

Modélisation graphique d'événements extrêmes

Rita Maatouk^{1,2}, Thomas Opitz², Mike Pereira¹

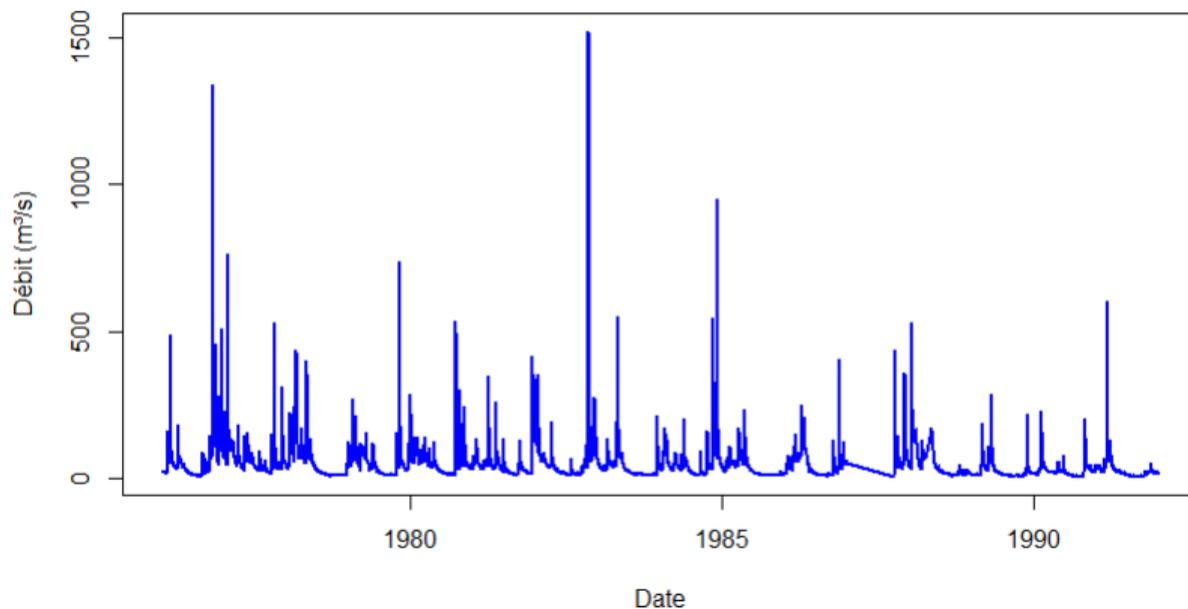
¹ Geostatistique/Centre de Géosciences de l'Ecole de Mines Paris - PSL

² BIOSP/INRAE



Comment étudier... ?

Évolution du débit à une station de la Garonne entre janvier 1976 et décembre 1991



Comment étudier... ?

Step 1 :

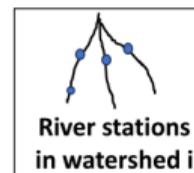


Comment étudier... ?

Step 1 :



Step 2 :

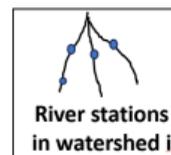


Comment étudier... ?

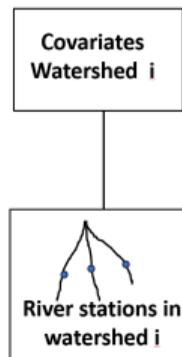
Step 1 :



Step 2 :



Step 3 :

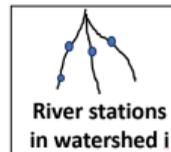


Comment étudier... ?

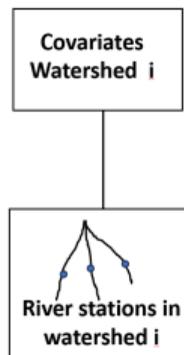
Step 1 :



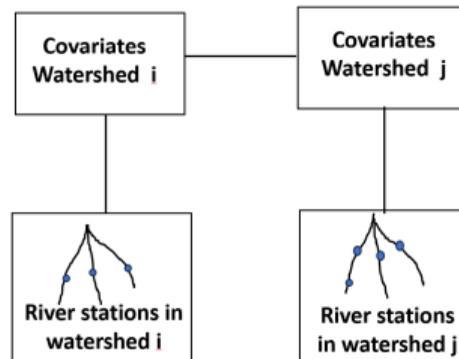
Step 2 :



Step 3 :



Step 4 :



Élément clé : Les événements extrêmes multivariés

$$\mathbf{X} = (X_1, \dots, X_d) \xrightarrow{\text{Transformation des marginales}} \mathbf{X}^* = (X_1^*, \dots, X_d^*) \longrightarrow \mathbf{Y} = \left\{ \frac{\mathbf{X}^*}{u} \mid \max(\mathbf{X}^*) > u \right\}$$

\mathbf{Y} suit une MGPD
lorsque u tend vers l'infini

MGPD : Distribution multivariée de Pareto généralisée

Élément clé : Modèle graphique Hüsler-Reiss $HR(\Gamma)$

Ingrédients :

- **Graphe** $G = (V, E)$. $\mathbf{Y} \in \mathbb{R}^d$ suit une MGPD, Y_i les sommets du graphe

Élément clé : Modèle graphique Hüsler-Reiss $HR(\Gamma)$

Ingrédients :

- **Graphe** $G = (V, E)$. $\mathbf{Y} \in \mathbb{R}^d$ suit une MGPD, Y_i les sommets du graphe
- **Paramètre** Γ la matrice de variogramme extrémal
 $\Gamma_{ij} = 0.5 * \mathbb{V}[\log(X_i^*) - \log(X_j^*) | X_j^* > u]$ où \mathbb{V} est la variance.

Élément clé : Modèle graphique Hüsler-Reiss $HR(\Gamma)$

Ingrédients :

- **Graphe** $G = (V, E)$. $\mathbf{Y} \in \mathbb{R}^d$ suit une MGPD, Y_i les sommets du graphe
- **Paramètre** Γ la matrice de variogramme extrémal
 $\Gamma_{ij} = 0.5 * \mathbb{V}[\log(X_i^*) - \log(X_j^*) | X_j^* > u]$ où \mathbb{V} est la variance.
- **Matrice de précision** Θ l'inverse généralisé de $P(\frac{-1}{2}\Gamma)P$, avec $P = I - d^{-1}\mathbf{1}\mathbf{1}^T$.

Élément clé : Modèle graphique Hüsler-Reiss $HR(\Gamma)$

Ingrédients :

- **Graphe** $G = (V, E)$. $\mathbf{Y} \in \mathbb{R}^d$ suit une MGPD, Y_i les sommets du graphe
- **Paramètre** Γ la matrice de variogramme extrémal
 $\Gamma_{ij} = 0.5 * \mathbb{V}[\log(X_i^*) - \log(X_j^*) | X_j^* > u]$ où \mathbb{V} est la variance.
- **Matrice de précision** Θ l'inverse généralisé de $P(\frac{-1}{2}\Gamma)P$, avec $P = I - d^{-1}\mathbf{1}\mathbf{1}^T$.



Élément clé : Modèle graphique Hüsler-Reiss $HR(\Gamma)$

Ingrédients :

- **Graphe** $G = (V, E)$. $\mathbf{Y} \in \mathbb{R}^d$ suit une MGPD, Y_i les sommets du graphe
- **Paramètre** Γ la matrice de variogramme extrêmeal
 $\Gamma_{ij} = 0.5 * \mathbb{V}[\log(X_i^*) - \log(X_j^*) | X_j^* > u]$ où \mathbb{V} est la variance.
- **Matrice de précision** Θ l'inverse généralisé de $P(\frac{-1}{2}\Gamma)P$, avec $P = I - d^{-1}\mathbf{1}\mathbf{1}^T$.

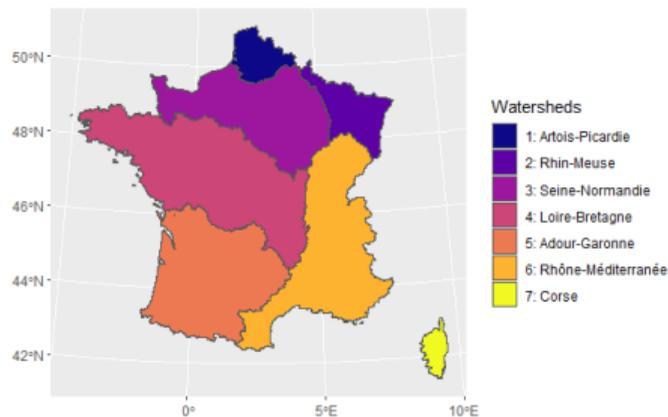


Résultat :

Modélisation d'événements extrêmes multivariés $\mathbf{Y} \in \mathbb{R}^d$, en utilisant un graphe indiquant la dépendance conditionnelle entre les variables quand au moins une variable dépasse un seuil élevé.

Notion d'indépendance conditionnelle des extrêmes

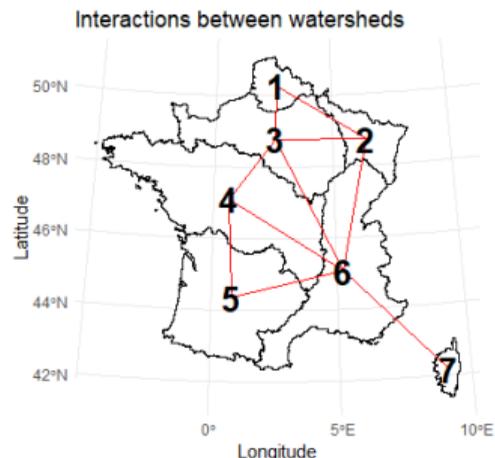
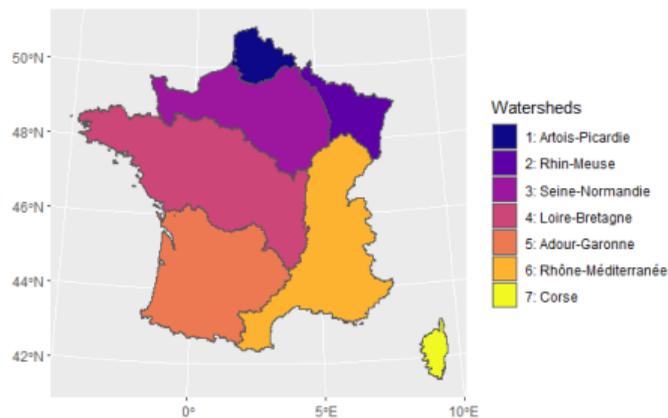
Exemple : Bassins versants en France



Notion d'indépendance conditionnelle des extrêmes

pas d'arête entre Y_i et $Y_j \iff Y_i \perp_e Y_j \mid \mathbf{Y}_{V \setminus \{i,j\}} \iff \Theta[i,j] = 0$

Exemple : Bassins versants en France



En pratique

**Transformation
des marginales
en loi exponen-
tielle de Laplace**

$$\mathbf{X}^* = (X_1^*, \dots, X_d^*)$$

En pratique

**Transformation
des marginales
en loi exponen-
tielle de Laplace**

$$\mathbf{X}^* = (X_1^*, \dots, X_d^*)$$

Sélection des extrêmes

Au moins une
variable dépasse
un seuil élevé

$$\mathbf{Y} = \left\{ \frac{\mathbf{X}^*}{u} \mid \max(\mathbf{X}^*) > u \right\}$$

qui suit une MGPD

En pratique

**Transformation
des marginales
en loi exponen-
tielle de Laplace**

$$\mathbf{X}^* = (X_1^*, \dots, X_d^*)$$

**Sélection des
extrêmes**

Au moins une
variable dépasse
un seuil élevé

$$\mathbf{Y} = \left\{ \frac{\mathbf{X}^*}{u} \mid \max(\mathbf{X}^*) > u \right\}$$

qui suit une MGPD

**Modélisation
des événements
extrêmes \mathbf{Y}**

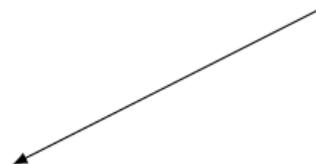
En utilisant le
modèle graphique
Hüsler-Reiss $HR(\Gamma)$

En pratique

Modélisation des événements extrêmes \mathbf{Y}



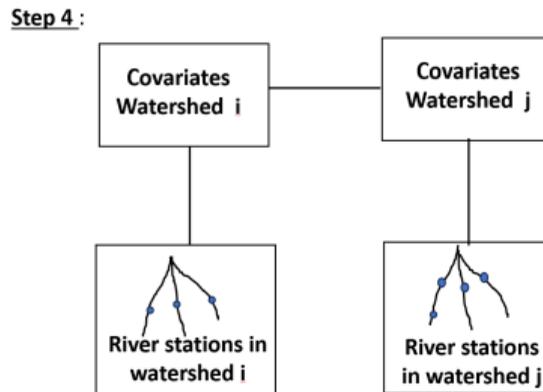
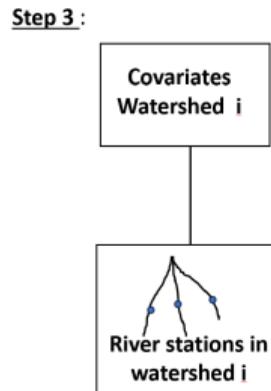
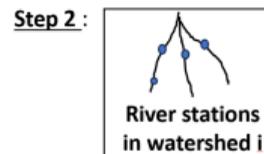
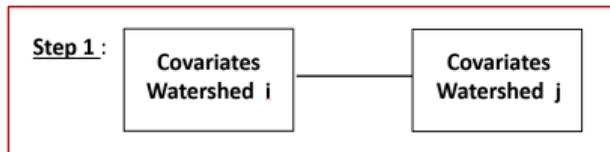
Estimation des paramètres du modèle Γ et Θ



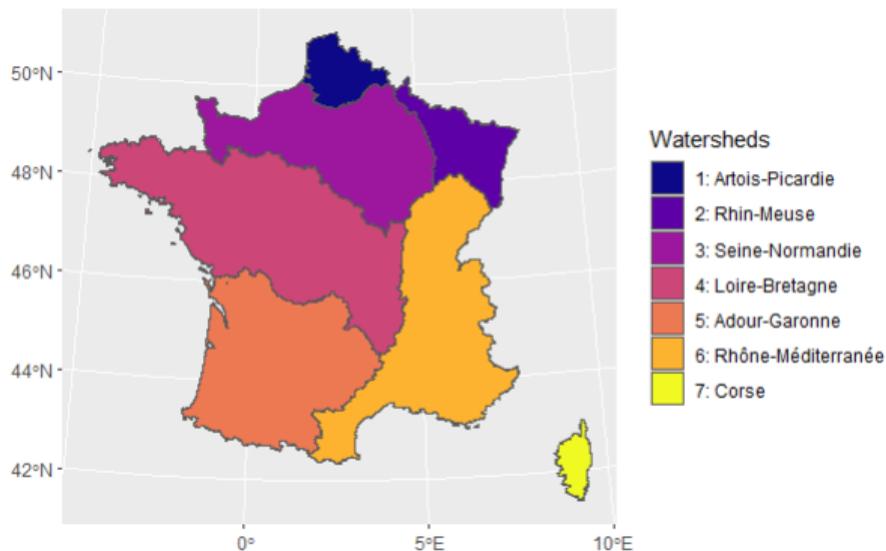
Approche utilisant un graphe imposé

Approche qui apprend des données
et renvoie un graphe

Première étape

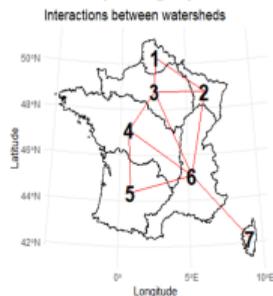


Etude des précipitations extrêmes entre les bassins versants de la France



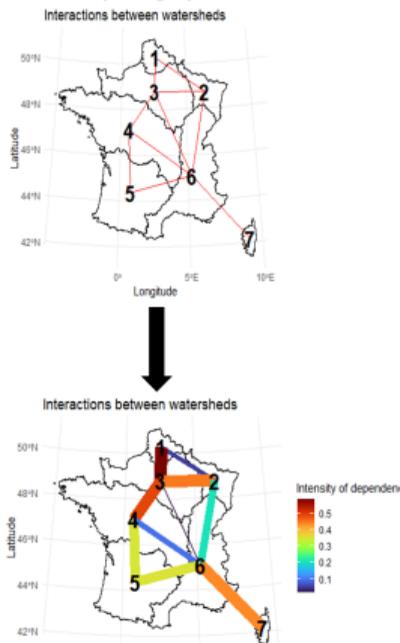
Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les précipitations

Approach 1 : Impose graph structure



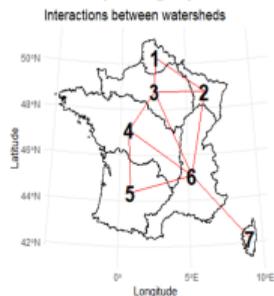
Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les précipitations

Approach 1 : Impose graph structure

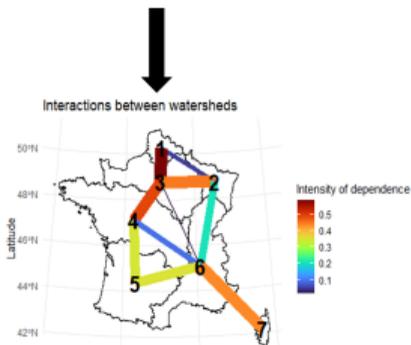
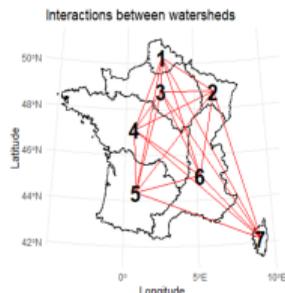


Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les précipitations

Approach 1 : Impose graph structure

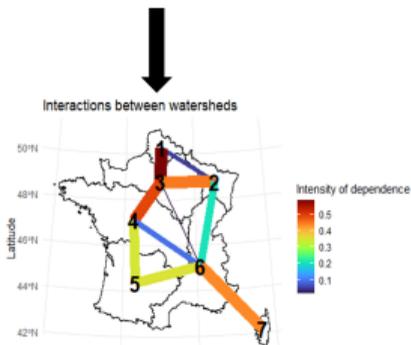
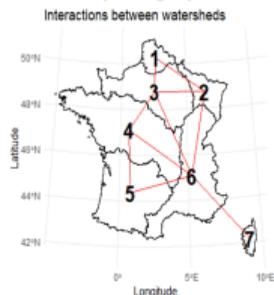


Approach 2 : Learn graph structure

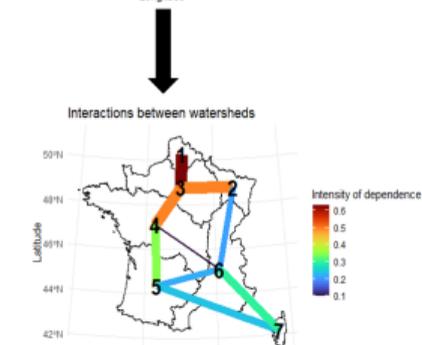
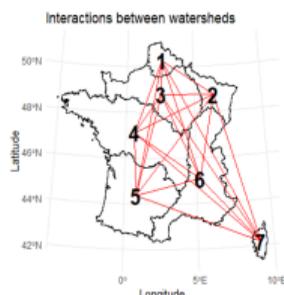


Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les précipitations

Approach 1 : Impose graph structure

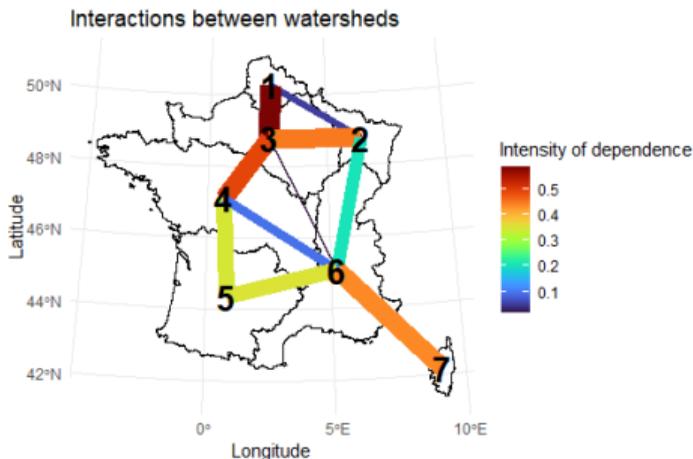


Approach 2 : Learn graph structure

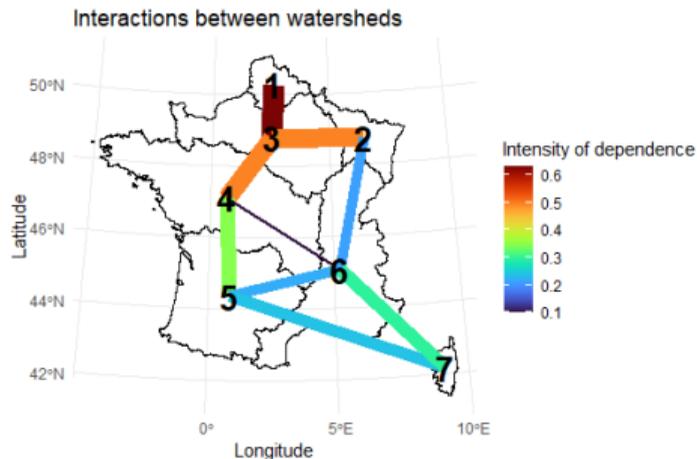


Resultats : Comparaison des graphes des 2 approches

Approach 1 : Impose graph structure

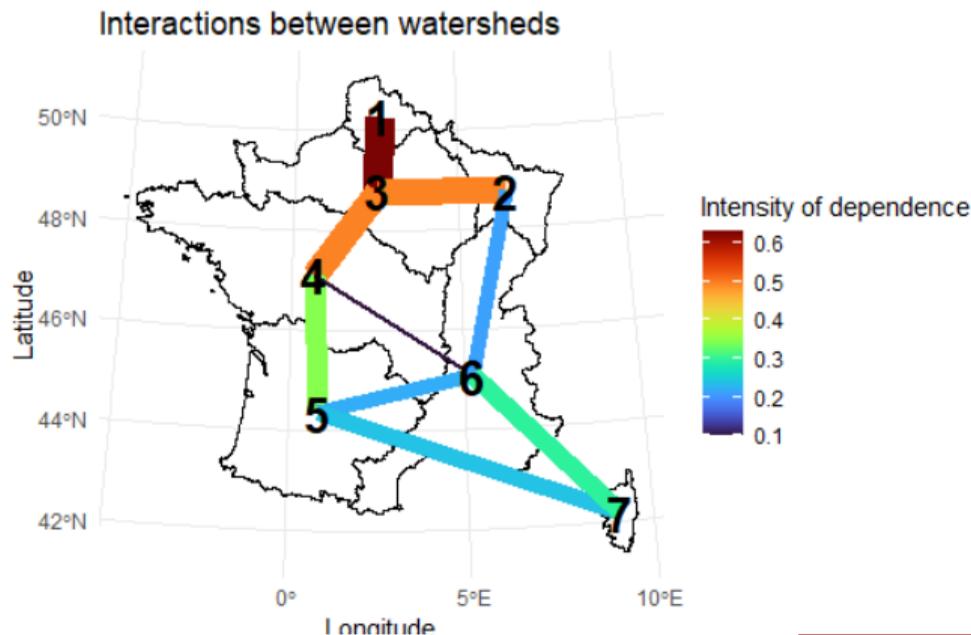


Approach 2 : Learn graph structure



Et le gagnant de la première étape est ...

L'approche 2 : Graphe donné par le modèle maximise la log-vraisemblance sur les données test

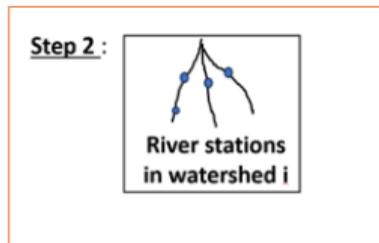


Seconde étape

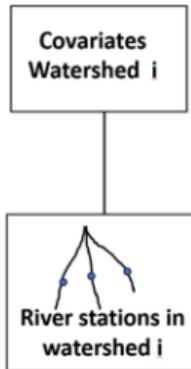
Step 1:



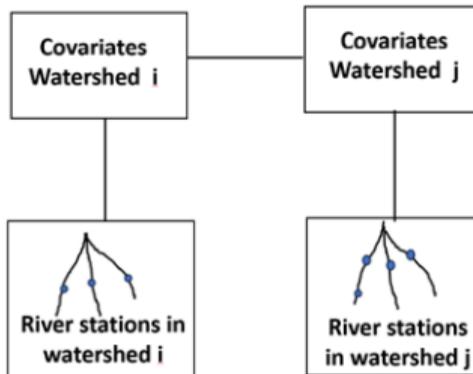
Step 2:



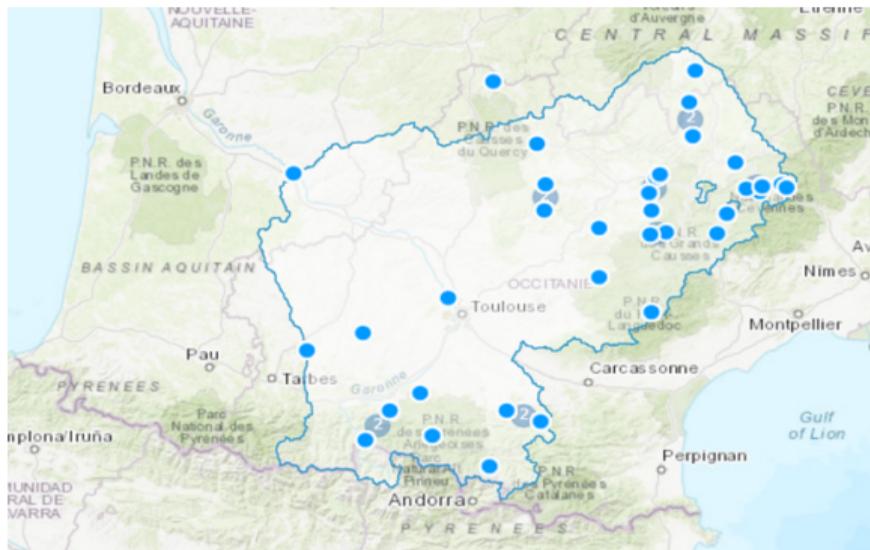
Step 3:



Step 4:

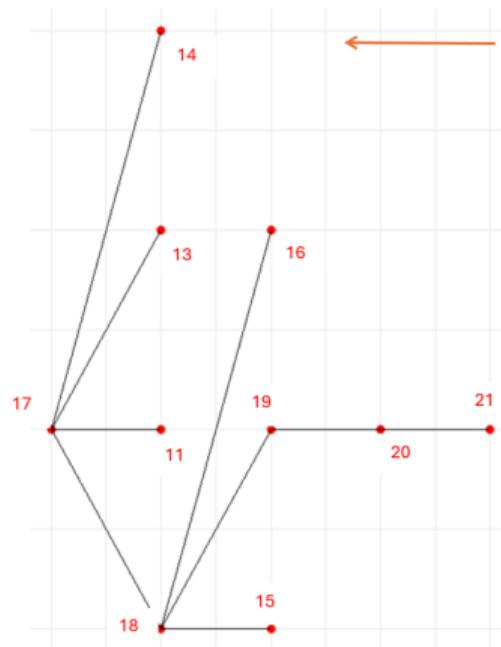
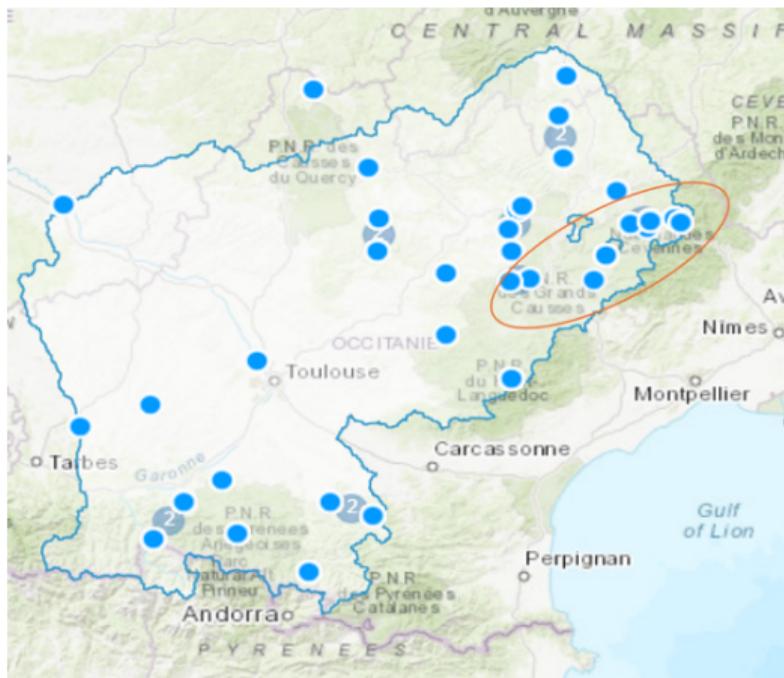


Bassin versant de la Garonne



Les points bleus représentent les stations de mesure des débits de rivière dans la Garonne.

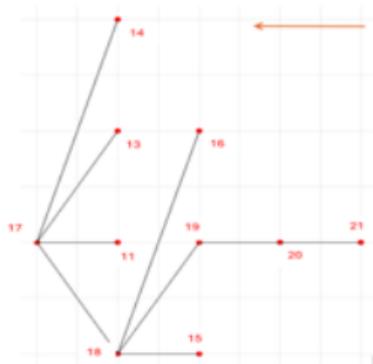
Dix stations sélectionnées dans le bassin versant de la Garonne



Graphe qui représente le flux d'eau

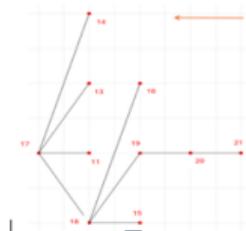
Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les débits de rivières

Approche 1 : Graphe imposé



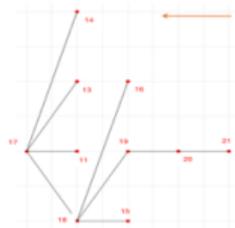
Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les débits de rivières

Approche 1 : Graphe imposé



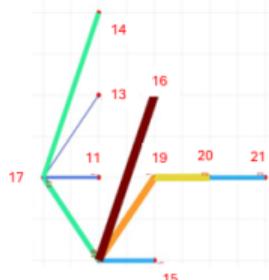
Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les débits de rivières

Approche 1 : Graphe imposé

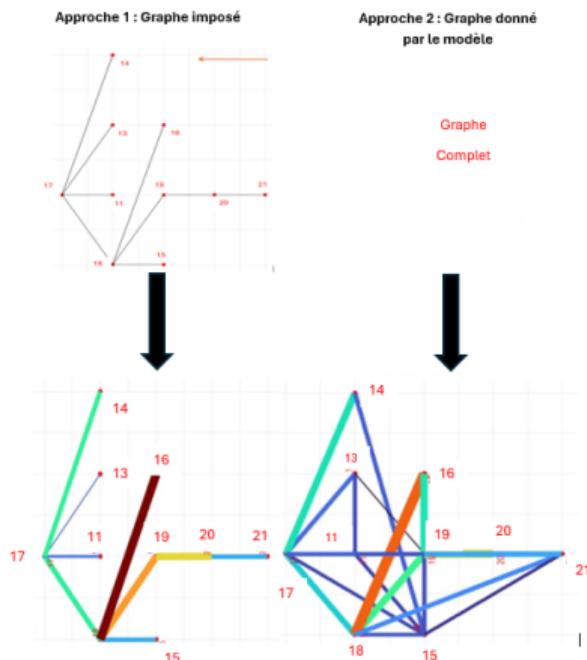


Approche 2 : Graphe donné par le modèle

Graphe
Complet

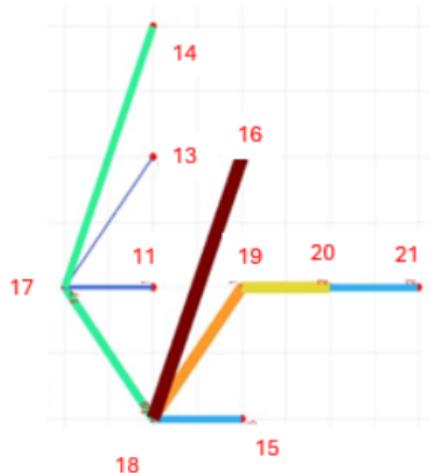


Application du modèle graphique de Hüsler-Reiss $HR(\Gamma)$ sur les débits de rivières

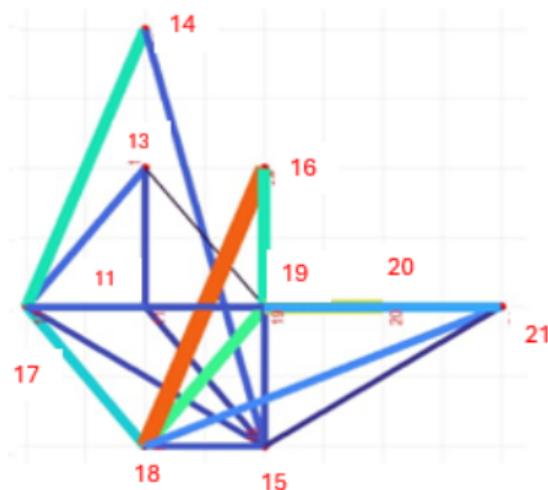


Resultats : Comparaison des graphes des 2 approches

Approche 1 : Graphe imposé

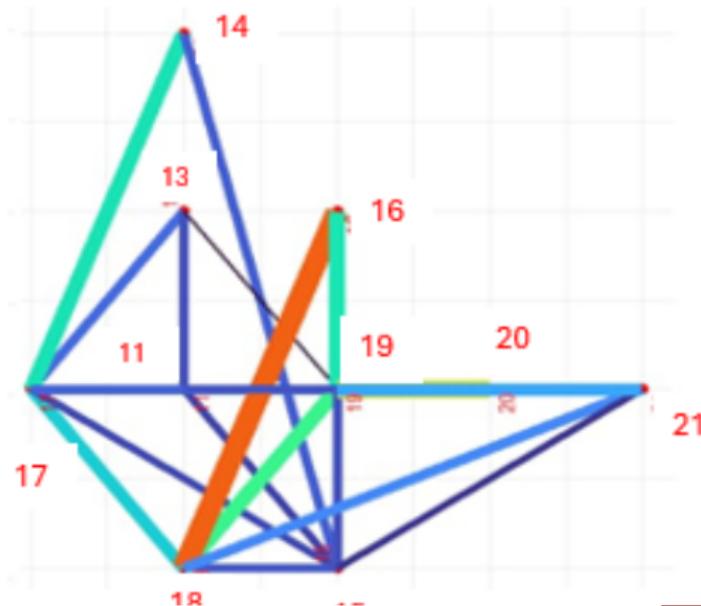


Approche 2 : Graphe donné par le modèle

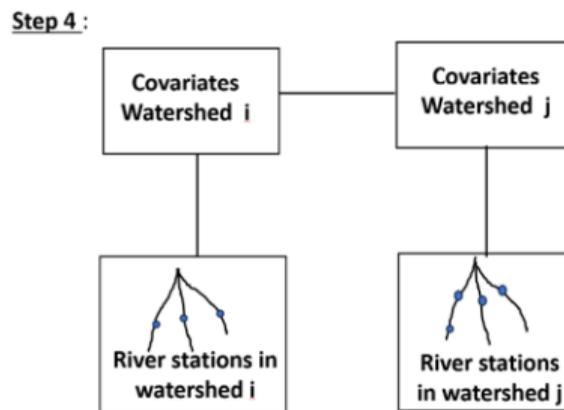
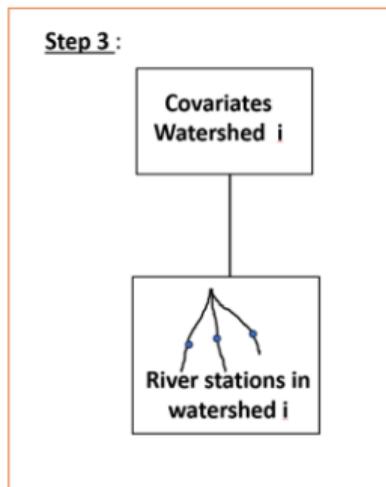
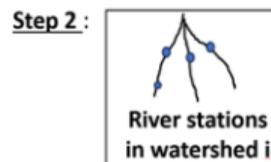
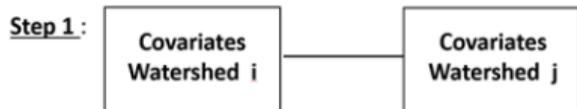


Et le gagnant de la deuxième étape est ...

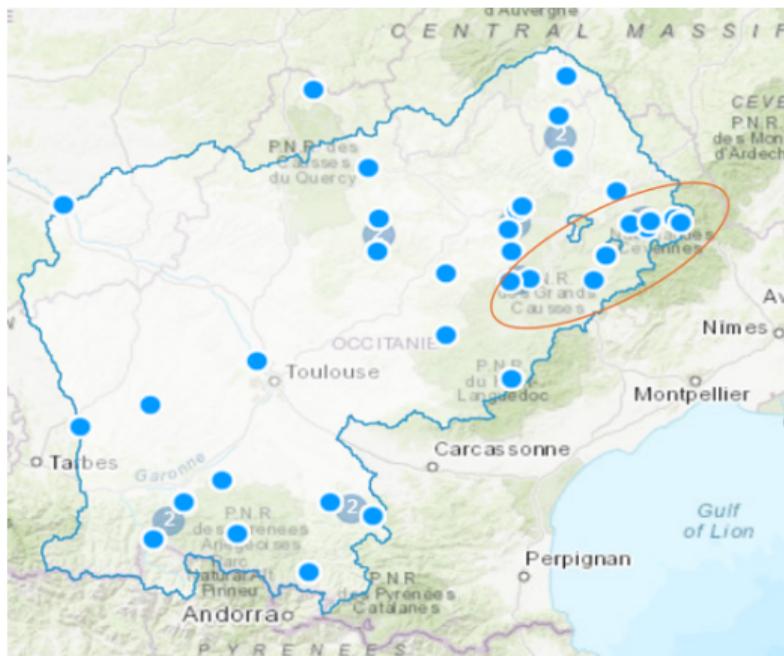
Approche 2 : Graphe donné par le modèle maximise la log-vraisemblance sur les données test



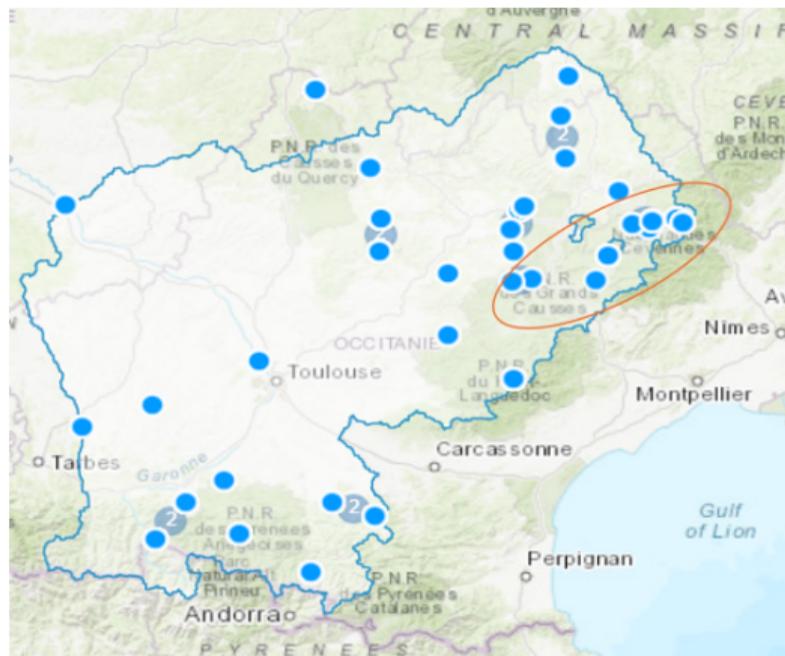
Troisième étape



Etude du lien entre les précipitations cumulées sur 7 jours et les débits de rivière extrêmes



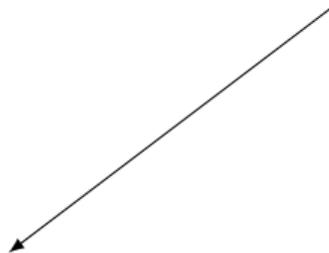
Etude du lien entre les précipitations cumulées sur 7 jours et les débits de rivière extrêmes



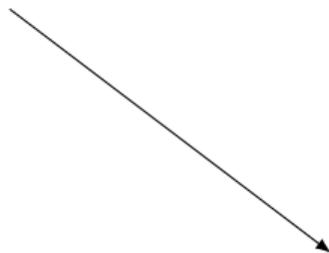
Est-ce que la quantité de pluie qui tombe sur la Garonne les jours où au moins une des 10 stations sélectionnées dépasse un seuil élevé de débit **est élevée ou non ?**

Etude du lien entre les précipitations cumulées sur 7 jours et les débits de rivière extrêmes

La quantité de pluie est-elle élevée ?



Par rapport à ce qui se passe
globalement dans la région la Garonne



Par rapport à ce qui se passe
localement en chaque pixel de la Garonne

Construction des composites maps

La Garonne divisée en pixels



Données de précipitations cumulées sur 7 jours par pixel



Transformation en Exponentielle de Laplace



Approche Globale :

Ajustement unique basé sur toutes les données des pixels



Approche Locale :

Ajustement de la distribution séparément pour chaque pixel

Construction des composites maps

- Sélection des jours où au moins une des 10 stations de débit sélectionnées dépasse un seuil élevé.

Construction des composites maps

- Sélection des jours où au moins une des 10 stations de débit sélectionnées dépasse un seuil élevé.
- Extraction des précipitations cumulées transformées (Exponentielle-Laplace) pour ces jours pour chaque pixel de la Garonne.

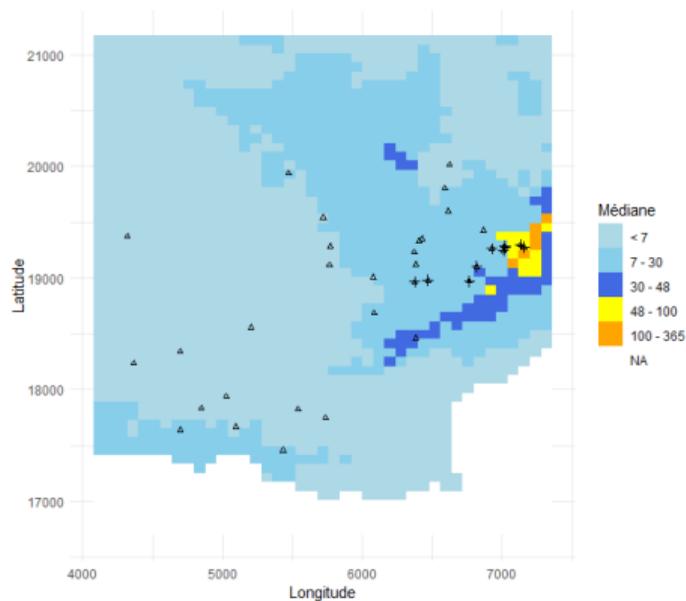
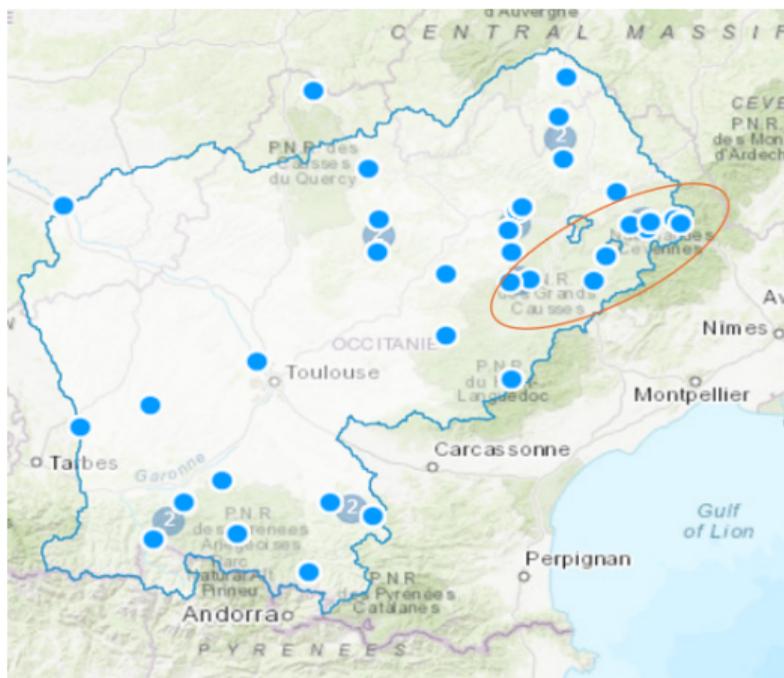
Construction des composites maps

- Sélection des jours où au moins une des 10 stations de débit sélectionnées dépasse un seuil élevé.
- Extraction des précipitations cumulées transformées (Exponentielle-Laplace) pour ces jours pour chaque pixel de la Garonne.
- Calcul de la médiane et la moyenne sur ces jours.

Construction des composites maps

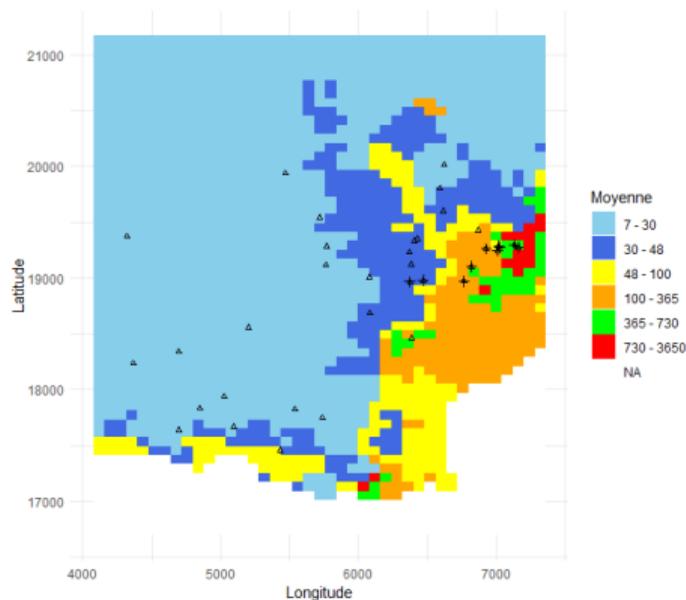
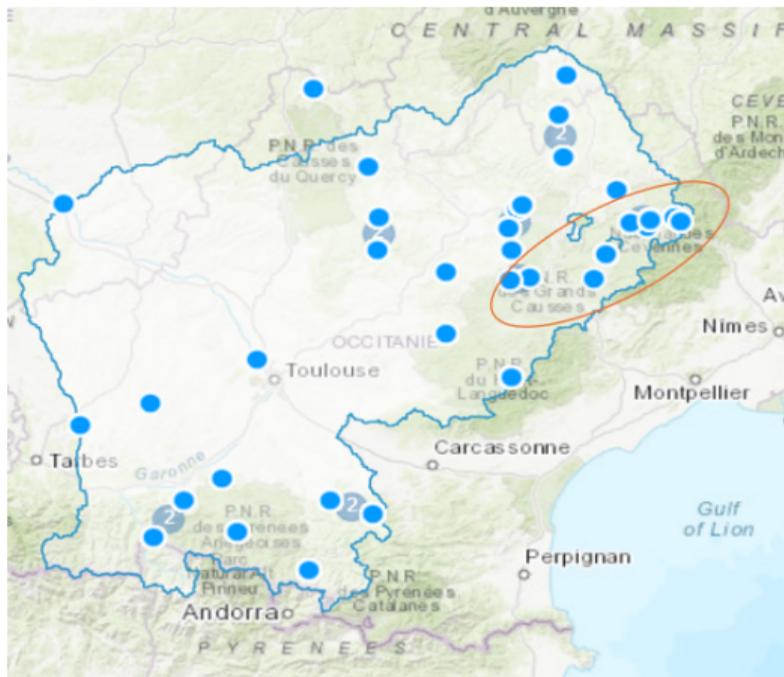
- Sélection des jours où au moins une des 10 stations de débit sélectionnées dépasse un seuil élevé.
- Extraction des précipitations cumulées transformées (Exponentielle-Laplace) pour ces jours pour chaque pixel de la Garonne.
- Calcul de la médiane et la moyenne sur ces jours.
- Visualisation sous forme de cartes (composite maps).

Comparaison globale : Médiane



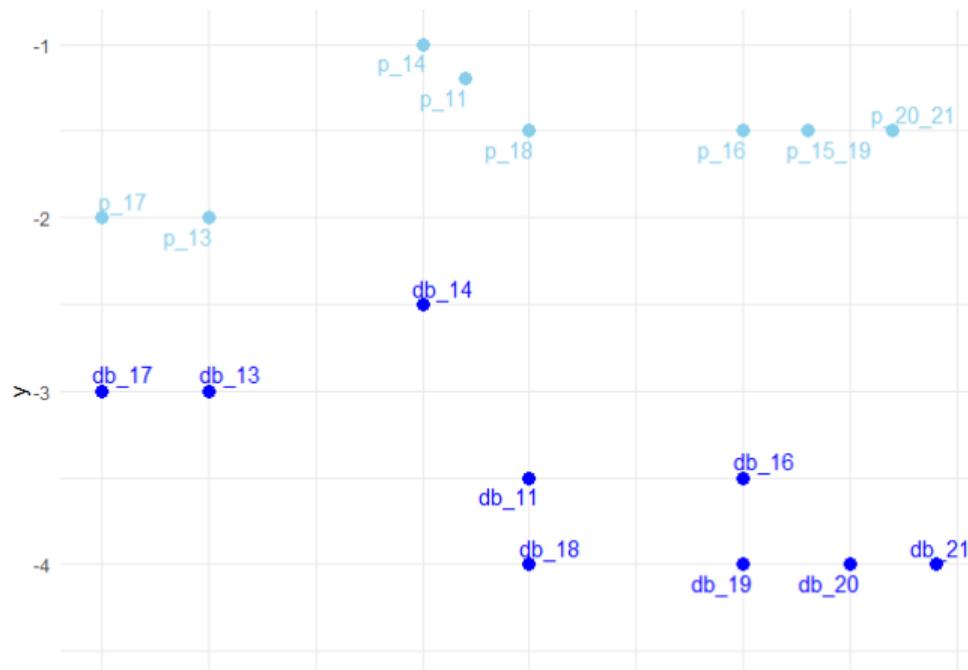
L'échelle de couleur représente le temps de retour de précipitations

Comparaison globale : Moyenne

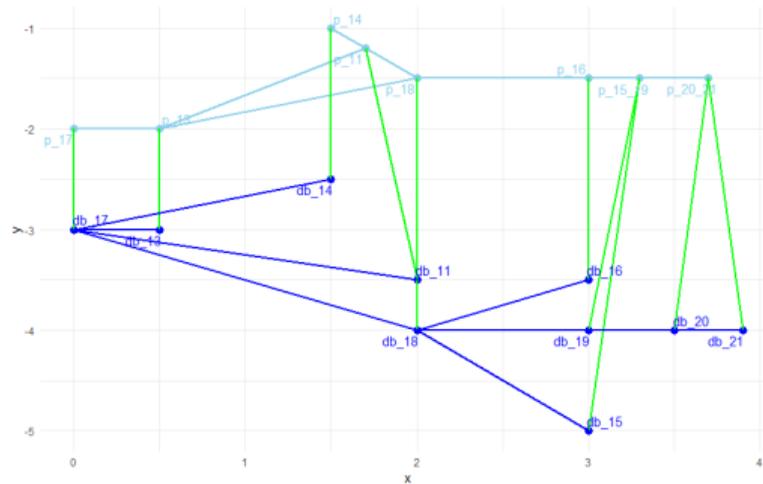


L'échelle de couleur représente le temps de retour de précipitations

Application du modèle graphique Hüsler-Reiss $HR(\Gamma)$ sur les précipitations et les débits de rivières

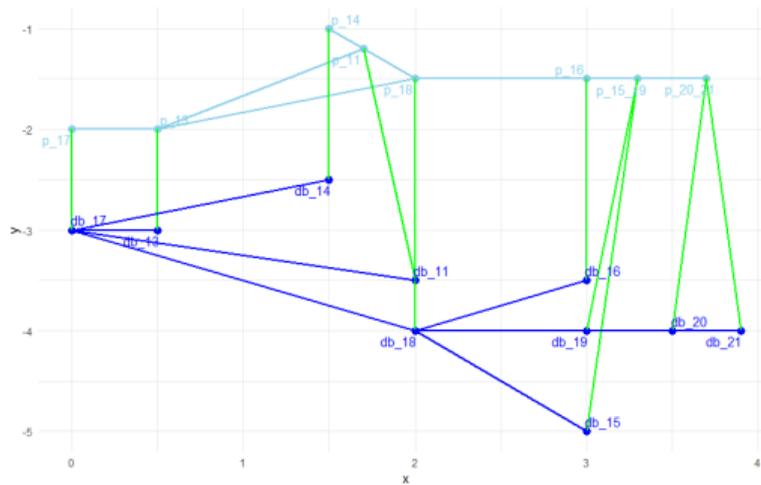


Application du modèle graphique Hüsler-Reiss $HR(\Gamma)$ sur les précipitations et les débits de rivières

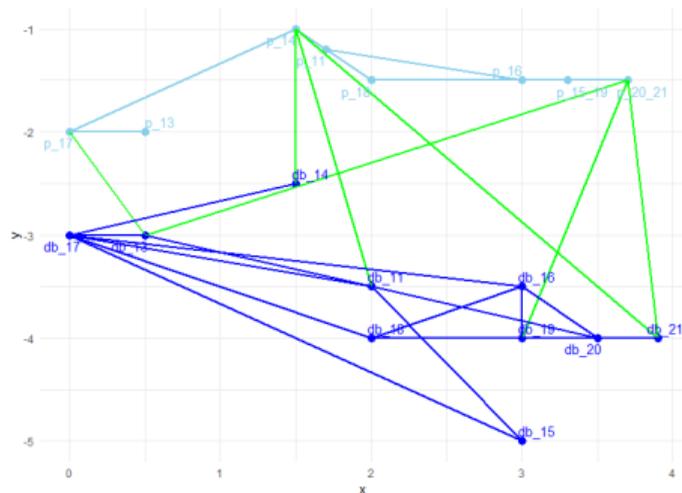


Modèle avec l'approche qui utilise un graphe imposé

Application du modèle graphique Hüsler-Reiss $HR(\Gamma)$ sur les précipitations et les débits de rivières



Modèle avec l'approche qui utilise un graphe imposé



Modèle avec l'approche qui apprend des données et renvoie un graphe

Perspectives

- Prendre des précipitations cumulées de 4 ou 5 jours au lieu de 7 jours
- Etudier le cas de plus de 10 stations sélectionnées
- Etudier d'autres covariables

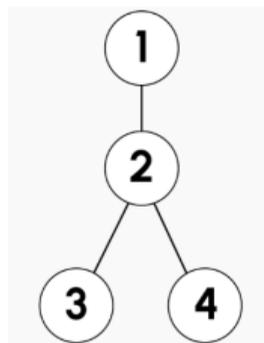
Merci pour votre attention!

“L’incertitude est la seule certitude qui existe, et savoir vivre dans l’insécurité est la seule sécurité”

Bibliographie

- 1974 Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. The Annals of probability, 2(5) :792–804.
- 1975 Pickands III, J. (1975). Statistical inference using extreme order statistics. the Annals of Statistics, pages 119–131.
- 2006 Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized pareto distributions. Bernoulli, 12(5) :917–930.
- 2006 Haan, L. and Ferreira, A. (2006). Extreme value theory : an introduction, volume 3. Springer.
- 2015 Asadi, P., Davison, A. and Engelke, S. (2015). Extremes on river networks
- 2020 Gardes, L. (2020). Théorie des valeurs extrêmes. Université de Strasbourg.
- 2022 de Fondeville, R. and Davison, A. C. (2022). Functional peaks-over-threshold analysis. Journal of the Royal Statistical Society Series B : Statistical Methodology, 84(4) :1392–1422.
- 2024 Engelke, S., Hentschel, M., Lalancette, M., and Röttger, F. (2024). Graphical models for multivariate extremes. arXiv preprint arXiv :2402.02187.

Notion d'indépendance des extrêmes



$$Y_i \perp_e Y_j \mid \mathbf{Y}_{V \setminus \{i,j\}} \iff \text{pas d'arête entre } Y_i \text{ et } Y_j \iff \Theta[i,j] = 0$$

But

- Le vecteur des extrêmes $\mathbf{Y} = \left\{ \frac{\mathbf{X}^*}{u} \mid \max(\mathbf{X}^*) > u \right\}$ suit une MGPD (Distribution multivariée de Pareto généralisée).
- Modéliser \mathbf{Y} par un modèle graphique des extrêmes $HR(\Gamma)$ de paramètre Γ .
- Le modèle permettra de simuler de nouveaux événements extrêmes.

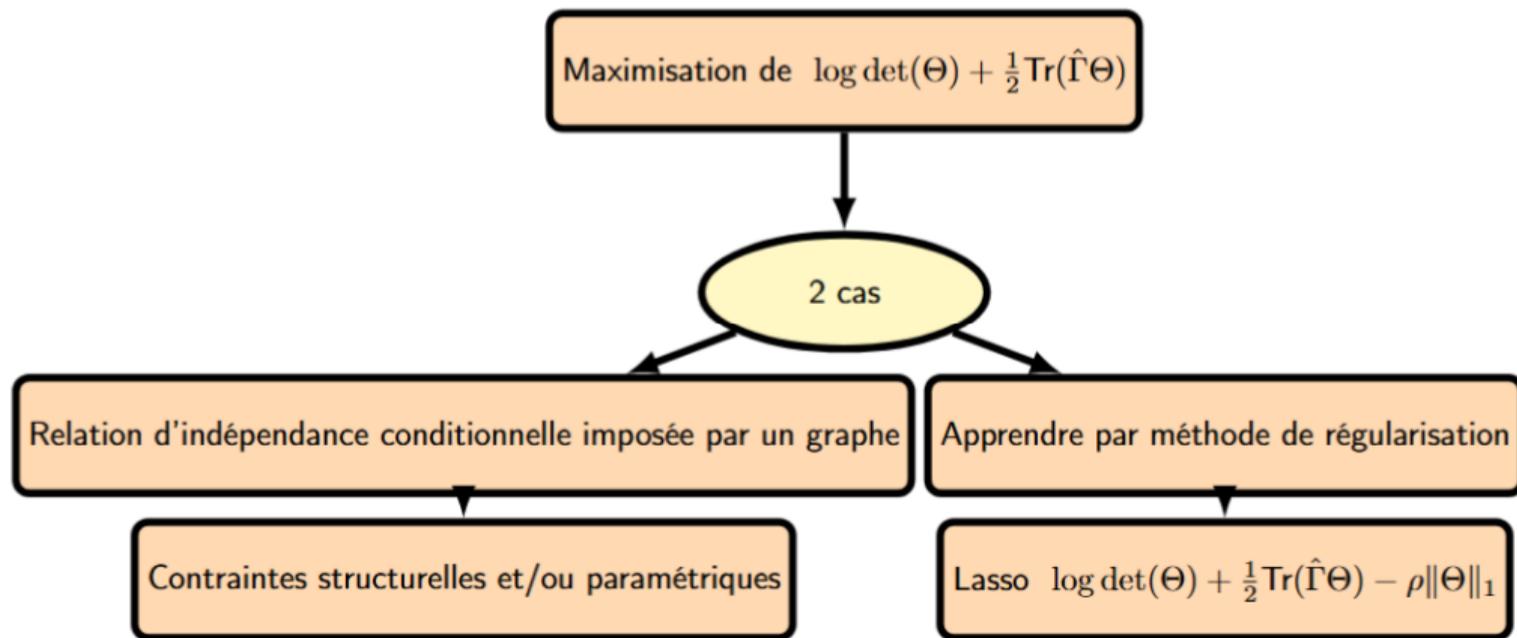
Comment?

Estimer la matrice Γ (paramètre du modèle Hüsler-Reiss $HR(\Gamma)$).

Calculer $\hat{\Gamma}$, l'estimateur empirique de Γ

Trouver Θ qui maximise la log vraisemblance sur les données $\log \det(\Theta) + \frac{1}{2} \text{Tr}(\hat{\Gamma}\Theta)$

Estimer Γ à l'aide de Θ



Critères de sélection de modèles pour la méthode de Lasso

- Maximiser la Log vraisemblance sur données test

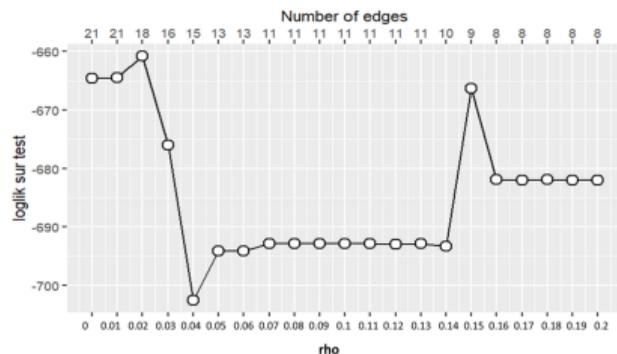


Figure: Log-vraisemblance sur les données de test en fonction des valeurs de ρ pour le seuil $u = 49$. Le nombre d'arêtes du graphe associé est précisé pour chaque valeur de ρ .

- Minimiser le BIC (critère d'information bayésien)

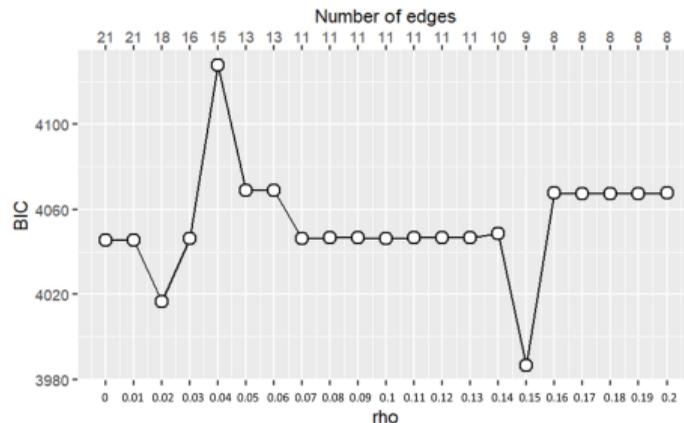


Figure: BIC en fonction des valeurs de ρ pour le seuil $u = 49$. Le nombre d'arêtes du graphe associé est indiqué pour chaque valeur de ρ .