



**HAL**  
open science

## Prédiction des épidémies de cercosporiose de la betterave par des approches d'apprentissage automatique.

Lucile Vallet, François Brun, Olivier Gauriau, François Joudelat, Céline Gouwie, Ghislain Malatesta, Juliette Adrian, Fabienne Maupas

### ► To cite this version:

Lucile Vallet, François Brun, Olivier Gauriau, François Joudelat, Céline Gouwie, et al.. Prédiction des épidémies de cercosporiose de la betterave par des approches d'apprentissage automatique.. VÉGÉ-PHYL -13ème Conférence internationale sur les maladies des plantes, Dec 2022, Orléans, France. pp.1-3. hal-03918080

**HAL Id: hal-03918080**

**<https://hal.science/hal-03918080>**

Submitted on 2 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**VÉGÉPHYL – 13<sup>e</sup> CONFÉRENCE INTERNATIONALE SUR LES MALADIES DES PLANTES**  
**ORLÉANS – 6, 7 et 8 DÉCEMBRE 2022**  
**PREDICTION DES EPIDEMIES DE CERCOSPORIOSE DE LA BETTERAVE PAR DES APPROCHES**  
**D'APPRENTISSAGE AUTOMATIQUE.**

L. VALLET<sup>(1)</sup>, F. BRUN<sup>(1)</sup>, O GAURIAU, F. JOUDELAT<sup>(2)</sup>, C. GOUWIE<sup>(2)</sup>, G. MALATESTA<sup>(2)</sup>, J. ADRIAN<sup>(2)</sup>, F. MAUPAS<sup>(2)</sup>

<sup>(1)</sup> Acta – les instituts techniques agricoles, Toulouse-Paris, France, francois.brun@acta.asso.fr

<sup>(2)</sup> ITB – Institut Technique de la Betterave, Paris, France

Mots-clés : apprentissage automatique, betterave, Bulletin de Santé du Végétal, *Cercospora beticola*, cercosporiose, OAD.

## RESUME

### INTRODUCTION

La cercosporiose constitue une maladie majeure de la betterave avec jusqu'à 30 % de perte de rendement et une baisse de la richesse en sucre de 1 à 2 points, en cas d'épidémie non contrôlée.

Ces travaux ont été menés dans le cadre du Projet CERCOCAP : pilotage de la CERcosporiose de la betterave par COuplage entre modèles agro-climatiques et CAPteurs connectés (2020-2023, financement CASDAR, AAP RT) porté par l'ITB, avec comme partenaires l'ACTA et l'université d'Angers.

L'objectif de nos travaux était de construire des outils d'aide à la décision prédictifs en valorisant les nombreuses données sur la cercosporiose de la betterave issues des réseaux d'épidémiosurveillance pour le Bulletin de Santé du Végétal et collectées depuis 2009 dans la partie nord de la France, en mobilisant une démarche d'apprentissage automatique (machine learning).

Premièrement, nous avons cherché à prédire la date d'apparition des premiers symptômes afin d'avoir des informations sur le début de la saison épidémique en amont (mi-juin).

Deuxièmement, nous souhaitons prédire la dynamique épidémique en cours de campagne en prédisant l'évolution de la maladie sur la semaine à venir (J+7) pour mieux piloter la lutte contre la cercosporiose.

### MATERIELS ET METHODES

Le jeu de données collecté dans le cadre du BSV concernant la cercosporiose de la betterave comporte 17359 observations d'incidence de la maladie (% de feuilles atteintes) réparties sur 2349 parcelles et couvrant de manière homogène les 11 années de 2009 à 2020 (année 2017 non disponible).

A partir d'une analyse bibliographique (Vallet, 2021) sur l'influence des facteurs sur la cercosporiose et le recueil de l'expertise, nous avons défini différents indicateurs météorologiques et variables agronomiques à prendre en compte. A partir d'une base de données météorologiques spatialisées (SAFRAN, Météo-France), nous avons calculé différents indicateurs agrégés de manière bimensuelle.

Dans la démarche d'apprentissage automatique (ou *machine learning*), pour prédire la date d'apparition des symptômes, nous avons utilisé des méthodes classiques de régression linéaire, de régression pénalisée (lasso, elastic net) et enfin des méthodes basées sur des prédictions d'ensemble avec des arbres (forêt aléatoire et *gradient boosting*). Pour prédire la dynamique de la maladie en cours de campagne à partir des données passées météorologiques, nous avons utilisé, testé et optimisé différentes structures de réseaux de neurones (convolutif ou perceptron multicouche).

Pour éviter de surévaluer les performances, nous avons gardé 10% des données pour le test final et, sur les 90% pour l'apprentissage, nous avons défini un schéma de validation croisée par année pour évaluer la qualité de prédiction de manière réaliste par rapport à l'utilisation opérationnelle visée.

## RESULTATS

La mobilisation de méthodes d'apprentissage automatique (*machine learning*) donne une information intéressante pour prédire la date d'apparition des premiers symptômes avec les deux méthodes retenues ( $R^2$ , part de la variance expliquée, de 0.32 pour le modèle linéaire ou  $R^2$  de 0.26 pour le gradient boosting), même si les performances restent modestes.

Ainsi cette prédiction se fait juste au 15 juin, en amont de la période à risque (en général entre mi-juin et fin juillet) en mobilisant des données météorologiques passées et permet de proposer l'information de la campagne en cours en la comparant aux campagnes précédentes (Figure 1).

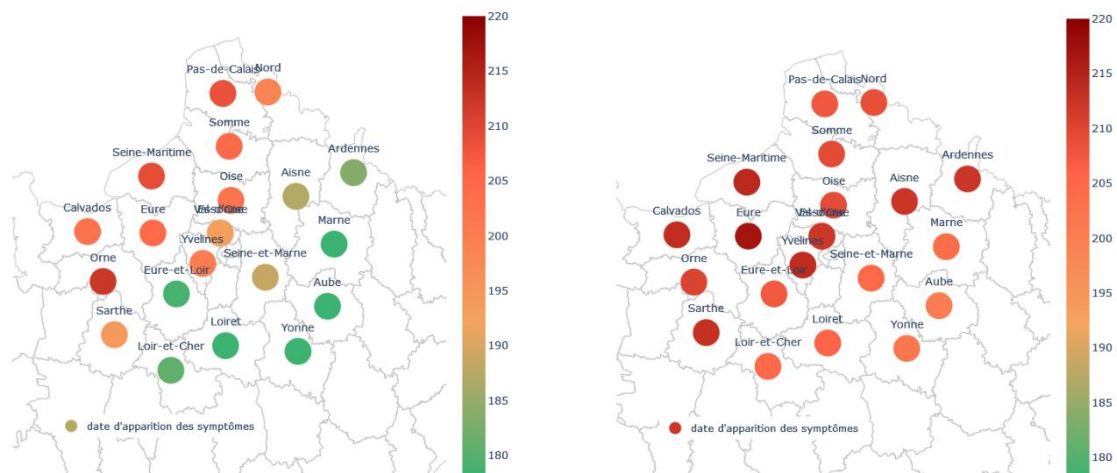


Figure 1 : Carte des prédictions de l'apparition des premiers symptômes de cercosporiose sur betterave, en 2018 (à gauche) et 2021 (à droite).

(Map of predicted first symptoms in 2018 (left) and 2021 (right))

Enfin, les méthodes d'apprentissage profond (*deep learning*), avec des réseaux de neurones à convolution (CNN), autorisent l'estimation de la dynamique épidémique en cours de campagne à J+7, d'une semaine à l'autre, en s'appuyant sur la météorologie des trois dernières semaines. Les performances montent alors jusqu'à 60% de la variance expliquée. Ces performances sont aussi nettement améliorées si l'on utilise en plus l'observation au jour J avec une variance expliquée de 76%.

## DISCUSSION

Ces performances sont très prometteuses pour l'aide à la décision et nous incitent à continuer les tests opérationnels en cours de campagne avec les acteurs de terrain pour consolider ces résultats.

Cela montre aussi tout le potentiel des données du Bulletin de Santé du végétal comme déjà envisagé précédemment avec d'autres approches (Van de Kerckhove et al, 2018, Chen et al, 2020), tout en valorisant les connaissances issues de la littérature et des experts sur les effets météorologiques et agronomiques plus ou moins documentés sur cette maladie relativement peu étudiée.

## BIBLIOGRAPHIE

Chen M, Brun F, Raynal M, Makowski D. 2020. Forecasting severe grape downy mildew attacks using machine learning. PLoS ONE 15(3): e0230254. <https://doi.org/10.1371/journal.pone.0230254>

Vallet L. 2021. Modélisation des épidémies de cercosporiose sur la betterave par des approches d'apprentissage automatique, mémoire de stage de M2. Acta, ITB, INSA Toulouse.

Van de Kerckhove S, Brun F, Chen M, Raynal M et al. 2018. Une démarche générique pour le développement d'outils d'analyse et de prédiction des dynamiques épidémiques à partir des données des réseaux d'épidémiosurveillance des cultures. 12ième CIMA (Végéphy).