

# Modèles interprétables issus des données et de la connaissance en agro-environnement



# « Le défi des données »

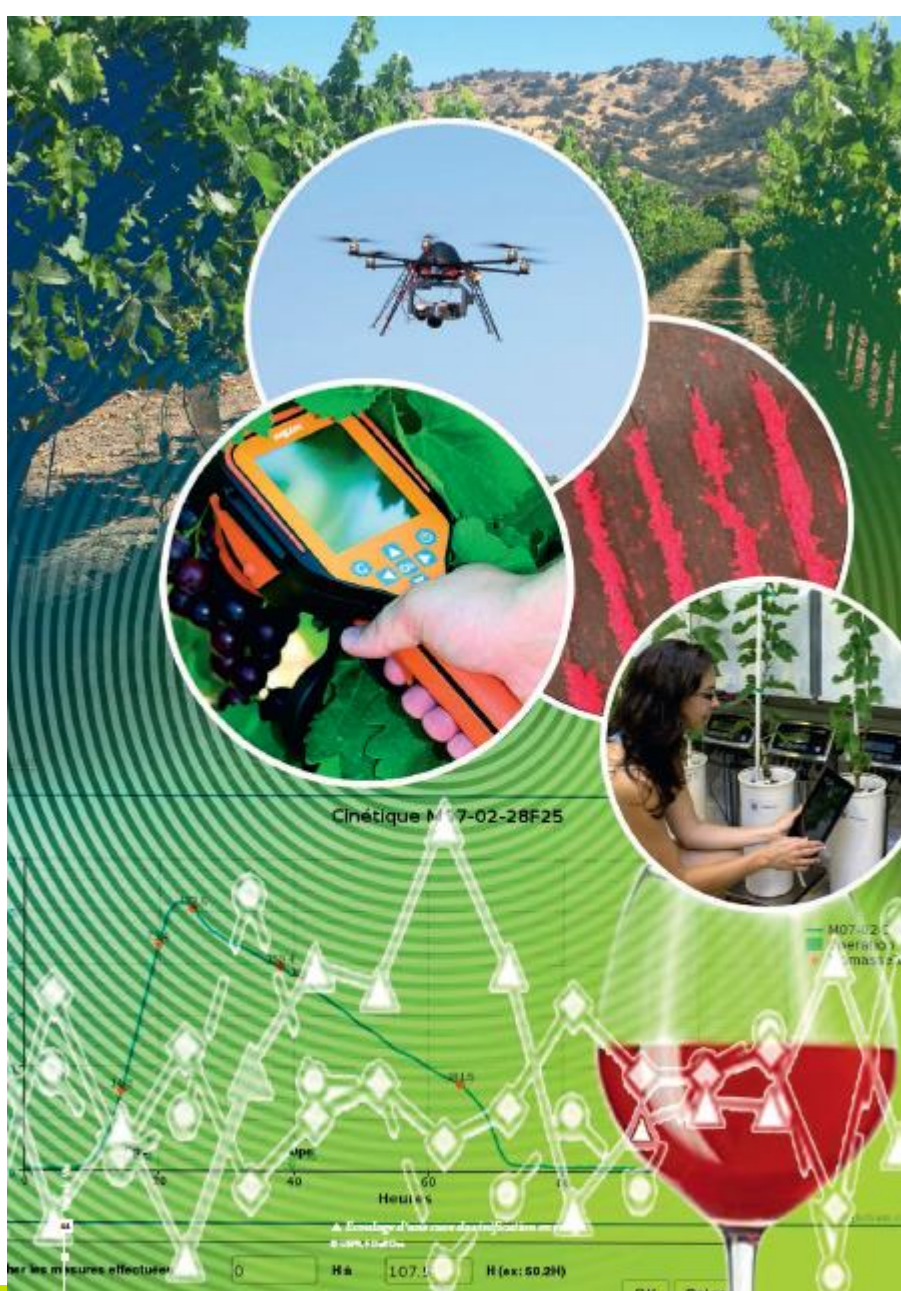
En « Agriculture et Transformation »

## Caractéristiques :

- complexes
- temporelles
- à différentes échelles
- incertaines...

**Exemples :** phénotypage, bioprocédés, filières

**Enjeu dans ce défi :** *maîtriser ce qui se passe aux champs et durant la transformation pour mieux piloter la qualité du produit final.*

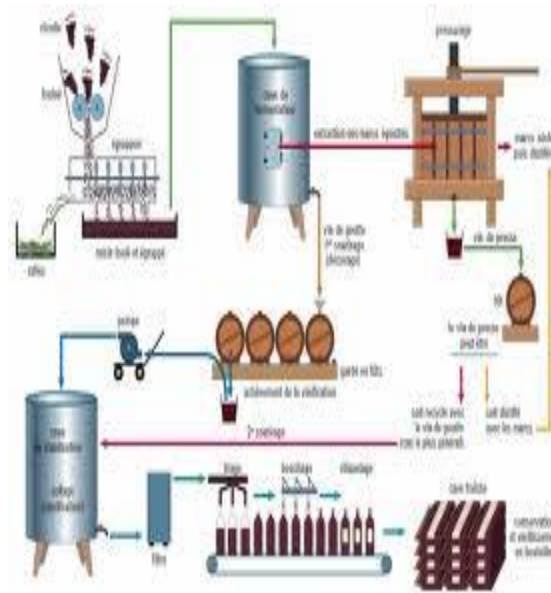


# Données et connaissance

- **Contexte** : phénomènes évoluant dans le temps (filières, expérimentations)
- **Objectifs** : comprendre, aider à la décision, atteindre des profils cible  
~~Modèles mécanistes : échelle très fine~~
- **Illustration** : filière vigne et vin



Vigne



Cave



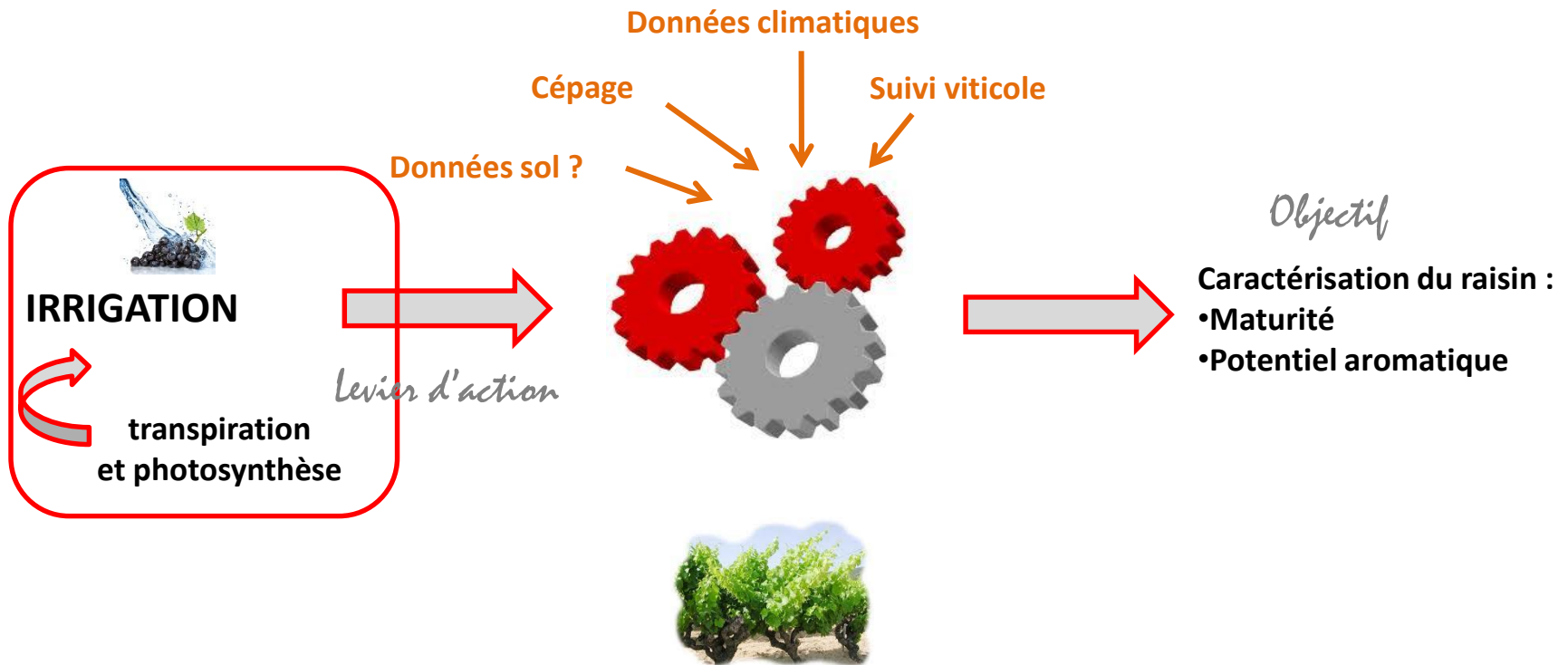
Vin

→  
temps

# Comprendre l'effet des opérations à la vigne et en cave

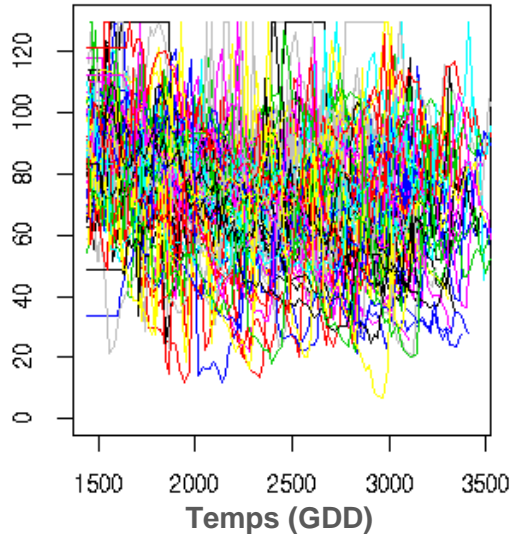
**Hypothèse** : le parcours hydrique est fondamental

→ **Objectif** : Mettre en évidence une relation de causalité entre **parcours hydrique** et **caractérisation du raisin**



# Statistique pour données fonctionnelles

Vécus hydriques



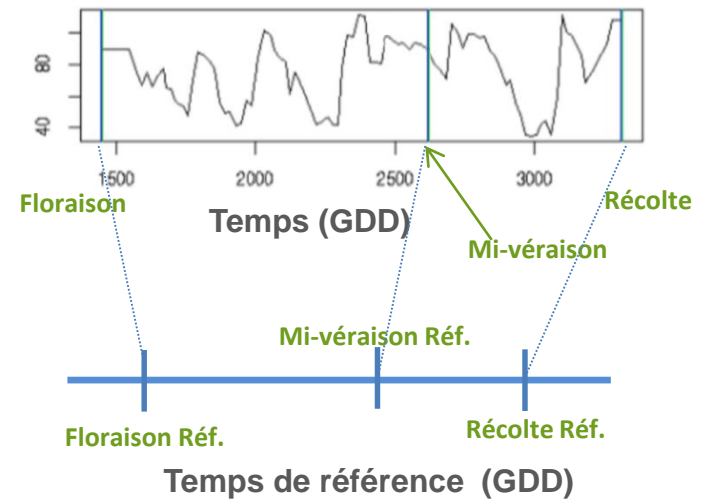
## → Verrous :

1. Distorsion de courbes
2. Prise en compte de l'incertitude et de la connaissance a priori

Difficultés : courbes = objets de dimension infinie (pas classique)

- Adapter, développer des méthodes statistiques pour données fonctionnelles
- Interprétabilité des quantités estimées ?

Vécu hydrique d'un cep de vigne





# Questions méthodologiques

## Inférence de connaissances

Un modèle linéaire pour données fonctionnelles

$$y_i = \mu + \int_{t_0}^{t_f} X_i(t)\theta(t)dt + \varepsilon_i$$

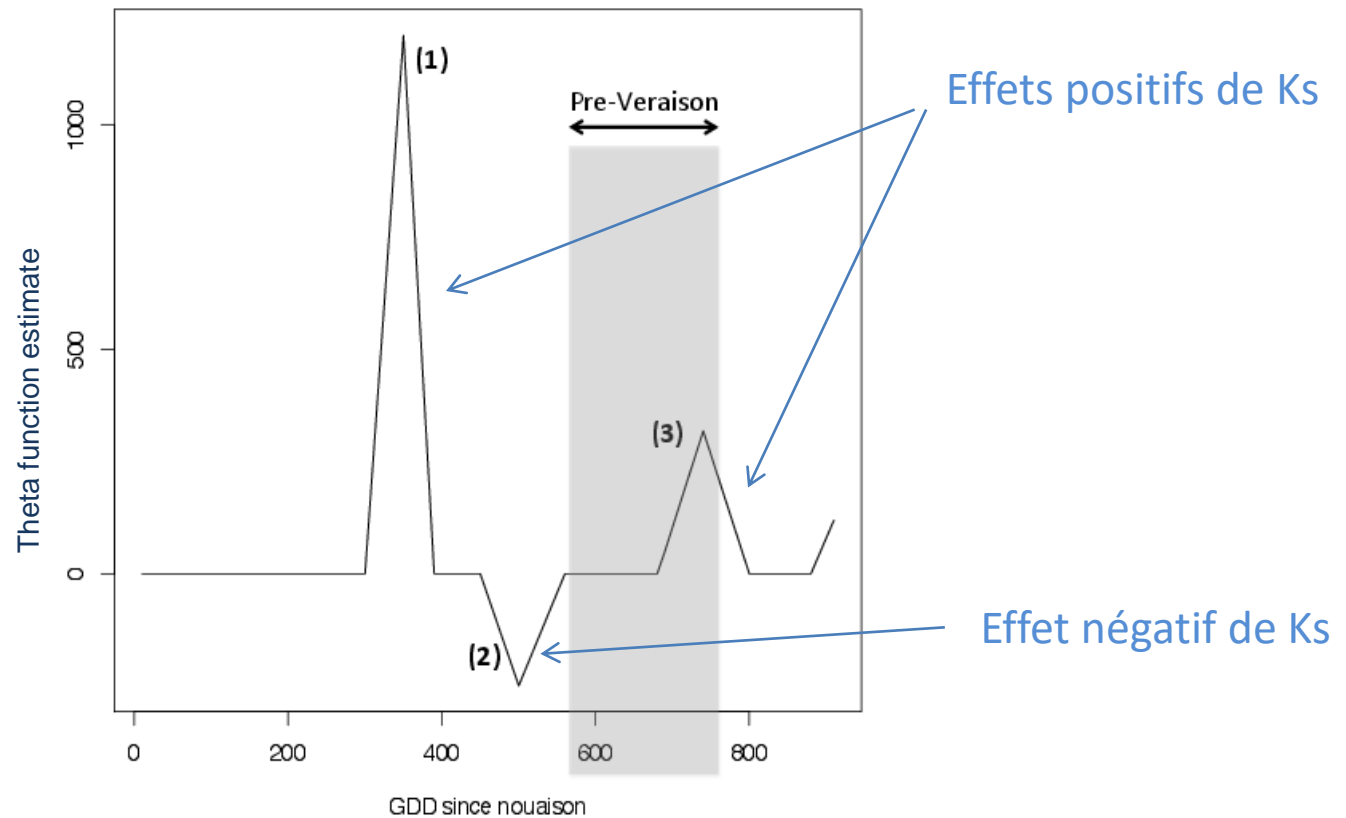
Où  $y$  : concentration en sucre du raisin à la vendange

$X(t)$  : vécu hydrique de la vigne (notée  $Ks(t)$  dans la suite)

Quelles sont les périodes d'influence du vécu hydrique sur la concentration de sucre ?

# Illustration : effet du vécu hydrique sur la quantité de sucre finale dans le raisin

Sucre : effet du vécu hydrique  $K_s(t)$



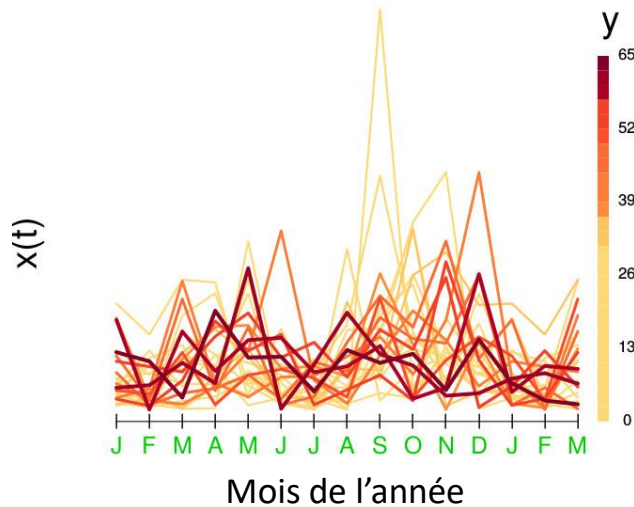
# Thèse de P. Grollemund (2017)

## Effet de la pluviométrie sur la production annuelle de truffe

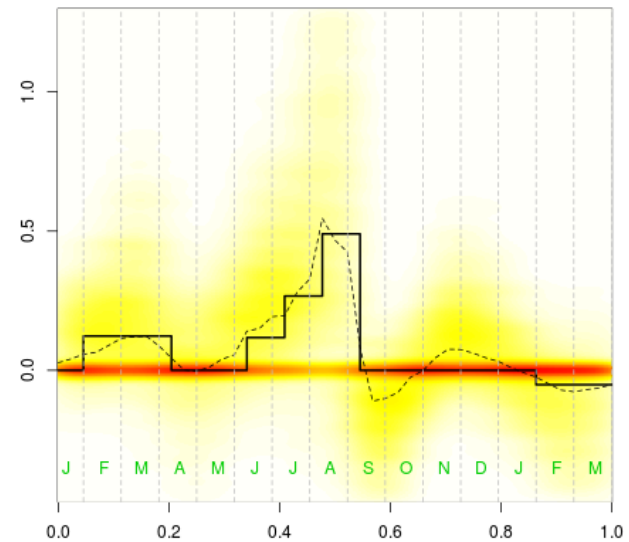


1. Quelles sont les périodes de pluie pertinentes ?
2. Quel est l'impact des pluies pendant ces périodes ?

Précipitations et rendements



Estimation de la fonction  $\theta$





# Extensions : 1/ Prise en compte de la connaissance experte

Comment construire un a priori qui prend en compte l'avis des experts ?

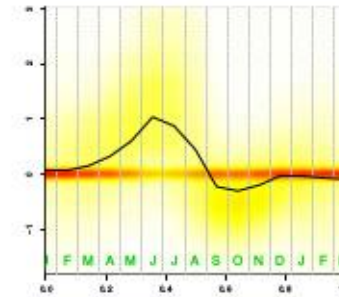
*Exemple :*

- L'effet des précipitations en fin d'été est bien connu sur le rendement
- Un effet des précipitations en fin d'hiver est soupçonné (confiance plus faible)

→ Utilisation de pseudo-données pour ajuster le modèle

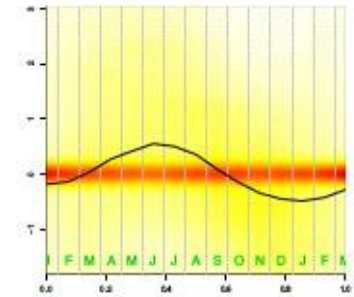
Estimation à partir des données

(a)

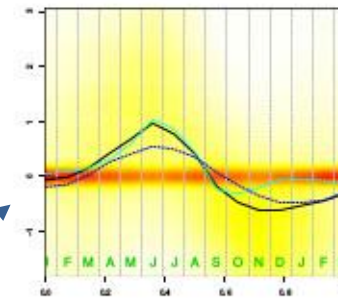


Estimation à partir des pseudo-données

(b)



(c)



Estimation qui tient compte des données et des pseudo-données

# Extensions : 2/ Modèle avec plusieurs variables explicatives

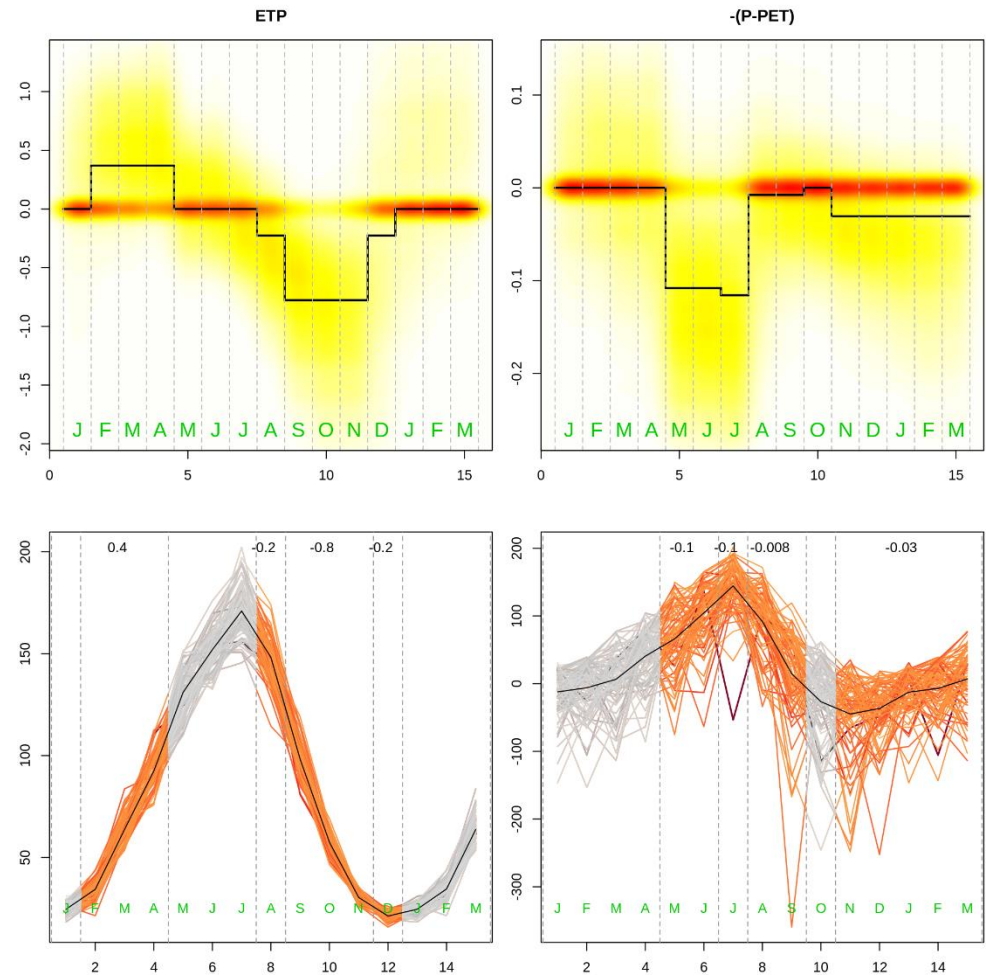
Effet de

- L'évapotranspiration potentielle mensuelle : ETP
- Du bilan climatique mensuel : - ( P – ETP)

→ Un modèle additif :

$$y_i = \mu + \int_{t_0}^{t_f} X_{1i}(t)\theta_1(t)dt + \int_{t_0}^{t_f} X_{2i}(t)\theta_2(t)dt + \varepsilon_i$$

→ **R - Package BLISS**





# Questions en suspens...

- ❖ Quel autre type de modèle multivarié ? Comment prendre en compte des interactions de variables explicatives ?
- ❖ Comment quantifier les contributions respectives de chaque variable explicative ?
- ❖ Comment tenir compte, dans les modèles, de la périodicité des variables climatiques ?
- ❖ Comment adapter ces modèles au phénotypage de plantes ?
  - ➔ Extension à des modèles contenant des variables qualitatives (+ effets aléatoires)

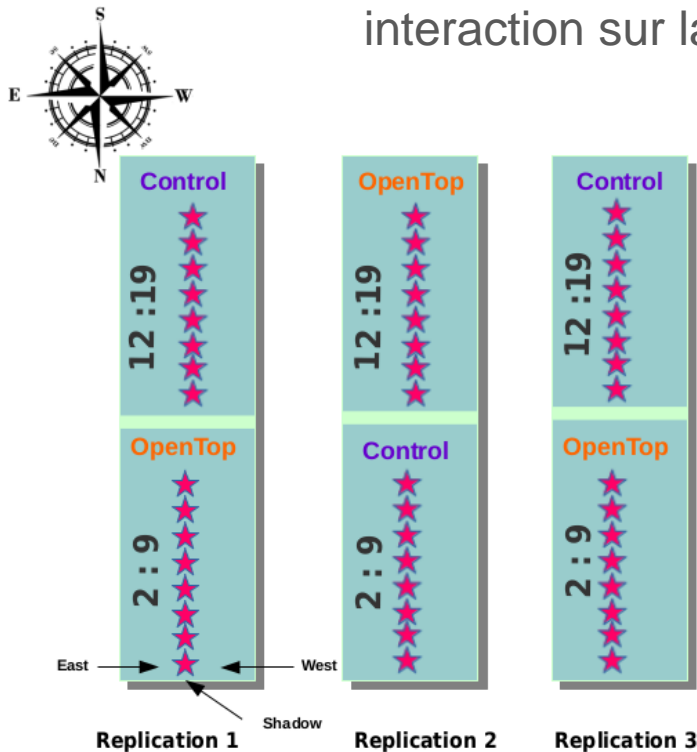
# Thèse de G. Gnanguenon-Guesse (2<sup>e</sup> année)

*Modélisation et visualisation des liens entre cinétiques de variables agro-environnementales et qualité des produits dans une approche parcimonieuse*

#DigitAg

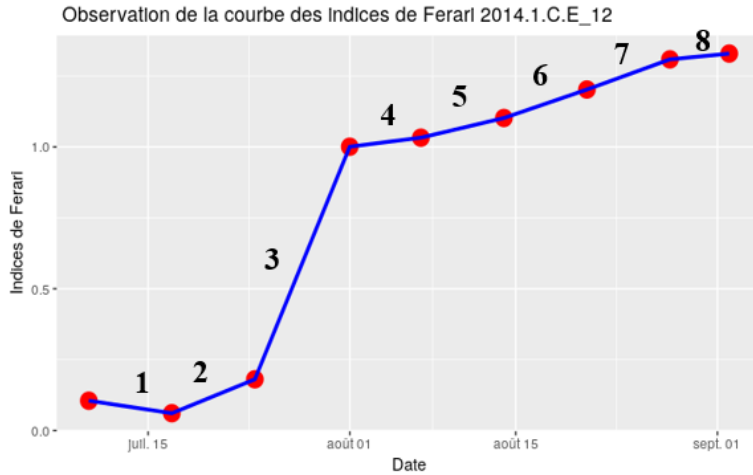
## Exemple des données du projet InnoVine

Comprendre l'effet de la température, de l'irradiance et de leur interaction sur la qualité de la baie.



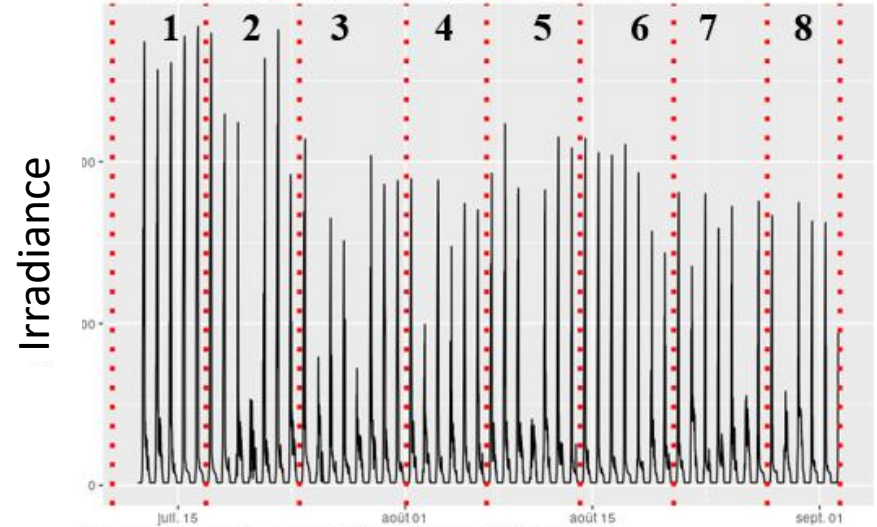
# Données observées sur 1 cep / 1 orientation

## Indice de Ferari

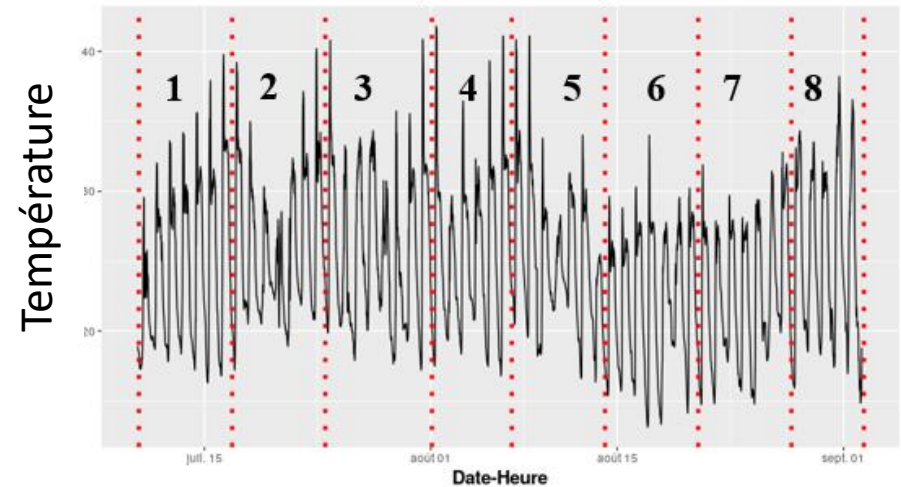


L'indice de Ferari est une mesure indirecte des anthocyanes

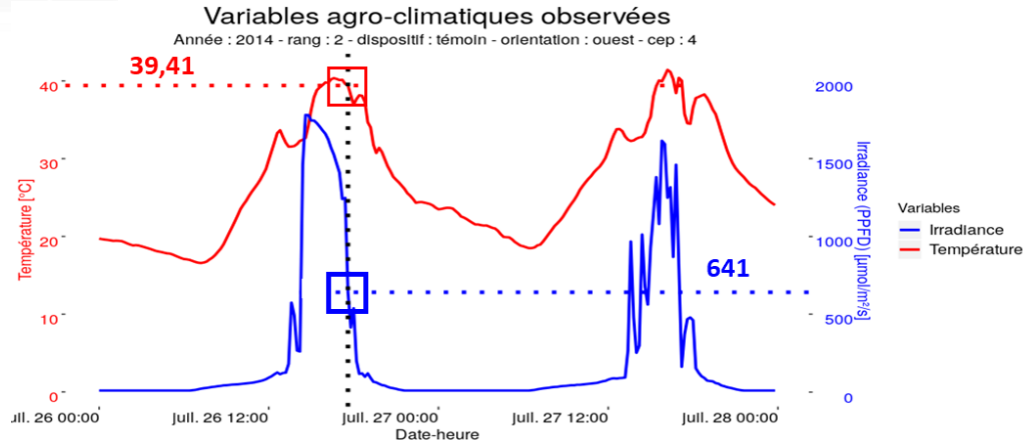
Observation de la courbe de Radiance 2014.1.C.E



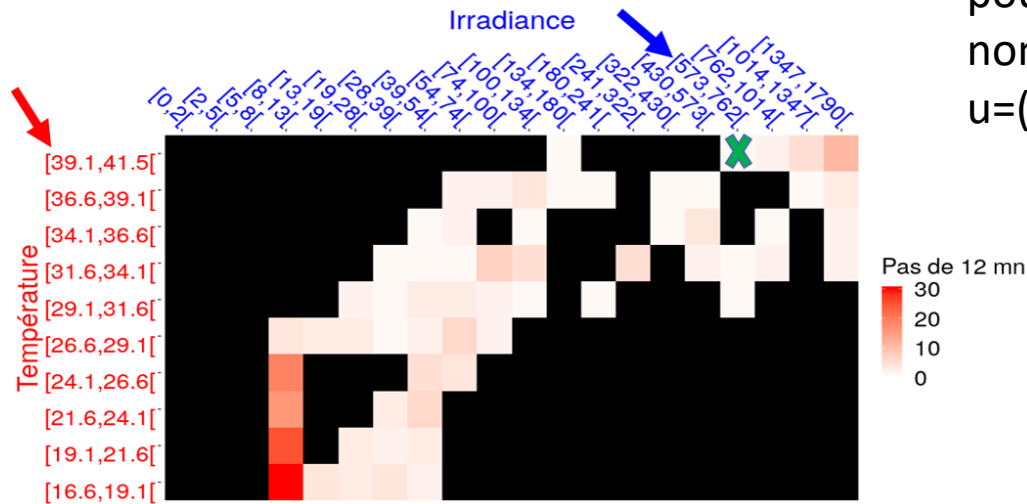
Observation de la courbe de température 2014.1.C.E\_12



# Un tableau de contingence des valeurs (Temp(t), Irr(t)) pour chaque individu



Matrice de comptage (individu 2014.2.C.W\_4)  
Éléments comptés = t,  $n_t = 10$ ,  $n_l = 20$ ,  $\alpha = 0.16$



pour un couple de  
nombres de breaks  
 $u=(n_{Temp}, n_{Irr})$  fixé



# Un modèle à grande dimension

$$y_i = \sum_{v=1}^{n_{\tau}} \sum_{w=1}^{n_{\mathcal{I}}} X_{i,(v,w)}^u \beta_{v,w}^u + \epsilon_i$$

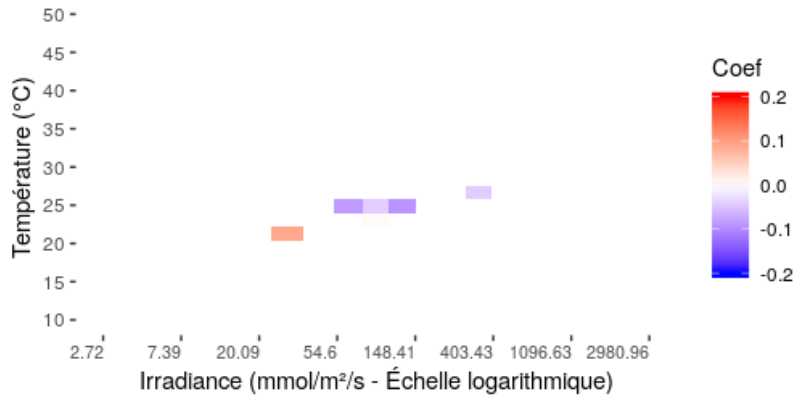
$y_i$  : une différence d'indice de Ferari (entre 2 dates)

- Estimation parcimonieuse de  $\beta^u$  pour chaque couple  $u$ , et donc avec chaque tableau de contingence vectorisé  $X_i^u$  candidat (ex : Lasso avec Cross Validation)
- Choix du couple  $u^*$  optimal (AIC)

# Résultats sur la 2<sup>e</sup> semaine : Lasso

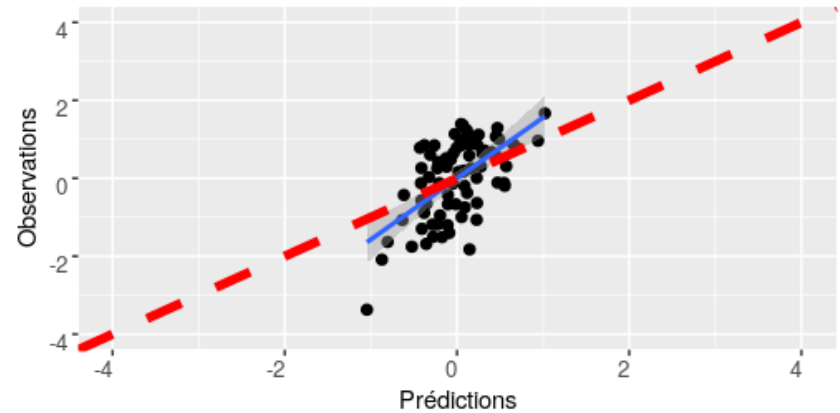
Approche : 1 - Itérations : 1 - L

Individus : tous - Lever du soleil à 12H - Semaine : 2



Approche : 1 - Itérations : 1 - L

Individus : tous - Lever du soleil à 12H - Semaine : 2



Résultats obtenus avec 1000 couples  $u$  candidats

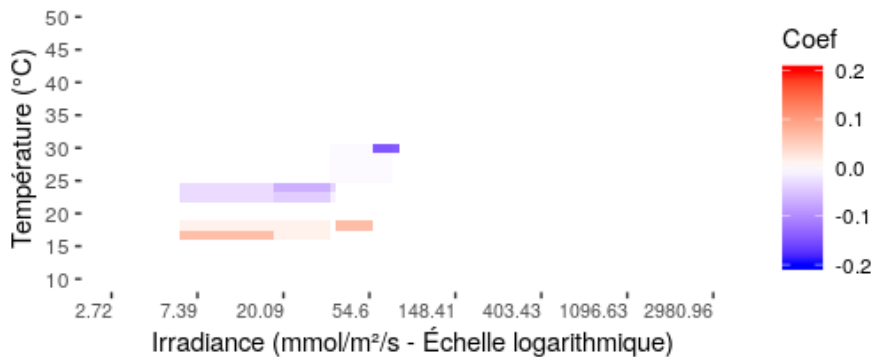
# Résultats sur la 3<sup>e</sup> semaine : GFL + Lasso

Etape 1 : GFL

Etape 2 : Lasso sur les résidus de l'étape 1

Approche : 1 - Itérations : 2 - GFL+L

Individus : IF3 < 0.4 - Lever du soleil à 12H - Semaine : 3



Approche : 1 - Itérations : 2 - GFL+L

Individus : IF3 < 0.4 - Lever du soleil à 12H - Semaine : 3

