

Projet ANR-08-RISK-03-01

Prédétermination des valeurs extrêmes de pluies et de crues (EXTRAFLO)

Programme RISKNAT 2008

Tâche II : Mise au point d'une stratégie commune d'inter-comparaison et de validation

Rapport II.2 « *Détails techniques des implémentations
comparées dans les actions 2, 4, et 6* »

Date : Septembre 2012

Rapport réalisé par :

⁽¹⁾ Irstea, Centre de Lyon HHLY

⁽²⁾ Irstea, Centre d'Aix-en-Provence, OHAX

Auteurs :

B. Renard¹, P. Arnaud², K. Kochanek¹, M. Lang¹



Sommaire

1	<i>Approches statistiques locales</i>	3
1.1	Généralités et notations	3
1.2	Estimation par la méthode des moments	3
1.3	Estimation par la méthode des L-moments	3
1.4	Estimation par maximum de vraisemblance	4
1.5	Estimation Bayesienne	4
1.6	Estimation des incertitudes	5
1.6.1	Bootstrap (moments et L-moments)	5
1.6.2	Approximation asymptotique (maximum de vraisemblance)	5
1.6.3	Approche Bayesienne	5
2	<i>Approches statistiques régionales</i>	6
2.1	Généralités et notations	6
2.2	Indice de crue et généralisations	6
2.3	Estimations locales	7
2.4	Régressions	7
2.5	Krigeage des résidus	8
2.6	Estimation de la loi régionale	9
2.7	Distribution estimée au site cible	9
3	<i>Approches statistiques mixtes locales-régionales</i>	9
3.1	Généralités et notations	9
3.2	Combinaison des informations locales et régionales	10
4	<i>La méthode SHYREG-Débit</i>	10
4.1	Calage de la méthode SHYREG-débit	11
4.1.1	Calage de la méthode SHYREG en local : Action 2 et Action 6	12
4.1.2	Calage de la méthode SHYREG en régional : Action 4	14
5	<i>Références</i>	17
6	<i>Annexes</i>	19
6.1	Annexe 1 : propriétés des distributions utilisées dans les actions	19
6.1.1	Loi Normale	19
6.1.2	Loi de Pearson III	20
6.1.3	Loi de Gumbel	21
6.1.4	Loi GEV	22
6.2	Annexe 2 : la base SHYREG-Pluie	23

1 Approches statistiques locales

1.1 Généralités et notations

En un site donné, les observations disponibles sont notées $y=(y_1, \dots, y_n)$. Ces observations sont considérées comme des réalisations indépendantes et identiquement distribuées (*iid*) d'une variable aléatoire Y . La loi de cette variables aléatoire est paramétrée par un vecteur de paramètres $\theta=(\theta_1, \dots, \theta_k)$, qui sont inconnus et qu'il faut donc estimer. La densité de probabilité de cette loi, calculée en une valeur z , est notée $p(z | \theta)$. Les distributions ayant été utilisées dans ce projet sont les suivantes : loi de Gumbel, loi généralisée des valeurs extrêmes (GEV), loi de Pearson III (PIII), loi de Log-Pearson III (LPIII) et loi Log-normale.

1.2 Estimation par la méthode des moments

Le principe de la méthode des moments est d'estimer les paramètres de sorte que les moments théoriques de la distribution soient égaux aux moments empiriques calculés sur les données. Pour une distribution à k paramètres, on devra égaliser k moments théoriques et empiriques. Les moments théoriques et empiriques d'ordre k sont définis de la façon suivante :

$$\begin{aligned} m_k^{(th)} &= E[Y^k] \\ m_k^{(emp)} &= \frac{1}{n} \sum_{i=1}^n y_i^k \end{aligned} \quad (1)$$

L'application de la méthode des moments réclame donc de résoudre un système de k équations à k inconnues. La résolution de ces équations peut parfois être explicite, mais réclame parfois de recourir à des solveurs numériques (voir annexe 1 pour un formulaire des estimateurs des moments).

Précisons que pour les distributions Log-normale et Log-Pearson III, les paramètres ont été estimés en prenant le logarithme des données et en calculant l'estimateur des moments pour des lois Normale et Pearson III. Ceci n'est pas équivalent à calculer l'estimateur des moments directement sur les données brutes. Nous avons choisi la première option car elle nous semblait plus représentative de la pratique opérationnelle.

1.3 Estimation par la méthode des L-moments

Le principe est similaire à celui de l'estimation par la méthode des moments, mais on substitue aux moments les L-moments, qui sont définis de la façon suivante (on ne donne que les 3 premiers L-moments empiriques):

$$\begin{aligned} l_k^{(th)} &= k^{-1} \sum_{i=0}^{k-1} (-1)^k \binom{k-1}{i} E[Y_{(k-i:k)}] \\ l_1^{(emp)} &= \binom{n}{1}^{-1} \sum_{i=1}^n y_{(i)} \\ l_2^{(emp)} &= \frac{1}{2} \binom{n}{2}^{-1} \sum_{i=1}^n \left(\binom{i-1}{1} - \binom{n-i}{1} \right) y_{(i)} \\ l_3^{(emp)} &= \frac{1}{3} \binom{n}{3}^{-1} \sum_{i=1}^n \left(\binom{i-1}{2} - 2 \binom{i-1}{1} \binom{n-i}{1} + \binom{n-i}{2} \right) y_{(i)} \end{aligned} \quad (2)$$

Dans ces équations, la notation $y_{(i)}$ signifie « la i^{eme} donnée triée par ordre croissant ».

L'annexe 1 fournit un formulaire des estimateurs des L-moments pour les distributions utilisées dans ce projet. Comme précédemment, pour les distributions Log-normale et Log-Pearson III, l'estimation est basée sur le logarithme des données.

1.4 Estimation par maximum de vraisemblance

Dans le cadre d'observations *iid* comme supposé en section 1.1, la fonction de vraisemblance, que l'on notera $p(\mathbf{y} | \boldsymbol{\theta})$, est définie de la façon suivante :

$$p(\mathbf{y} | \boldsymbol{\theta}) = \prod_{i=1}^n p(y_i | \boldsymbol{\theta}) \quad (3)$$

Le calcul de la vraisemblance consiste donc simplement à effectuer le produit des densités de probabilité évaluées en chacune des observations. L'estimateur du maximum de vraisemblance est alors défini comme le vecteur de paramètres qui maximise la vraisemblance :

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{ArgMax}} \{ p(\mathbf{y} | \boldsymbol{\theta}) \} \quad (4)$$

Dans certains cas, il est possible de calculer explicitement cet estimateur (cf. Annexe 1). En l'absence de résolution explicite, on aura recours à une optimisation numérique. Comme précédemment, pour les distributions Log-normale et Log-Pearson III, l'estimation est basée sur le logarithme des données.

1.5 Estimation Bayésienne

L'estimation Bayésienne est également basée sur la fonction de vraisemblance, mais utilise en plus une autre source d'information, nommée la distribution a priori. Cette distribution, dont la densité est notée $p(\boldsymbol{\theta})$, contient toute connaissance sur les paramètres à estimer qui peut être mobilisée sans utiliser les données au site d'étude. Une telle connaissance peut provenir d'expertise ou d'information régionale (ce dernier cas sera détaillé ultérieurement en section 3). Dans le cas d'une estimation locale, on utilisera un a priori non-informatif pour tous les paramètres, à l'exception du paramètre de forme de la loi GEV, pour lequel on utilisera un a priori Gaussien centré en zéro et d'écart type 0.25. Cet a priori est assez vague mais reflète une expertise sur les valeurs typiques de ce paramètre pour les données hydrologiques (voir [Martins and Stedinger, 2000] pour une discussion sur ce type d'a priori).

La distribution a priori et la fonction de vraisemblance sont combinées d'après le théorème de Bayes pour calculer la distribution a posteriori des paramètres, dont la densité est notée $p(\boldsymbol{\theta} | \mathbf{y})$:

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (5)$$

L'équation (5) stipule donc que l'on obtient la densité a posteriori en multipliant simplement la densité a priori et la fonction de vraisemblance (cf. équation (3)). Précisons que le symbole ' \propto ' signifie '*est proportionnel à*' : la densité a posteriori n'est donc connue qu'à une constante multiplicative près. Ceci n'est pas problématique puisque l'estimateur utilisé correspond au vecteur de paramètres qui maximise la densité a posteriori (ce maximum étant inchangé lorsque la densité a posteriori est multipliée par une constante) :

$$\hat{\theta} = \underset{\theta}{\text{ArgMax}} \{p(\theta | y)\} \quad (6)$$

En général il n'est pas possible de calculer explicitement cet estimateur, on a donc recours à une méthode d'optimisation numérique.

1.6 Estimation des incertitudes

La méthode utilisée pour la quantification des incertitudes varie en fonction de l'estimateur.

1.6.1 Bootstrap (moments et L-moments)

La méthode du Bootstrap est utilisée pour quantifier l'incertitude dans le cas des estimations par moments ou L-moments. L'algorithme peut être décrit de la manière suivante :

Pour $i = 1 : N_{sim}$

1. **Rééchantillonnage** : Tirer au hasard et avec remise n valeurs dans les données disponibles $y=(y_1, \dots, y_n)$. Certaines valeurs apparaîtront donc plusieurs fois, d'autres seront absentes.
2. **Estimation** : calculer l'estimateur $\hat{\theta}^{(i)}$ sur ces données rééchantillonnées.

Fin

L'ensemble des valeurs $(\hat{\theta}^{(i)})_{i=1:N_{sim}}$ représente ainsi l'incertitude sur les paramètres estimés.

L'incertitude résultante sur les quantiles peut aisément être quantifiée en appliquant la fonction quantile de la distribution supposée à l'ensemble de ces paramètres.

1.6.2 Approximation asymptotique (maximum de vraisemblance)

Dans le cas de l'estimation par maximum de vraisemblance, il existe une théorie asymptotique bien développée qui fournit une approximation sur la distribution d'échantillonnage des estimateurs. En effet, on peut montrer (voir par exemple [Coles, 2001]) que lorsque n (la taille de l'échantillon) tend vers l'infini, la distribution d'échantillonnage de l'estimateur du maximum de vraisemblance $\hat{\theta}$ est Gaussienne, de moyenne θ_0 (la vraie valeur du paramètre : absence de biais) et de matrice de covariance $V(\hat{\theta})$, où $V(\theta)$ est la matrice d'information de Fisher :

$$V(\theta) = \begin{pmatrix} -\frac{\partial^2}{\partial \theta_1^2} \log p(y | \theta) & \dots & -\frac{\partial^2}{\partial \theta_1 \partial \theta_k} \log p(y | \theta) \\ \vdots & \ddots & \vdots \\ -\frac{\partial^2}{\partial \theta_k \partial \theta_1} \log p(y | \theta) & \dots & -\frac{\partial^2}{\partial \theta_k^2} \log p(y | \theta) \end{pmatrix} \quad (7)$$

Dans l'équation (7), les dérivées partielles sont généralement approchées numériquement avec un schéma de différences finies.

1.6.3 Approche Bayésienne

Dans l'approche Bayésienne, la distribution a posteriori (équation (5)) fournit directement une quantification de l'incertitude. Néanmoins, l'utilisation directe de cette distribution est

délicate pour deux raisons : (i) il s'agit d'une distribution multi-dimensionnelle ; (ii) elle n'est connue qu'à une constante près.

Pour contourner cette difficulté, on fait appel à des simulateurs de Monte Carlo par Chaînes de Markov (MCMC). Les algorithmes MCMC désignent une famille de méthodes qui permettent de simuler des réalisations à partir d'une densité de probabilité arbitraire, connue éventuellement seulement à une constante près : ceci correspond exactement au contexte de la distribution a posteriori. L'ensemble des valeurs de paramètres simulées par MCMC peut ainsi être utilisé pour dériver aisément des incertitudes sous la forme d'intervalles par exemple.

Une présentation détaillée des algorithmes MCMC dépasse largement le cadre de ce rapport : les détails techniques de l'algorithme utilisé dans ce projet sont donnés dans [Renard et al., 2006].

2 Approches statistiques régionales

2.1 Généralités et notations

En un site donné i , les observations disponibles sont à présent notées $\mathbf{y}^{(i)} = (y_1^{(i)}, \dots, y_{n_i}^{(i)})$. Ces observations sont considérées comme des réalisations indépendantes et identiquement distribuées (*iid*) d'une variable aléatoire $Y^{(i)}$. La loi de cette variable aléatoire est supposée identique sur tous les sites, mais avec des paramètres variant de site en site, notés $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \dots, \theta_k^{(i)})$. Deux distributions ont été comparées pour les approches régionales : la loi de Gumbel et la loi GEV. L'objectif des approches régionales est d'estimer les paramètres d'un site cible c , $\boldsymbol{\theta}^{(c)} = (\theta_1^{(c)}, \dots, \theta_k^{(c)})$, pour lequel aucune donnée locale n'est disponible.

2.2 Indice de crue et généralisations

La méthode de l'indice de crue [Dalrymple, 1960] constitue certainement l'approche régionale la plus utilisée. Le schéma général est le suivant :

1. **Estimation locale** : sur chaque site de calage i (pour lequel on dispose de données), on estime un « indice de crue », que l'on notera v_i . L'indice de crue utilisé ici est simplement la moyenne des données locales.
2. **Régionalisation** : une régression est mise en place pour lier les valeurs de l'indice de crue avec les caractéristiques des bassins versants.
3. **Estimation d'une loi régionale** : on émet l'hypothèse que les « données normalisées » $\mathbf{y}^{(i)} / v_i$ de l'ensemble des sites proviennent toutes d'une même distribution, dont on estime les paramètres $\boldsymbol{\theta}^{(R)} = (\theta_1^{(R)}, \dots, \theta_k^{(R)})$.
4. **Transfert au site cible non jaugé** : La distribution au site cible c peut alors être estimée en combinant l'indice de crue v_c (estimé grâce à la régression) et la distribution régionale.

La démarche de l'indice de crue peut être généralisée en recherchant des régressions directement sur les paramètres, plutôt que sur l'indice de crue. Cette approche permet de s'affranchir de l'étape d'estimation d'une loi régionale. Ce schéma régional peut être résumé de la façon suivante :

1. **Estimation locale** : sur chaque site de calage i (pour lequel on dispose de données), on estime les paramètres de la distribution $\theta^{(i)} = (\theta_1^{(i)}, \dots, \theta_k^{(i)})$.
2. **Régionalisation** : Pour chacun des k paramètres, une régression est mise en place pour lier les valeurs du paramètre estimé localement avec les caractéristiques des bassins versants.
3. **Transfert au site cible non jaugé** : La distribution au site cible c peut alors être estimée en appliquant les régressions pour prédire les paramètres du site cible.

Ces deux approches (indice de crue et régressions sur paramètres) seront implémentées et comparées dans les actions de comparaison.

2.3 Estimations locales

La première étape consiste à estimer l'indice de crue v_i ou les paramètres $\theta^{(i)}$ pour l'ensemble des sites de calibration, i.e. des sites sur lesquels des données locales sont disponibles. Une estimation Bayésienne est utilisée, comme décrit en section 1.5.

2.4 Régressions

Les régressions implémentées visent à relier l'estimation locale au site i , w_i , avec les caractéristiques des bassins versants (covariables), $x_i^{(1)}, \dots, x_i^{(N_{\text{cov}})}$. w_i peut représenter soit l'indice de crue, soit un des paramètres de la distribution. La formule de régression est la suivante:

$$\log(w_i) = \beta_0 + \sum_{j=1}^{N_{\text{cov}}} \beta_j x_i^{(j)} + \varepsilon_i, \varepsilon_i \sim N(0, \sqrt{\sigma_\varepsilon^2 + \sigma_i^2}) \quad (8)$$

Dans l'équation (8), l'écart-type σ_i^2 dénote l'incertitude liée à l'estimation locale de w_i : cette valeur est connue avant d'effectuer la régression et correspond à l'écart-type a posteriori de w_i à l'issue de l'étape d'estimation locale Bayésienne. Par contre, l'écart-type σ_ε^2 est inconnu et représente l'incertitude résiduelle liée à l'imperfection de la formule de régression.

Le choix des covariables se base sur un travail effectué précédemment pour régionaliser la crue décennale [Cipriani et al., 2012]. Il s'agit de : (i) superficie du bassin versant ; (ii) altitude moyenne du bassin ; (iii) moyenne des pluies décennales dans le bassin, à partir des quantiles estimés par [Benichou and Le Breton, 1987]; (iv) l'indice IDPR (Indice de Développement et de Persistance de Réseaux) qui caractérise la capacité d'infiltration [Mardhel et al., 2004]. Enfin, des régressions distinctes ont été implémentées pour chacune des hydro-écorégions définies d'après la classification de [Wasson et al., 2004]. La Figure 1 fournit une cartographie de ces régions et de ces covariables.

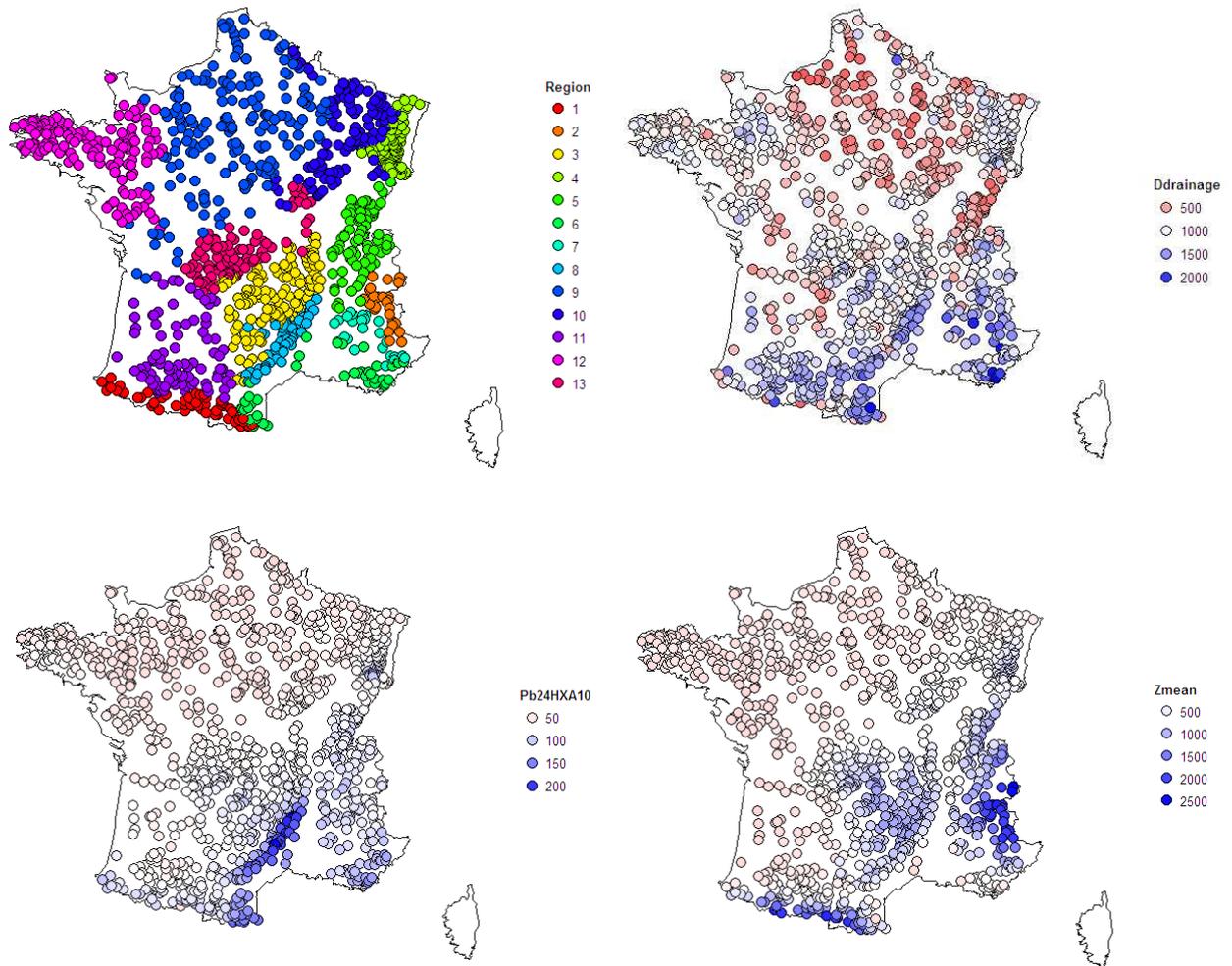


Figure 1. Cartographie des covariables : région, densité de drainage (indice IDPR), pluie décennale en 24 heures et altitude moyenne.

Pour la distribution GEV, le traitement du paramètre de forme est différent de celui décrit dans l'équation (8) : en effet, la régression ne peut pas porter sur le logarithme du paramètre car le paramètre de forme peut être positif ou négatif. De plus, des analyses préliminaires ont montré que l'effet des covariables n'était pas détectable. La régression est donc simplifiée en considérant un paramètre de forme constant au sein de la région, ce qui revient à utiliser la régression suivante :

$$\xi_i = \beta_0 + \varepsilon_i, \varepsilon_i \sim N(0, \sqrt{\sigma_\varepsilon^2 + \sigma_i^2}) \quad (9)$$

2.5 Krigeage des résidus

Une étape optionnelle de krigeage des résidus a été implémentée. Il s'agit d'un krigeage ordinaire, où la distance entre bassins versants est calculée comme la distance des centres de gravité, et où un variogramme exponentiel a été utilisé. Ce krigeage a été effectué à l'aide du package 'geoR' disponible dans le langage R.

2.6 Estimation de la loi régionale

Une distribution régionale adimensionnelle est estimée pour chaque région, en mélangeant les données normalisées de tous les sites de la région (i.e. les données brutes divisées par l'indice de crue). Précisons que cette étape n'est nécessaire que pour l'approche par l'indice de crue : la régression directe sur les paramètres de la distribution ne nécessite pas d'estimer cette distribution régionale. La distribution régionale est estimée par une approche Bayésienne, comme décrit en section 1.5.

2.7 Distribution estimée au site cible

Le transfert des informations au site cible dépend de l'approche régionale utilisée.

Pour la méthode de l'indice de crue, on commence par estimer l'indice de crue du site cible en utilisant la régression :

$$\hat{v}_c = \exp\left(\hat{\beta}_0 + \sum_{j=1}^{N_{cov}} \hat{\beta}_j x_c^{(j)} + \varepsilon_c\right) \quad (10)$$

Dans l'équation (10), ε_c correspond au résidu krigé si l'option du krigeage a été utilisée, et est égal à zéro sinon. La seconde étape consiste à estimer les paramètres du site cible en combinant cet indice de crue et la loi régionale (section 2.6) de la façon suivante :

1. Loi de Gumbel :
 - a. position : $\hat{\mu}_c = \hat{v}_c \times \hat{\mu}_R$
 - b. échelle: $\hat{\lambda}_c = \hat{v}_c \times \hat{\lambda}_R$
2. Loi GEV:
 - a. position : $\hat{\mu}_c = \hat{v}_c \times \hat{\mu}_R$
 - b. échelle: $\hat{\lambda}_c = \hat{v}_c \times \hat{\lambda}_R$
 - c. forme: $\hat{\xi}_c = \hat{\xi}_R$

Pour la méthode utilisant une régression directe sur les paramètres, les paramètres du site cible sont obtenus directement en appliquant chacune des régressions mises en place :

$$\text{Paramètres de position/échelle : } \hat{\theta}_c = \exp\left(\hat{\beta}_0 + \sum_{j=1}^{N_{cov}} \hat{\beta}_j x_c^{(j)} + \varepsilon_c\right) \quad (11)$$

$$\text{Paramètre de forme : } \hat{\xi}_c = \hat{\beta}_0 + \varepsilon_c$$

Comme précédemment, le résidu ε_c correspond au résidu krigé si l'option du krigeage a été utilisée, et est égal à zéro sinon.

3 Approches statistiques mixtes locales-régionales

3.1 Généralités et notations

Les notations sont identiques à celles utilisées pour les approches régionales (section 2.1). La seule différence est que des données $\mathbf{y}^{(c)} = (y_1^{(c)}, \dots, y_{n_c}^{(c)})$ sont également disponibles au site cible c .

3.2 Combinaison des informations locales et régionales

Le principe de la combinaison d'informations locales et régionales est inspiré du formalisme Bayésien décrit par [Ribatet et al., 2006]: l'information régionale est utilisée pour spécifier la distribution a priori, tandis que les données locales permettent de calculer la vraisemblance.

Plus précisément, la distribution a priori de chaque paramètre découle des régressions implémentées par l'approche régionale de la manière suivante :

1. Approche par l'indice de crue
 - a. Position : $\log(\mu_c) \sim N(\hat{v}_c + \log(\hat{\mu}_R), \hat{\sigma}_\varepsilon^{(\mu)})$
 - b. Echelle : $\log(\lambda_c) \sim N(\hat{v}_c + \log(\hat{\lambda}_R), \hat{\sigma}_\varepsilon^{(\lambda)})$
 - c. [GEV seulement] Forme : $\xi_c \sim N(\hat{\xi}_R, \hat{\sigma}_R^{(\xi)})$
2. Approche par régression directe sur les paramètres
 - a. Position : $\log(\mu_c) \sim N(\hat{\mu}_c, \hat{\sigma}_\varepsilon^{(\mu)})$
 - b. Echelle : $\log(\lambda_c) \sim N(\hat{\lambda}_c, \hat{\sigma}_\varepsilon^{(\lambda)})$
 - c. [GEV seulement] Forme : $\xi_c \sim N(\hat{\xi}_c, \hat{\sigma}_\varepsilon^{(\xi)})$

Dans les spécifications 1a-b et 2a-b-c, l'écart type correspond soit à l'écart-type de krigeage (si l'option est activée) soit à l'écart-type résiduel (pas d'option krigeage). Pour la spécification 1c, l'écart-type correspond à l'écart-type a posteriori issu de l'estimation Bayésienne de la loi régionale.

Les données du site cible $\mathbf{y}^{(c)} = (y_1^{(c)}, \dots, y_{n_c}^{(c)})$ sont ensuite utilisées pour calculer la vraisemblance, comme décrit dans l'équation (3). Le reste de l'analyse est une estimation Bayésienne classique comme décrit en section 1.5.

4 La méthode SHYREG-Débit

La méthode SHYPRE a initialement été développée pour simuler des scénarios de crues sur un bassin versant donné (Arnaud and Lavabre 2002). Elle associe un générateur stochastique de pluies horaires à une modélisation simple de la pluie en débit (Cernesson 1993; Arnaud 1997). La méthode génère ainsi un ensemble d'hydrogrammes de crue, à partir duquel on peut déduire empiriquement les quantiles de débits maximums et moyens sur différentes durées, et ceci pour différentes périodes de retour. La mise en œuvre de la méthode nécessite :

- la connaissance locale des pluies horaires sur le bassin pour caler le générateur de pluie,
- la connaissance des débits à l'exutoire du bassin pour le calage d'un modèle hydrologique.

La méthode SHYREG a été développée à la suite de SHYPRE et repose sur les mêmes hypothèses. Elle a pour objectif la connaissance régionale des débits de crue extrêmes. Sa mise en œuvre passe par :

1. **La régionalisation du générateur de pluies horaires, sur la base de caractéristiques de pluies journalières.** Ce travail a abouti à la cartographie des quantiles de pluies en France Métropolitaine (Arnaud, Lavabre et al. 2006; Arnaud, Lavabre et al. 2008) et dans les DOM. Le travail en milieu tropical humide (La Réunion, Les Antilles) nous a permis de bien contrôler les performances du modèle pour des événements courants sous ces climats mais exceptionnels en Métropole (Arnaud, Fine et al. 2006).

2. **La régionalisation du modèle hydrologique.** On choisit de travailler à une maille de 1 km² pour pouvoir générer des quantiles de débits spécifiques qui seront ensuite agglomérés à l'échelle des bassins versants. Ce passage au pixel, imposé par la nature ponctuelle des pluies horaires modélisées par le générateur de pluies, nous permet de simplifier le modèle hydrologique et de reporter toute la variabilité du comportement hydrologique des bassins versant sur un seul paramètre (l'état initial moyen en début de crue du réservoir de production, qui est assimilable à un coefficient d'écoulement non linéaire). Les problèmes d'abattement des pluies et de routage hydrologique ne sont donc pas abordés dans un premier temps car le travail de modélisation s'effectue sur des pixels de 1 km², identifiés comme des bassins versants virtuels. SHYREG génère donc, dans un premier temps, des débits spécifiques géo-référencés obtenus par la pluviométrie locale et un paramètre hydrologique local.
3. **L'agrégation de l'information sur les bassins.** Cette information pixellisée est ensuite utilisée pour estimer les quantiles de débits d'un bassin versant donné, par le biais d'une Fonction de Transfert Statistique (FTS). Cette fonction prend en compte de façon globale, l'abattement des pluies et l'abattement hydraulique opérants sur des bassins versants de différentes tailles. La FTS est une fonction de la superficie du bassin versant et reste paramétrée de façon unique sur le territoire pour une durée de quantiles donnée.

La méthode SHYREG est donc une version régionalisée de la méthode SHYPRE. Elle s'appuie fortement sur la connaissance de la pluviométrie pour estimer les quantiles de crues courants à extrêmes. La pluviométrie est estimée régionalement suite au travail de cartographie des paramètres d'un générateur de pluies horaires de SHYPRE. C'est la base SHYREG-pluie. Ce travail, synthétisé en annexe 2, est présenté en détail dans un guide Méthodologique (Arnaud and Lavabre 2010).

4.1 Calage de la méthode SHYREG-débit

La méthode SHYREG-débit participe à l'ensemble des actions du projet ExtraFlo, que ce soit pour l'étude des méthodes locales, régionales ou locale-régionales, aussi bien pour la prédétermination des quantiles de pluies que pour la prédétermination des quantiles de débits. Concernant la partie « débit », pour la méthode SHYREG-débit, les calages ont été faits de la manière suivante :

- Calage local : sur chaque bassin versant, le paramètre hydrologique est calé avec une partie des années de la chronique de débit disponible. A partir de ces années, on extrait les débits maximums annuels qui vont servir à caler une loi GEV pour déterminer les quantiles 2, 5 et 10 ans. Ces quantiles servent alors à caler le paramètre hydrologique. Puis on en déduit une distribution de fréquences fournies par la méthode, qui est alors comparée aux observations issues des années de validation.
- Calage régional : sur chaque bassin de l'échantillon de calage, on cale le paramètre hydrologique de la même façon que précédemment, mais avec l'ensemble des données de la chronique. La procédure de régionalisation du paramètre hydrologique est alors réalisée pour déterminer les valeurs de ce paramètre pour les bassins de validation. On compare alors les distributions de fréquences fournies par la méthode aux observations issues des stations de validation.
- Calage local-régional : c'est la même procédure que pour le calage local qui est utilisée pour la méthode SHYREG. Dans cette action, c'est l'échantillonnage qui change.

4.1.1 Calage de la méthode SHYREG en local : Action 2 et Action 6

Le calage de la méthode SHYREG-débit consiste à choisir une paramétrisation du modèle hydrologique qui permette de retrouver au mieux la distribution de fréquence des débits observés aux stations jaugées. Différentes modélisations ont été testées dans diverses études (Fouchier and Lavabre 2002; Arnaud 2005; Graff 2005; Aubert 2011). On présente ici la dernière version utilisée dans le cadre du projet ANR ExtraFlo, suite au travail de thèse d'Aubert (2011). Le principe du calage et la version de la modélisation retenue sont présentés schématiquement sur la Figure 2.

1. **La première étape consiste à générer des quantiles de crue à la maille du pixel.** Pour cela, des simulations sont faites à l'échelle du pixel, pour différentes valeurs de l'unique paramètre qu'il faudra régionaliser : $S0/A$. Ces simulations consistent à générer, à chaque pixel de la zone étudiée, des événements de pluies horaires qui sont transformés en événements de crues par un modèle hydrologique du type GR (Edijatno and Michel 1989). Ce modèle a préalablement été simplifié parce qu'il est utilisé à une maille kilométrique, mais aussi pour faciliter sa régionalisation. Ce modèle est composé de 2 réservoirs (A le réservoir de production et B, le réservoir de transfert) et d'un hydrogramme unitaire. Les paramètres de ces opérateurs sont fixés, sauf le niveau de remplissage initial du premier réservoir. L'hydrogramme unitaire répartit la pluie horaire sur deux pas de temps suivant la répartition 70% au pas de temps de la pluie et 30% au pas de temps suivant. La taille du réservoir A est imposée par l'hydrogéologie (Aubert 2011). La taille du réservoir B est imposée à 50 mm pour les événements d'été et à 100 mm pour les événements d'hiver, et son niveau de remplissage initial est imposé à 30% de la capacité de B. Le niveau de remplissage du réservoir A ($S0$) est donc le seul paramètre variable (entre 0 et 100% de la capacité de A), et des simulations sont réalisées pour différentes valeurs de ce paramètre. Au final, pour chaque valeur de $S0/A$ et à chaque pixel on peut simuler des événements de crue d'où sont extraits de façon empirique les quantiles de crue : $q_i(d, T, S0/A)$. A l'ensemble des débits générés, on rajoute un débit de base (dit $Q0$) qui correspond à l'estimation du débit mensuel moyen fourni de façon régionale par la méthode LOIEAU (Folton and Lavabre 2006; Folton and Lavabre 2007). La variabilité spatiale des débits, pour une même durée, une même période de retour et une même valeur de $S0/A$, est assurée par la pluviométrie (variabilité des paramètres du générateur) et à moindre échelle aussi à la variabilité de A et de $Q0$. La notion de durée caractéristique du bassin versant a été, sinon éludée sur cette modélisation au km^2 , tout au moins considérée comme identique pour tous les pixels. Le paramétrage régional ne porte donc que sur le rendement des pluies.

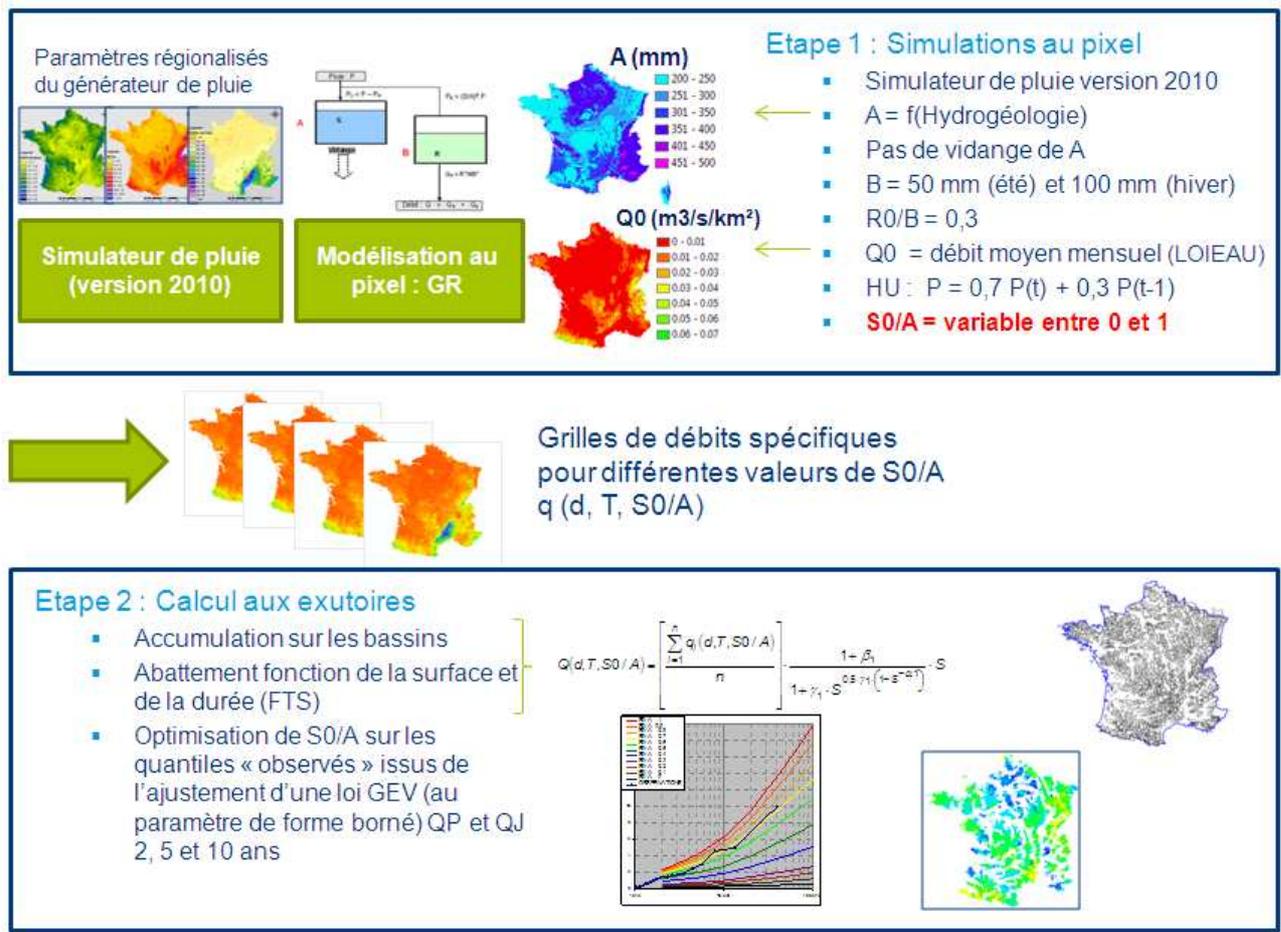


Figure 2 : Principe du calage de la méthode SHYREG-débit

2. La seconde étape consiste à calculer les quantiles de crues aux exutoires des bassins jaugés et à optimiser la valeur du paramètre S0/A. C'est l'étape proprement dite de calage de la méthode. Le choix des bassins versants de calage a été effectué pour ne retenir que les bassins non influencés, jugés « fiables » en hautes eaux par les gestionnaires, présentant une chronique d'observation suffisamment longue (10 ans de données au minimum) et pour lesquels on dispose d'un contour du bassin versant (Hydris 2009). Pour chaque bassin versant, un ajustement des débits de pointe et des débits journaliers maximums annuels a été réalisé par une loi de probabilité GEV dont le paramètre de forme a été borné (valeur devant être comprise entre 0 et 0,4). Cet ajustement est utilisé pour obtenir les quantiles de crue courants (T = 2, 5 et 10 ans), qui serviront au calage de la méthode. Ainsi, pour chaque bassin versant et pour chaque valeur de S0/A, on cumule les débits des pixels contenus dans le bassin. Ces débits sont ensuite réduits par une fonction de la surface et de la durée. Cette fonction nommée FTS⁽¹⁾ (Fonction de Transfert Statistique) permet de prendre

¹ La FTS crée une réduction des débits cumulés sur les bassins versants, en fonction de la surface du bassin. Elle s'exprime par les relations suivantes :

Pour $d \geq 24$ h

$$Q(d, T, S0/A) = \frac{\sum_{i=1}^n q_i(d, T, S0/A)}{n} \cdot \frac{1 + \beta_1}{1 + \gamma_1 \cdot S^{0.5 \cdot \gamma_1 (1 - S^{-\alpha_1})}} \cdot S$$

en compte à la fois l'abattement des pluies avec la surface, mais aussi un abattement hydraulique. On obtient pour chaque bassin versant les débits $Q(d,T,S0/A)$. Le calage de la méthode consiste donc à trouver la valeur de $S0/A$ appropriée pour respecter les quantiles de période de retour 2, 5 et 10 ans des débits de pointe et des débits journaliers. Pour chaque bassin on trouve donc la valeur du paramètre $S0/A$ qui minimise les écarts entre les 6 quantiles issus des observations et les 6 mêmes quantiles fournis par la méthode SHYREG-débit.

Ainsi, pour chaque bassin versant jaugé, la méthode SHYREG-débit peut-être calée en optimisant un seul paramètre. C'est sur ce paramètre que va porter la régionalisation nécessaire pour pouvoir appliquer la méthode en tout point du réseau hydrographique (milieu non jaugé).

Précision : Dans le cadre du projet ANR Extraflo, les données disponibles pour caler les méthodes étaient soit les débits journaliers, soit les débits de pointe. Dans chaque cas, la méthode SHYREG n'a alors été calée que sur les 3 quantiles de crues disponibles :

- Soit, les quantiles 2, 5 et 10 ans des débits journaliers pour une partie de l'action 2 et pour l'action 6.
- Soit, les quantiles 2, 5 et 10 ans des débits de pointe pour la seconde partie de l'action 2 (ou action 9).

Dans les deux cas, les quantiles de crues dits « observés » sont fournis par l'ajustement d'une loi GEV dont le paramètre de forme est borné entre les valeurs 0 et 0,4 : loi dite « GEV Bornée ».

Le paramètre d'optimisation est la moyenne pondérée des écarts relatifs entre les quantiles « observés » et les quantiles fournis par la méthode SHYREG. La pondération donne un poids de 1, 2 et 2 aux quantiles 2, 5 et 10 ans.

4.1.2 Calage de la méthode SHYREG en régional : Action 4

La version régionale de SHYREG est simplement basée sur l'utilisation d'une valeur régionalisée du paramètre $S0/A$. Dans l'action 4, comparant les méthodes régionales, le paramètre $S0/A$ est calculé sur les stations de calage, puis régionaliser sur la zone d'étude pour pouvoir être estimé sur les stations de validation.

La méthode de régionalisation choisie repose sur la recherche de régressions entre le paramètre optimisé sur chaque bassin versant de calage et des variables environnementales,

Pour $d < 24$ h

$$Q(d,T,S0/A) = Q(24h,T,S0/A) + \left[\frac{\sum_{i=1}^n q_i(d,T,S0/A)}{n} - \frac{\sum_{i=1}^n q_i(24h,T,S0/A)}{n} \right] \cdot \frac{1 + \beta_2}{1 + \gamma_2 \cdot S^{0,5} \gamma_2 (1 - S^{-\alpha_2})} \cdot S$$

Avec :

- n : le nombre de pixels de 1 km² contenus dans le bassin versant
- S : surface du bassin versant (km²)
- $Q(d,T,S0/A)$: débit moyen de durée d et de période de retour T, calculé à l'exutoire du bassin versant par la FTS (d=0 pour le débit de pointe), pour une valeur de $S0/A$.
- $q_i(d,T,S0/A)$: débit moyen de durée d et de période de retour T, simulé sur un pixel du bassin versant (d=0 pour le débit de pointe), pour une valeur de $S0/A$.

puis sur l'interpolation spatiale des résidus à ces régressions (Hydris 2009). Cette méthode est décrite par la Figure 3.

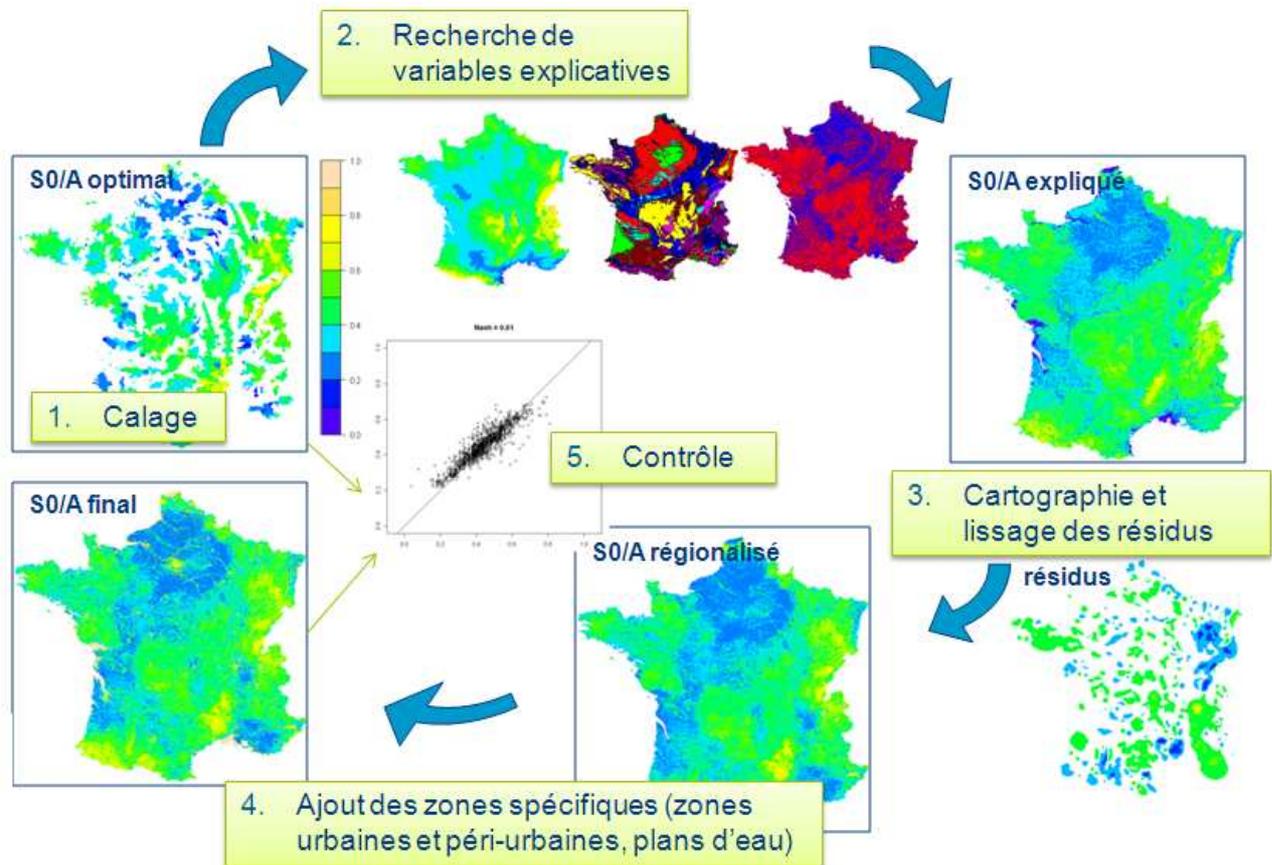


Figure 3 : Principe de la régionalisation du paramètre du modèle hydrologique.

Elle passe par plusieurs étapes :

- 1. Détermination des valeurs du paramètre optimal.** Cette valeur est issue de l'étape de calage présentée au paragraphe précédent. Dans ce cas, le paramètre est considéré comme constant pour l'ensemble des pixels d'un même bassin versant.
- 2. La recherche de variables explicatives** consiste à trouver des corrélations linéaires entre les valeurs optimales du paramètre à régionaliser et des valeurs de variables environnementales moyennées sur les bassins versants. Dans le cas de la régionalisation du paramètre S0/A, les variables permettant d'expliquer la variabilité du paramètre sont la densité du réseau de drainage, l'hydrogéologie et une variable de bilan hydrique. Bien que les corrélations trouvées soient faibles, avec un coefficient de corrélation multiple (r^2) de 0,28, on détermine une carte du paramètre dite « S0/A expliqué ».
- 3. Cartographie des résidus.** Les corrélations trouvées précédemment permettent d'expliquer une variabilité relativement faible du paramètre. Les résidus à cette corrélation sont alors calculés pour chaque bassin versant. Ces résidus sont alors cartographiés en attribuant à chaque pixel de la zone, la valeur du résidu du bassin le plus petit présent sur ce pixel (cas de bassins emboîtés). Une interpolation simple du type IDW (inverse de la distance au carré) est réalisée pour prendre en compte les tendances régionales à la surestimation ou la sous-estimation. On obtient alors une

carte dite des « résidus » qui peut éventuellement être lissée. La carte dite « S0/A régionalisée » est alors la somme des cartes « S0/A expliqué » et « résidus ».

4. **Intégration de zones spécifiques.** A partir de l'information issue de la base de données d'occupation des sols (Corine Land Cover), on a défini des pixels dont la valeur du paramètre est imposée. Ces pixels de 1 km² sont ceux concernés majoritairement par la présence de zones urbaines, périurbaines et plans d'eau que l'on considère comme fortement ou totalement imperméable. Sur ces pixels la valeur de S0/A est alors imposée dès l'étape d'optimisation.
5. **Le contrôle** consiste à vérifier la concordance entre les valeurs optimales du paramètre et celles fournies par la régionalisation. A noter que la présence de bassins emboîtés ne nous permet pas de retrouver exactement les valeurs des paramètres optimums même en prenant en compte les résidus.

Différentes procédures de régionalisation ont été testées. Ces procédures ont été comparées pour retenir celle qui conduisait à la meilleure restitution des quantiles de crues sur les échantillons de validation. La méthode retenue repose sur :

- L'utilisation de trois variables explicatives au maximum retenue pour les régressions
- Régionalisation des résidus par la méthode IDW2
- Pas de lissage de carte
- Prise en compte de l'hydrogéologie en attribuant un coefficient correcteur par grandes classes d'hydrogéologie.

Remarque : La méthode de régionalisation retenue est perfectible sur plusieurs points :

- La régression testée est linéaire (en log ce peut être mieux) et unique sur l'ensemble de la France. Or une régression par sous-région est sûrement préférable.
- La méthode d'interpolation des résidus n'est pas exacte, c'est-à-dire que l'on ne retrouve pas les valeurs optimales sur les stations ayant servi au calage. Ce problème est lié à la présence de bassins emboîtés.

5 Références

- Arnaud, P. (1997).** *Modèle de prédétermination de crues basé sur la simulation. Extension de sa zone de validité, paramétrisation du modèle horaire par l'information journalière et couplage des deux pas de temps.* Thèse de l'Université Montpellier II. : 258 p. + annexes.
- Arnaud, P. (2005).** *Simplification de gr3h pour la prédétermination des crues. Application sur des petits bassins versants.* Note interne.: 26 p.
- Arnaud, P. and J. Lavabre (2002).** "Coupled rainfall model and discharge model for flood frequency estimation." *Water Resources Research* 38(6).
- Arnaud, P. and J. Lavabre (2010).** *Estimation de l'aléa pluvial en France métropolitaine.* Editions QUAE, Editions QUAE: 158 pages.
- Arnaud, P., J. Lavabre, et al. (2006).** "Cartographie de l'aléa pluviographique de la France " *La houille blanche* 5: 102-111.
- Arnaud, P., J. Lavabre, et al. (2008).** "Regionalization of an hourly rainfall model in French territory for rainfall risk estimation." *Hydrological Sciences Journal* 53(1): 21p.
- Arnaud, P., J.-A. Fine, et al. (2006).** "An hourly rainfall generation model adapted to all types of climate." *Atmospheric Research* 85(2): 230-242.
- Aubert, Y. (2011).** *Estimation des valeurs extrêmes de débit par la méthode SHYREG : réflexions sur l'équifinalité dans la modélisation de la transformation pluie en débit. .* Thèse doctorat Université Paris VI 317 pages.
- Benichou, P., and O. Le Breton (1987),** Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques, *La Météorologie*, 7(19), 23-34
- Cernesson, F. (1993).** *Modèle simple de prédétermination des crues de fréquences courante à rare sur petits bassins versants méditerranéens.* Thèse de doctorat de l'Université Montpellier II: 240 p + annexes.
- Cipriani, T., T. Toilliez, and E. Sauquet (2012),** Estimating 10 year return period peak flows and flood durations at ungauged locations in France, *La houille blanche*
- Coles, S. (2001),** *An Introduction to Statistical Modeling of Extreme Values*, 210 pp., Springer-Verlag, London.
- Dalrymple, T. (1960),** Flood frequency analyses, in *Water-supply paper 1543-A*, edited, US Geological Survey.
- Edijatno and C. Michel (1989).** "Un modèle pluie-débit journalier à trois paramètres." *La Houille Blanche* 2: 113-121.
- Folton, N. and J. Lavabre (2006).** "Regionalization of a monthly rainfall-runoff model for the southern half of France based on a sample of 880 gauged catchments." *IAHS Publication Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment - MOPEX* vol. 4: 264-277.
- Folton, N. and J. Lavabre (2007).** "Approche par modélisation pluie-débit pour la connaissance régionale de la ressource en eau: application à la moitié du territoire français." *Houille-Blanche* n° 03-2007: 64-70.
- Fouchier, C. and J. Lavabre (2002).** *Synthèse des débits de crue sur les régions Provence-Alpes-Côte d'Azur et Languedoc-Roussillon. Phase II : régionalisation du modèle pluie-débit. .* Convention n° 57/2000 MATE/Météo France/CETE Méditerranée/ Cemagref. : 79 p.
- Graff, B. (2005).** *Recherche d'une modélisation pluie-débit invariante pour la prédétermination des crues. Application à de petits bassins versants.* Rapport d'étude Cemagref - Convention DPPR 2004: 61 pages + annexes.
- Hydris (2009).** *Cartographie des débits de crue en Métropole. Régionalisation du modèle pluie-débit. Détermination des débits de crues de référence par l'application de la*

méthode SHYPRE régionalisée sur la métropole. 1. Analyse des données hydrométriques.
Rapport d'étude: 39 p.

- Mardhel, V., P. Frantar, J. Uhan, and A. Mio (2004)**, Index of development and persistence of the river networks as a component of regional groundwater vulnerability assessment in Slovenia., paper presented at *Int. Conf. groundwater vulnerability assessment and mapping, Ustron, Poland*, 15-18 June 2004.
- Martins, E. S., and J. R. Stedinger (2000)**, Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resources Research*, 36(3), 737-744
- Renard B., Kochanek K., Lang M., Garavaglia F., Paquet E., Neppel L., Najib K., Carreau J., Arnaud P., Aubert Y., Borchì F., Soubeyrou J.M., Jourdain S., Veysseire J.M., Sauquet E., Cipriani T., Auffray A., 2013.** Data-based comparison of frequency analysis methods: A general framework. *Water Resources Research*, 49, 1-19.
- Renard, B., V. Garreta, and M. Lang (2006)**, An application of Bayesian analysis and MCMC methods to the estimation of a regional trend in annual maxima, *Water Resources Research*, 42(12)
- Ribatet, M., E. Sauquet, J. M. Gresillon, and T. B. M. J. Ouarda (2006)**, A regional Bayesian POT model for flood frequency analysis, *Stoch. Environ. Res. Risk Assess.*, 21(4), 327-339
- Wasson, J. G., A. Chandesris, H. Pella, and L. Blanc (2004)**, Les hydro-écorégions: une approche fonctionnelle de la typologie des rivières pour la directive cadre européenne sur l'eau, *Ingénieries*, 40, 3-10

6 Annexes

6.1 Annexe 1 : propriétés des distributions utilisées dans les actions

6.1.1 Loi Normale

Paramètres	Moyenne μ ; Ecart-type $\sigma > 0$
Support	$z \in]-\infty; +\infty[$
Densité	$p(z \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$
Estimateurs des moments	$\hat{\mu} = \bar{y}$ (moyenne empirique) $\hat{\sigma} = \sigma_y$ (écart-type empirique)
Estimateurs des L-moments	$\hat{\mu} = \bar{y}$ $\hat{\sigma} = l_2^{(emp)} / \sqrt{\pi}$
Estimateurs du maximum de vraisemblance	$\hat{\mu} = \bar{y}$ $\hat{\sigma} = \sigma_y$

6.1.2 Loi de Pearson III

Paramètres	Position μ ; Echelle $\lambda \neq 0$; Forme $\xi > 0$
Support	$(z - \mu) / \lambda \geq 0$
Densité	$p(z \mu, \lambda, \xi) = \begin{cases} \frac{((z - \mu) / \lambda)^{\xi-1} \exp(-(z - \mu) / \lambda)}{ \lambda \Gamma(\xi)} & \text{si } (z - \mu) / \lambda \geq 0 \\ 0 & \text{sinon} \end{cases}$
Estimateurs des moments	$\hat{\xi} = 4 / \text{asym}_y^2$ $\hat{\lambda} = \text{signe}(\text{asym}_y) \times (\sigma_y / \sqrt{\hat{\xi}})$ $\hat{\mu} = \bar{y} - \hat{\lambda} \times \hat{\xi}$ <p>asym_y désigne l'asymétrie empirique</p>
Estimateurs des L-moments	$\hat{\mu} \approx \bar{y} - 2(\tau / \eta)$ $\hat{\lambda} \approx 0.5 \tau \eta$ $\hat{\xi} \approx 4 / \eta^2$ <p>où les valeurs τ et η sont définies de la façon suivante :</p> $\eta = (2 / \sqrt{a}) \times \text{signe}(l_3^{(emp)})$ $\tau = l_2^{(emp)} \sqrt{\pi a} \frac{\Gamma(a)}{\Gamma(a + 0.5)}$ <p>avec :</p> <p>si $l_3^{(emp)} < 1 / 3$:</p> $w = 3\pi(l_3^{(emp)})^2$ $a = (1 + 0.2906w) / (w + 0.1882w^2 + 0.0442w^3)$ <p>sinon:</p> $w = 1 - l_3^{(emp)} $ $a = (0.36067w - 0.59567w^2 + 0.25361w^3) / (1 - 2.78861w + 2.56096w^2 - 0.77045w^3)$
Estimateurs du maximum de vraisemblance	Pas de formule explicite, optimisation numérique

6.1.3 Loi de Gumbel

Paramètres	Position μ ; Echelle $\lambda > 0$
Support	$z \in]-\infty; +\infty[$
Densité	$p(z \mu, \lambda) = \frac{1}{\lambda} \exp\left(-\frac{z-\mu}{\lambda} - \exp\left(-\frac{z-\mu}{\lambda}\right)\right)$
Estimateurs des moments	$\hat{\lambda} = \frac{\sqrt{6}}{\pi} \sigma_y$ $\hat{\mu} = \bar{y} - \gamma \hat{\lambda}, \text{ où } \gamma = 0.5772$
Estimateurs des L-moments	$\hat{\lambda} = l_2^{(emp)} / \log(2)$ $\hat{\mu} = \bar{y} - \gamma \hat{\lambda}$
Estimateurs du maximum de vraisemblance	Pas de formule explicite, optimisation numérique

6.1.4 Loi GEV

Paramètres	Position μ ; Echelle $\lambda > 0$; Forme $\xi \neq 0$
Support	$1 - \frac{\xi(z - \mu)}{\lambda} > 0$
Densité	$p(z \mu, \lambda, \xi) = \begin{cases} \frac{1}{\lambda} \left(1 - \frac{\xi(z - \mu)}{\lambda}\right)^{\frac{1}{\xi} - 1} \exp\left(-\left(1 - \frac{\xi(z - \mu)}{\lambda}\right)^{\frac{1}{\xi}}\right) & \text{si } 1 - \frac{\xi(z - \mu)}{\lambda} > 0 \\ 0 & \text{sinon} \end{cases}$ <p>Note : lorsque $\xi \rightarrow 0$, la densité tend vers celle d'une loi de Gumbel</p>
Estimateurs des moments	$\frac{\hat{\xi}}{ \hat{\xi} } \frac{[\Gamma(3\hat{\xi} + 1) - 3\Gamma(\hat{\xi} + 1)\Gamma(2\hat{\xi} + 1) + 2\Gamma^3(\hat{\xi} + 1)]}{[\Gamma(2\hat{\xi} + 1) - \Gamma^2(\hat{\xi} + 1)]^{3/2}} = asym_y \text{ (à résoudre numériquement)}$ $\hat{\lambda} = \hat{\xi} \sigma_y [\Gamma(2\hat{\xi} + 1) - \Gamma^2(\hat{\xi} + 1)]^{-1/2}$ $\hat{\mu} = \bar{y} - \frac{\hat{\lambda}}{\hat{\xi}} [1 - \Gamma(\hat{\xi} + 1)]$
Estimateurs des L-moments	$\hat{\xi} \approx 7.8590c + 2.9554c^2, \text{ avec } c = 2 / (3 + I_3^{(emp)}) - \log(2) / \log(3)$ $\hat{\lambda} = \left(\hat{\xi} \times I_2^{(emp)}\right) / \left((1 - 2^{-\hat{\xi}})\Gamma(1 + \hat{\xi})\right)$ $\hat{\mu} = \bar{y} + \left(\hat{\lambda} / \hat{\xi}\right) \left(1 - \Gamma(1 + \hat{\xi})\right)$
Estimateurs du maximum de vraisemblance	Pas de formule explicite, optimisation numérique

6.2 Annexe 2 : la base SHYREG-Pluie

Nous rappelons rapidement les grands principes de la mise en œuvre de la base SHYREG-pluie. L'élaboration de la base SHYREG-pluie est passée par différentes étapes, représentées sur la Figure 4 :

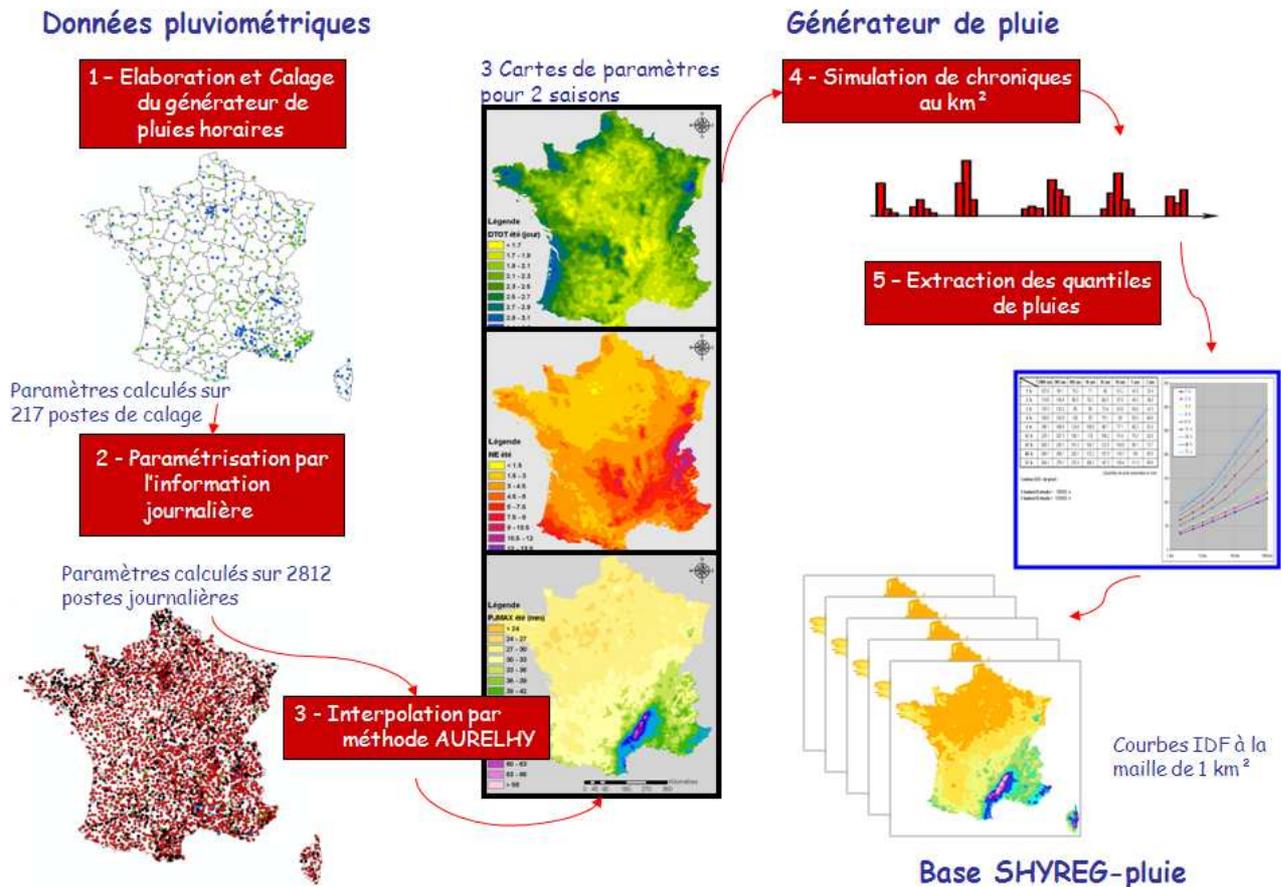


Figure 4 : Principe de l'élaboration de la base SHYREG-pluie

- 1. Le calage d'un générateur de pluie.** Le principe est de décrire les chroniques des pluies horaires par l'intermédiaire de différentes variables aléatoires. Ces variables sont au nombre de 8 (nombre d'événements pluvieux par an, nombre d'averses par événements, durée des averses, volumes des averses, etc...). Elles sont toutes associées à une loi de probabilité adaptée qui permet le tirage aléatoire de valeurs nécessaires pour simuler des chroniques de pluies horaires (Cernesson, Lavabre et al. 1996). La dépendance entre certaines variables a été mise en évidence et modélisée (Cantet 2009). Le calage du modèle a été réalisé sur 217 postes horaires de référence et contrôlé sur 207 autres postes. Ce calage est effectué pour deux saisons (l'été de juin à novembre et l'hiver de décembre à mai).
- 2. La paramétrisation du générateur par une information journalière.** Par le biais de corrélations linéaires, on arrive aisément à corrélérer les paramètres les plus importants du générateur de pluie avec des caractéristiques journalières de la pluie. D'autres paramètres sont fixés à des valeurs régionales. Sur la base de la définition d'un événement pluvieux comme étant la succession de pluies journalières de plus de 4 mm

et ayant au moins une pluie journalière de plus de 20 mm, les caractéristiques journalières sont : le nombre d'événements pluvieux par année, la moyenne des durées des événements pluvieux et la moyenne des pluies journalières maximales des événements pluvieux. L'intérêt de cette paramétrisation par l'information journalière est de disposer d'une information pluviométrique nettement plus dense au pas de temps journalier qu'au pas de temps horaire. De plus, cette paramétrisation est basée sur des caractéristiques moyennes qui garantissent la robustesse de la méthode face aux problèmes d'échantillonnage des observations.

3. **La régionalisation des paramètres journaliers** : les trois paramètres journaliers permettant de mettre en œuvre le générateur de pluie horaire ont été déterminés sur 2812 chroniques de pluies journalières observées sur la période 1977-2002. Ces trois paramètres caractérisant l'occurrence, la durée et l'intensité des pluies, ont alors été régionalisés à une maille de 1x1 km en utilisant une méthode de type AURELHY (Benichou and Le Breton 1987) par Météo-France (Sol and Desouches 2005).
4. **La simulation de chroniques de pluies** : à partir des 3 paramètres journaliers régionalisés sur une maille de 1 x 1 km, on génère au pas de temps horaire un échantillon d'événements pluvieux indépendants équivalant à un échantillon de plusieurs milliers d'années d'observation.
5. **Extraction des quantiles de pluies (courbes IDF)** : On analyse les événements pluvieux comme des observations en extrayant les valeurs de pluies maximales de différentes durées. Ces valeurs sont ensuite classées pour déterminer des distributions empiriques d'où sont extraits directement les quantiles empiriques de pluies de différentes durées, sans ajustement de lois. On obtient ainsi une base de données de courbes IDF à 1km², représentant la pluie ponctuelle.

La régionalisation du générateur de pluie permet son utilisation sur l'ensemble du territoire d'étude. Elle fournit une information pixellisée robuste et pertinente de l'aléa pluvial, prenant en compte l'information pluviométrique régionale et une information topographique. Cette base de données SHYREG-pluie servira de donnée d'entrée pour la constitution de la base SHYREG-débit.