

# Epidémiologie et traitement des données, quels apports dans l'étude des dépérissements ?

Christian LANNOU (CD SPE) et Samuel SOUBEYRAND ( <sup>Biostatistique</sup>  $\mathcal{B}90/\Pi$  )  
& Processus Spatiaux

INRAE



Epidémiosurveillance  
Santé Végétale



## Plateforme d'Epidémiosurveillance en Santé végétale

- ❖ Signature de la convention cadre en juillet 2018



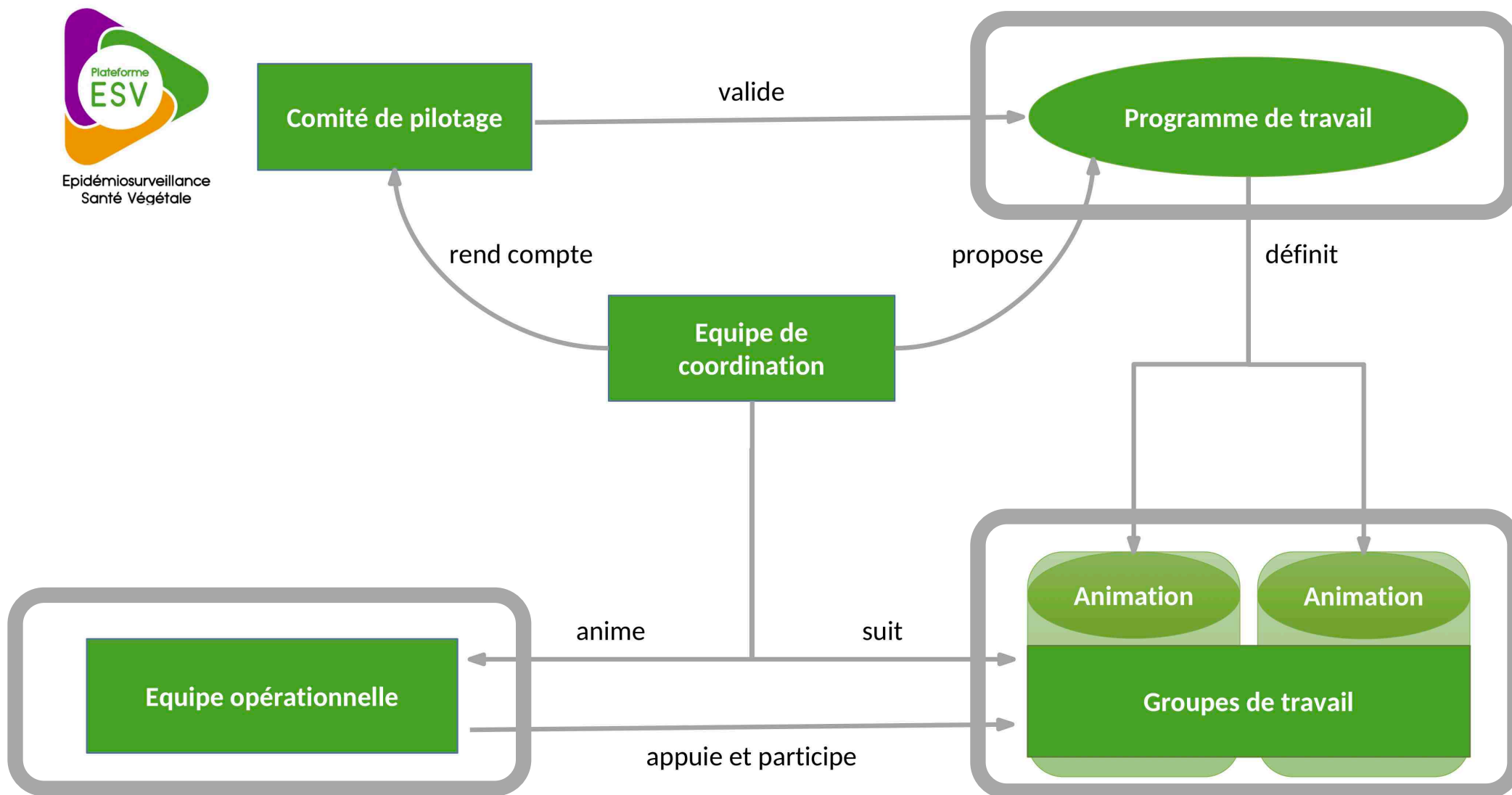
- ❖ Champ d'action de la Plateforme ESV : potentiellement tout danger sanitaire ou phénomène phytosanitaire ayant ou pouvant avoir un impact sur l'état sanitaire des végétaux et les effets non intentionnels des pratiques agricoles sur l'environnement



- ❖ Trois plateformes soeurs



## Structuration de la Plateforme ESV



## Les Groupes de Travail de la Plateforme ESV

- ❖ GT « Surveillance officielle des organismes nuisibles réglementés ou émergents »
- ❖ GT « Surveillance épidémiologique du dépérissement du vignoble »

**Année de démarrage : 2019**

**Objectif : améliorer la surveillance en vue d'une meilleure valorisation de la connaissance de l'état du vignoble et alimenter la recherche, en cohérence avec le plan national dépérissement du vignoble**

**Animateurs : Anastasia Rocque et Jacques Grosman**

- ❖ GT « Comité de rédaction des bilans sanitaires »
- ❖ GT « Surveillance de *Xylella fastidiosa* »
- ❖ GT « Surveillance du nématode du pin »
- ❖ GT « Veille sanitaire internationale »
- ❖ GT inter-plateformes « Qualité des données »

## L'équipe opérationnelle de la Plateforme ESV

### Ses missions principales :

- ❖ **Veille scientifique et phytosanitaire** sur les dangers touchant les végétaux et susceptibles d'avoir un impact sanitaire et/ou économique
- ❖ **Gestion d'un système d'information** centralisant les données de surveillance des dangers listés dans le programme de travail annuel de la plateforme plate-forme et intégrant des espaces de communication et de collaboration
- ❖ **Traitement et analyse des données** issues de divers dispositifs d'épidémiosurveillance
- ❖ **Communication scientifique** et diffusion d'informations via l'élaboration et la mise à disposition de bilans sanitaires, de données, de supports d'informations et de formations à destination de différents interlocuteurs
- ❖ **Evaluation des dispositifs de surveillance** et proposition d'améliorations

### Sa composition :



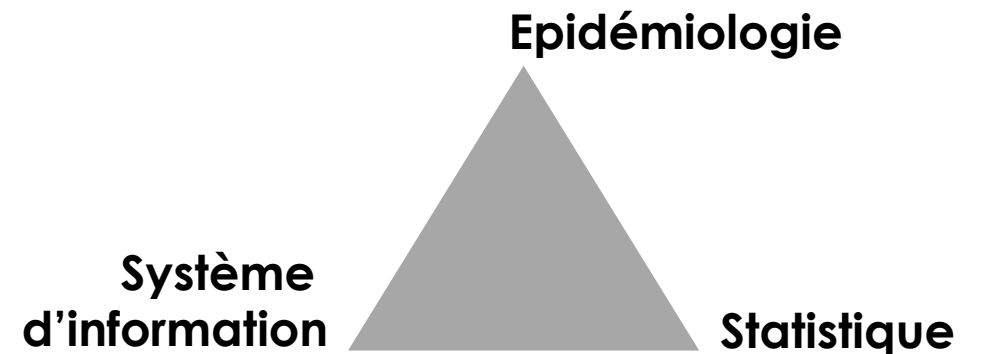
Lucie Michel, Marie Grosdidier, Marine Marjou,  
Stéphane Lechalier, Jean-Loup Gaussen, Audrey Foncelle



Anne Quillévéré-Hamard, Adler Moukouboulou,  
Charlotte Rüger

### Une équipe multidisciplinaire :

- ❖ **Epidémiologie, statistique, informatique, communication**



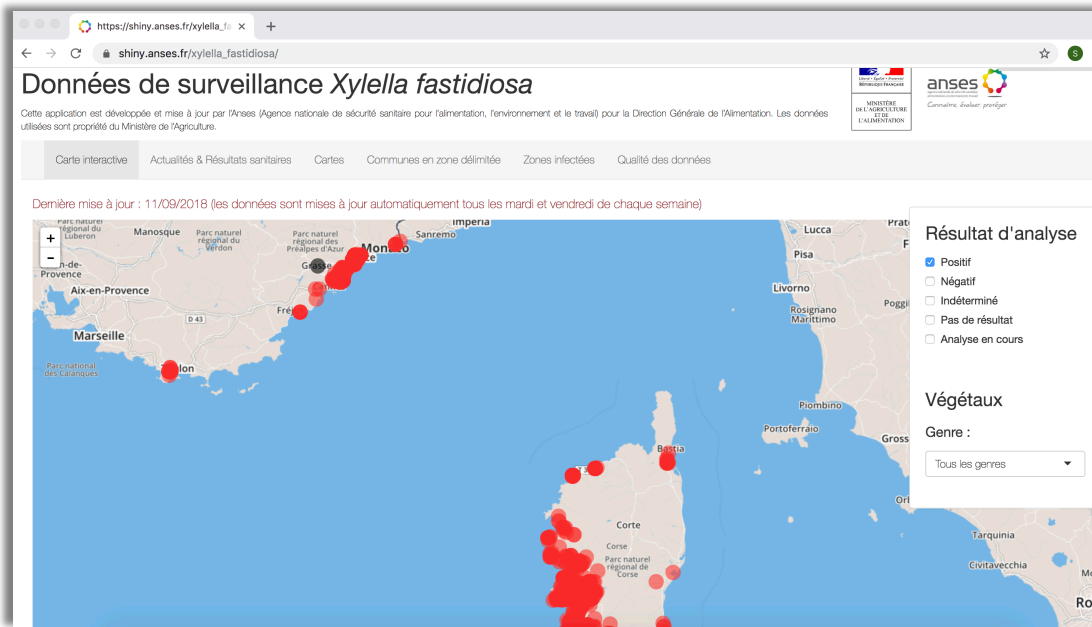
## L'exemple de *Xylella fastidiosa*



### *Xylella fastidiosa*, c'est quoi ?

La bactérie *Xylella fastidiosa*, transmise et véhiculée par des insectes vecteurs, s'attaque à un très large spectre de végétaux : vignes, oliviers, arbres fruitiers (Prunus), agrumes, caféiers, chênes, luzerne, etc.

02/03/2018 INFO +

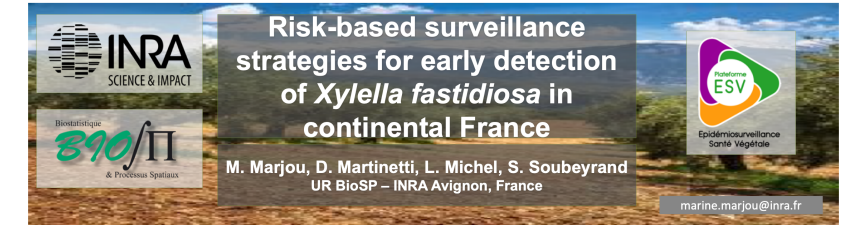


- ❖ Première détection *in situ* en Corse en 2015
- ❖ Campagne de surveillance / arrachage
- ❖ Centralisation des données par l'Anses et création d'un système d'information dédié

# Exemples de travaux menés ou en cours sur Xf

- ❖ Mise au point d'outils de diagnostic moléculaire sur plantes et sur vecteurs
- ❖ Identification des origines des introductions
- ❖ Datation et localisation des introductions
- ❖ Détermination des facteurs de risque (variables climatiques, usages du sol, réservoirs)
- ❖ Cartographie du risque et des enveloppes climatiques
- ❖ Caractérisation du réseau hôtes - pathogène - vecteur
- ❖ Design de stratégies d'échantillonnage (détection précoce, délimitation des zones infectées)
- ❖ Evaluation des mesures de lutte (arrachage)

Réflexion en cours autour de travaux semblables sur la flavescente dorée



**Abstract & Highlights**

- First cases of *Xylella fastidiosa* July 2015 in Corsica & October 2015 in PACA.
- Risk assessment for the presence of XF in the whole country.
  - ◊ South-eastern coast has highest risk, then temperate southern regions. Lower risk elsewhere.
- Design of surveillance strategies based on predicted risk.
  - ◊ Strategy benchmarking for sampling effort ranging from 100 to 10000 sampling sites
  - ◊ Comparison of surveillance strategies based on early-detection probability
  - ◊ Trade off between sampling in high-risk zones vs. good spatial coverage
- Marginal effect of detection probability, per added unit of sampling
  - ◊ It increases up to 1000 sampling units, after that, the marginal effect decreases rapidly

**Introduction**

Recent detections of *Xylella fastidiosa* in Corsica and the south-eastern region of Provence-Alpes-Côte d'Azur (PACA) raised the concern about the presence of the disease in the rest of the country. Extensive sampling campaigns on known susceptible host plants have been conducted both in the surroundings of detected foci, following EU quarantine directives, and in other remote areas. Although buffer zones have already been established, there is still need to continue surveilling and sampling to get more confidence in the exact delimitation of the infection, which seems to be confined to few coastal areas.

Optimal strategies should balance the need of circumscribing high-risk areas with the need to explore new locations in the attempt to increase the chances of early detection.

**Disease status (2013-2019)**

- 1288 positive samples
- PACA + Corsica

**Materials**

We divided metropolitan France with a 8-km grid. For each grid cell, we computed 100 bioclimatic variables + altitude + land use + presence/absence XF.

**Most relevant variables :**

- Precipitation of driest month
- Precipitation seasonality
- Precipitation of driest and coldest quarters
- Solar radiation in January, August and September
- Median altitude

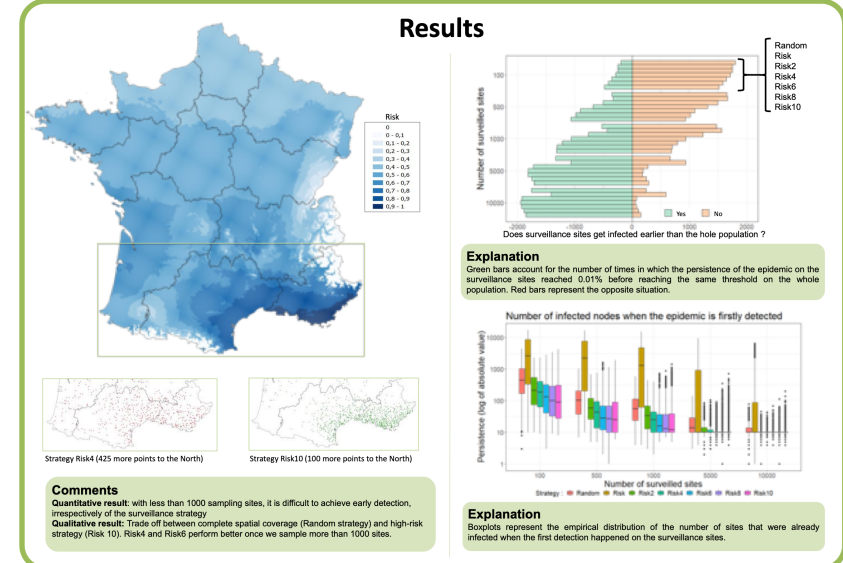
**Methods**

Step 1 : Variables selection with gain information + PCA.  
 Step 2 : Production of a risk map with MESS (Multivariate Environmental Similarity Surfaces) with selected variables. Rescale risk between 0 and 1.  
 Step 3 : Design of 7 risk-based surveillance strategies with increasing number of sampling sites : 100, 500, 1000, 5000 and 10000  
 Step 4 : Evaluate the efficacy of surveillance  
 Step 5 : Identify best strategies with respect to the number of sampling sites and the trade off between spatial coverage and high-risk

**Designing surveillance strategies**

Risk-based surveillance should balance two objectives: being close to high-risk areas and maximize spatial coverage.

- Random
- Risk
- Risk2 = Risk<sup>2</sup>
- Risk4 = Risk<sup>4</sup>
- Risk10 = Risk<sup>10</sup>



Machine learning pour estimer le risque  
Surveillance basée sur le risque

**Acknowledgement:** HORIZON 2020 XF-ACTORS Project SFS-09-2016; DGAL, Anses, SRAL, FREDON, LNR-LSV and certified laboratories for data collection and data availability.

**References**

Bain, J. & Arnaud, M. & Prévost, S. (2018). The art of modeling single-affected species. *Methods in Ecology and Evolution*, 9, 1-11.

Fernández de Arriba, G. & al (2014). Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E*, 90, 016111.

Gohar, A. & Schweitzer, F. & Havlin, S. (2012). A shell decomposition method for weighted networks. *New Journal of Physics*, 14, 043011.

Herrera, J. & al (2018). Disease Surveillance on Complex Social Networks. *PLoS Computational Biology*, 14, e1006111.

Martinez, D. & Soubeyrand, S. (2018). Identifying Locations for Epidemic-Surveillance: Application to the Emergence of *Xylella fastidiosa* in France. *Phytopathology*, 108, 1000-1008.

Zhang, J.-Y. & al (2016). Identifying a set of influential spreaders in complex networks. *Scientific Reports*, 6, 24111.

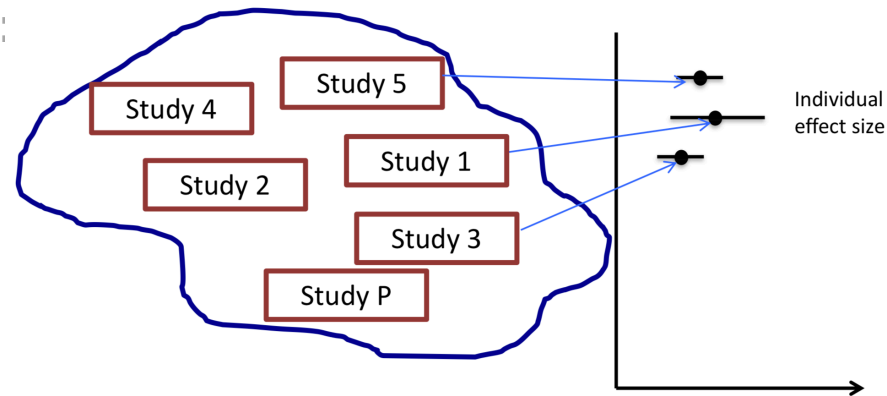


## La méta-analyse pour capitaliser sur l'existant

❖ La méta-analyse : une approche formelle pour synthétiser les résultats d'analyses quantitatives multiples

❖ Processus de la méta-analyse :

1. Scoping
2. Literature search
3. Paper selection
4. Data extraction
5. Statistical analysis
6. Bias and uncertainty



❖ Méta-analyse fondée sur la modélisation hiérarchique et l'approche bayésienne :

- ❖ Adaptation à la structuration des données
- ❖ Prise en compte de données manquantes
- ❖ Intégration d'a priori
- ❖ Gestion probabiliste des incertitudes

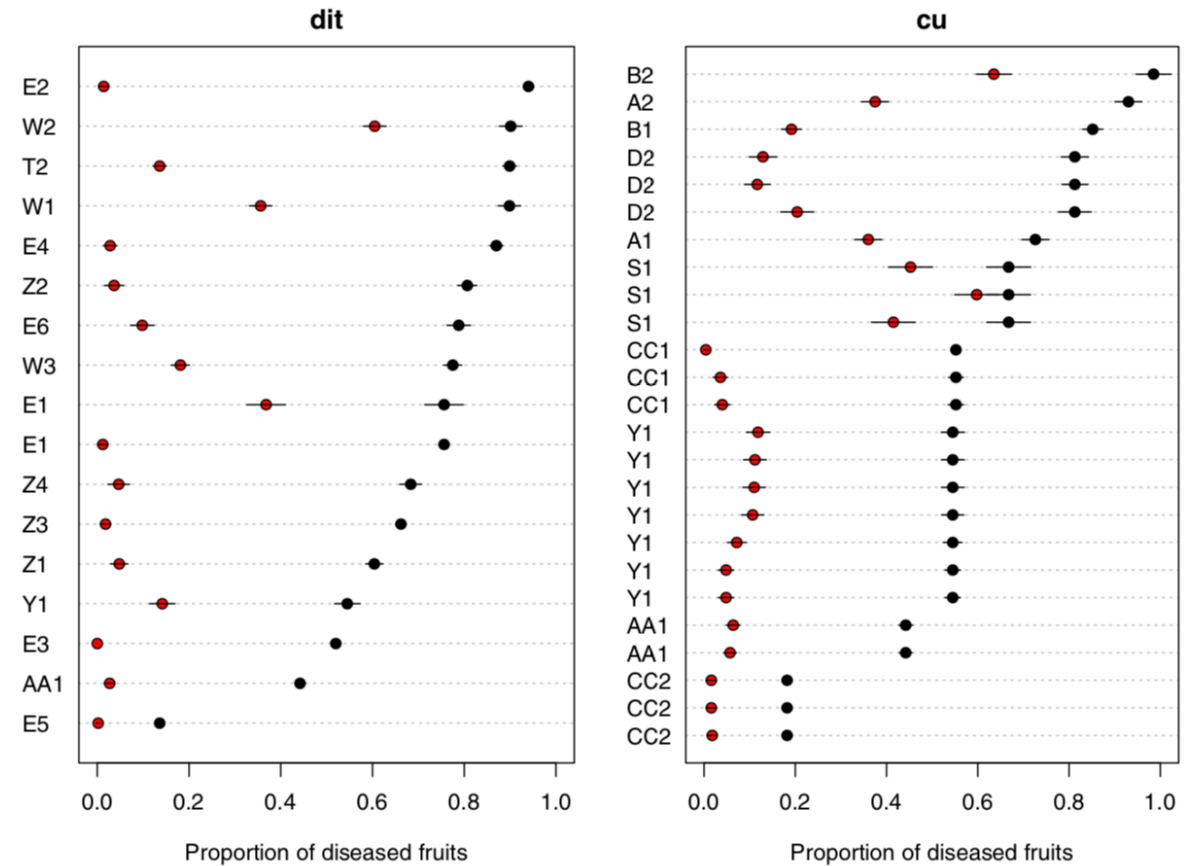




## Exemple en santé des plantes

- ❖ Effet de traitements différents sur l'incidence d'une maladie
  - ❖ Essais traitements-contrôles multiples
  - ❖ Etudes multiples
  - ❖ Résultats hétérogènes
- ❖ Synthèse des résultats par méta-analyse :

Type of fungicide	Probability of diseased fruit	
	Untreated fruits	Treated fruits
"cu"	0.74 (0.53-0.87)	0.16 (0.072-0.32)
"dit"	0.74 (0.62-0.83)	0.053 (0.022-0.12)



**Fig. 1** Measured disease incidences in treated fruits (red points) and in untreated fruits (black points). Bars indicate 95 % confidence intervals. Each letter indicates a specific paper, each letter/number combination indicates a specific plot, and each row indicates a specific subplot

Makowski, D., Vicent, A., Pautasso, M., Stancanelli, G., & Rafoss, T. (2014). Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*. *European journal of plant pathology*, 139(1), 79-94.



## Le traitement de données pour l'épidémiologie

Intégration  
de données hétérogènes

Machine / deep learning  
pour expliquer et classifier

Systeme  
d'information

Visualisation /  
Data science

Intégration  
d'analyses existantes

Algorithmes d'optimisation  
pour la surveillance et la lutte

Appréhender la performance  
de mesures prophylactiques  
en environnements multiples

Développer une vision dynamique  
et à grande échelle spatiale  
du microbiome