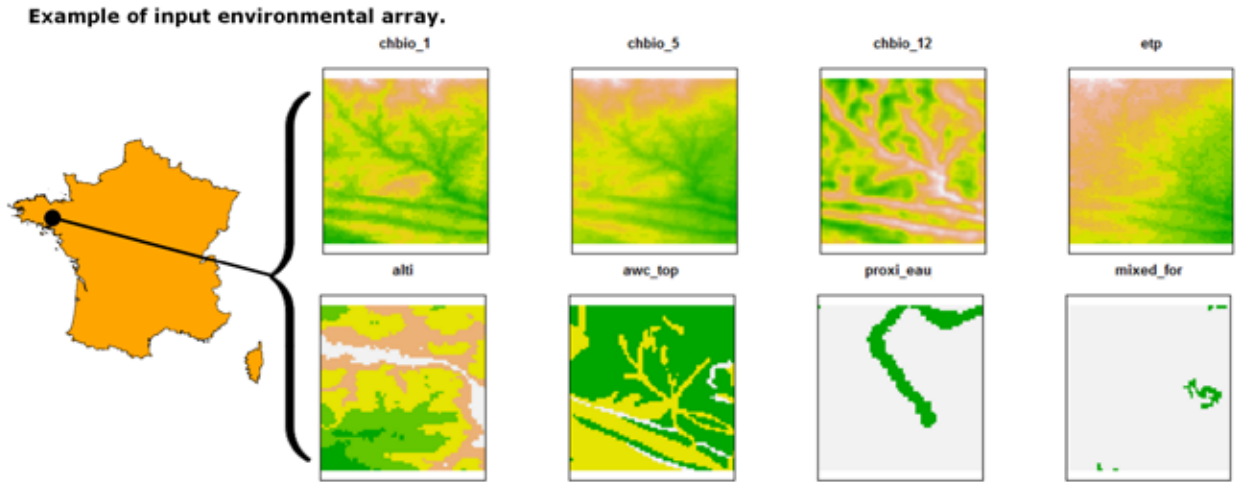
modèles DNN & CNN

DNN schématique

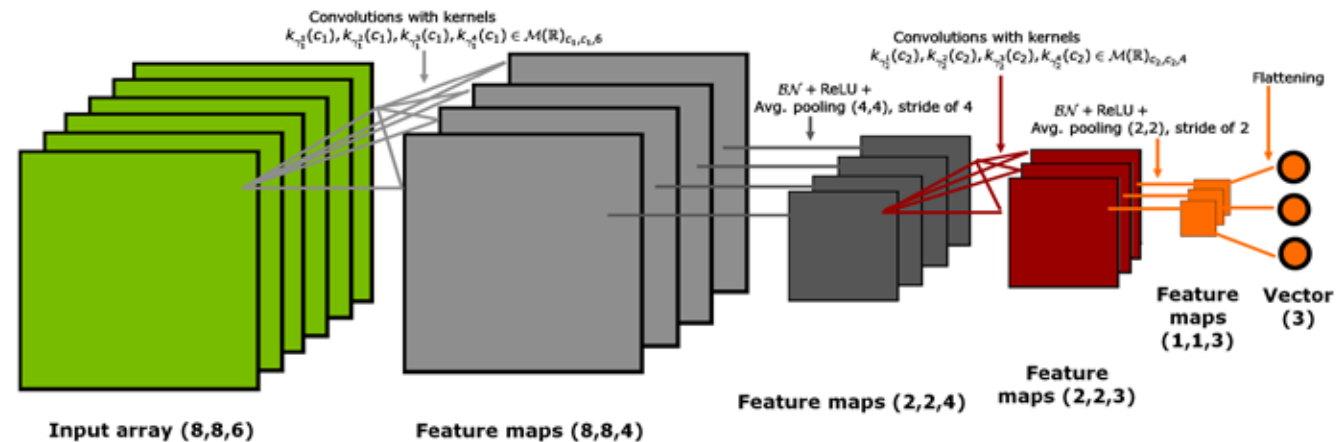


- Ajustement des modèles sous R.
- Package **mxnet**.
- Calcul sur un poste mono-GPU
Nvidia quadro M1000M
- Algo: SGD+momentum:0.9
- Batch-size: 50.
- Environ 7 semaines de travail.

CNN schématique

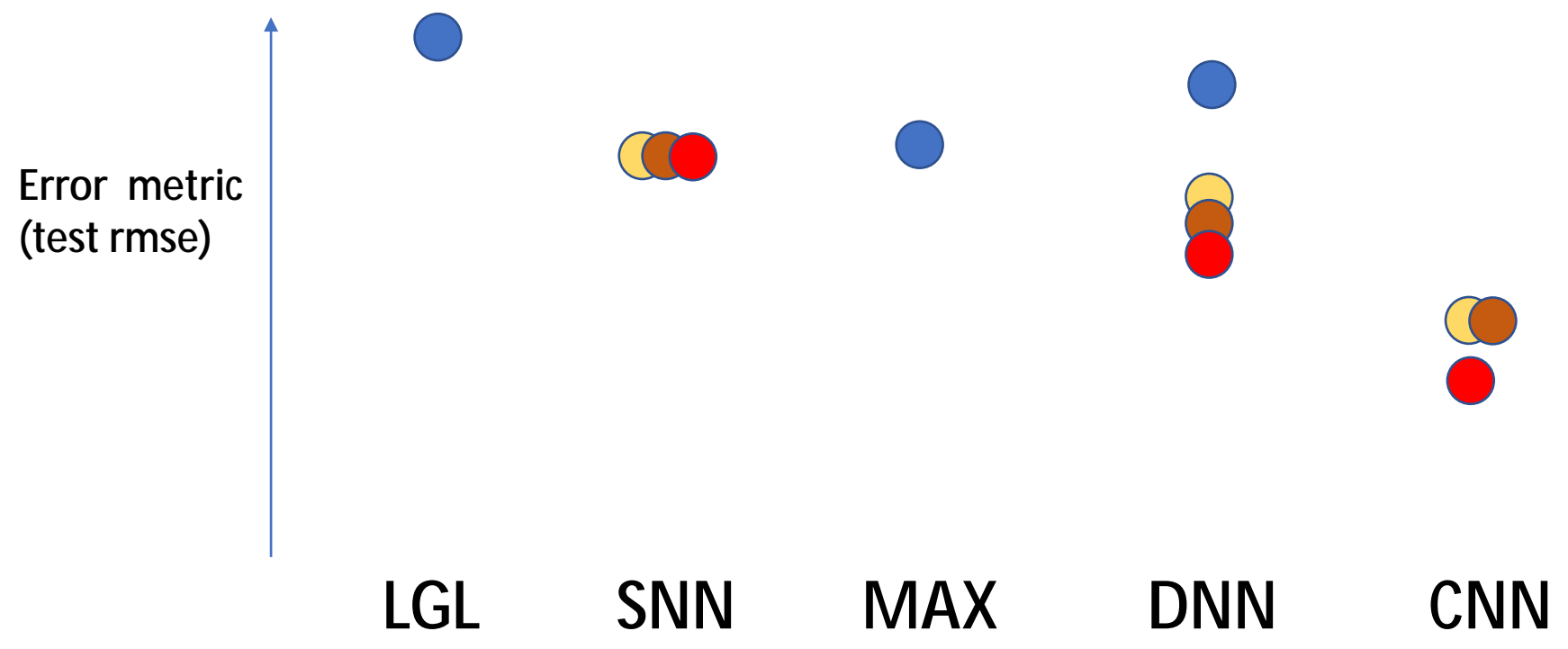


Schematic structure of convolutional layers.





Résultats (en bref).





• Verrous actuels :

- Biais des données:
 - Intérêts hétérogènes sur les espèces entre les CBN.
 - Capacités d'observation hétérogènes entre les CBN.
 - Zones difficiles d'accès et non prospectées.
- Optimisation:
 - Choix du learning rate initial
 - Divergence impromptue.
 - Règle d'arrêt
 - Durée (DNN:10min, CNN: 4h)
- Choix et ajustement de la régularisation à améliorer.
- Interprétabilité.
- Incertitude de prediction
(Ex: conditions environnementales originales).

• Intérêts:

- Apprentissage automatique de features haut niveau:
 - Meilleures prédictions sur espèces rares.
 - Etudier les co-abondances entre espèces.
 - Ré-utiliser pour d'autres modèles.
- Des cartes pour 3000 espèces en 4h.

• Perspectives:

- Etudier les propriétés DNN & CNN.
- Méthodes de selection et d'interpretation écologique des features.
- "Data-augmentation": points uniformément répartis par sites.
- Fine-tuning à partir PI@ntNet sur INPN.

