

# Analyse de sensibilité pour données climatiques

Ronan Trépos (master 2 de Franck Boizard)

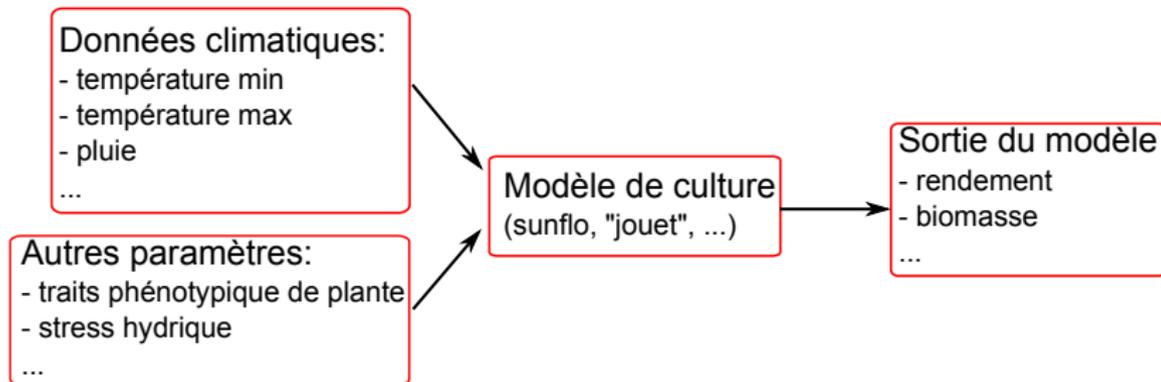
1<sup>er</sup> octobre 2015

Collaborations : F. Boizard, R. Faivre, R. Trépos et pôle transverse  
du CATI IUMA (S. Roux, S. Buis).



- 1 Présentation du stage
- 2 Analyse de sensibilité par climat
- 3 Classification non supervisée des séries climatiques
  - Classification experte
  - Classification automatique
- 4 Visualiser les résultats d'AS en utilisant la classification
- 5 Analyse de sensibilité conjointe

- Modèle de culture (dynamique et "boîte noire") :



- Objectif : calculer conjointement des indices de sensibilités pour les entrées climatiques et autres paramètres.

# Présentation du modèle "jouet"

- Entrées climatiques :  $x = (x_{rad}(t), x_{tmoy}(t), x_{et0}(t), x_{rain}(t))$
- Paramètres :  $\theta = (t_1, t_2, \tau_{tmoy}, \tau_{FTSW}, k_c, TTSW)$
- Equations :

$$y = \sum_{t_1}^{t_2} \Delta b_t(x, \theta)$$

$$\Delta b_t = RUE_t(x, \theta) \cdot x_{rad}(t)$$

$$RUE_t = (x_{tmoy}(t) < \tau_{tmoy}) \cdot \min\left(1, \frac{FTSW_{t-1}}{\tau_{FTSW}}\right)$$

$$ATSW_t^{tmp} = ATSW_{t-1} + x_{rain}(t) - k_c \cdot x_{et0}(t) \cdot \min\left(1, \frac{FTSW_{t-1}}{\tau_{FTSW}}\right)$$

$$ATSW_t = \min(TTSW, \max(0, ATSW_t^{tmp}))$$

$$FTSW_t = \frac{ATSW_t}{TTSW}$$

# Présentation du modèle "jouet" (données)

- Domaine de variation des entrées climatiques : séries de Montpellier (début : 14 nov, durée : 365 jours) de 1972 à 2014 avec les variables climatiques  $(x_{rad}(t), x_{tmoy}(t), x_{et0}(t), x_{rain}(t)) \rightarrow 42$  séries climatiques
- Domaine de variation des paramètres :

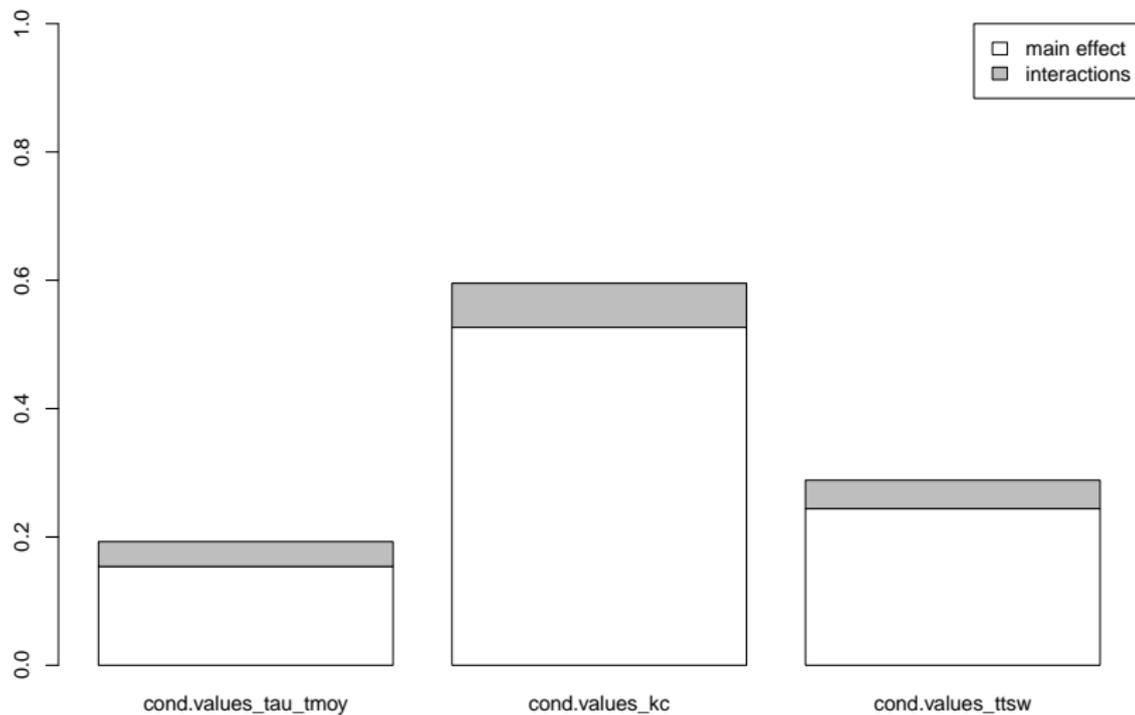
paramètre	min	max
$\tau_{tmoy}$	20	30
$k_c$	0.2	0.8
$TTSW$	50	250

- Autres paramètres et valeurs initiales fixés :

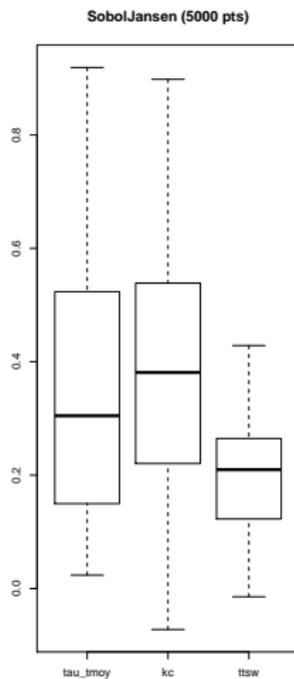
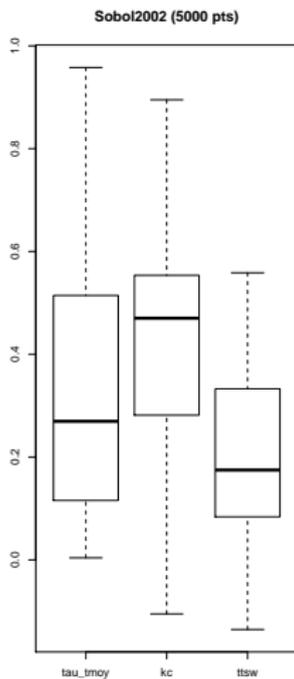
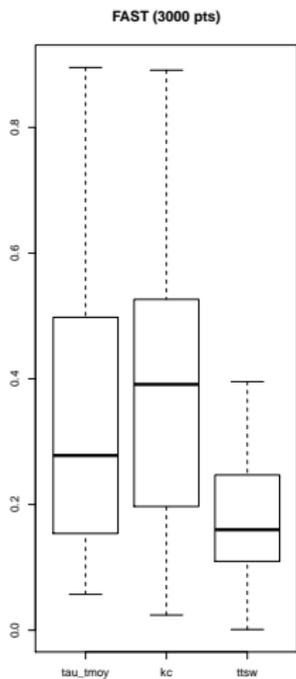
$ATSW(0)$	50
$t_1$	180
$t_2$	280
$\tau_{FTSW}$	0.4

- 1 Présentation du stage
- 2 Analyse de sensibilité par climat**
- 3 Classification non supervisée des séries climatiques
- 4 Visualiser les résultats d'AS en utilisant la classification
- 5 Analyse de sensibilité conjointe

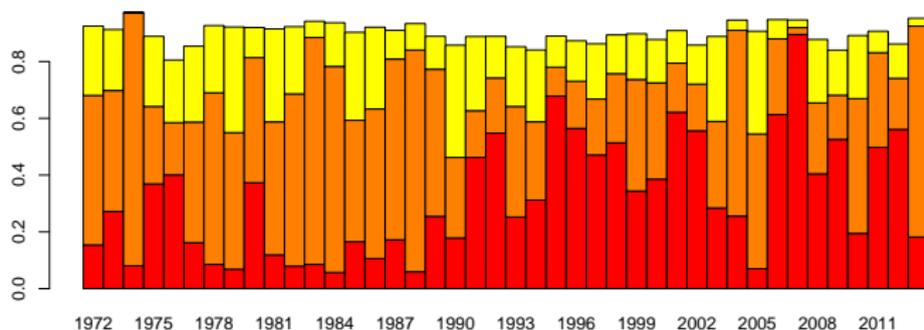
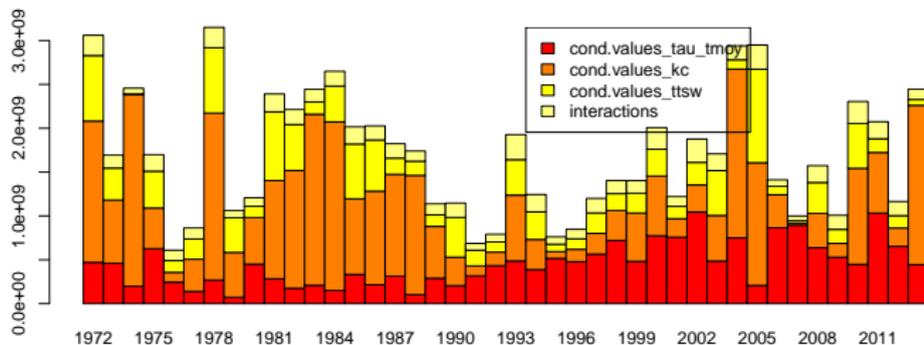
# AS pour la série de 1972 (FAST, 3000 pts)



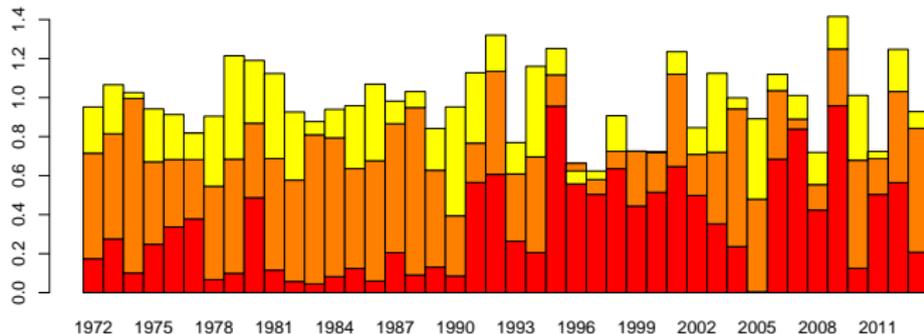
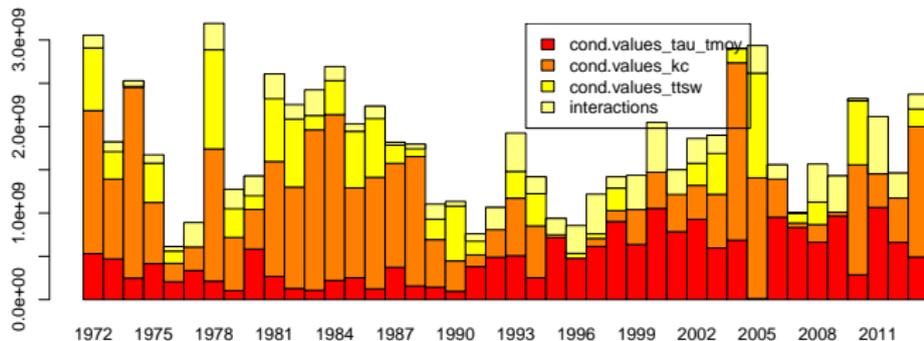
# Variabilité des indices de 1er ordre selon la série



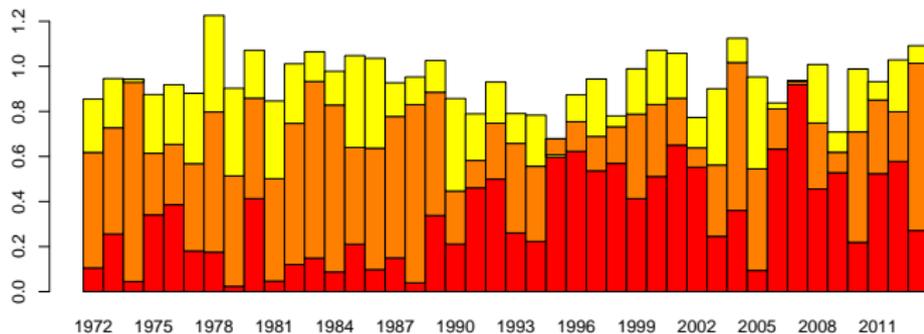
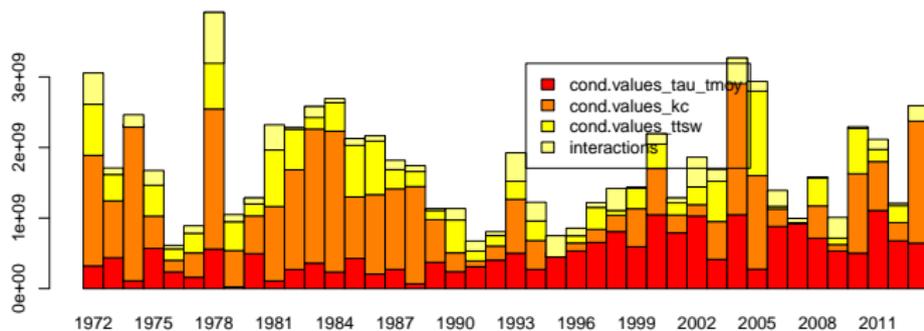
# Détails des AS (FAST)



# Détails des AS (Sobol 2002)



# Détails des AS (Sobol Jansen)



- 1 Présentation du stage
- 2 Analyse de sensibilité par climat
- 3 Classification non supervisée des séries climatiques**
  - Classification experte
  - Classification automatique
- 4 Visualiser les résultats d'AS en utilisant la classification
- 5 Analyse de sensibilité conjointe

Une première classification experte est basée directement sur la somme des pluies en période de production de biomasse (entre  $t_1 = 180$  et  $t_2 = 280$ ).

Classe1 : 1972 1978 1981 1982 1983 1984 1985\* 1986 1989 1993 2002 2005 2010  
2013

→ "peu humide"

Classe2 : 1974 1975 1987 1988 1990 1994 1997 1999 2000 2003\* 2004 2008 2009  
2012

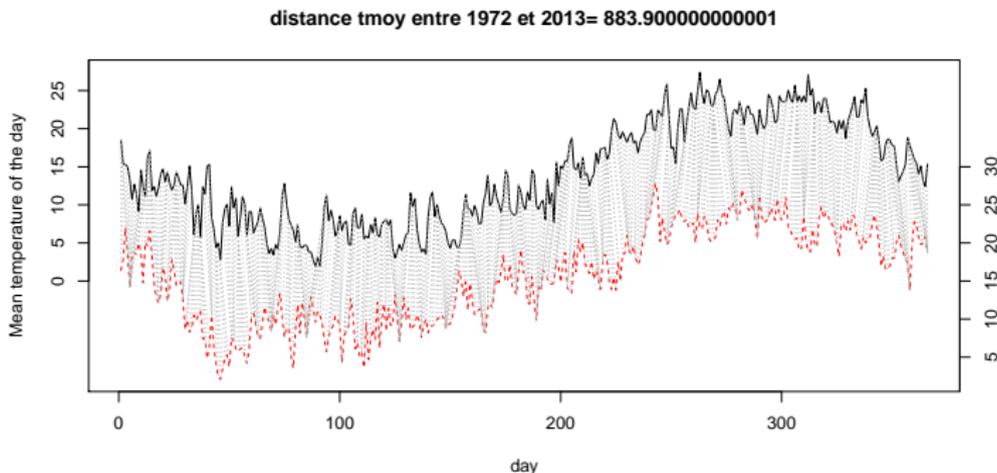
→ "moyennement humide"

Classe3 : 1973 1976 1977 1979 1980 1991 1992 1995 1996 1998 2001 2006\* 2007  
2011

→ "très humide"

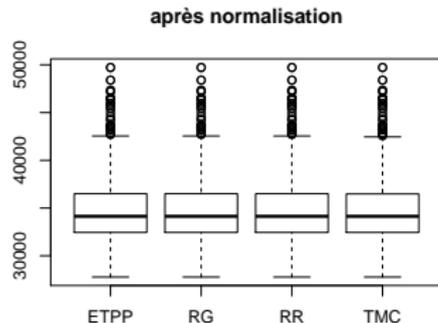
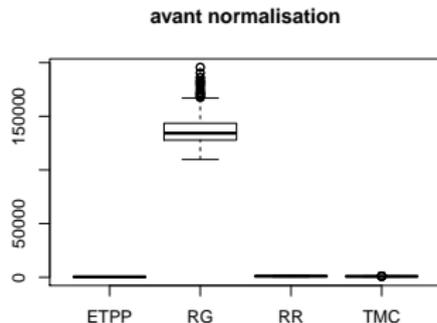
- 1 Présentation du stage
- 2 Analyse de sensibilité par climat
- 3 Classification non supervisée des séries climatiques**
  - Classification experte
  - Classification automatique
- 4 Visualiser les résultats d'AS en utilisant la classification
- 5 Analyse de sensibilité conjointe

- Définition d'une distance entre séries temporelles (dtw)



- Normalisation quantile et somme des distances sur les variables climatiques → distance entre 2 séries climatiques
- Classification ascendante hiérarchique (CAH)
- Détermination des centres de classes à partir de la CAH pour initialiser une méthode de K-medoids

- Définition d'une distance entre séries temporelles (dtw)
- **Normalisation quantile et somme des distances sur les variables climatiques** → distance entre 2 séries climatiques



- Classification ascendante hiérarchique (CAH)
- Détermination des centres de classes à partir de la CAH pour initialiser une méthode de K-medoïdes

- Définition d'une distance entre séries temporelles (dtw)
- Normalisation quantile et somme des distances sur les variables climatiques → distance entre 2 séries climatiques
- **Classification ascendante hiérarchique (CAH)**
- Détermination des centres de classes à partir de la CAH pour initialiser une méthode de K-medoïdes

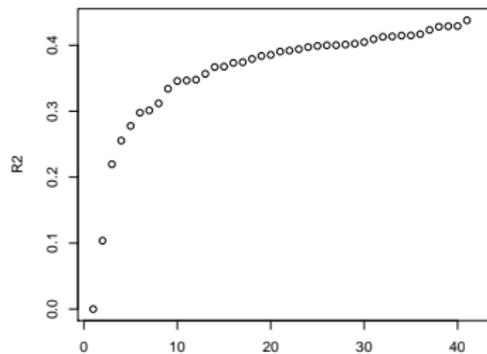
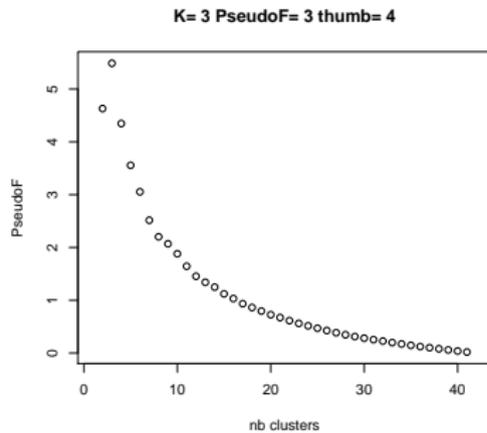
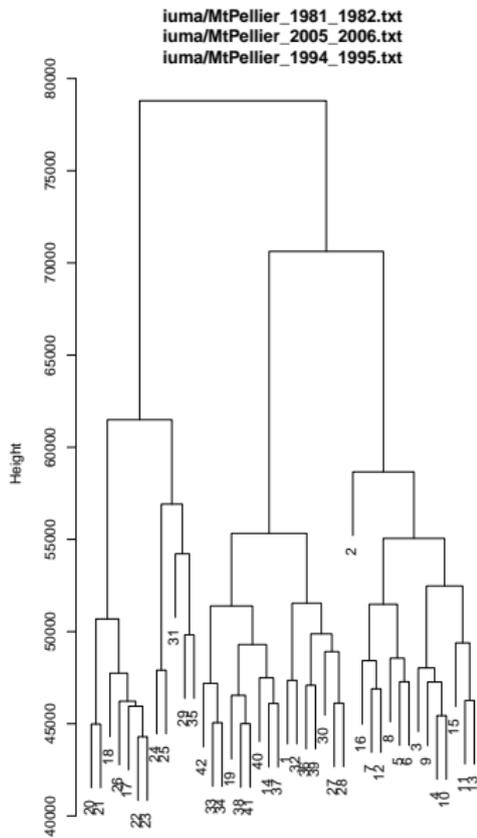
- Définition d'une distance entre séries temporelles (dtw)
- Normalisation quantile et somme des distances sur les variables climatiques → distance entre 2 séries climatiques
- Classification ascendante hiérarchique (CAH)
- **Détermination des centres de classes à partir de la CAH pour initialiser une méthode de K-medoïdes**

$$R2(k) = \frac{I_{inter}(k)}{I_{tot}}$$

$$F(k) = \frac{R2(k)}{1-R2(k)} * \frac{n-k}{k-1}$$

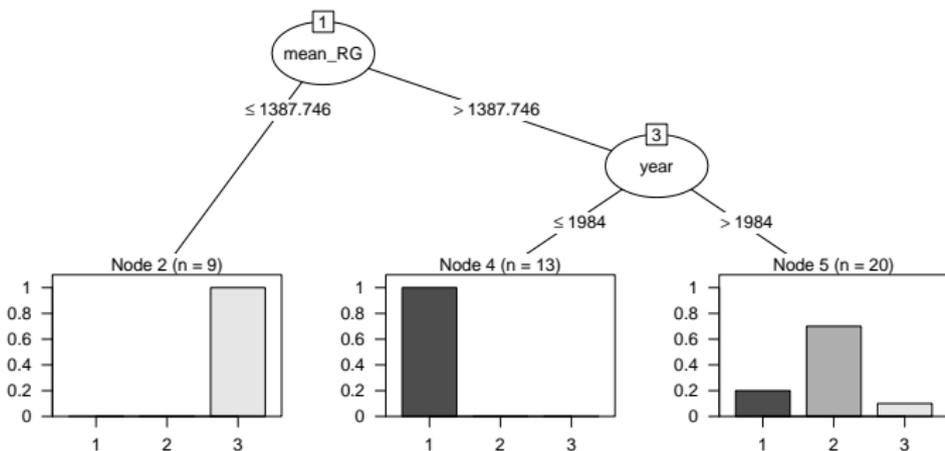
$$k_t = \sqrt{\frac{n}{2}}$$

# Résultats de classification



# Caractérisation des classes (apprentissage supervisé)

Arbre de décision : moyenne des vars clim. + année vs classe.



Classe1 : 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981\* 1982 1983 1984  
1987 2006 2007 2010

→ “très lumineuse avant 84”

Classe2 : 1985 1986 1990 1998 1999 2000 2001 2002 2004 2005\* 2008 2009 2012  
2013

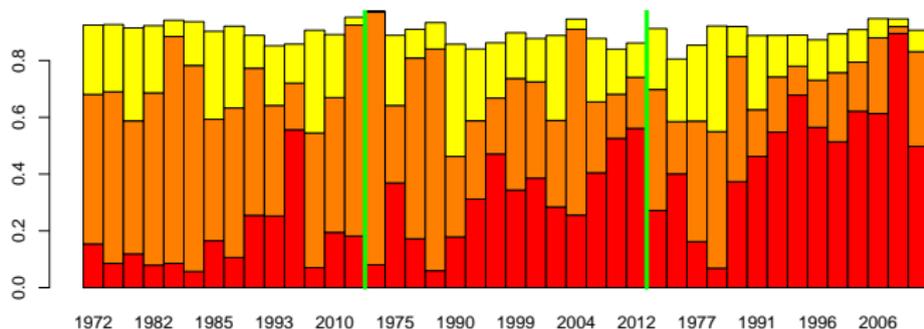
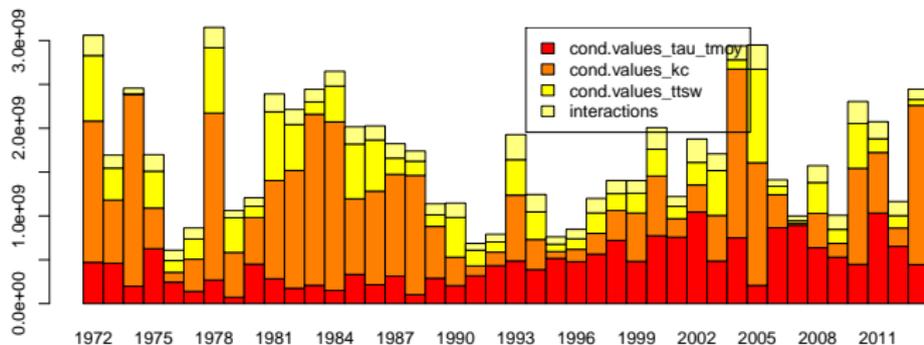
→ “très lumineuse après 84”

Classe3 : 1988 1989 1991 1992 1993 1994\* 1995 1996 1997 2003 2011

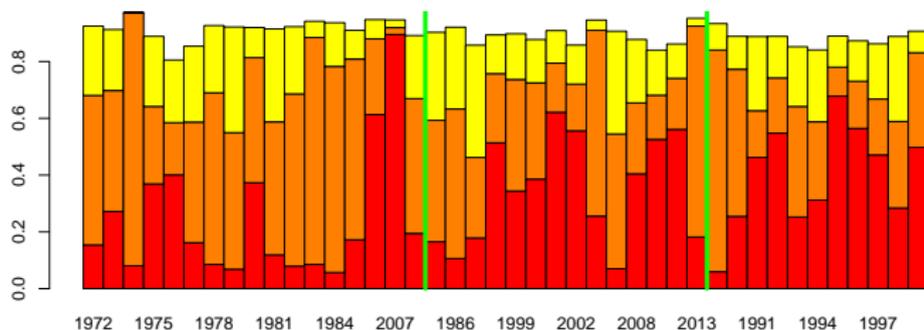
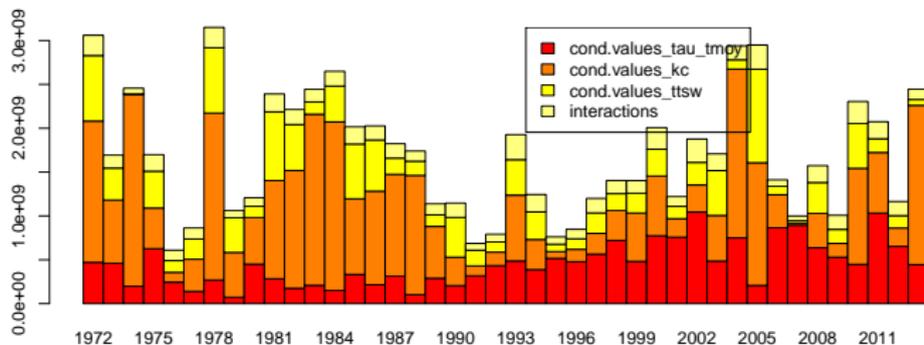
→ “peu lumineuse”

- 1 Présentation du stage
- 2 Analyse de sensibilité par climat
- 3 Classification non supervisée des séries climatiques
- 4 Visualiser les résultats d'AS en utilisant la classification**
- 5 Analyse de sensibilité conjointe

# Réarrangement des indices (FAST, classif experte)

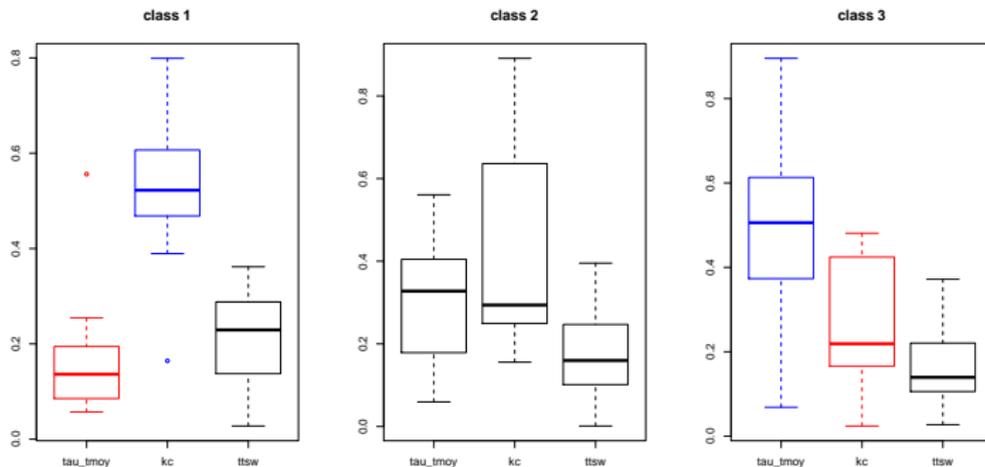


# Réarrangement des indices (FAST, classif automatique)



# Analyse de variance des indices de sensibilités

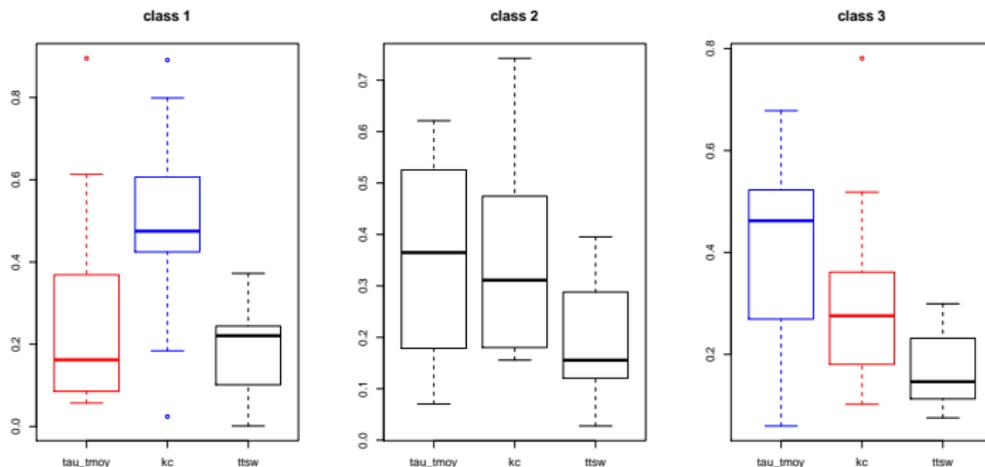
Test de Turkey ( $p\_value$  à 0.2) sur une aov des indices de sensibilités issus de FAST pour la classification experte.



- *tau\_tmoy* moins influent sur la biomasse en situation peu humide ?
- *kc* moins influent sur la biomasse en situation très humide ?

# Analyse de variance des indices de sensibilités

Test de Turkey ( $p\_value$  à 0.2) sur une aov des indices de sensibilités issus de FAST pour la classification automatique.



- *tau\_tmoy* moins influent sur la biomasse en situation très lumineuse avant 84 ?
- *kc* moins influent sur la biomasse en situation peu lumineuse ?

- 1 Présentation du stage
- 2 Analyse de sensibilité par climat
- 3 Classification non supervisée des séries climatiques
- 4 Visualiser les résultats d'AS en utilisant la classification
- 5 Analyse de sensibilité conjointe**

- Problématique : intégrer dans l'AS la série climatique comme un paramètre d'entrée discret non ordonné :  
 $C \in \{1972, 1973, \dots, 2013\}$
- Morris et fast non adaptées pour ce type de données → utilisation du *pick and freeze* de Sobol.

L'utilisateur fournit 2 matrices :  $A := (a_{i,j})$ ,  $B := (b_{i,j})$

$1 \leq i \leq n$ ,  $1 \leq j \leq M$ , où  $n$  est choisi et  $M$  est le nbre de params.

Plan simulé :

$$X = \begin{bmatrix} A \\ B \\ C^1 \\ \vdots \\ C^M \end{bmatrix}$$

où :

$$C^j = \begin{bmatrix} a_{1,1} & \cdots & a_{1,j-1} & b_{1,j} & a_{1,j+1} & \cdots & a_{1,M} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & b_{n,j} & a_{n,j+1} & \cdots & a_{n,M} \end{bmatrix}$$

# Introduction de la variable climat

- la méthode proposée consiste à ajouter une colonne ds  $A$  et  $B$  pour la variable climat.
- on cherche à construire ces deux colonnes (dans  $A$  et  $B$ ) de manière “orthogonale”.

On part de  $A$  et  $B$  définies comme précédemment (sur les  $M$  paramètres continus) en utilisant un hypercube latin de taille  $n$ .

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,M} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,M} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{1,1} & \cdots & b_{1,M} \\ \vdots & & \vdots \\ b_{n,1} & \cdots & b_{n,M} \end{bmatrix}$$

Par la suite, on identifie une série climatique par un indice  $k$ ,  $1 \leq k \leq K = 42$ .

# Introduction de la variable climat

$$A^k := \begin{bmatrix} k & a_{1,1} & \cdots & a_{1,M} \\ \vdots & \vdots & & \vdots \\ k & a_{n,1} & \cdots & a_{n,M} \end{bmatrix}$$

$$B^k := \begin{bmatrix} k & b_{1,1} & \cdots & b_{1,M} \\ \vdots & \vdots & & \vdots \\ k & b_{n,1} & \cdots & b_{n,M} \end{bmatrix}$$

$$\mathbf{A}^k := \left. \begin{bmatrix} A^k \\ \vdots \\ A^k \end{bmatrix} \right\} (K-1) * n$$

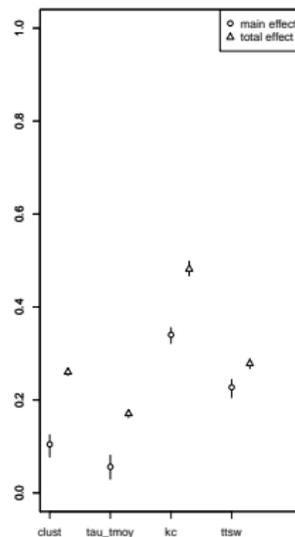
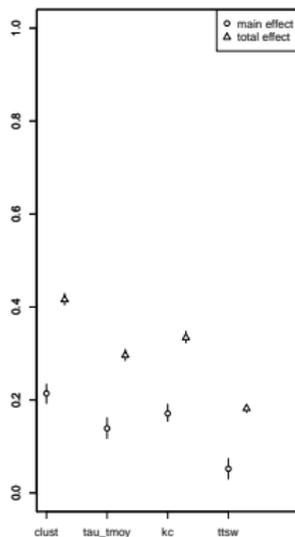
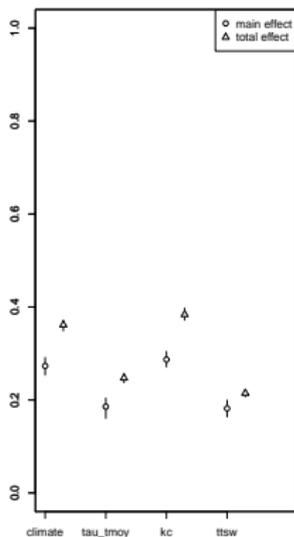
$$\mathbf{B}^{-k} := \left. \begin{bmatrix} B^1 \\ \vdots \\ B^{k-1} \\ B^{k+1} \\ \vdots \\ B^K \end{bmatrix} \right\} (K-1) * n$$

$$\mathbf{A} := \left. \begin{bmatrix} \mathbf{A}^1 \\ \vdots \\ \mathbf{A}^K \end{bmatrix} \right\} K * (K-1) * n$$

$$\mathbf{B} := \left. \begin{bmatrix} \mathbf{B}^{-1} \\ \vdots \\ \mathbf{B}^{-K} \end{bmatrix} \right\} K * (K-1) * n$$

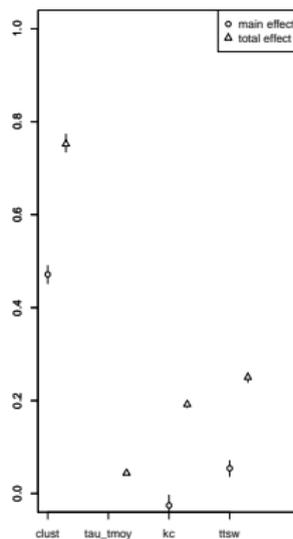
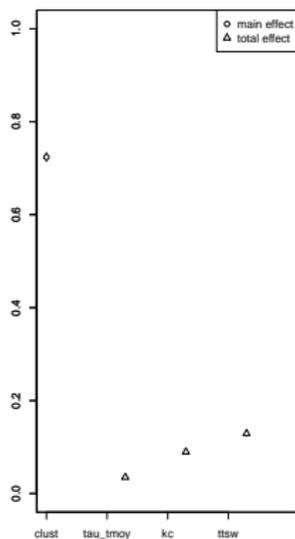
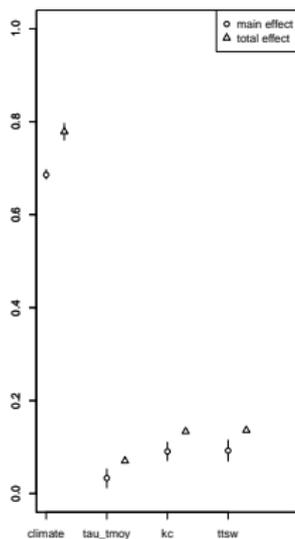
# Résultats : comparaison avec et sans classif (jansen)

- a sans classif ( $K = 42, M = 4$ ),  $n = 10 \rightarrow K * (K - 1) * n * (M + 2) = 103320$
- b classif experte ( $K = 3, M = 4$ ),  $n = 2000 \rightarrow K * (K - 1) * n * (M + 2) = 72000$
- c classif auto ( $K = 3, M = 4$ ),  $n = 2000 \rightarrow K * (K - 1) * n * (M + 2) = 72000$



# Résultats : impact de l'intervalle des variables

Expe : intervalle de variation des paramètres  $\{tau\_tmoy, kc, ttsw\}$  divisé par 2.



- plus d'effet du climat, lorsque les autres paramètres varient moins ?

- Aspects méthodologiques :
  - reproduction des AS pour plus de robustesse (plusieurs graines aléatoires)
  - vérifications à apporter sur l'utilisation de Sobol 2002 (et Sobol Jansen)
  - justification méthodologique plus poussée pour l'AS conjointe
  - classification : choix d'un représentant ou tirage dans la classe de climats
  - étude interactions climat / autres paramètres
- Aspects informatiques :
  - génération de tutoriels R pour les plateformes du CATI IUMA (OpenFluid, Sol virtuel, OpenAlea, Capsis, RECORD) dans le cadre du pari scientifique : "Approches multi-plateformes pour la résolution de problèmes avancés en Analyse de Sensibilité"
  - intégration avec le paquet R mtk (Mexico Toolkit) déjà sur le CRAN.