

Projet ANR-08-RISK-03-01

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

Programme RISKNAT 2008

Tâche III : Inter-comparaison des méthodes probabilistes

Rapport III.3 « *Comparaison des méthodes purement
régionales pour l'estimation des pluies extrêmes* »

Date : Mars 2012

Rapport réalisé par :

⁽¹⁾ HydroSciences Montpellier

Avec la participation de :

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

Auteurs :

J. Carreau¹, L. Neppel¹, P. Arnaud²



Sommaire

1.	Présentation des modèles.....	4
1.1.	Loi régionale : LR.....	4
1.1.1.	Définition du voisinage associé au site i	5
1.1.2.	Estimation de la loi régionale.....	5
1.1.3.	Estimation de l'index value.....	6
1.1.4.	Estimation des intervalles de confiance et distribution prédictive.....	6
1.2.	Interpolation spatiale des paramètres d'une GEV locale : SIGEV.....	7
1.3.	Modèle de génération stochastique d'événements pluvieux : Shypre.....	7
1.3.1.	Mise en œuvre régionale.....	9
2.	Méthode de comparaison.....	10
2.1.	Les critères de comparaison.....	10
2.2.	Les échantillons C/V.....	10
3.	Résultats de la comparaison.....	12
3.1.	Comparaison à l'échelle de toute la zone d'étude.....	12
3.2.	Comparaison par zones climatiques.....	15
3.3.	Hiérarchisation des modèles par zone climatique.....	18
4.	Cartographie des différences de quantiles Shypre – loi régionale.....	18
5.	Conclusions / Perspectives.....	20
6.	Bibliographie.....	20

1. Présentation des modèles

Les modèles d'estimation purement régionales possèdent les propriétés suivantes :

- Ils sont totalement régionalisés, i.e. ils ne nécessitent aucune donnée locale en mode prédictif.
- ils permettent donc le transfert de l'estimation vers des sites non jaugés.

On cherchera à comparer la capacité de trois méthodes à estimer la distribution des pluies en un site non jaugé :

- la loi régionale des pluies maximales annuelles
- l'interpolation spatiale des paramètres d'une GEV calée localement
- le modèle SHYPRE

1.1. Loi régionale : LR

On note X_i la variable aléatoire des pluies maximales annuelles au site i et $F(x_i)$ une estimation de la fonction de répartition (fdr) de X_i . Le principe général des approches régionales consiste à estimer la fdr en un site i à partir des N_v observations situées dans un voisinage de X_i considéré comme homogène. On augmente ainsi l'échantillon permettant d'estimer $F(x_i)$. La méthode choisie est celle des stations années. On transforme les N_v variables aléatoires initiales $(X_j)_{j=1 \text{ à } N_v}$:

$$X_j \rightarrow Y_j = \frac{X_j}{\bar{X}_j}$$

où \bar{X}_j est la moyenne des pluies maximales annuelles en j , cette norme est appelée index value.

Cette transformation assure que toutes les variables Y_j aux N_v stations du voisinage ont :

- même moyenne :

$$E[Y_j] = \frac{\bar{X}_j}{\bar{X}_j} = 1$$

- même variance et moments d'ordre 3 si les coefficients de variation CV et de dissymétrie γ des X_j sont les mêmes au N_v stations :

$$[VAR(Y_j)] = \frac{[VAR(X_j)]}{\bar{X}_j^2} = CV(X_j)^2$$

$$M_3(Y_j) = M_3(X_j) / \mu_j^3 = \gamma(X_j) \cdot VAR(X_j)^{1.5} / \mu_j^3 = \gamma(X_j) \cdot CV(X_j)^3$$

Un voisinage homogène est donc défini par un ensemble de stations de même CV et γ . Ainsi si les Y_j suivent la même distribution, les paramètres seront les mêmes ce qui autorise la constitution d'un échantillon régional des N_v observations normalisées.

Cet échantillon est utilisé pour l'estimation de la fdr de Y_i au site i , $F(y_i)$ appelée loi régionale. Pour estimer au site i un quantile $x_i(p)$ de probabilité de non-dépassement p , on estime d'abord $\hat{F}_i(p)$ avec la loi régionale, puis la transformation inverse permet ensuite de revenir à $x_i(p)$, en multipliant $\hat{F}_i(p)$ par l'index value

La mise en œuvre de la méthode se décline en 4 étapes :

- Définition du voisinage associé au site i : V_i
- Estimation de la loi régionale : $\hat{F}(y_i)$
- Estimation de l'index value au site i : \bar{X}_i^*
- Estimation des intervalles de confiance des quantiles au site i

1.1.1. Définition du voisinage associé au site i

La sélection du voisinage se fait suivant la méthode ROI *Region of Influence* (Burn, 1990). Comme les précipitations présentent une corrélation spatiale, la région d'influence est continue et déterminée dans l'espace géographique. Elle est définie par un cercle de rayon RV_i centré sur le site i . La taille du voisinage dépend de la position du site i . Le rayon initial est fixé à 50 km, il est ensuite optimisé de la façon suivante :

- i) Les tests d'homogénéité d'Anderson et Darling (Viglione et al., 2007) et de Hosking (Hosking et Wallis, 1997) sont utilisés pour juger de l'homogénéité du voisinage
- ii) Si l'un des deux tests conclut à l'inhomogénéité du voisinage, on réduit de rayon RV_i de 5 km et l'on revient à l'étape i)
- iii) Si les deux tests concluent à l'homogénéité on retient le voisinage défini par le rayon RV_i

1.1.2. Estimation de la loi régionale

On échantillonne les pluies journalières maximales annuelles indépendantes spatialement aux N_V stations à l'intérieure du voisinage défini précédemment. On considère que plusieurs max annuels à la même date d'une même année sont indépendants spatialement si ils sont distants d'au moins 60 km. Cette distance correspond à la distance de décorrélation des précipitations journalières estimée à partir d'une analyse géostatistique des pluies (Lebel et Laborde, 1988, Neppel et al., 1997). Dans le cas contraire seul le max X_j le plus important est sélectionné.

On estime ensuite à chaque station l'index value, puis on constitue l'échantillon régional des variables normalisées $\{y_j\}_{j=1 \text{ à } N}$ où N est la somme des effectifs des max annuels échantillonnés à chaque station du voisinage.

Selon la théorie des valeurs extrêmes la distribution du maximum annuel tend asymptotiquement vers une loi généralisée des valeurs extrêmes (GEV). La loi régionale au site i est de type GEV :

$$F(y_i) = \exp \left[- \left(1 - k_R \left(\frac{y_i - \alpha_R}{\sigma_R} \right) \right)^{\frac{1}{k_R}} \right]$$

où α_R , σ_R et k_R sont respectivement les paramètres de position, d'échelle et de forme de la loi régionale.

Les cas $k_R \rightarrow 0$, $k_R < 0$ and $k_R > 0$ correspondent respectivement aux lois de Gumbel, de Fréchet et de Weibull. Ces paramètres sont estimés par la méthode des L-moments (Hosking, 1990) avec l'échantillon régional associé au site cible i .

1.1.3. Estimation de l'index value

On dispose de N index value estimés aux différents sites de l'ensemble de calibration. On veut estimer \bar{X}_i^* , l'index value du site cible, sans mesures. \bar{X}_i^* est estimé par interpolation spatiale à partir d'une régression dont les variables explicatives potentielles sont constituées de 22 covariables décrivant l'espace géographique : latitude, longitude, altitude, 12 composantes principales du relief, et 7 distances à la mer, au Rhône, à l'océan, etc ... (Sol et Desouches, 2005).

Les covariables sont centrées et réduites et une ACP basée sur la méthode Slice Inverse Regression (Li, 1991) permet de résumer ces 22 covariables avec 8 composantes principales $(CP_i)_{i=1 \text{ à } 8}$ qui expliquent 90% de la variance des index value. On retient ces composantes principales comme variables explicatives.

$$\bar{X}_i^* = \sum_{j=1}^8 \lambda_{ij} CP_{ij} + \varepsilon_i$$

Où λ_{ij} est le poids de la composante CP_j dans l'estimation de l'index value en i et ε_i est le résidu de la régression en i .

Les tests réalisés n'ont pas permis de montrer qu'une régression par zone est plus performante qu'une régression unique. On utilise une régression linéaire avec les mêmes variables explicatives pour tout le domaine. Les résidus de la régression sont lissés dans l'espace géographique par une méthode non paramétrique à noyau (Bishop, 2006, section 6.3.1). L'estimation du résidu en un point donné est une moyenne pondérée des valeurs observées dans un voisinage du point. Le voisinage se trouve dans l'espace géographique et sa forme est donnée par un noyau gaussien (une densité gaussienne centrée à zéro d'écart-type σ). Ainsi, des points rapprochés du point d'intérêt recevront une pondération élevée. La zone d'influence du voisinage est fixée par l'écart-type de la Gaussienne que nous avons fixé à 10 km. Cet estimateur à noyau a de bonnes propriétés de convergence.

1.1.4. Estimation des intervalles de confiance et distribution prédictive

Les intervalles de confiance à 90% des quantiles sont obtenus par une technique de bootstrap non paramétrique. Pour un site cible fixé, le voisinage est déterminé et fixé. Soit D la matrice des données avec en lignes les années, chaque colonne correspondant à une station du voisinage. Le bootstrap consiste à ré échantillonner les années donc les lignes de la matrice D . Pour chacun des NB échantillons bootstrapés on estime les paramètres de la GEV régionale suivant la méthode du §1.1.2



Figure 1 : illustration du ré échantillonnage pour la construction des intervalles de confiance de la loi régionale.

Les incertitudes sur l'index value résident surtout dans l'estimation locale de l'index value par la moyenne des max annuels observés, aussi on réalise un bootstrap spatial comme précédemment mais sur toutes les stations disponibles plutôt que seulement sur les stations du voisinage. Pour chacun des NB échantillons on reconstruit une régression avec les mêmes covariables initiales et les stations sélectionnées. On en déduit une nouvelle régression de l'index value.

Pour chacune des NB GEV bootstrapée et des NB Index Value on calcule le quantile de probabilité de non-dépassement p . On obtient ainsi un ensemble de NB quantiles $\{x_i(p)\}_{j=1 \text{ à } NB}$ le 5^{ème} et 95^{ème} percentile de cet échantillon représentent respectivement la borne inférieure et supérieure du quantile $x_i(p)$.

La distribution prédictive est construite en prenant la médiane des quantiles bootstrapés pour différentes valeurs de p .

1.2. Interpolation spatiale des paramètres d'une GEV locale : SIGEV

On dispose d'un ensemble N de postes de mesures, sur lesquels la pluie maximale annuelle est échantillonnée : $\{X_j\}_{j=1 \text{ à } N}$. Chaque variable X_j est modélisée par une GEV dont les paramètres α_j , σ_j et k_j sont estimés par la méthode des L-moments. On interpole spatialement les paramètres de la GEV au site cible sans mesure. Pour chaque paramètre on utilise la même technique d'interpolation que pour l'index value présentée ci-dessus : une régression à partir d'une combinaison linéaire des 22 covariables décrivant l'espace géographique et un lissage des résidus par une fonction à noyau.

L'intervalle de confiance est construit par bootstrap non paramétrique suivant la même procédure que pour l'index value. La médiane des quantiles bootstrapés permet de construire la distribution prédictive.

1.3. Modèle de génération stochastique d'événements pluvieux : Shypre

En absence d'information horaire, le générateur peut être calé par une information journalière. Dans ce cas, certains paramètres sont fixés (car peu variables ou peu sensibles) et d'autres sont estimés à partir de variables issues de pluies journalières.

Localement, on peut donc déterminer ces variables journalières pour caler le générateur de pluies horaires. Ces variables journalières caractérisent les événements pluvieux normalement sélectionnés pour être analysés par SHYPRE lorsque l'on dispose de chroniques horaires (cf *Figure 2*). En l'absence de pluies horaires on retient donc uniquement les caractéristiques journalières.

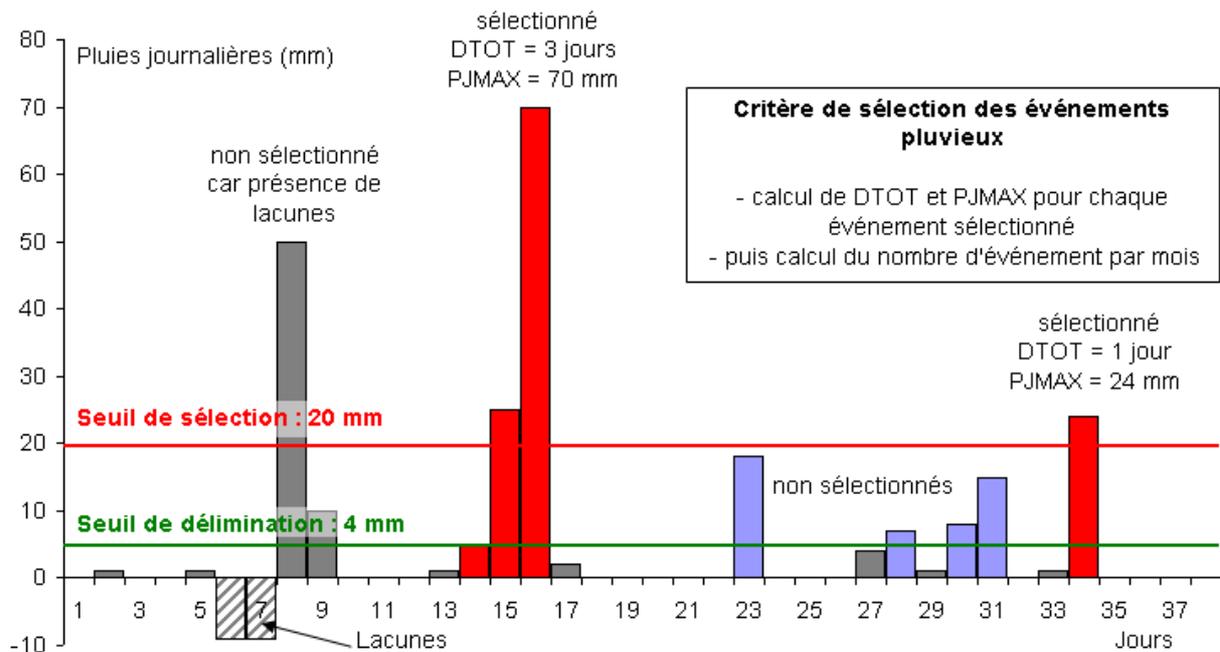


Figure 2 : critère de sélection des événements pluvieux et calcul de leur caractéristique.

La procédure mise en œuvre pour caler la méthode SHYPRE est donc la suivante:

- ♦ Pour chaque mois de chaque poste disponible, on détermine le nombre d'événements pluvieux définis au sens de SHYPRE⁽¹⁾. Un événement pluvieux est associé au mois de son premier jour.
- ♦ Pour chaque événement pluvieux on calcule sa durée (DTOT en jour) et sa pluie journalière maximale (PJMAX en mm).
- ♦ Pour chaque mois de chaque poste disponible, on détermine le nombre de jours en lacune.
- ♦ Un mois ayant dix jours de lacune ou plus est considéré en lacune, ainsi que tous les événements qui pourraient y être associé.

On dispose alors de caractéristiques mensuelles des chroniques de pluies : nombre d'événements pluvieux du mois (0 si le mois est considéré en lacune), la durée de chaque épisode et la pluie journalière maximale de l'événement.

On calcule alors pour chaque poste, et sur les années choisies² pour les différents tests d'échantillonnage, les caractéristiques suivantes :

- ♦ La moyenne des DTOT des événements des mois de juin à novembre : μ DTOT été
- ♦ La moyenne des DTOT des événements des mois de décembre à mai : μ DTOT hiver

¹ Un événement pluvieux est défini par une succession de pluies journalières supérieures à 4 mm (non bornées par des lacunes) avec la présence d'au moins une pluie journalière dépassant les 20 mm.

² On rappelle que par convention, l'année N est caractérisée par les données des mois de juin à décembre de l'année N et des mois de janvier à mai de l'année N+1.

- ◆ La moyenne des PJMAX des événements des mois de juin à novembre : μ_{PJMAX} été
- ◆ La moyenne des PJMAX des événements des mois de décembre à mai : μ_{PJMAX} hiver
- ◆ Le nombre d'événements retenus sur les mois de juin à novembre, par an : NE été
- ◆ Le nombre d'événements retenus sur les mois de décembre à mai, par an : NE hiver

Le calage du générateur de pluies est réalisé par le calcul de ses 3 paramètres pour les deux saisons définies : l'été de juin à novembre et l'hiver de décembre à mai.

On peut alors simuler des chroniques de pluies horaires (séries d'événements non datés) sur les deux saisons définies. On extrait alors de ces simulations les caractéristiques des pluies horaires générées : les pluies maximales en 1, 2, 3 ... 72 heures de chaque événement (PMD). On trace ensuite les distributions empiriques de ces caractéristiques pour en extraire certains quantiles.

On rappelle ici que les distributions de fréquences issues de SHYPRE sont des distributions empiriques associées aux caractéristiques des événements pluvieux horaires générés. C'est donc un produit de contrôle des capacités du générateur à reproduire des pluies horaires dont les caractéristiques statistiques sont proches des chroniques observées. En aucun cas ces distributions ne sont issues d'un ajustement d'une loi statistique sur les mêmes caractéristiques observées.

Les simulations effectuées correspondent à la simulation d'une centaine d'échantillons de 500 ans. La distribution moyenne des cent distributions déduites des 500 ans de simulation nous permet d'obtenir une distribution centrale relativement peu soumise à l'échantillonnage des simulations. Ce point a déjà été abordé dans différentes études qui montrent qu'une centaine de simulations reste un minimum pour la stabilité des estimations par SHYPRE (Muller 2006, Arnaud et Lavabre, 2010).

1.3.1. Mise en œuvre régionale

La mise en œuvre en régionale de la méthode SHYPRE consiste à l'appliquer en des points ne présentant pas d'observation pour le calage. Cette mise en œuvre « régionale » consiste donc à déterminer les paramètres journaliers en tout point d'une zone, par interpolation / extrapolation des valeurs observées.

La régionalisation de la méthode SHYPRE est réalisée par une méthode type « Aurhély ». (Benichou et Lebreton, 1987) Elle consiste en deux points :

- ◆ Trouver des variables environnementales connues en tout point, permettant d'expliquer une partie de la variabilité spatiale des différents paramètres. Généralement, ce travail se fait par régressions linéaires multiples.
- ◆ Interpoler les résidus. Les résidus sont les écarts entre les valeurs observées et les valeurs expliquées par les variables environnementales. L'interpolation choisie est généralement le krigeage.

Le calage de SHYPRE en mode régional est réalisé de la façon suivante, à partir des postes de l'échantillon de calage :

- ◆ Calcul des paramètres μ_{PJMAX} , μ_{DTOT} et NE (pour les deux saisons) sur les postes journaliers de l'échantillon de calage.

- ◆ Extraction des valeurs des variables environnementales au niveau des coordonnées des postes de l'échantillon de calage.
- ◆ Détermination des coefficients des régressions linéaires multiples. Ces régressions pourront être établies sur des grandes zones considérées comme climatiquement homogènes.
- ◆ Calcul des résidus aux niveaux des postes de l'échantillon de calage.
- ◆ Interpolation des résidus par krigeage (on pourra tester l'influence de la méthode d'interpolation choisie).
- ◆ Cartographie des paramètres μ PJMAX, μ DTOT et NE (pour les deux saisons) en sommant la carte des résidus et la carte des valeurs expliquées par les régressions.

Un calcul des quantiles SHYPRE a été réalisé au préalable avec une série de paramètres suivant un découpage de l'hyper espace des paramètres. Pour des paramètres estimés, on recherche les paramètres les plus proches et on interpole les quantiles à estimer en interpolant les quantiles préalablement calculés associés à ces plus proches paramètres.

La validation consistera alors à estimer les valeurs des quantiles issus des paramètres régionalisés, au niveau des postes de l'échantillon de validation.

2. Méthode de comparaison

2.1. Les critères de comparaison

Comme pour les autres actions, la performance des méthodes est jugée sur :

- La justesse : critères NT10, NT100 et FF pour la queue de la distribution
- La robustesse : SPAN10, SPAN100 et SPAN1000
- La stabilité des intervalles de confiance : COVER100

Pour faciliter la comparaison ces critères sont normalisés entre 0 et 1, correspondant respectivement au moins bon et au meilleur score.

2.2. Les échantillons C/V

Pour les critères de justesse, la décomposition C/V est spatiale (*Figure 3*) :

- On fixe le pourcentage p de stations utilisées pour le calage, le complément étant réservé à la validation. Suivant la densité spatiale des jeux régionaux disponibles, on prendra typiquement $p = 50\%$, 33% , 25%
- Pour pousser les approches régionales dans leur retranchement on construit également un jeu dit « méchant » (M) et « très méchant » (TM) :
 - M : On conserve en calage 388 stations avec des séries courtes entre 15 et 35 ans, en validation 650 stations avec de longues séries d'au moins 50 ans ; on pénalise ainsi le calage en n'utilisant que des séries courtes
 - TM : On pénalise encore davantage le calage en ne conservant que 273 stations avec des séries de 15 à 25 ans ; l'échantillon de validation est identique au cas M

Dans chaque cas C/V, les stations de calage seront tirées au hasard, en respectant un minimum de couverture spatiale.

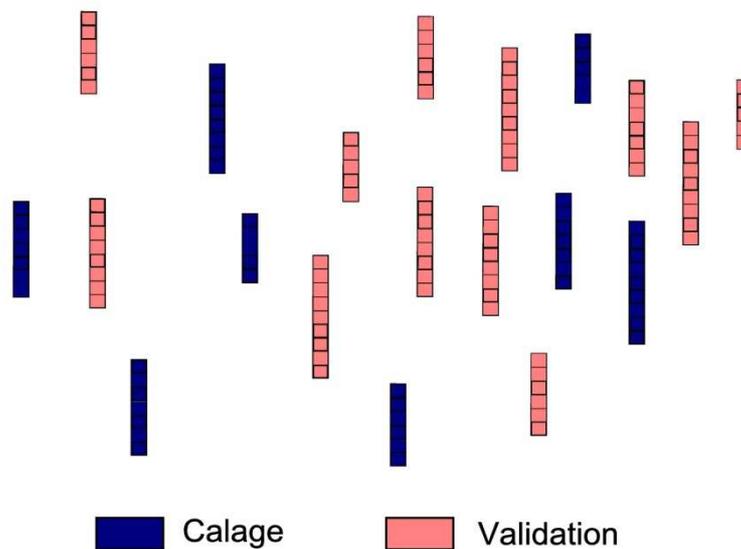


Figure 3 : illustration des jeux de données calage / validation

Pour la robustesse et la stabilité des intervalles de confiance, le découpage consiste simplement à choisir aléatoirement 50% des stations pour constituer le jeu C1, le reste pour le jeu C2 (Figure 4), en veillant à ce que la répartition spatiale de chaque jeu soit homogène.

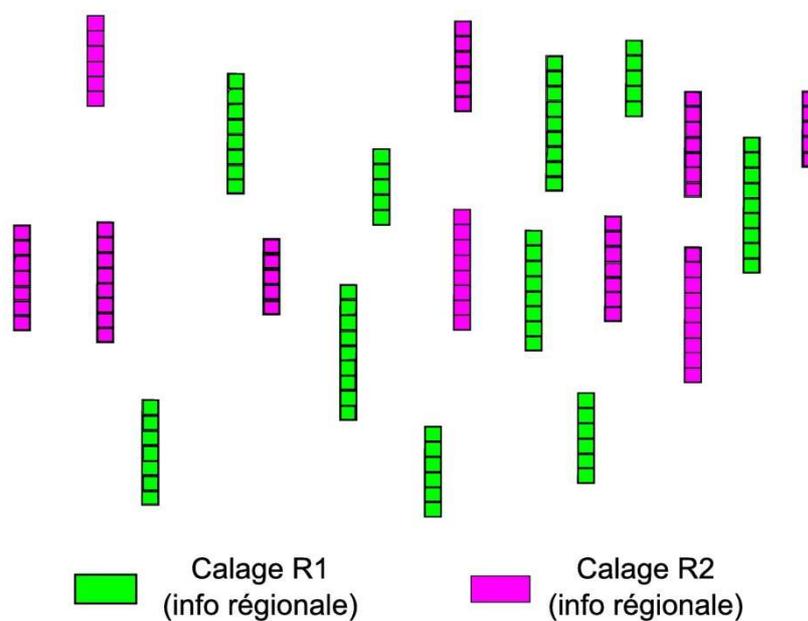


Figure 4. Décomposition C1/C2 pour les méthodes régionales.

Afin de mettre en évidence d'éventuelle influence climatique, les critères sont calculés en décomposant l'ensemble des stations en trois zones définies ci-dessous (Figure 5) :

- zone de montagnes : concerne toutes les stations de plus de 1000 m d'altitude
- zone méditerranéenne : délimitée arbitrairement en fonction de la cartographie des pluies journalières décennales
- zone continentale et océanique : les stations complémentaires aux deux autres zones.

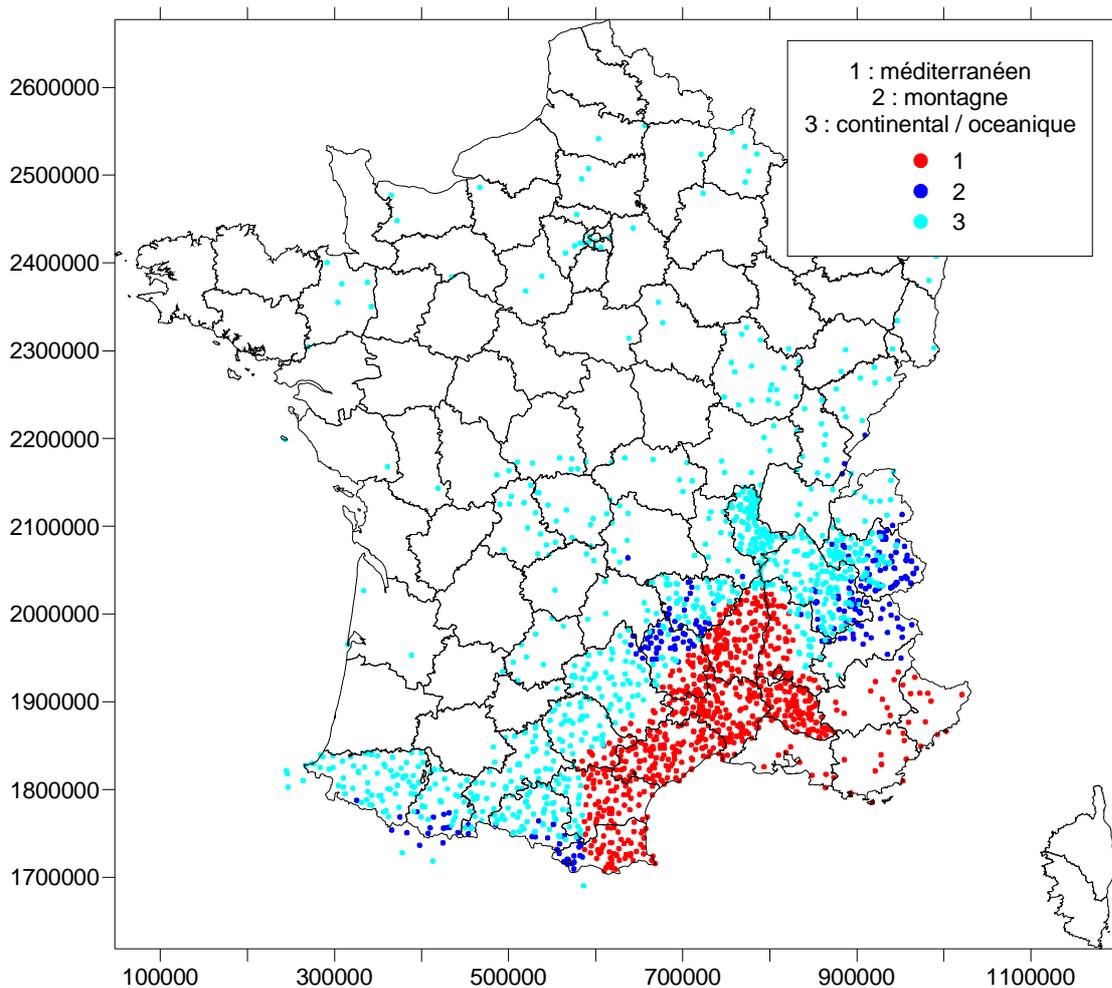


Figure 5 : classification des postes en trois zones climatiques.

3. Résultats de la comparaison

3.1. Comparaison à l'échelle de toute la zone d'étude

On constate que Shypre présente globalement les meilleures performances, même si la loi régionale et la fdr centrale de SIGEV ont des performances comparables pour la fiabilité dans la queue de distribution (test FF).

Shypre et la loi régionale ont des performances équivalentes en robustesse pour des périodes de retour de 10 ans. A partir de 100 ans, la robustesse de Shypre est meilleure que la loi régionale qui régresse.

Pour Shypre les performances en justesse et en robustesse semblent peu sensibles à l'échantillonnage C/V. La loi régionale et SIGEV affichent des meilleures performances sur le NT10 lorsque la densité spatiale de poste en calage augmente, ainsi que pour le SPAN pour la loi régionale uniquement. Le critère FF est peu sensible à l'échantillonnage pour la loi régionale.

On constate globalement des performances semblables entre distribution prédictive et centrale pour la loi régionale et SIGEV. Les performances de la loi régionale prédictive sont sensiblement améliorées pour FF et NT10 et pour SIGEV c'est la fdr centrale qui est un peu plus performante pour la fiabilité dans les extrêmes.

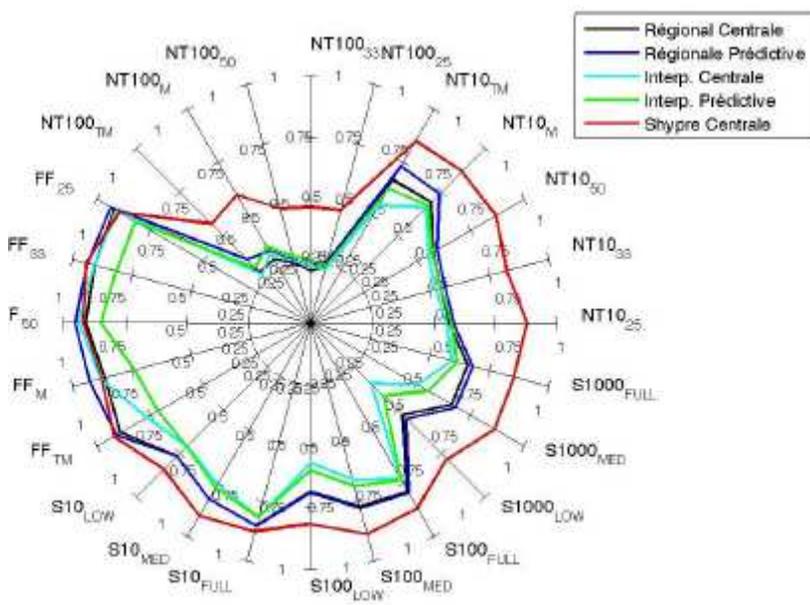


Figure 6 : Récapitulatif des critères pour la loi régionale centrale et prédictive, la méthode SIGEV centrale et prédictive et Shypre centrale. S est le critère SPAN pour densité faible (Low), moyenne (medium) et forte (full).

Analyse détaillé des critères FF en validation :

L'allure des courbes FF peut révéler des informations qui sont masquées si l'on ne considère que les aires sous la courbe. On compare dans un premier temps les critères FF correspondant aux FDR centrales des modèles loi régionale, SIGEV et Shypre. Les comportements de la loi régionale et de Shypre sont comparables : l'asymptote horizontale montre que les modèles tendent à sur estimer les fréquences des plus fortes valeurs observées, cela se produit pour des fréquences rares (>0.98 à 0.99). Pour la loi régionale et le jeu C/V TM, on observe un coefficient de forme de la loi régionale positif (domaine de Weibull, extrêmes bornées) pour toutes les stations où FF=1. Par contre toutes les stations avec un coefficient de forme positif n'ont pas nécessairement un critère FF=1 : si la valeur maxi observée à la station est inférieure à la borne de la loi, le modèle attribue une fréquence élevée mais pas de 1. Pour SIGEV cette surestimation des fréquences se produit pour des événements plus

courants (fréquences > 0.7) ce qui traduit une performance moindre pour SIGEV à assigner « la bonne » fréquence au plus fort événement observé dans une série.

La comparaison entre critère FF des modèles central et prédictif dépend des modèles. Pour SIGEV on observe une nette dégradation avec la fdr prédictive, qui tend à sur-estimer davantage les quantiles les plus forts. Pour la loi régionale, les fdr prédictive et centrale sont assez proches, si la fdr prédictive tend à surestimer les quantiles, la fdr centrale tend à les sous-estimer.

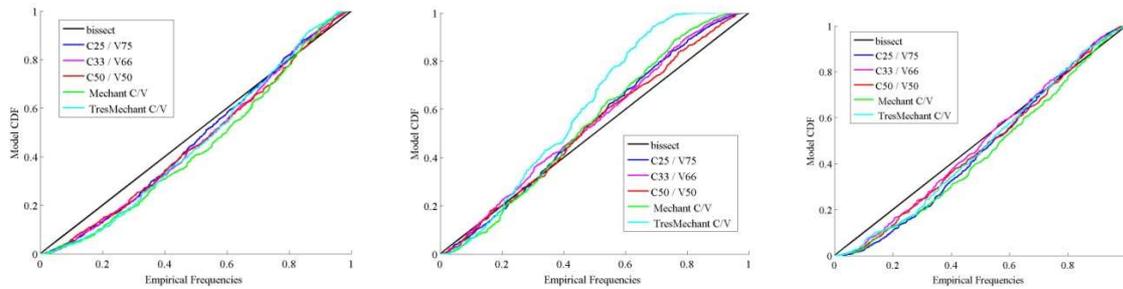


Figure 7: critère FF en validation pour les fdr centrale de la loi régionale, de SIGEV et Shypre.

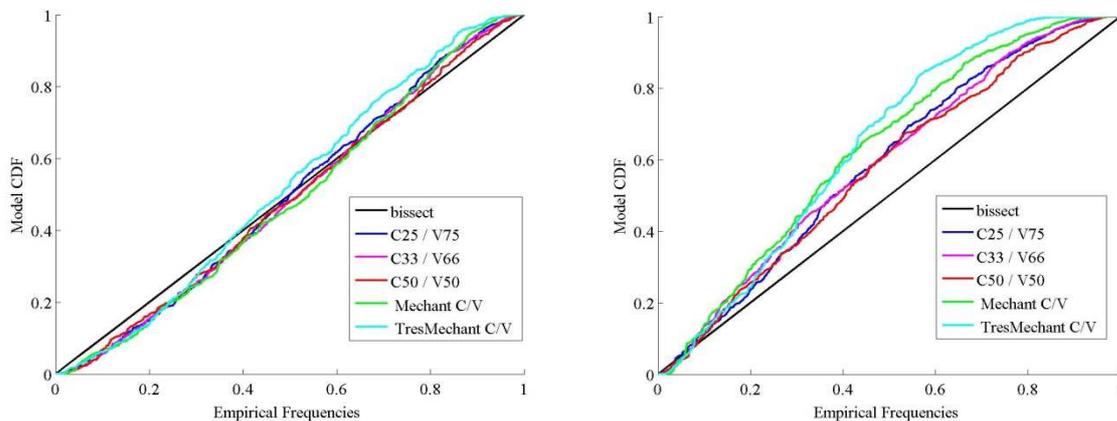


Figure 8 : critère FF en validation pour les fdr prédictives pour la loi régionale et SIGEV

La stabilité des intervalles de confiance est très nettement meilleure pour la loi régionale que pour SIGEV : pour le quantile centennal (Figure 9), pour 80% des stations on a un recouvrement des intervalles de confiance estimés sur deux périodes différentes qui excède 80% pour la loi régionale, alors qu'il n'y en a que 20% pour SIGEV avec un recouvrement supérieur à 80%. De plus pour la loi régionale les performances vis-à-vis de ce critère sont peu sensibles à l'échantillonnage alors qu'elles le sont plus pour SIGEV.

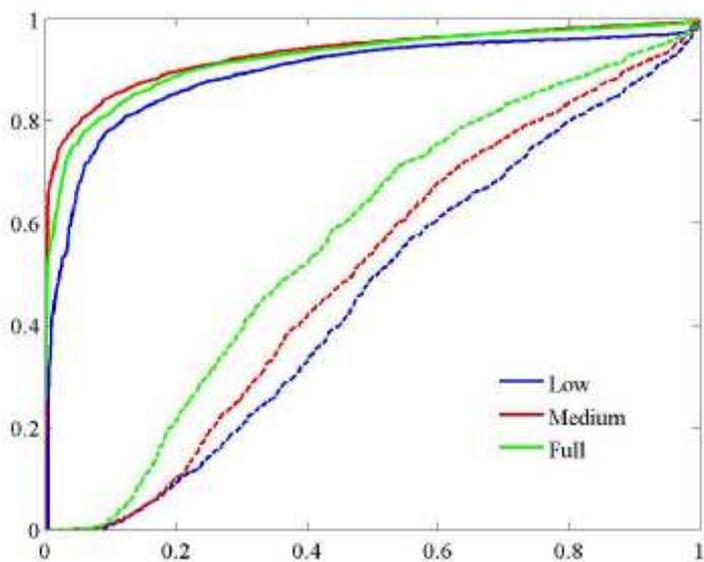


Figure 9 : critère COVER pour le quantile centennal pour la loi régionale et SIGEV.

3.2.Comparaison par zones climatiques

Les performances de loi régionale centrale sont similaires dans les 3 zones. On notera toutefois une légère baisse de performance sur le NT en zone méditerranéenne et une légère amélioration de la justesse des extrêmes dans la zone de montagne (Figure 10). Aucune différence n'est à souligner pour la loi régionale prédictive.

Les modèles SIGEV prédictif et central sont assez peu sensibles à la zone géographique. Seules la robustesse diminue en zone de montagne et la justesse (NT10) en zone méditerranéenne (Figure 11).

Comme les autres modèles, les critères de performance de Shypre varient très peu suivant la zone considérée. On souligne uniquement une légère baisse de performance globale en zone de montagne pour la justesse dans les extrêmes et pour la robustesse. La justesse (NT10) semble aussi légèrement supérieure en zone océanique et continentale par rapport aux deux autres zones (Figure 12).

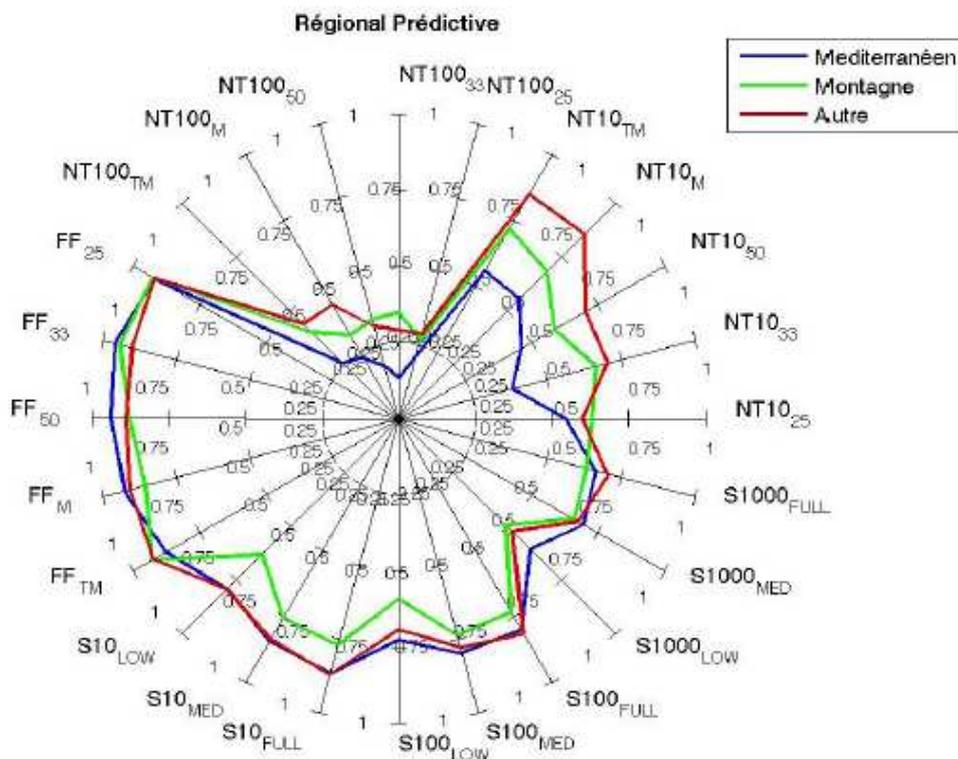
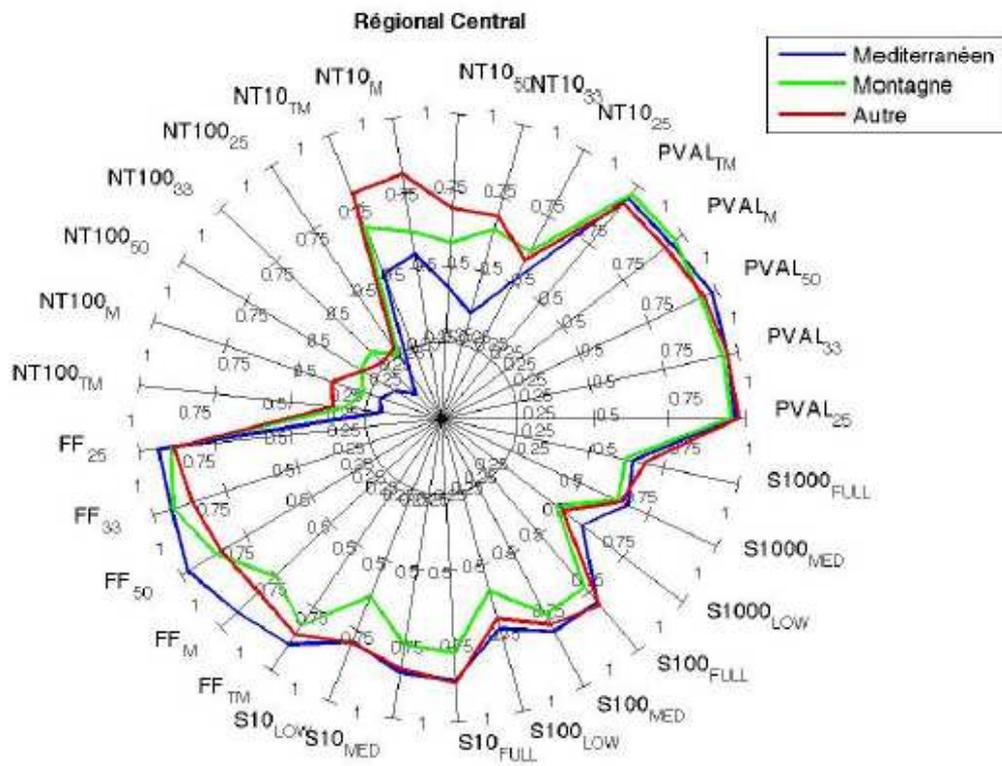


Figure 10 : récapitulatif des critères pour la loi régionale centrale (en haut) et prédictive (en bas) suivant les 3 zones climatiques.

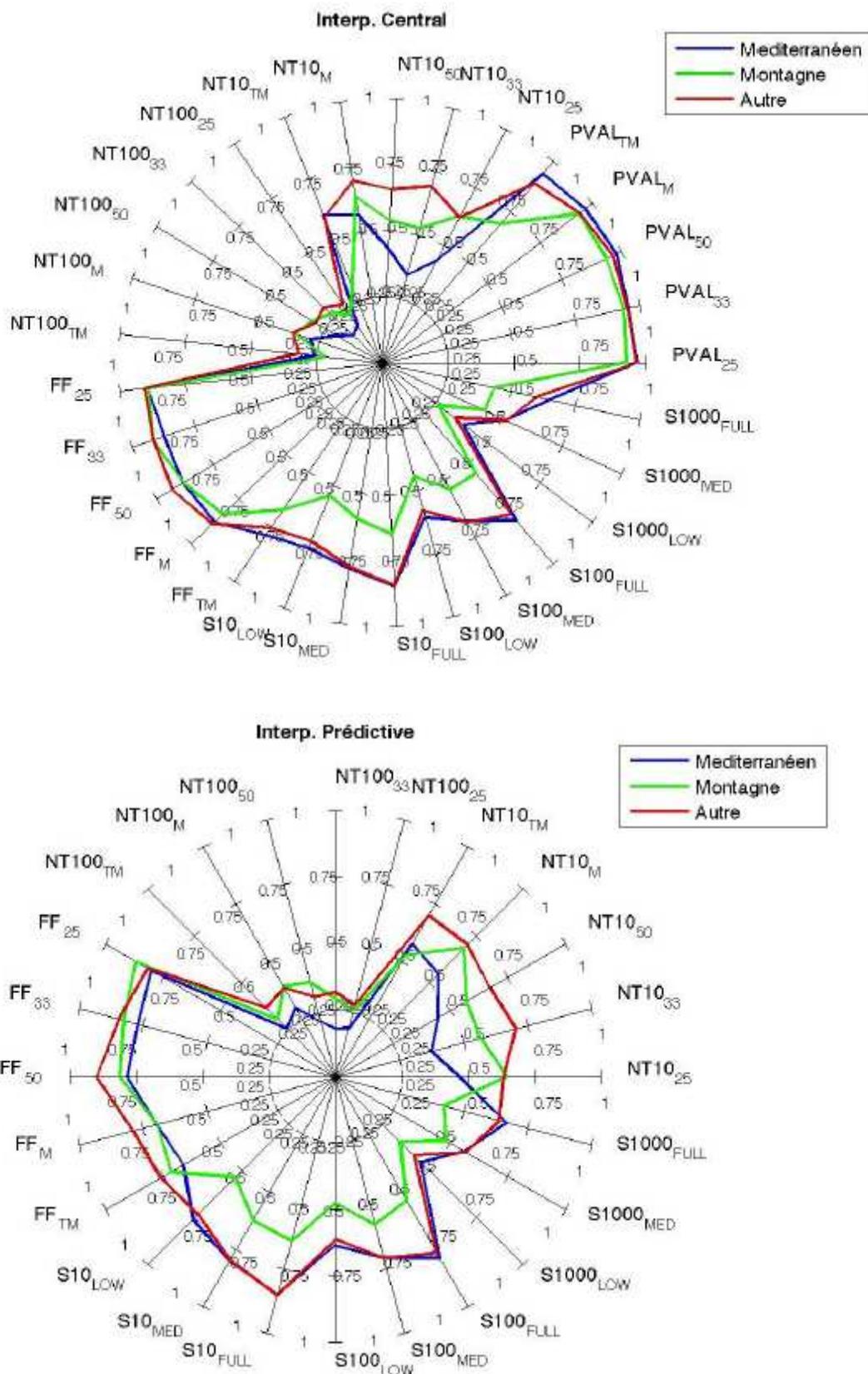


Figure 11 : récapitulatif des critères pour la méthode SIGEV centrale (en haut) et prédictive (en bas) suivant les 3 zones climatiques.

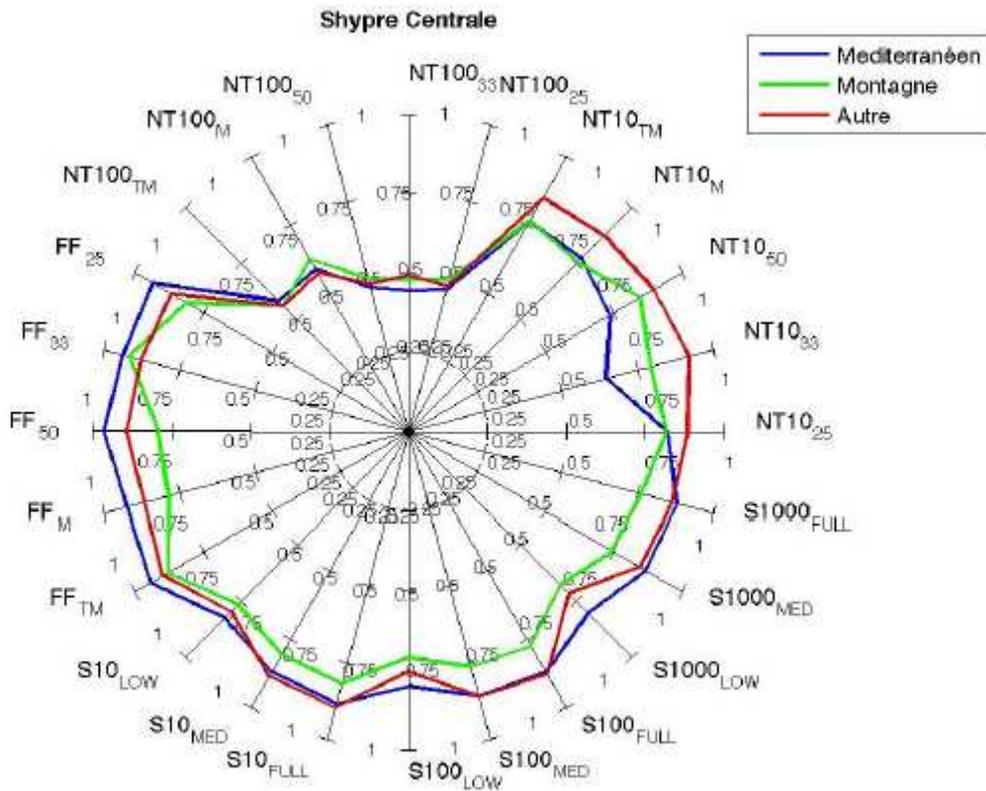


Figure 12 : récapitulatif des critères pour la méthode Shypre centrale suivant les 3 zones climatiques.

3.3. Hiérarchisation des modèles par zone climatique

La hiérarchisation des modèles n'est pas sensible à la zone géographique. Les conclusions de 3.1 restent valables pour toutes les zones. Les écarts entre performances des modèles sont cependant un peu plus marqués dans la zone méditerranéenne que dans les autres zones.

4. Cartographie des différences de quantiles Shypre – loi régionale

On représente les différences entre quantiles obtenus par la méthode Shypre et ceux de la loi régionale calculée sur tous les points de grilles du domaine. Pour plus de clarté on porte sur les Figure 13 et Figure 14 les points avec une différence absolue supérieure à 20 mm pour la pluie décennale et supérieure à 50 mm pour la pluie centennale. On a 6% des différences absolues qui dépassent 20 mm pour le quantile décennal et 3% qui dépassent 50 mm pour le quantile centennal. Pour ces deux quantiles on observe que les différences positives (les estimations Shypre sont les plus grandes) sont localisées sur les reliefs cévenols, des Pyrénées et des Alpes. Les différences négatives (les estimations de la loi régionale sont les plus grandes) sont localisées sur le pourtour méditerranéen et sur les piedmonts. La forme circulaire du voisinage de la loi régionale peut en être à l'origine : en plaine le voisinage inclut parfois des stations de montagnes ce qui tend à produire des quantiles plus forts et l'inverse se produit sur les stations des reliefs qui peuvent contenir dans leur voisinage des stations de plaines.

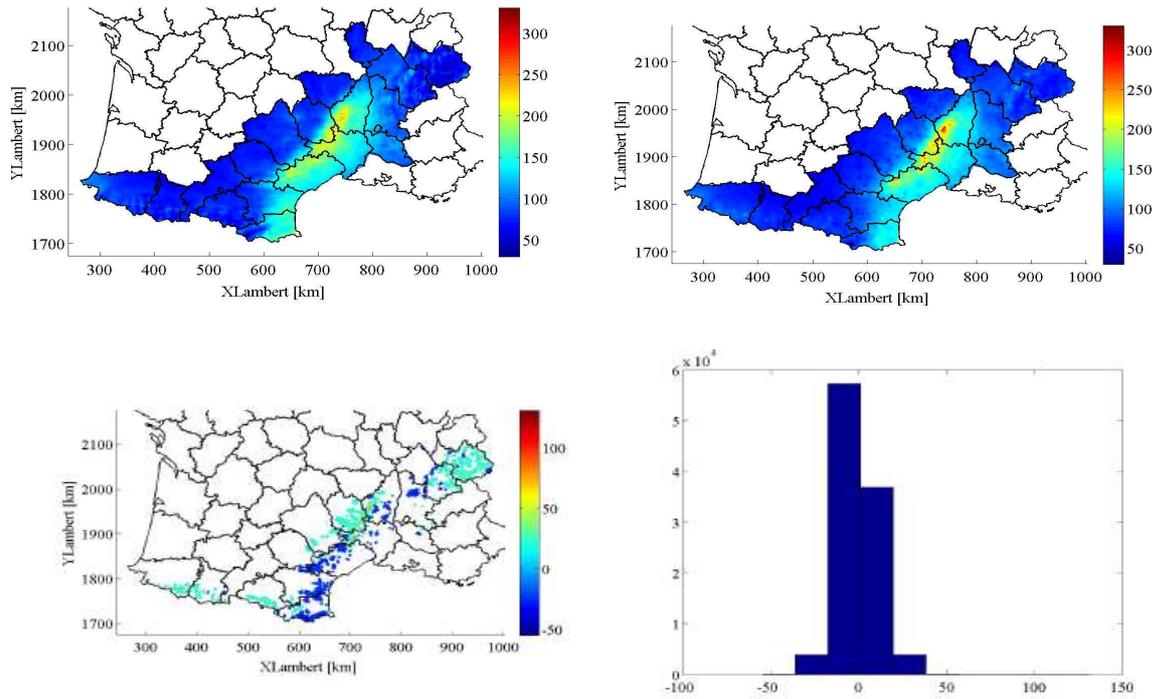


Figure 13 : Cartes du quantile décennal par la méthode régionale et par Shypré (respect. en haut à gauche et à droite), des différences absolues Shypré – loi régionale supérieure à 20 mm et histogramme des différences.

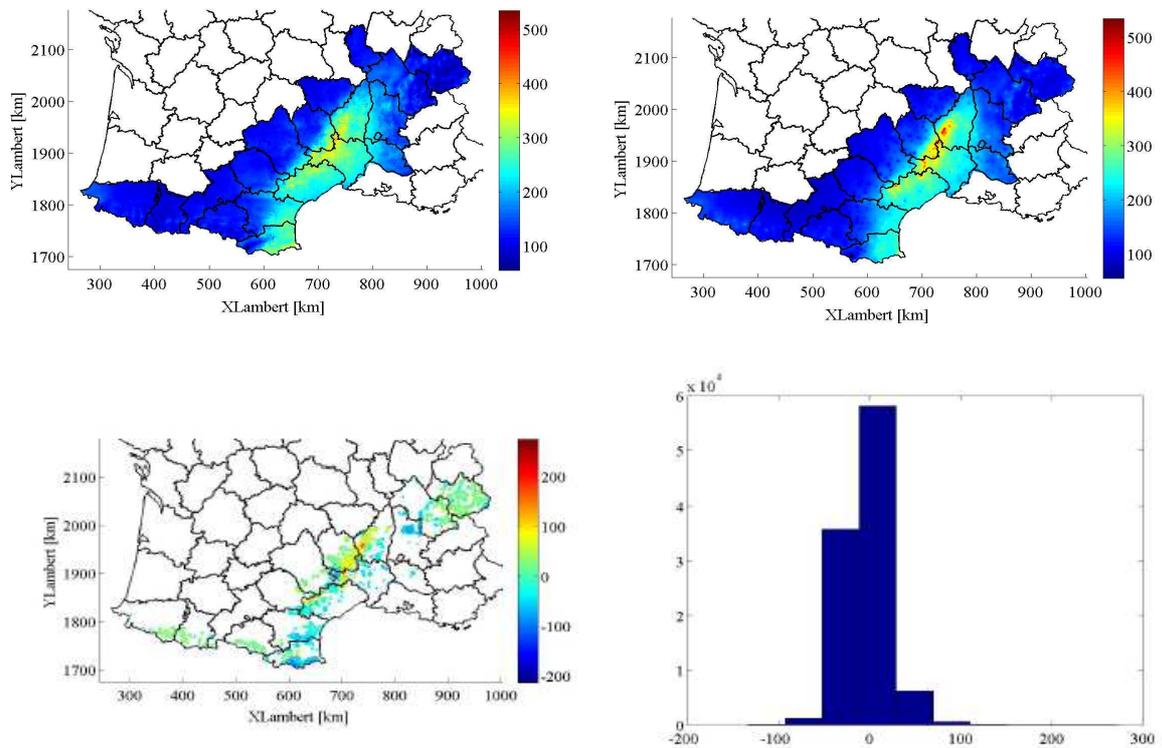


Figure 14 : Cartes du quantile centennial par la méthode régionale et par Shypré (respect. en haut à gauche et à droite), des différences absolues Shypré – loi régionale supérieure à 20 mm et histogramme des différences

5. Conclusions / Perspectives

Shypre présente globalement les meilleures performances en justesse et robustesse, il est peu sensible au découpage C/V et donc à l'échantillonnage. La loi régionale a des performances comparables pour la justesse dans les extrêmes et la robustesse pour des périodes de retour inférieures à 100 ans, ensuite la robustesse se dégrade plus vite que Shypre. Le modèle SIGEV présente de moins bonnes performances en robustesse et justesse dans les extrêmes et est le plus sensible à l'échantillonnage.

Les intervalles de confiances de la loi régionale sont plus stables que pour SIGEV. Shypre ne fournit pas d'intervalle de confiance.

Les différences entre distributions prédictives et centrales ne sont pas remarquables, et les performances des modèles sont peu sensibles aux zones climatiques.

6. Bibliographie

Arnaud P, Lavabre J., 2010. *Estimation de l'aléa pluvial en France Métropolitaine.* Guide Pratique, Edition Quae, 157 pages.

Benichou, P., Le Breton, O., 1987. Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *Météorologie* 7 (19), 23±34.

Bishop, M C, 2006. *Pattern recognition and machine learning.* Edition Springer.

Burn D., 1990. Evaluation of regional flood frequency analysis with a region of influence approach.. *Water Resources Research*, 26(10), p.2257–2265.

Hosking JRM, Wallis J.R, 1997. *Regional Frequency analysis.* Cambridge university press, 224p.

Lebel, T. & Laborde, J., 1988. A Geostatistical approach for areal rainfall statistics assessment.. *Stochastic Hydrology and Hydraulics*, 2(4), p.245-261.

Li, K-C, 1991. Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association*, vol 86, 414, p316-327.

Muller A., 2006. *Comportement asymptotique de la distribution des pluies extrêmes en France.* Thèse de l'Université Montpellier 2 ,246 pages.

Neppel, L., Desbordes, M. & Masson, J.M., 1997. Spatial extension of extreme rainfall events: return period of isohyets area and influence of rain gauges network evolution. *Atmospheric Research*, 45(3), p.183–199.

Sol,B., Desouches, C., 2005. *Spatialisation à résolution kilométrique sur la France de paramètres liés aux précipitations.* Rapport d'étude Météo-France, Convention Météo-France DPPR n°03/1735, 41 pages.

Viglione A, Laio F, Claps P, 2007. A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, vol 43 (3), DOI: 10.1029/2006WR005095.