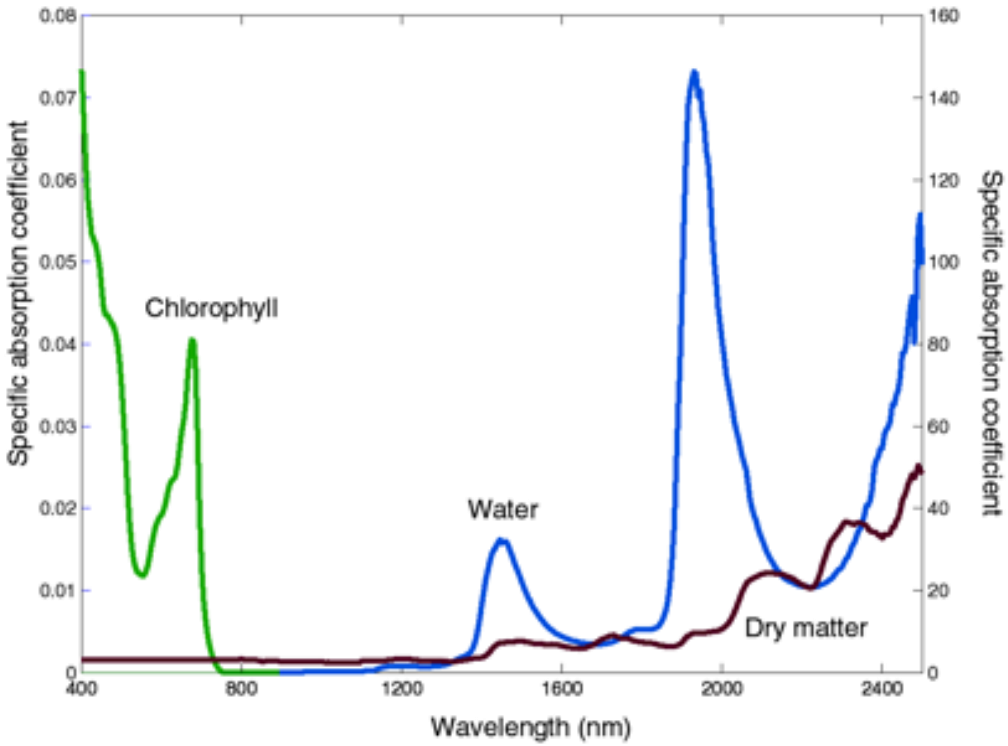
# Estimation de la chimie foliaire par spectroscopie, apprentissage automatique et modélisation physique

Florian de Boissieu, Dino Ienco, Jean-Baptiste Féret

Financemé par le projet TOSCA HYPERTROPIK

# Contexte

- Deux traits foliaires d'intérêt sont spécifiquement étudiés ici:
  - Teneur en eau (Equivalent Water Thickness, EWT, mg/cm$^2$)
  - Masse sèche (Leaf Mass per Area, LMA, mg/cm$^2$)



# Contexte

- Deux traits foliaires d'intérêt sont spécifiquement étudiés ici:
  - Teneur en eau (Equivalent Water Thickness, EWT, $mg/cm^2$)
  - Masse sèche (Leaf Mass per Area, LMA, $mg/cm^2$)
- Estimation de EWT et LMA a partir des propriétés optiques foliaires:
  - Inversion du modèle physique PROSPECT par optimisation itérative:
    - Bonnes performances pour EWT, variables pour LMA selon domaine spectral utilisé
    - Capacité de généralisation maximale
    - Temps de calcul important : passage à l'échelle compliquée en hyperspectral
  - Modèle statistique / apprentissage automatique:
    - Bonnes performances mais potentiellement surévaluées
    - Nécessite données d'apprentissage en quantité
    - Modèles peu généralisables (sites, protocole expérimental, matériel…)
    - Risque de sur-apprentissage

# Objectifs et interrogations

- Déterminer une méthode hybride optimale:

  Apprentissage automatique sur modèle physique

  - Mise au point d'une base d'apprentissage simulée par PROSPECT
  - Détermination de l'algorithme d'apprentissage automatique approprié
  - Performances et capacité de généralisarion pour l'estimation de LMA et EWT

- Peux-ton améliorer les performances d'estimation avec ce type de modèle?

- Le choix de l'intervalle spectral et l'ajout d'un bruit gaussien aux propriétés optiques d'apprentissage permet-il d'améliorer les performances?

J.-B. Féret *et al.*, "Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2742–2750, 2011.

# Matériels

- Données expérimentales:
  - ANGERS (308 échantillons) et LOPEX (320 échantillons)
  - $R(\lambda)$ et $T(\lambda)$ avec $\lambda \in [900, 2400]$ nm
- Simulations PROSPECT
  - $10^4$ combinaisons aléatoires de EWT, LMA et structure foliaire (N)

| | EWT ($g.cm^{-2}$) | LMA ($g.cm^{-2}$) | N |
|---|---|---|---|
| Min | 0 | 0 | 0.8 |
| Max | 60 | 40 | 3.8 |

  - Simulation de $\hat{R}(\lambda)$ et $\hat{T}(\lambda)$ avec $\lambda \in [900, 2400]$ nm
  - Bruit additif : $\varepsilon(\lambda) \in N(0, \sigma)$ avec $\sigma = \{0, 0.5, 1, 2, 3, 4, 5\}\ 10^{-2}$

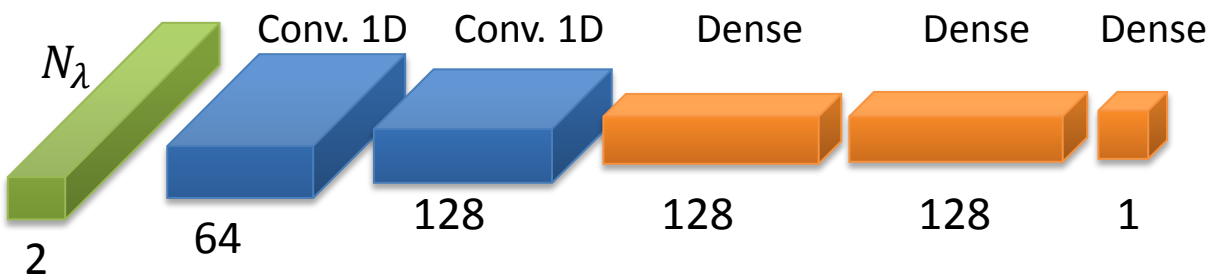
# Méthodes 1/2

- Inversion PROSPECT: Optimisation itérative avec contrainte (SQP)

$$\min_{N,EWT,LMA} f(N, EWT, LMA) = \sum_{\lambda=\lambda_1}^{\lambda_n} \left[ \left(R_\lambda - \hat{R}_\lambda\right)^2 + \left(T_\lambda - \hat{T}_\lambda\right)^2 \right]$$

$$N \in [0.5, 5]\ ;\ EWT \in [10^{-3}, 150]\ \mathrm{mg \cdot cm^{-2}}\ ;\ LMA \in [10^{-3}, 40]\ \mathrm{mg \cdot cm^{-2}}$$

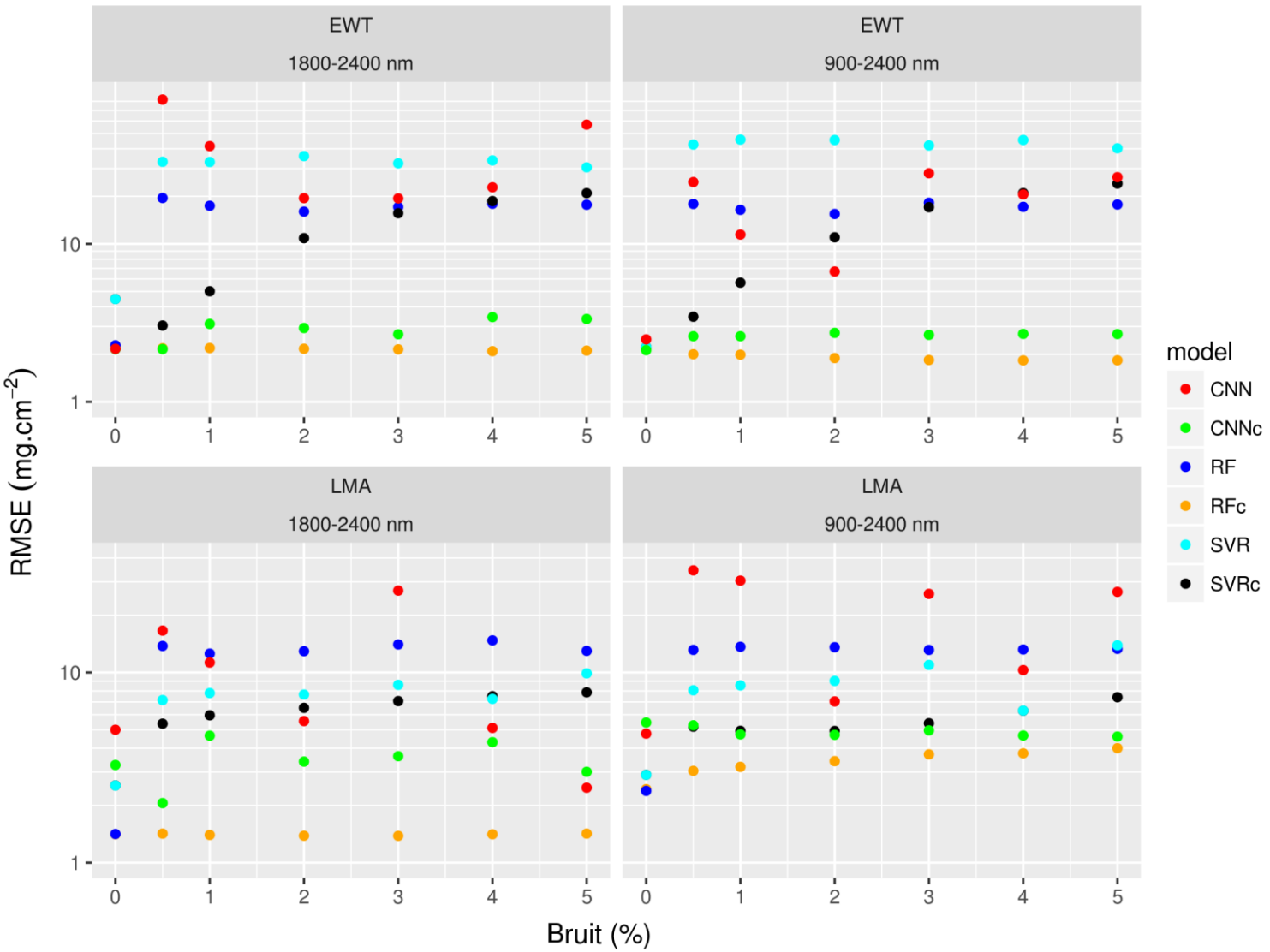
- Algorithmes de régression par apprentissage automatique:
  - ϵ-SVR: RBF, $C = 1, \gamma = 1/2N_\lambda$
  - RF: 400 arbres
  - CNN1D:



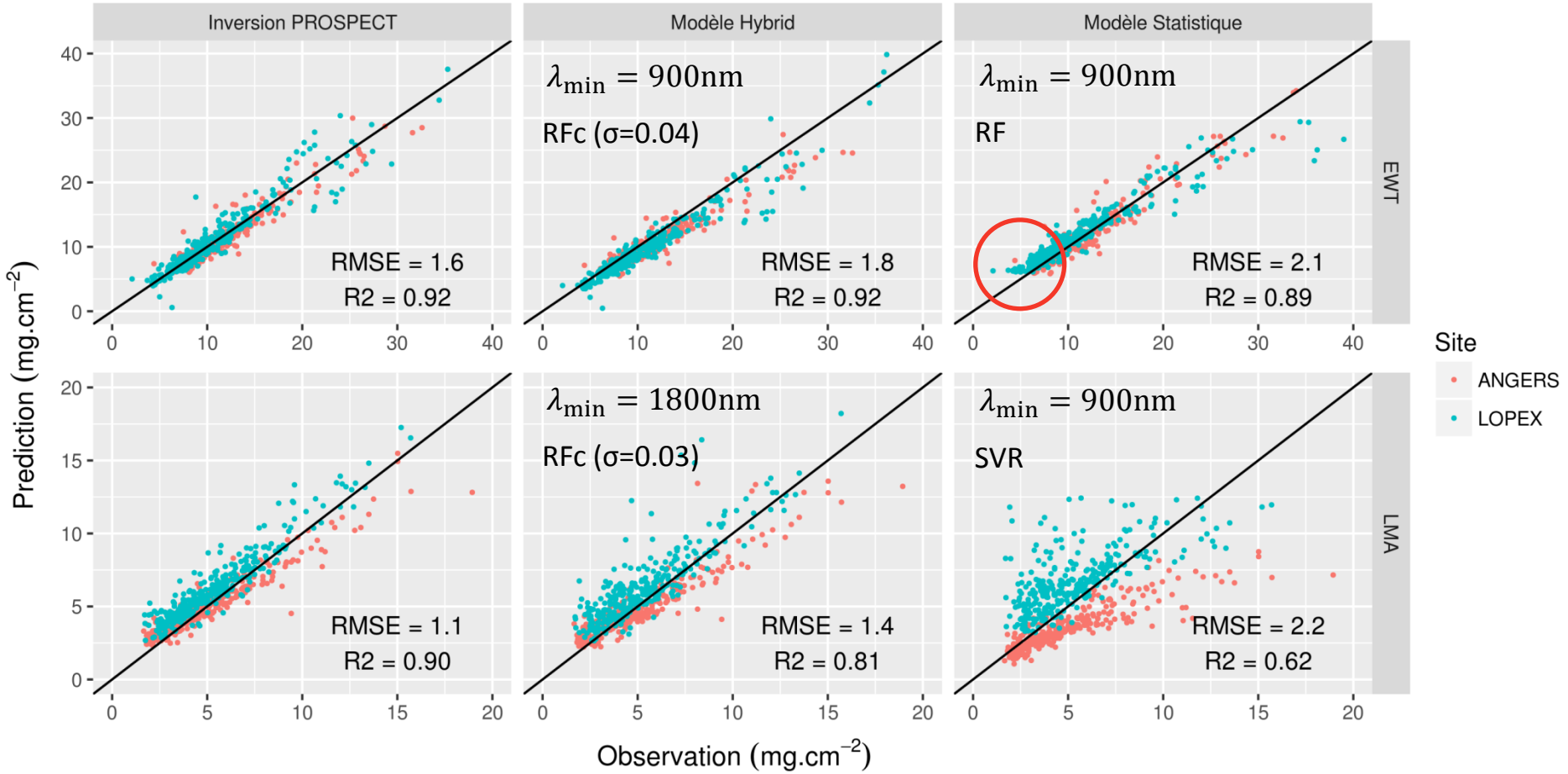
# Méthode 2/2

- Modèles par apprentissage automatique sur données expérimentales:
  - Algorithmes: SVR, RF
  - Apprentissage sur un site (cross-validation)
  - Test sur l'autre site
- Modèles hybrides:
  - Algorithmes: SVR, RF, CNN
  - Apprentissage sur données simulées
    - 1 distribution de bruit
    - Diversité de bruits (« c »:  cumulé )
  - Test sur les données expérimentales
- $\lambda$: [900, 2400]nm et [1800, 2400]nm

# Résultats : modèles hybrides



# Résultats: meilleurs modèles



# Conclusions & Perspectives

- Performances RF similaires au modèle physique
- Adaptation de la plage de $\lambda$ améliore les performances pour LMA
- L'ajout de bruit gaussien aux données d'apprentissage simulées a peu d'effet sur les performances des modèles hybrides, contrairement aux modèles statistiques simples ajustes a partir d'indices spectraux

Work in progress …

- Modèles hybrides encore améliorables
- A tester sur des images hyperspectrales
- Inversion PROSPECT sur GPU