



# *funHDDC* : Extension du package R Clustering de courbes fonctionnelles multivariées

---

Orateur : **Amandine Schmutz**<sup>1,2,3,4,5</sup>

Encadrants : Julien Jacques<sup>3</sup>, Charles Bouveyron<sup>6</sup>, Laurence Chèze<sup>4</sup> & Pauline Martin<sup>1,2,5</sup>



# Introduction

Package

Exemples pratiques

Présentation de la méthode

Conclusion



# ❖ Présentation des données fonctionnelles

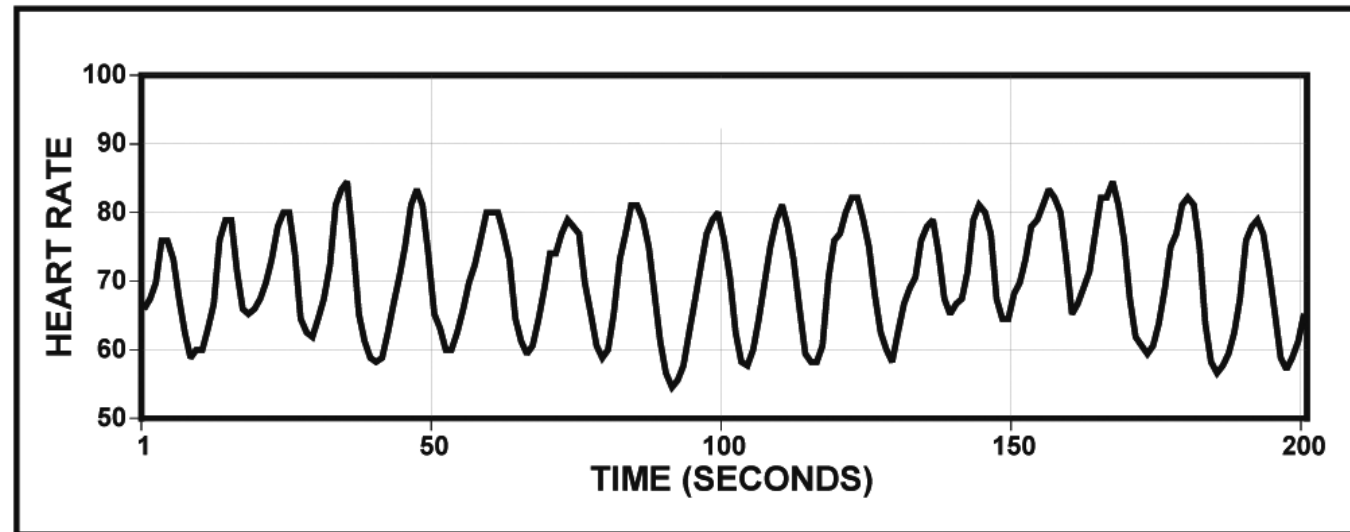
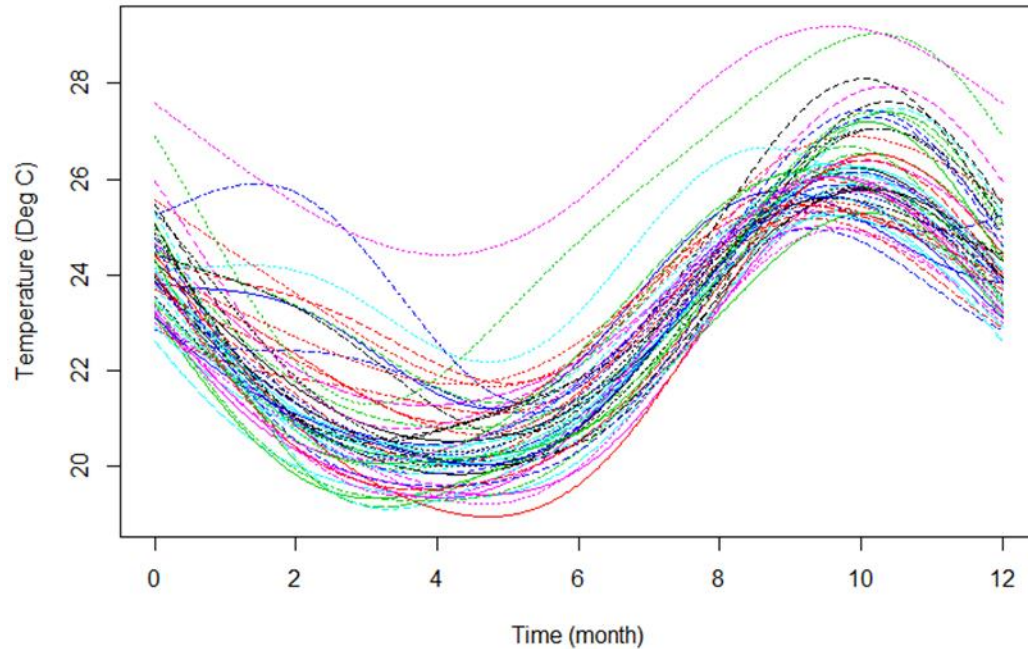
---

Objets connectés collectent des **mesures continues**



# ❖ Présentation des données fonctionnelles

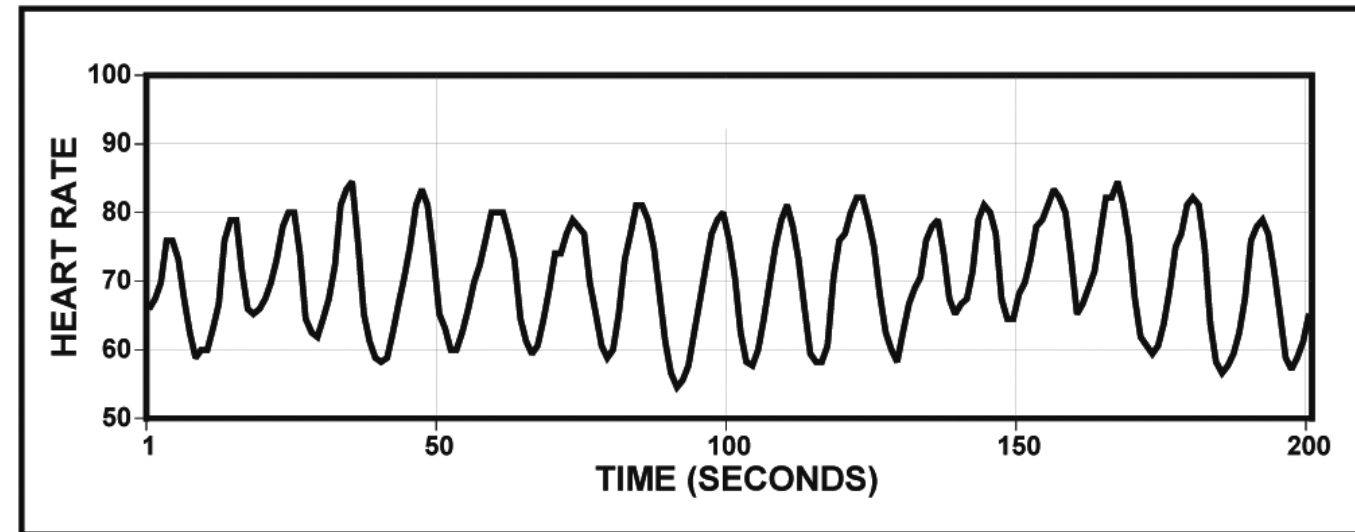
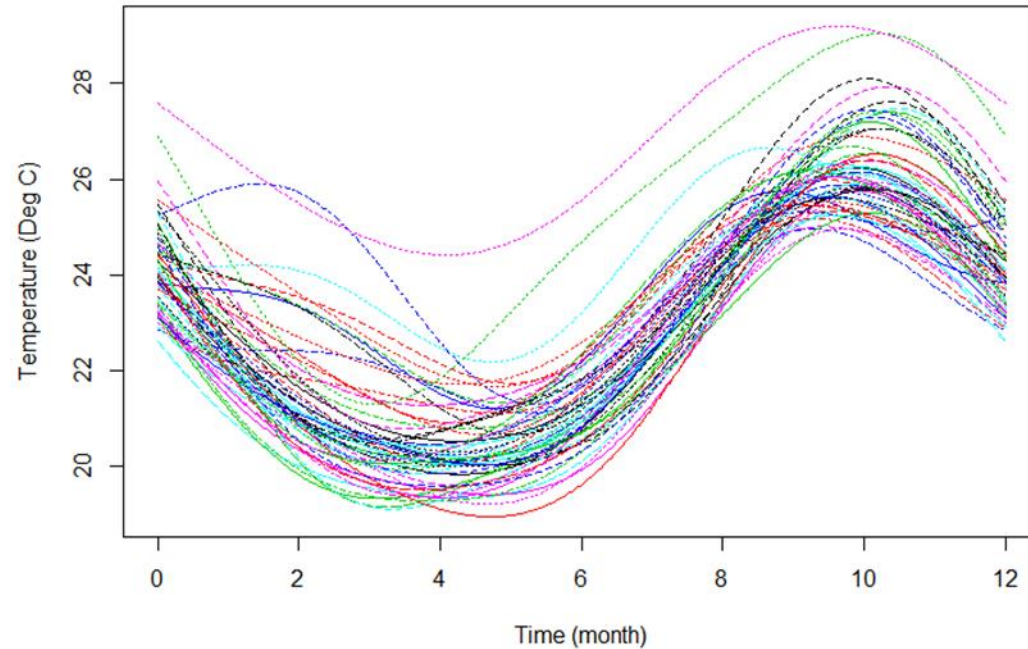
Objets connectés collectent des **mesures continues**





# ❖ Présentation des données fonctionnelles

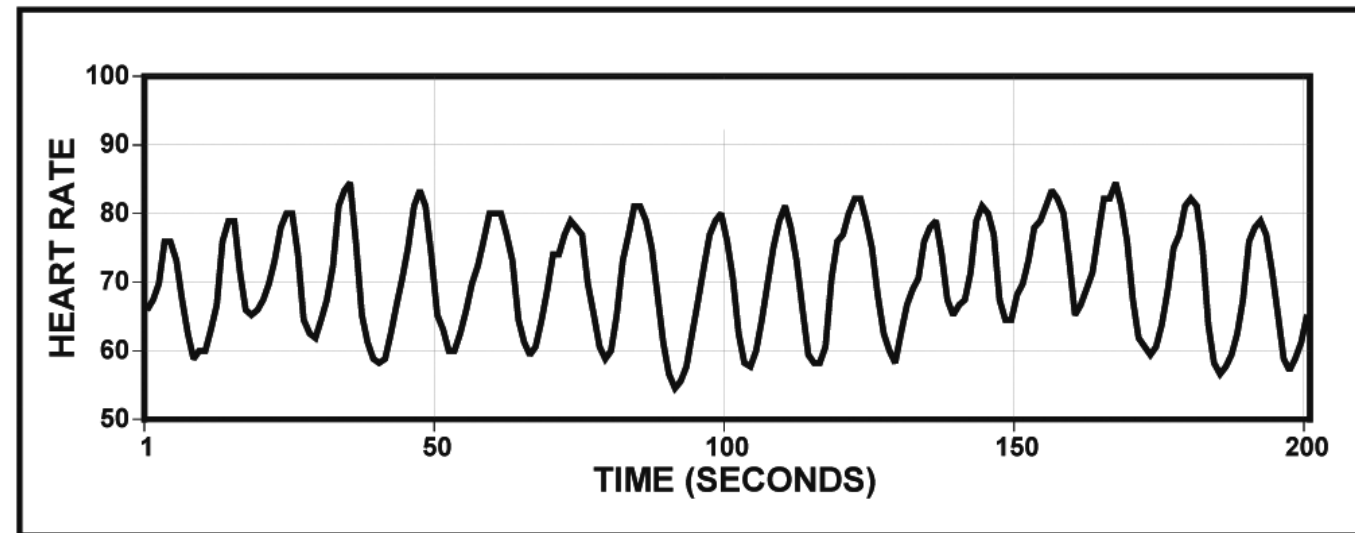
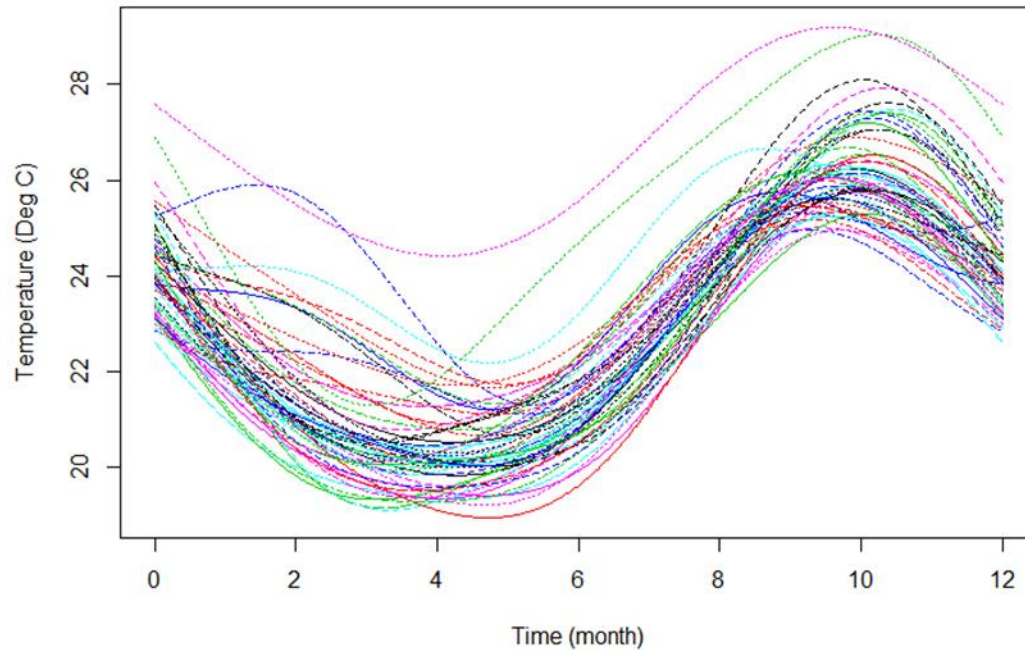
Objets connectés collectent des **mesures continues**



Données fonctionnelles : **1 individu = 1 courbe**

# ❖ Présentation des données fonctionnelles

Objets connectés collectent des **mesures continues**



Données fonctionnelles : 1 individu = 1 courbe

→ *Dépendance conservée* entre les points



# ❖ Plan

---

Introduction

**Package**

Exemples pratiques

Présentation de la méthode

Conclusion





# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*funHDDC(data, K, init, ...)*



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*funHDDC(data, K, init, ...)*



1 jeu de données fonctionnel  
ou une liste de jeux de  
données



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*funHDDC*(*data*, *K*, *init*, ...)

↓  
Nombre  
de  
groupes



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*funHDDC*(*data*, *K*, *init*, ...)

→ Type d'initialisation de l'algorithme EM



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*slopeheuristic(mod)*



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*slopeheuristic(mod)*

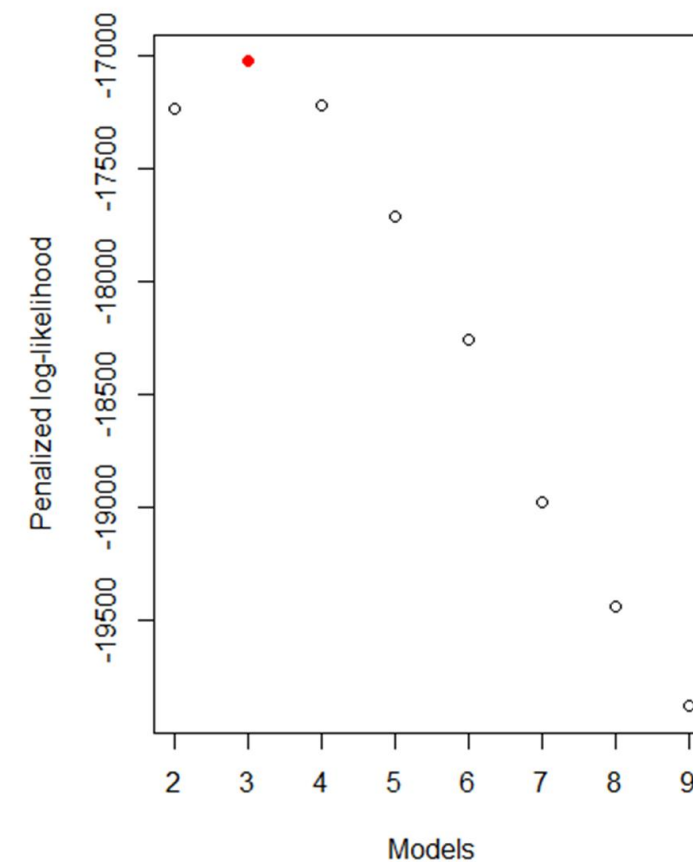
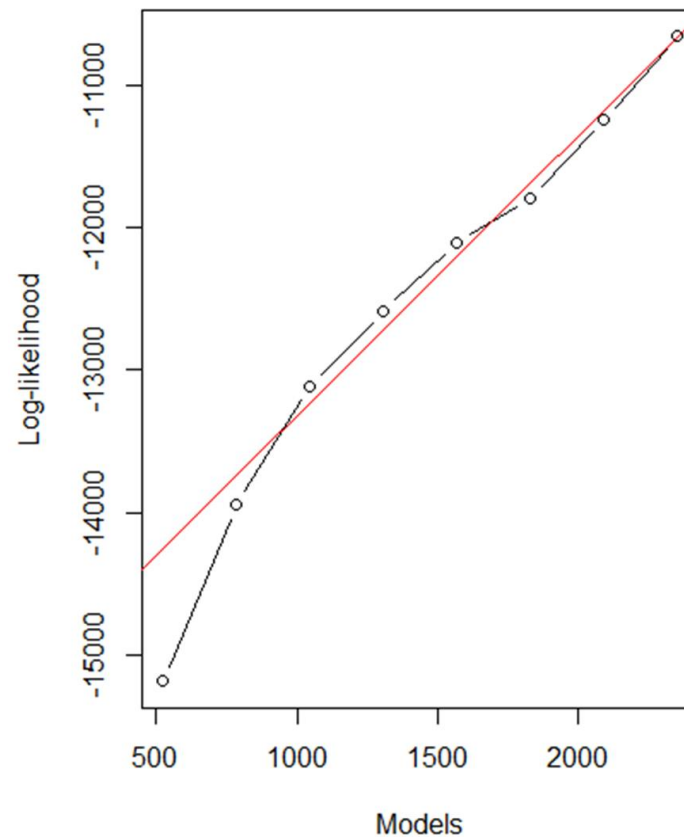
→ Un objet de  
classe `funHDDC`



# ❖ Présentation du package R

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

*slopeheuristic(mod)*





# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

```
mfpca(data)
```





# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

`mfpca(data)`

1 jeu de données  
fonctionnel ou une  
liste de jeux de  
données



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

```
plot.mfpca(x, nharm, threshold)
```



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

`plot.mfpca(x, nharm, threshold)`

↙  
Un objet de  
classe `mfpca`



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

```
plot.mfpca(x, nharm, threshold)
```

↙  
Nombre d'axes  
factoriels pour  
lesquels on souhaite  
les graphes



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

`plot.mfpca(x, nharm, threshold)`

Le seuil  
d'affichage des  
graphes



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

```
predict(mod, newdata)
```



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

```
predict(mod, newdata)
```



Un objet de classe `funHDDC`



# ❖ Présentation du package R

---

- `funHDDC`: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces

`predict(mod, newdata)`

Un jeu de données ou une liste de jeux de données





# ❖ Plan

---

Introduction

Package

**Exemples pratiques**

Présentation de la méthode

Conclusion

## ❖ Station météo : Météo du Canada

- 35 villes réparties sur le territoire
- Température et pluviométrie sur 1 an



## ❖ Station météo : Météo du Canada

- 35 villes réparties sur le territoire
- Température et pluviométrie sur 1 an



*res1* ← *funHDDC(list(temp,pluvio), K=2, model='AkjBkQkDk')*

## ❖ Station météo : Météo du Canada

- 35 villes réparties sur le territoire
- Température et pluviométrie sur 1 an



```
res1 ← funHDDC(list(temp,pluvio), K=2, model='AkjBkQkDk')
```

```
slopeheuristic(K=c(2:8),nbparam=c(res1$complexity,...,res7$complexity),  
loglik=c(max(res1$loglik),...,max(res7$loglik)))
```

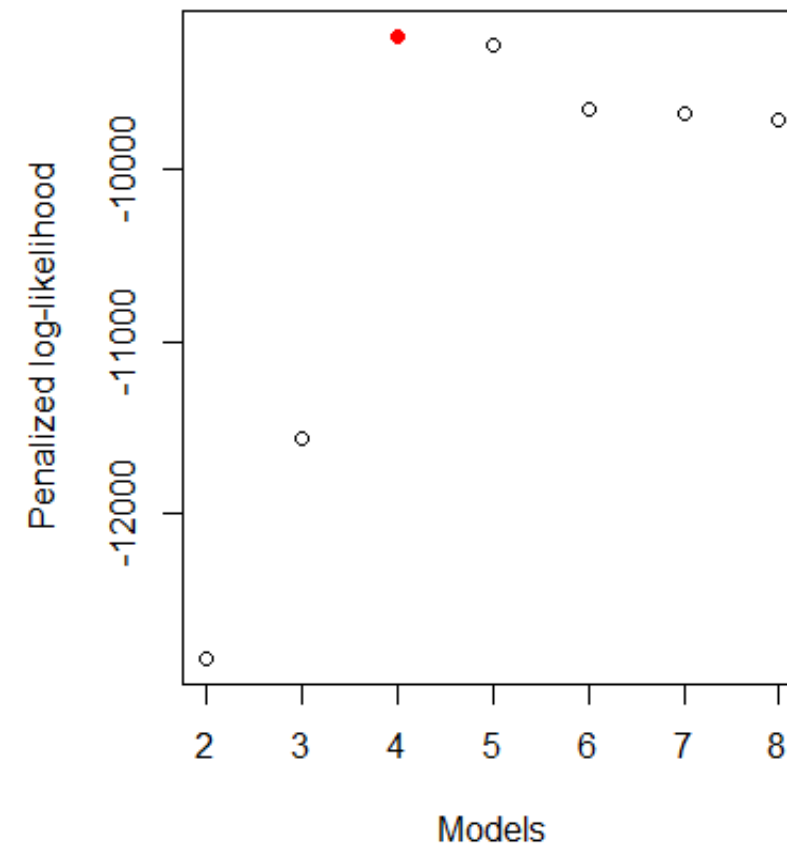
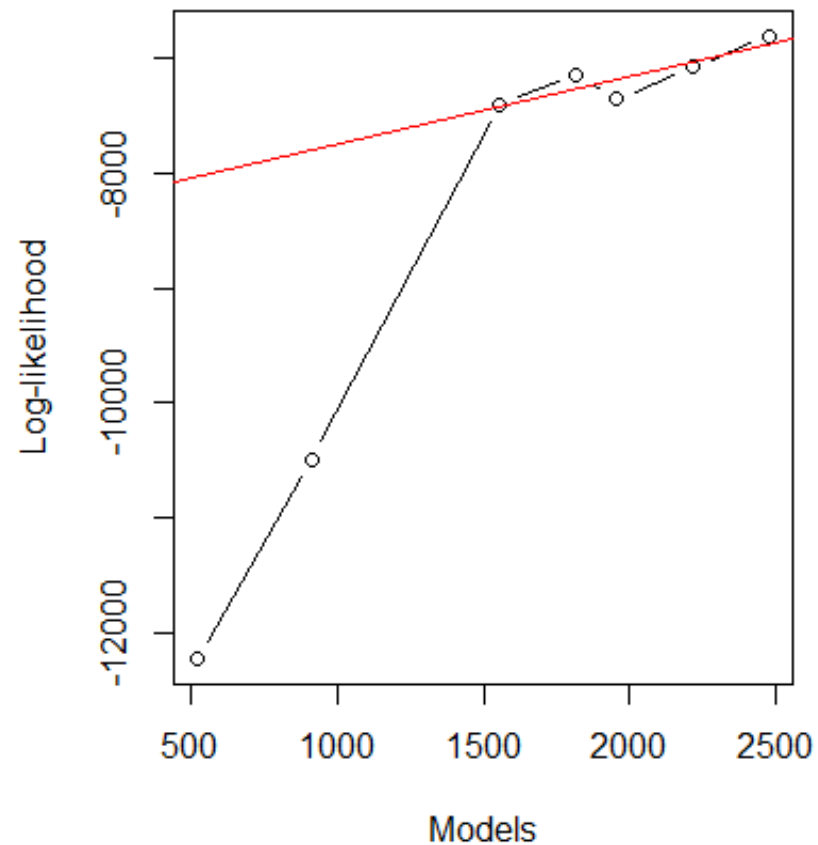


## ❖ Station météo : Météo du Canada

- 35 villes réparties sur le territoire
- Température et pluviométrie sur 1 an

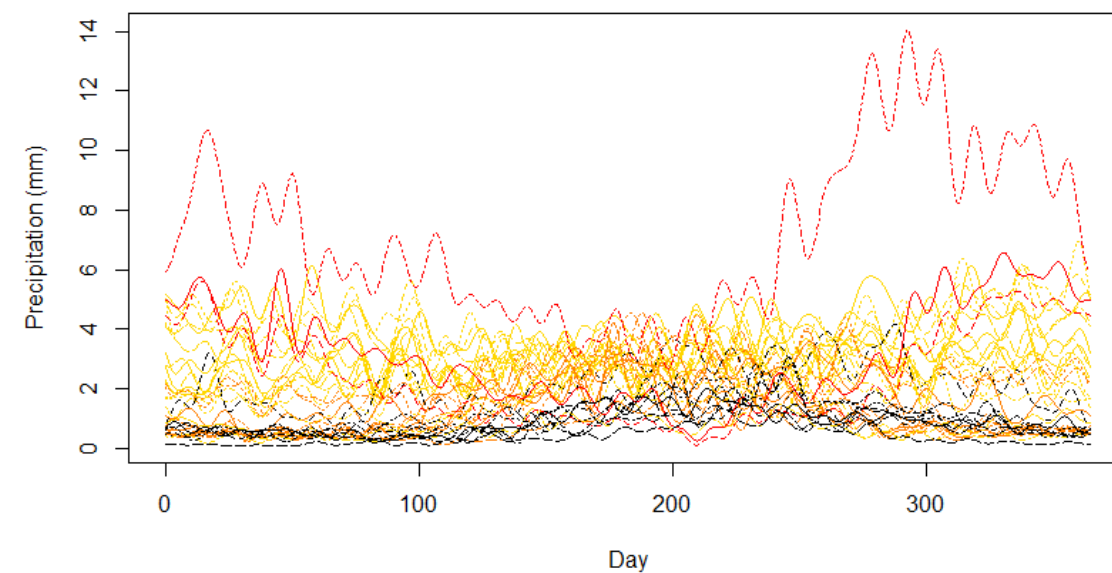
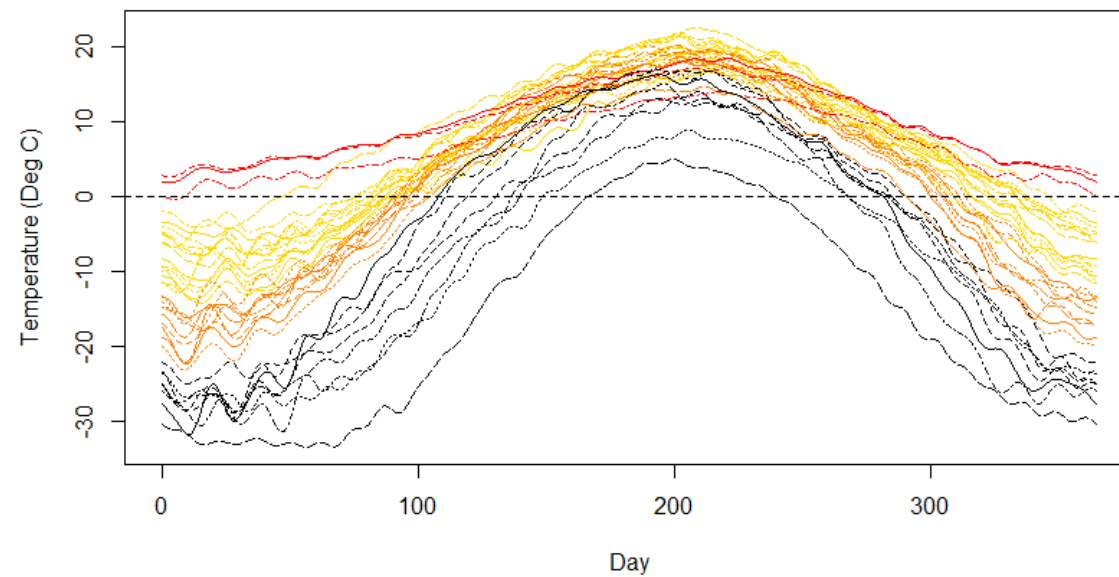
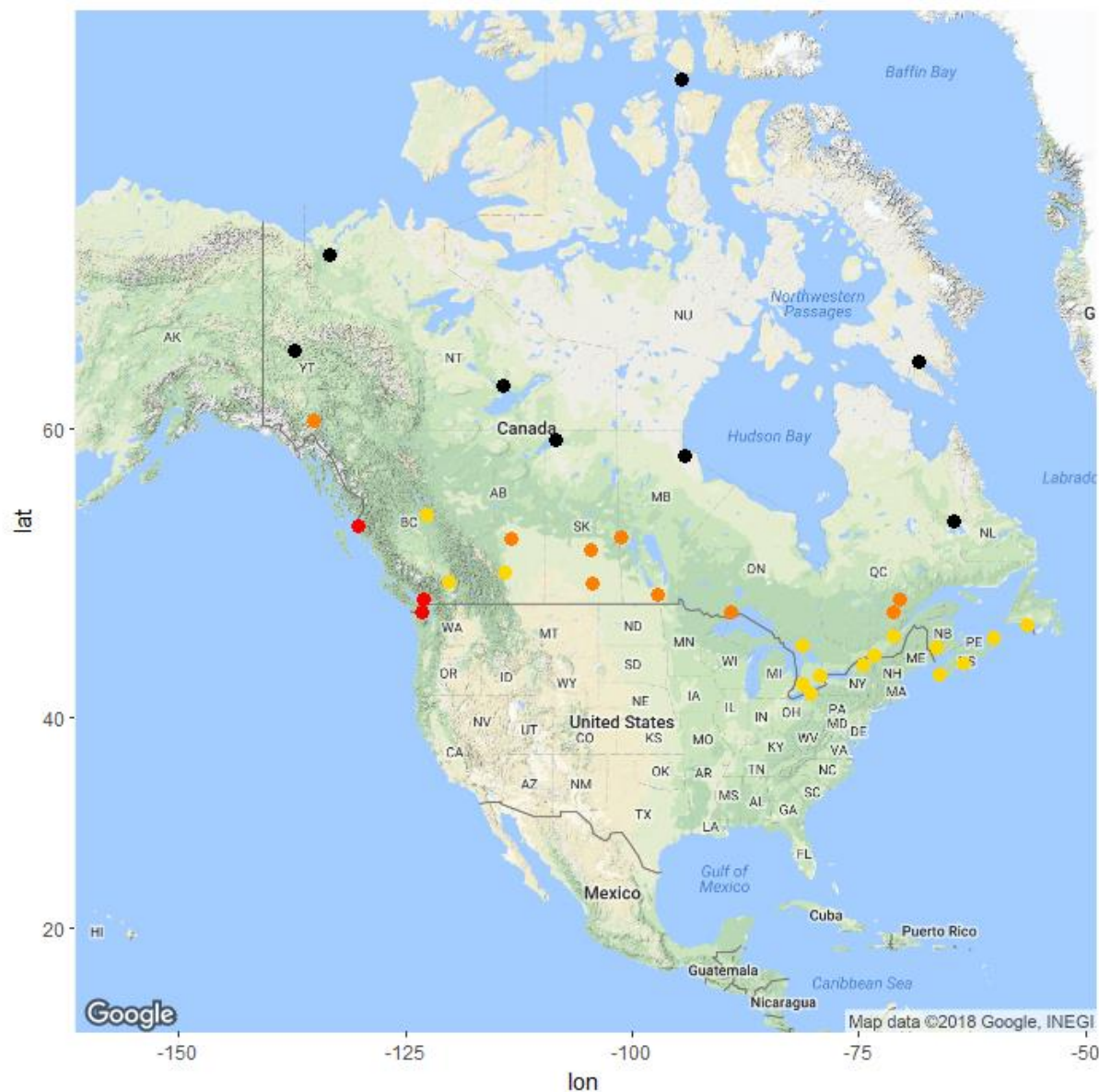


- Sélection de modèle





# ❖ Météo du Canada



## ❖ Objets connectés pour le sport

- Émergence des montres connectées pour les runners (Garmin, Polar...)
- Raquette connectée pour le tennis (Babolat...)
- Cyclisme, natation...



## ❖ Objets connectés pour le sport

- Émergence des montres connectées pour les runners (Garmin, Polar...)
- Raquette connectée pour le tennis (Babolat...)
- Cyclisme, natation...



**Absence dans le monde de l'équitation**



## ❖ Objets connectés pour le sport

- Émergence des montres connectées pour les runners (Garmin, Polar...)
- Raquette connectée pour le tennis (Babolat...)
- Cyclisme, natation...



## Absence dans le monde de l'équitation





# ❖ La discipline du saut d'obstacle en équitation





## ❖ **Objet connecté : Prédiction de la vitesse du cheval de sport**

- Outil d'aide à l'entraînement pour l'équitation





## ❖ **Objet connecté : Prédiction de la vitesse du cheval de sport**

- Outil d'aide à l'entraînement pour l'équitation

Accéléromètre & gyroscope





## ❖ **Objet connecté : Prédiction de la vitesse du cheval de sport**

- Outil d'aide à l'entraînement pour l'équitation

Accéléromètre & gyroscope

Une antenne de communication  
Bluetooth®



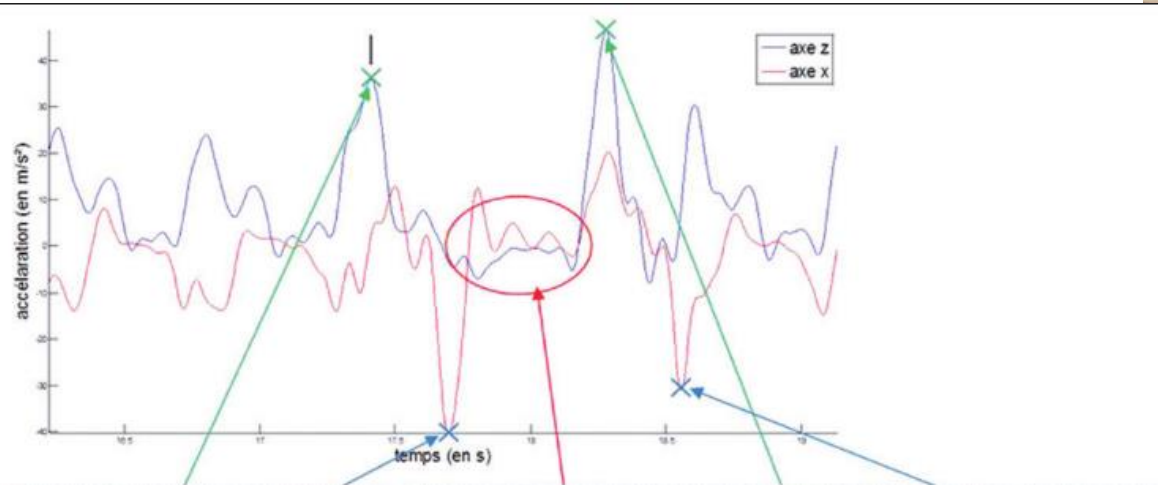


# ❖ Objet connecté : Prédiction de la vitesse du cheval de sport

- Outil d'aide à l'entraînement pour l'équitation

Accéléromètre & gyroscope

Une antenne de communication  
Bluetooth®





## ❖ **Objet connecté : Prédiction de la vitesse du cheval de sport**

- Outil d'aide à l'entraînement pour l'équitation

Accéléromètre & gyroscope

Une antenne de communication  
Bluetooth®





## ❖ Prédiction de la vitesse

---

- **Clustering** des foulées :  $funHDDC(list(az,gy),K=2)$





## ❖ Prédiction de la vitesse

---

- **Clustering** des foulées :  $funHDDC(list(az,gy),K=2)$
- **SVM** par cluster pour prédire la vitesse



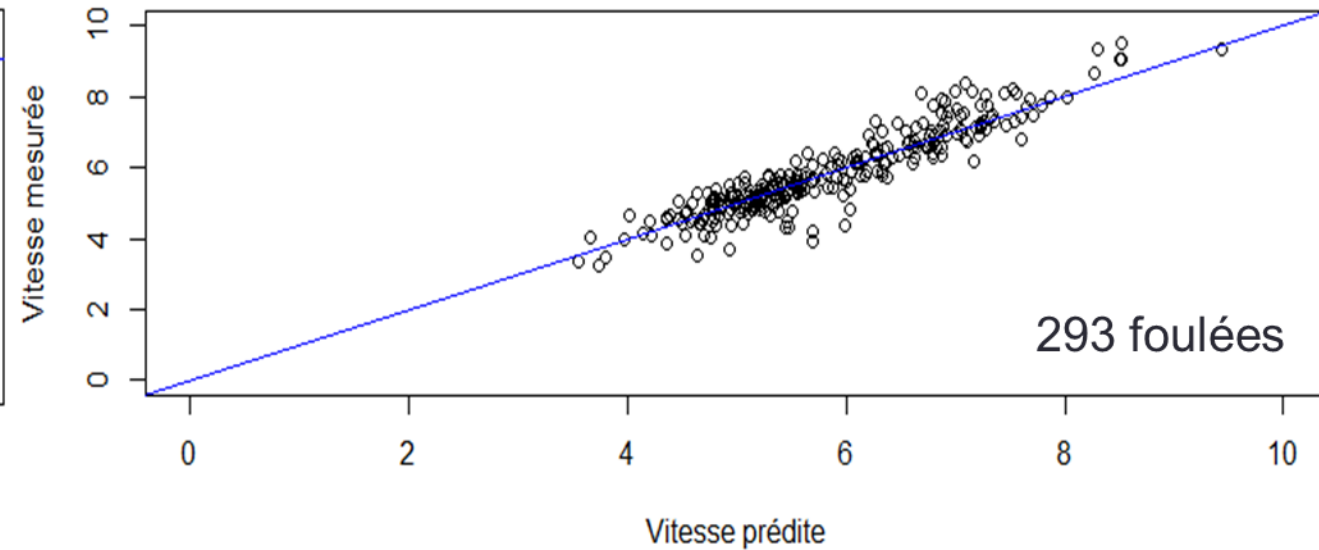
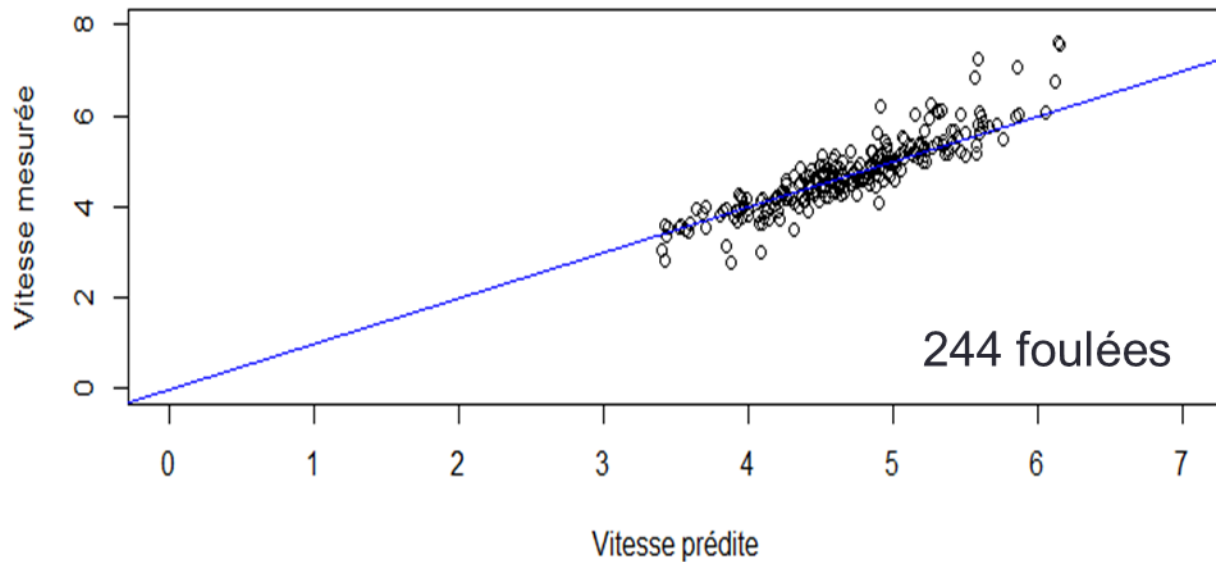
## ❖ Prédiction de la vitesse

---

- **Clustering** des foulées :  $funHDDC(list(az,gy),K=2)$
- **SVM** par cluster pour prédire la vitesse
- Evaluation du pourcentage d'erreur *supérieur à 0,6 m/s*
  - Training dataset (80%), Test dataset (20%)

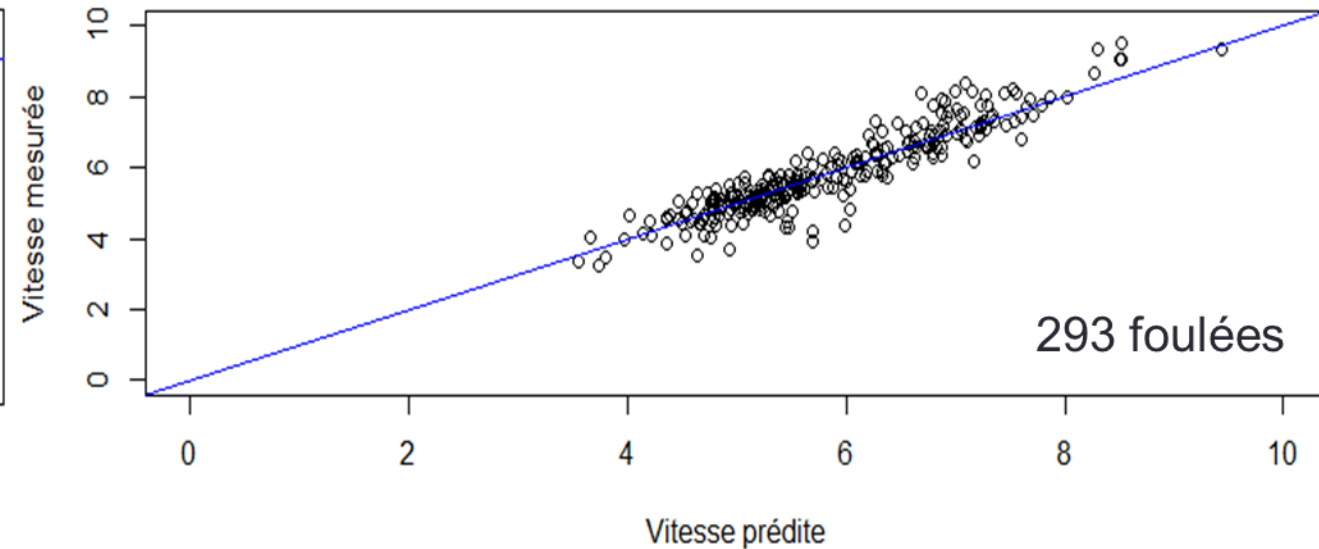
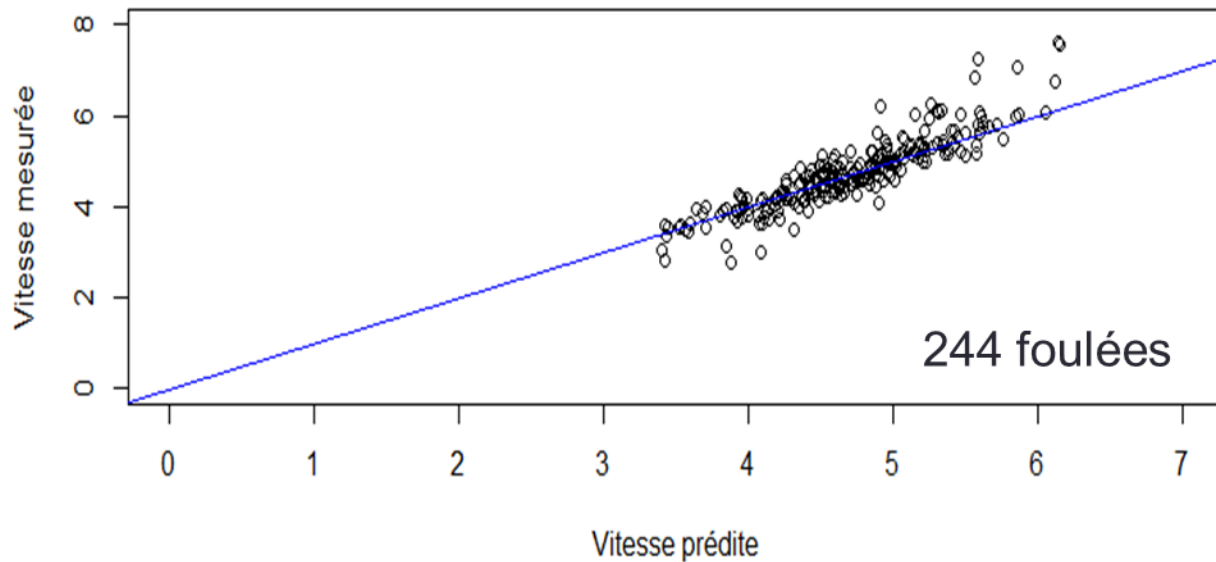
## ❖ Prédiction de la vitesse

- **Clustering** des foulées :  $funHDDC(list(az,gy),K=2)$
- **SVM** par cluster pour prédire la vitesse
- Evaluation du pourcentage d'erreur *supérieur à 0,6 m/s*
  - Training dataset (80%), Test dataset (20%)



## ❖ Prédiction de la vitesse

- **Clustering** des foulées :  $funHDDC(list(az,gy),K=2)$
- **SVM** par cluster pour prédire la vitesse
- Evaluation du pourcentage d'erreur *supérieur à 0,6 m/s*
  - Training dataset (80%), Test dataset (20%)



**10,7% d'erreurs supérieures à 0,6 m/s**

## ❖ Automatisation dans une application mobile

Objectif : **Automatiser** les calculs pour fournir un **outil d'aide à l'entraînement** pour les cavaliers



## ❖ Automatisation dans une application mobile

Objectif : **Automatiser** les calculs pour fournir un **outil d'aide à l'entraînement** pour les cavaliers



- Utilisation fonction *predict* :

```
model ← funHDDC(list(az_tot, gy_tot), K=2, model='AkjBkQkDk')
```

```
prediction ← predict(model, list(new_az, new_gy))
```

## ❖ Automatisation dans une application mobile

Objectif : **Automatiser** les calculs pour fournir un **outil d'aide à l'entraînement** pour les cavaliers



- Utilisation fonction *predict* :

```
model ← funHDDC(list(az_tot, gy_tot), K=2, model='AkjBkQkDk')
```

```
prediction ← predict(model, list(new_az, new_gy))
```

- Modèle **SVM** par cluster pour **prédire la vitesse**



# ❖ L'application client

The screenshot displays a video player interface for an equestrian event. At the top, a dark red navigation bar contains a back arrow, a timer at 00:05, a play/pause button, a progress bar with red vertical markers, a forward button, a horse silhouette icon, and a menu icon. The main video area shows a wide shot of an outdoor arena with a large white tent, a building, and various equestrian equipment. In the foreground, a brown horse is visible on the left. Overlaid at the bottom are five red circular widgets with white text and icons:

- Régularité**: 89%
- Vitesse**: 278 (with a right-pointing arrow icon below)
- Abord**: (with a horse silhouette icon)
- Symétrie**: (with a gauge icon labeled 'G' and 'D')
- Reprise**: OK





# ❖ Plan

---

Introduction

Package

Exemples pratiques

**Présentation de la méthode**

Conclusion

## ❖ Clustering fonctionnel multivarié

---

Objectif : **Séparer n courbes p-variées en K clusters**

- Les courbes sont exprimées comme une combinaison linéaire de bases de fonctions :

$$X_i^j(t) = \sum_{r=1}^{R_j} c_{ir}^j(X) \Phi_r^j(t)$$

## ❖ Clustering fonctionnel multivarié

Objectif : **Séparer n courbes p-variées en K clusters**

- Les courbes sont exprimées comme une combinaison linéaire de bases de fonctions :

$$X_i^j(t) = \sum_{r=1}^{R_j} c_{ir}^j(X) \Phi_r^j(t)$$

Matrice de coefficients

Une base de fonctions

## ❖ Clustering fonctionnel multivarié

Objectif : **Séparer n courbes p-variées en K clusters**

- Les courbes sont exprimées comme une combinaison linéaire de bases de fonctions :  
$$X_i^j(t) = \sum_{r=1}^{R_j} c_{ir}^j(X) \Phi_r^j(t)$$

Matrice de coefficients

Une base de fonctions
- Projection des courbes dans des sous-espaces de dimension réduite:

## ❖ Clustering fonctionnel multivarié

Objectif : **Séparer n courbes p-variées en K clusters**

- Les courbes sont exprimées comme une combinaison linéaire de bases de fonctions :
 
$$X_i^j(t) = \sum_{r=1}^{R_j} c_{ir}^j(X) \Phi_r^j(t)$$

Matrice de coefficients
Une base de fonctions
- Projection des courbes dans des sous-espaces de dimension réduite:
  - Calcul des scores à l'aide d'une ACPF par groupe :

$$X_k(t) = \mu_k(t) + \sum_{j=1}^R C_{kj} \psi_{kj}(t)$$

## ❖ Clustering fonctionnel multivarié

Objectif : **Séparer n courbes p-variées en K clusters**

- Les courbes sont exprimées comme une combinaison linéaire de bases de fonctions :
 
$$X_i^j(t) = \sum_{r=1}^{R_j} c_{ir}^j(X) \Phi_r^j(t)$$

Matrice de coefficients

Une base de fonctions
- Projection des courbes dans des sous-espaces de dimension réduite:
  - Calcul des scores à l'aide d'une ACPF par groupe :

Variable aléatoire fonctionnelle

$$X_k(t) = \mu_k(t) + \sum_{j=1}^R c_{kj} \psi_{kj}(t)$$

Fonction moyenne de X

Scores

Système orthonormal de fonctions propres de l'opérateur de covariance de X

## ❖ Clustering fonctionnel multivarié

---

- Application d'un modèle de mélange latent fonctionnel :

$$p(\delta) = \sum_{k=1}^K \pi_k N(\delta; \mu_k, \Delta_k)$$

## ❖ Clustering fonctionnel multivarié

- Application d'un modèle de mélange latent fonctionnel :

$$p(\delta) = \sum_{k=1}^K \pi_k N(\delta; \mu_k, \Delta_k)$$

Scores issus de l'ACPF

Probabilité prior du k-ième groupe

Fonction de densité Gaussienne



## ❖ Clustering fonctionnel multivarié

- Application d'un modèle de mélange latent fonctionnel :

$$p(\delta) = \sum_{k=1}^K \pi_k N(\delta; \mu_k, \Delta_k)$$

Scores issus de l'ACPF

Probabilité prior du k-ième groupe

Fonction de densité Gaussienne

- Définition d'un algorithme EM:

- Etape E : Calcul de la probabilité postérieure d'appartenance à la classe
- Etape M : Estimation des paramètres du modèle qui maximisent l'espérance de la vraisemblance complète conditionnellement aux probabilités postérieures



# ❖ Plan

---

Introduction

Package

Exemples pratiques

Présentation de la méthode

**Conclusion**

## ❖ Conclusion

- **Nouveau modèle** de clustering fonctionnel *univarié* et *multivarié*
- **Développement** d'un package R associé disponible sur :  
<https://cran.r-project.org/web/packages/funHDDC/index.html>
- **Modèle applicatif**
  - Prédiction de la vitesse du cheval avec une précision de 0,6 m/s
  - Données simulées sur HAL



## ❖ Bibliographie

---

- Schmutz A., Jacques J., Bouveyron C., Chèze L., Martin P. (2017). Clustering multivariate functional data in group-specific functional subspaces. Preprint HAL n°01652467.
- Yamamoto M., Terada Y. (2014). Dimension-reduced clustering of functional data via subspace separation. *Journal of Classification*, 34, 294-326.
- Jacques J., Preda C. (2014). Model based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, 71, 92-106.
- Ieva F., Paganoni A., Pigoli D., Vitelli V. (2013). Multivariate functional clustering for the morphological analysis of ECG curves. *Journal of the Royal Statistical Society, Series C*, 62(3), 401-418.