

Modélisation des débits extrêmes dans l'Est du Québec à partir de données multi-modèles

Nicolas Lafon, Christian Genest, Johanna Neslehova, Jonathan Jalbert

Statistiques au sommet de Rochebrune 2026



Pourquoi modéliser les extrêmes de débit ?

- Dommages humains et matériels importants
- Evaluer les risques d'inondation à différentes échelles
- Dimensionnement des ouvrages hydrauliques et planification urbaine



Crue de 2019 sur la rivière Chaudière

Objectif

Décrire les phénomènes rares de débit et estimer leur probabilité afin d'aider à la gestion durable du risque hydrologique.

Modélisation 1D et niveau de retour

Définition

Le **niveau de retour** pour une période T est la valeur du débit $R(T)$ qui a une probabilité $1/T$ d'être dépassée au cours d'une année donnée :

$$\mathbb{P}(X > R(T)) = \frac{1}{T}.$$

Estimation locale des niveaux de retour

Idee de base : théorie des valeurs extrêmes 1D

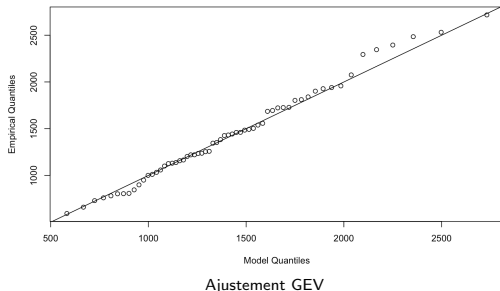
Soit $M^{(k)}$ le **maximum annuel** de débit en une localité k :

$$M^{(k)} = \max(X_1, X_2, \dots, X_{365}),$$

Alors $M^{(k)}$ suit approximativement une loi GEV:

$$G(x) = \exp\left(-[1 + \xi((x - \mu)/\sigma)]_+^{-1/\xi}\right)$$

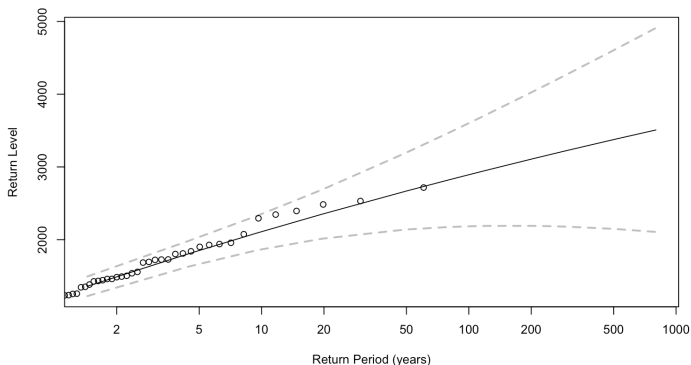
où μ : position, $\sigma > 0$: échelle, ξ : forme.



Estimation locale des niveaux de retour

- **Niveau de retour** = quantile de cette loi :

$$R(T) = \mu + \frac{\sigma}{\xi} \left[(-\log(1 - 1/T))^{-\xi} - 1 \right].$$



Estimation du niveau de retour et IC

Modélisation multidimensionnelle : approche copule

Limites de l'approche locale

L'analyse locale des maxima annuels ne tient pas compte de la **dépendance spatiale entre les extrêmes**.

- Crues **simultanées ou corrélées** en plusieurs points du bassin.
- Ignorer la structure spatiale → **sous-estimation du risque global**.

Vers le cadre multidimensionnel

Soit le vecteur de maxima annuels :

$$\mathbf{M} = (M^{(1)}, M^{(2)}, \dots, M^{(d)}),$$

Approximativement:

- Lois marginales **GEV**
- Dépendance décrite par **une copule de valeurs extrêmes**

Copule de valeurs extrêmes

Définition générale

Une **copule** est une fonction de répartition multidimensionnelle $C : [0, 1]^d \rightarrow [0, 1]$ dont les lois marginales sont uniformes sur $[0, 1]$:

$$C(u_1, \dots, u_d) = \Pr(U_1 \leq u_1, \dots, U_d \leq u_d), \quad U_i \sim \mathcal{U}(0, 1).$$

Toute loi jointe F peut s'écrire comme :

$$F(x_1, \dots, x_d) = C_F(F_1(x_1), \dots, F_d(x_d)),$$

ce qui permet de **séparer les lois marginales** F_i de la **structure de dépendance** C_F .

Exemple : copule de Gumbel 2d

$$C_\theta(u, v) = \exp\left\{-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{1/\theta}\right\}.$$

$\theta = 1$: indépendance ; $\theta \rightarrow \infty$: dépendance parfaite.

Copules de valeurs extrêmes

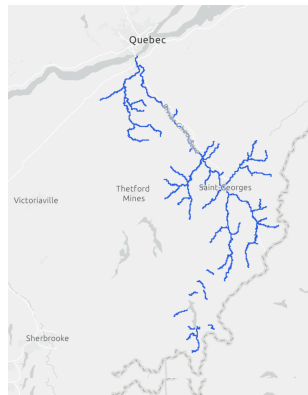
Une copule C est dite de **valeurs extrêmes** si elle est **max-stable** :

$$C(u_1, \dots, u_d) = [C(u_1^{1/n}, \dots, u_d^{1/n})]^n, \quad \forall n > 0.$$

Réseau hydrologique de la rivière Chaudière

Approche conceptuelle (idéale)

- 1 **Étape 1** : estimer les paramètres (μ_i, σ_i, ξ_i) des lois GEV pour chaque segment $i = 1, \dots, 194$
- 2 **Étape 2** : modéliser la **copule de valeurs extrêmes** en dimension 194

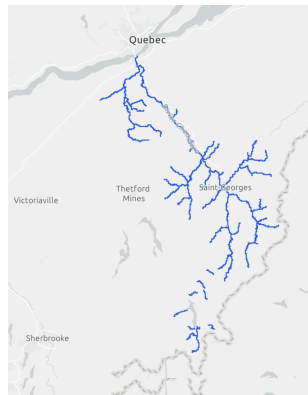


Réseau hydrologique de la rivière
Chaudière : 194 segments

Réseau hydrologique de la rivière Chaudière

Approche conceptuelle (idéale)

- 1 **Étape 1** : estimer les paramètres (μ_i, σ_i, ξ_i) des lois GEV pour chaque segment $i = 1, \dots, 194$;
- 2 **Étape 2** : modéliser la **copule de valeurs extrêmes** en dimension 194



Réseau hydrologique de la rivière
Chaudière : 194 segments

Données multi-modèles

Stations hydrométriques observées

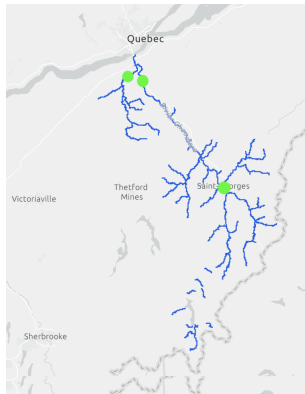
- **Trois stations de mesure** en activité continue depuis 60 ans.
- **Bruit d'observation important**

Complément par modélisation numérique

- Combinaison des observations disponibles et de simulations hydrologiques.
- **Six modèles hydrologiques distincts** pour estimer le débit sur le réseau entier.
- **Estimations très différentes d'un modèle à l'autre**

Enjeu majeur

Intégrer ces différentes estimations hydrologiques dans une modélisation cohérente des extrêmes.



Stations de mesure de débit actives depuis 60 ans

Approche développée : Estimation puis agrégation

Étape 1 – Estimation

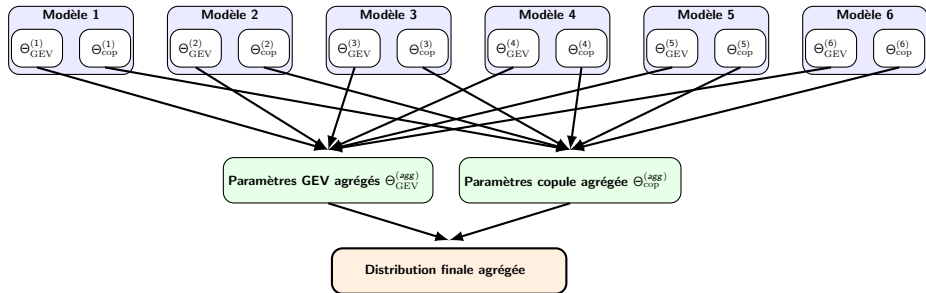
Estimation d'une distribution multidimensionnelle pour chaque **modèle hydrologique** :

- Estimation de **lois de valeurs extrêmes (GEV)** sur chaque segment ;
- Estimation d'une **copule de valeurs extrêmes** décrivant la dépendance spatiale.

Étape 2 – Agrégation

Combinaison des 6 distributions pour produire une **distribution synthétique des extrêmes** sur l'ensemble du bassin.

Approche développée : estimation puis agrégation

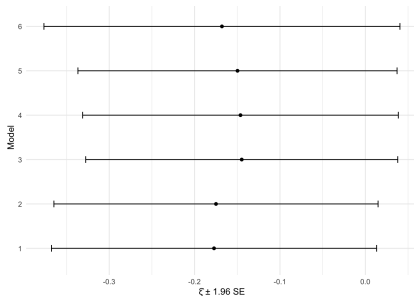


Estimation — Paramètres des lois marginales

- **Choix structurel** : paramètre de forme commun aux 6 modèles pour chaque segment.
- **Loi marginale** pour le segment i sous le modèle m :

$$G_{i,m}(x) = \text{GEV}(x \mid \mu_{i,m}, \sigma_{i,m}, \xi_i).$$

- **Estimation des paramètres par maximum de vraisemblance**, calcul des matrices de variance-covariance.



Estimation $\xi_{i,m}$ et IC associés.

Estimation — Copule archimédienne hiérarchique

Dépendance spatiale représentée par une **copule archimédienne hiérarchique (HAC)**.

Exemple illustratif (HAC Gumbel à 7 dimensions)

$$C(u_1, \dots, u_7) = C_{\theta_{11}} \left\{ C_{\theta_{10}} \left[C_{\theta_9}(u_1, u_2), C_{\theta_8}(u_3, u_4, u_5, u_6) \right], u_7 \right\}.$$



HAC définie par un unique triplet $(\mathcal{S}, \mathcal{A}, \Theta)$:

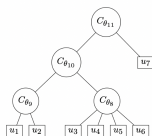
- $\mathcal{S} = \mathcal{F} \cup \mathcal{N} = \{1, 2, 3, 4, 5, 6, 7\} \cup \{8, 9, 10, 11\}$
- $\mathcal{A} = \{(1; 8), (2; 8), \dots, (10; 11)\}$
- $\Theta = \{\theta_8, \theta_9, \theta_{10}, \theta_{11}\}$

Estimation — Copule archimédienne hiérarchique

Dépendance spatiale représentée par une **copule archimédienne hiérarchique (HAC)**.

Exemple illustratif (HAC Gumbel à 7 dimensions)

$$C(u_1, \dots, u_7) = C_{\theta_{11}} \left\{ C_{\theta_{10}} \left[C_{\theta_9}(u_1, u_2), C_{\theta_8}(u_3, u_4, u_5, u_6) \right], u_7 \right\}.$$



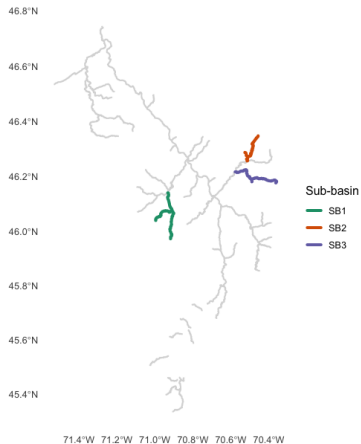
Propriétés fondamentales

- (1) Une HAC est une **copule de valeurs extrêmes** \iff toutes les copules imbriquées doivent être de Gumbel.
- (2) La dépendance entre deux variables est décrite par la **copule du plus jeune ancêtre commun** dans l'arbre

Estimation — Pertinence d'une HAC sur un réseau hydrographique

Topologie du réseau \Rightarrow structure en arbre

- **Entre sous-bassins versants :** la dépendance d'un segment à l'autre est **globalement homogène.**
- Homogénéité \longrightarrow **structure d'arbre**



Estimation — Copule archimédienne hiérarchique

Hypothèse d'arbre commun

Même **arbre** ($\mathcal{S} = \mathcal{F} \cup \mathcal{N}$, \mathcal{A}) pour les HACs issues de chaque modèle.

Seuls les paramètres des copules $\{\theta_{k,m}\}_{k \in \mathcal{N}}$ varient d'un modèle à l'autre.

Estimation des paramètres de copule

Pour chaque modèle $m = 1, \dots, 6$, estimation par maximum de vraisemblance des paramètres et des variances associées.

Six copules HAC distinctes :

$$(\mathcal{S}, \mathcal{A}, \Theta_m = \{\theta_{k,m}\}_{k \in \mathcal{N}}), \quad m \in \{1, \dots, 6\}$$

Agrégation des marges par Vincentisation

Poids d'agrégation : $w_{i,m} \geq 0$, $\sum_{m=1}^6 w_{i,m} = 1$.

Vincentisation des lois marginales (segment i) :

$$R_{i,agg}(T) = \sum_{m=1}^6 w_{i,m} R_{i,m}(T)$$

Equivalent à

$$G_{i,agg}(x) = \text{GEV} \left(x \mid \sum_{m=1}^6 w_{i,m} \mu_{i,m}, \sum_{m=1}^6 w_{i,m} \sigma_{i,m}, \xi_i \right)$$

⇒ Chaque segment i possède une **loi GEV "moyenne"**

Agrégation de la structure de dépendance

Poids d'agrégation : $\lambda_{k,m} \geq 0$, $\sum_{m=1}^6 \lambda_{k,m} = 1$.

Agrégation des copules paramètre par paramètre (arbre fixé de noeuds \mathcal{N}):

$$\theta_{k,agg} = \sum_{m=1}^6 \lambda_{k,m} \theta_{k,m}, \quad \forall k \in \mathcal{N},$$

Copule HAC agrégée C_{agg} :

$$\text{HAC} : (\mathcal{S}, \mathcal{A}, \Theta_{agg} = \{\theta_{k,agg}\}_{k \in \mathcal{N}})$$

Distribution finale agrégée :

$$\mathbf{F}_{agg}(\mathbf{x}) = C_{agg}(G_{1,agg}(x_1), \dots, G_{194,agg}(x_{194})),$$

Choix des poids d'agrégation

(1) Poids pour les lois marginales

Pour chaque segment i ,

$$\operatorname{argmin}_{w_{i,m} \geq 0, \sum w_{i,m} = 1} \operatorname{Var} [R_{i,agg}(T)]$$

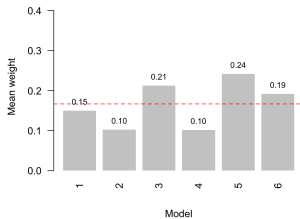
pour une période de retour T fixée (ex. 100 ans).

(2) Poids pour la copule HAC

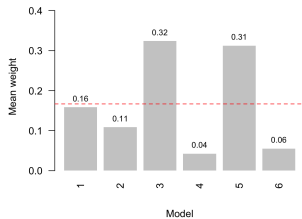
Pour chaque noeud k ,

$$\operatorname{argmin}_{\lambda_{m,k} \geq 0, \sum \lambda_{m,k} = 1} \operatorname{Var}(\theta_{k,agg})$$

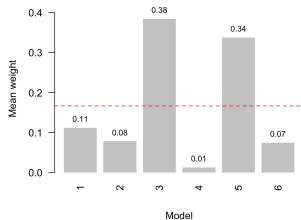
Résultats : poids d'agrégation pour différents T



$T = 10$

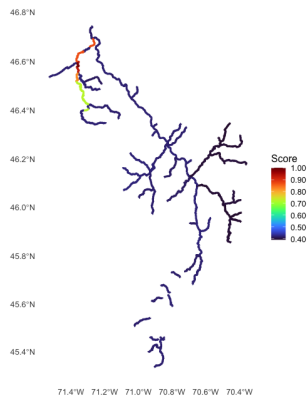
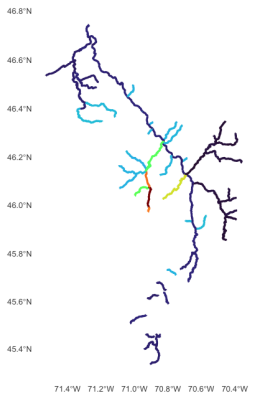
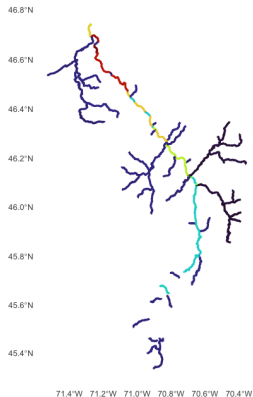


$T = 100$



$T = 1000$

Résultats : Carte d'excès conditionnels à $T = 100$



Conclusion

- Introduction du **cadre multi-modèles** pour la modélisation des extrêmes en hydrologie.
- Utilisation d'une **copule archimédienne hiérarchique** pour représenter la structure de dépendance :
 - approche nouvelle dans le contexte des extrêmes hydrologiques ;
 - modèle **parcimonieux**, et **efficace** en haute dimension.
- Agrégation calibrée pour les extrêmes
- Modèle adapté au calculs rapide d'indicateurs hydrologiques :
 - estimation multidimensionnelle de niveaux de retour,
 - cartographie d'étendues d'inondation.

- Dépendance symétrique
- Sur-estimation de la dépendance
- Manque de flexibilité