



HAL
open science

Estimation régionale des débits décennaux et durées caractéristiques de crue en France

T. Cipriani, T. Toilliez, Eric Sauquet

► **To cite this version:**

T. Cipriani, T. Toilliez, Eric Sauquet. Estimation régionale des débits décennaux et durées caractéristiques de crue en France. *La Houille Blanche - Revue internationale de l'eau*, 2012, 4-5, p. 5 - p. 13. 10.1051/lhb/2012024 . hal-00773509

HAL Id: hal-00773509

<https://hal.science/hal-00773509>

Submitted on 14 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation régionale des débits décennaux et durées caractéristiques de crue en France

Thomas CIPRIANI¹, Tristan TOILLIEZ², Eric SAUQUET³

¹ Irstea, Lyon – e-mail : thomas.cipriani@irstea.fr

² Irstea, Lyon, désormais à HGM Environnement, Marseille – e-mail : t.hgm@orange.fr

³ Irstea, Lyon – e-mail : eric.sauquet@irstea.fr

RÉSUMÉ. – Appréhender le fonctionnement actuel des bassins versants français est un élément essentiel pour répondre aux exigences de la Directive Cadre européenne sur l'Eau. La connaissance du régime hydrologique – a minima au travers de descripteurs statistiques - en sites non jaugés est de fait nécessaire. Cet article présente les résultats de différentes techniques d'estimation de deux variables descriptives des crues : le débit de pointe maximal annuel de fréquence décennale QIXA10 et une durée caractéristique issue d'une sélection d'hydrogrammes tronqués de l'écoulement de base. Sont présentées la procédure de sélection du jeu de stations hydrométriques fiables, le mode d'extraction des échantillons de données et le calcul des variables aux sites instrumentés à spatialiser. Le point principalement abordé concerne l'établissement de formules empiriques pour l'estimation en bassins non jaugés et les incertitudes associées. Leurs performances sont mises en perspective par rapport à celles d'autres méthodes exploitant la proximité géographique (voisinage glissant et approche géostatistique).

Mots clefs : Prédétermination des crues, débit de pointe, durée caractéristique, formulations empiriques

Estimating 10 year return period peak flows and flood durations at ungauged locations in France

ABSTRACT. – Knowing the river flow regime along the French river network is required to answer basic questions related to the European Water Framework directive. The river flow regime should be at least estimated at ungauged locations through relevant statistics. A comparison of different approaches applied to estimate two flood regime descriptors is presented: the instantaneous peak discharges with 10 year return period and a flood duration derived from a selection of hydrographs removed from baseflow. This paper focuses on the framework developed to establish the reference dataset of gauging stations, the sampling technique applied to extract variables from an extended set of time series and the procedure to compute the two variables. Several empirical formulas at both national and regional levels are presented. An analysis of the predictive performance is performed for comparing the accuracy of the tested approaches including the empirical formulas, an approach using geographical neighbourhood and a geostatistical method.

Key words : Flood frequency analysis, peak flow, flood duration, regression analysis

I. INTRODUCTION

Les hydrologues français ont été et sont encore fortement imprégnés des méthodes d'estimation SOCOSE et CRUPEDIX [Ministère de l'Agriculture et al., 1980] visant à approcher la valeur d'une caractéristique statistique à partir de descripteurs des bassins versants. Ces méthodes établies à l'échelle nationale il y a près de 30 ans sont encore enseignées et font foi dans nombre d'études d'ingénierie. Chacune propose une estimation du débit de pointe maximal annuel décennal QIXA10 à partir de caractéristiques géomorphologiques (*a minima* la surface du bassin versant) et climatiques (*a minima* un quantile de pluie extrême). Elles n'ont pas depuis leur création été réactualisées, malgré les progrès dans l'acquisition et le traitement des données environnementales (notamment via la géomatique), la diffusion de nouvelles techniques d'interpolation dans la communauté hydrologique, (notamment le krigeage) et les progrès des moyens de calcul. Notons cependant des adaptations sur des zooms géographiques : sud-est de la France

dans [Faure *et al.*, 1991] et versant français des Pyrénées [Peteuil *et al.*, 2010].

Un exercice récent lié aux besoins de l'application de la Directive Cadre européenne sur l'Eau (DCE) a été réalisé pour des besoins de compréhension de l'évolution des milieux [Cipriani *et al.*, 2012]. L'objectif visé a été l'élaboration de cartographies de descripteurs du régime des crues, dont le QIXA10 et une durée caractéristique. Le point abordé principalement dans cet article concerne l'établissement de formules empiriques pour l'estimation en sites non jaugés et les incertitudes associées.

II. CARACTÉRISATION DU RÉGIME DES CRUES

II.1. Dynamique des crues

Plusieurs définitions de durée de crue sont possibles. Celle que nous avons retenue se prête à une procédure

automatique. Elle vise à appréhender la réponse du bassin aux pluies fortes et – plus spécifiquement – la dynamique de ruissellement. L'obtention de la durée se déroule en plusieurs étapes. Tout d'abord un pré-traitement pour soustraire l'écoulement de base à la chronique des débits bruts Q est réalisé. Les débits de base Q_b sont estimés selon l'algorithme développé par le Centre for Ecology and Hydrology décrit dans [Tallaksen et Van Lanen, 2004]. La durée est obtenue à partir d'une sélection d'hydrogrammes privés de la composante de base selon [Sauquet *et al.*, 2008]. Pour chaque station, les épisodes autour de chaque maximum annuel de la chronique $Q-Q_b$ sont extraits, puis rendus adimensionnels en divisant chacun d'entre eux par le débit de pointe. Tous les hydrogrammes normés ont donc un débit maximal égal à 1. La durée choisie pour caractériser le ruissellement θ_R est la médiane des durées de dépassement de 0,5 des hydrogrammes échantillonnés et l'écart-type σ renseigne la dispersion des valeurs. A partir de cet échantillon de crues normées, nous construisons également un hydrogramme supposé représentatif du régime de crue. Pour cela, les hydrogrammes sont discrétisés à un pas de temps régulier (l'heure) et centrés sur l'instant du pic. Un hydrogramme représentatif est formé des valeurs médianes des débits normés de chaque pas de temps (Figure 1). Sur cet exemple la durée θ_R est égale à 19 heures. Il faut souligner que la durée d'un événement varie fortement d'un épisode à un autre (liée à la dynamique des précipitations, à l'état hydrique des sols) : le rapport σ/θ_R pour la station de la Figure 1 est de 0,63. Une option aurait été de calculer la durée de crue sur un sous-échantillon d'événements conditionnés par le dépassement d'un débit de référence à l'instar de la durée de SOCOSE. Néanmoins le faible nombre d'événements aurait rendu la valeur obtenue encore plus fortement tributaire à l'échantillon et cette définition se prête moins à une mise en œuvre automatique.

II.2. Intensité des crues

Les échantillons de débits de pointe sont obtenus par extraction des valeurs maximales d'un ensemble d'événements ayant dépassé un seuil donné [Lang et Lavabre, 2008]. L'échantillonnage sup-seuil permet de rassembler un plus grand nombre d'événements sélectionnés que l'échantillonnage par maxima annuels, et si le seuil est suffisamment élevé, seuls les épisodes majeurs seront capturés. L'extraction des valeurs selon l'approche sup-seuil est délicate en une station. L'échantillon doit satisfaire

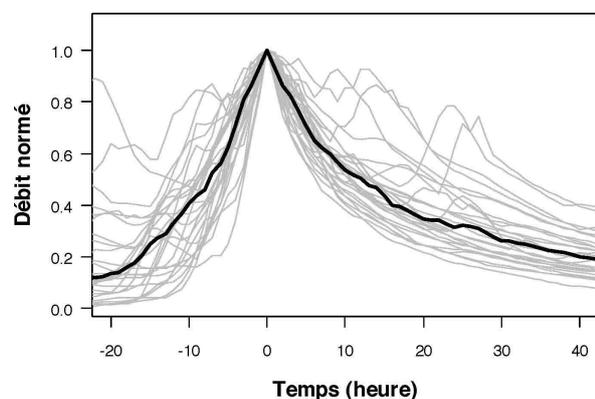


Figure 1 : Hydrogrammes normés (en gris) et hydrogramme médian (en noir) extraits de la série de débits privés de la composante de base de la Savoureuse à Giromagny

entre autres, la condition « *i.i.d.* », *i.e.* formé de valeurs indépendantes et identiquement distribuées. L'extraction repose bien souvent sur une expertise hydrologique et statistique qu'il est impossible de mettre en œuvre tel quel sur un ensemble conséquent de chroniques. Il s'agit donc de s'approcher au mieux des résultats attendus en limitant la place de l'expertise.

La procédure usuelle décrite par [Lang *et al.*, 1999] nécessite trois paramètres propres à la chronique : le nombre moyen d'événement par an μ , le paramètre de redescende α (le débit entre deux pointes successives doit être redescendu sous le seuil et sous $\alpha\%$ de la valeur maximale de l'épisode précédent) et D^* (durée minimale entre deux valeurs échantillonnées). La difficulté majeure est d'attribuer des valeurs pertinentes à ces trois paramètres. Il n'est en effet pas judicieux d'imposer une contrainte temporelle « laxiste » - une heure, par exemple - pour un bassin versant de temps de réponse de l'ordre de plusieurs jours. De même trop contraindre le critère de redescende perturbe la sélection des événements lorsque le régime présente une forte composante souterraine ou nivale. Les options suivantes ont été finalement choisies. Le nombre moyen d'événements par an μ pour chaque station est ici égal à 2 sauf cas particuliers, ce qui garantit un seuil élevé. Nous avons fixé D^* à θ_R préalablement défini à partir d'un échantillonnage des maxima annuels. Le coefficient de redescende α est de fait conditionné par la contribution de l'écoulement de base dans les épisodes de crue. Ce paramètre est fixé en premier lieu par le minimum des hydrogrammes adimensionnels médians extraits par maxima annuels de la série des débits bruts (en noir sur la Figure 2 pour un exemple de bassins à forte composante souterraine). Le test d'adéquation du nombre d'épisodes à la loi de Poisson [Cunnane, 1979] oriente la valeur numérique définitive pour α , les autres paramètres μ et θ_R n'étant pas modifiés.

L'étape après extraction des valeurs de pointe consiste à ajuster une loi, ici suggérée par la Théorie des Valeurs Extrêmes et adaptée à l'échantillonnage sup-seuil : la loi exponentielle ou la loi de Pareto Généralisée. Cette dernière offre plus de souplesse grâce à un troisième paramètre, cependant entaché de trop fortes incertitudes. Les deux lois proposent des valeurs similaires sur la gamme des crues fréquentes. Afin de manipuler des estimations robustes, nous avons décidé d'étudier exclusivement les quantiles décennaux ajustés par la loi exponentielle. Notons que si la même étude devait être réalisée pour la centennale, alors le choix de la distribution deviendrait un point central.

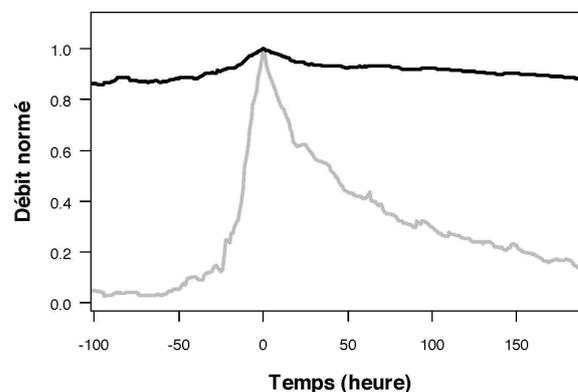


Figure 2 : Hydrogrammes de crue normés médians obtenus sur les débits (en noir) et sur les débits privés de la composante de base (en gris) de la Barbuise à Pouan-les-Vallées

III. CONSTITUTION DU JEU DE STATIONS DE RÉFÉRENCE

Les séries de mesure sélectionnées sont issues de la banque HYDRO (<http://hydro.eaufrance.fr>) et du réseau de mesure de DTG Grenoble d'EDF. La démarche de sélection n'a pas été linéaire : des aller-retours entre recherche de méta-données et critique des données ont été opérés pour isoler les stations atypiques et les origines de ces comportements.

III.1. Sélection selon la disponibilité des données et le caractère influencé du régime

Un premier ensemble formé des 1454 stations disponibles au pas de temps variable en France métropolitaine a été constitué. Certaines stations ont été exclues du jeu de données de référence si elles ne répondaient pas à au moins un des critères suivants :

- présenter un fonctionnement hydrologique naturel ou peu influencé. Sur la base des informations fournies par les gestionnaires, sont écartées les stations sans signification hydrologique (21 stations) ou fortement influencées (70 stations) ;
- mesurer les Hautes Eaux de manière fiable. 139 stations aux enregistrements de qualité « douteuse » en hautes eaux ont été exclues ;
- posséder une chronique de longueur effective supérieure à 16 ans afin de garantir la robustesse des estimations des caractéristiques de crue. 346 stations sont éliminées sur ce critère.

Les deux premiers critères se sont fondés principalement sur les commentaires fournis par les gestionnaires. Il reste à ce stade 878 stations.

III.2. Sélection selon la stationnarité des chroniques de maxima annuels

Nous avons mis en œuvre des tests statistiques afin de déceler des stations au comportement suspect. L'étude se base sur les tests de Pettitt et de Mann-Kendall appliqués aux séries de maxima annuels. L'hypothèse est faite que l'existence de tendances (Mann-Kendall) ou de ruptures (Pettitt) dans les séries peut provenir de biais météorologiques ou de la présence d'aménagements hydrauliques, non mentionnés par les gestionnaires. Ces tests doivent être considérés comme des outils d'aide à la décision : l'hydrologue rejettera ou non la série vraisemblablement non stationnaire sur la base des conclusions des tests et de son expertise.

Les stations acceptées par ces tests sont de fait conservées. Quant aux autres, nous avons recherché des informations complémentaires sur la station et son bassin versant et avons effectué une comparaison entre sites proches pour vérifier si les anomalies pouvaient être dues à un échantillonnage temporel malheureux. Pour certaines stations, nous avons pu relier la présence d'une discontinuité à un événement précis (changement dans la gestion de la station, implantation d'ouvrages, etc.) ; il a alors été décidé de supprimer une partie de l'enregistrement pour ne conserver que les données exploitables. Sur les 109 stations rejetées par au moins un des tests, nous décidons d'en conserver 50.

III.3. Sélection selon les L-Moments

L'objectif est d'identifier les formes des distributions statistiques très particulières pour ensuite conclure sur le rejet

des stations associées après expertise. Cette recherche se fonde sur les valeurs des L-moments [Hosking et Wallis, 1997]. Ces statistiques très simples à calculer sont couramment utilisées en Grande-Bretagne en analyse régionale pour guider le choix de la loi statistique à ajuster : l'échantillon est caractérisé par les quatre premiers L-Moments qui définissent un espace à trois dimensions. Les échantillons constitués sur un jeu de stations sont projetés dans cet espace et il est possible de repérer les « outliers » dans cet espace sur la base d'une distance au centre de gravité du nuage. Ces échantillons ont fait l'objet d'une dernière et ultime expertise. Sur les 51 stations suspectes, 31 ont été définitivement exclues.

IV. ANALYSE SPATIALE

Au terme de la procédure de sélection, 788 stations ont été retenues, jaugeant des bassins de 3 à 2240 km². La répartition du quantile décennal des débits observés, ramenés en valeurs spécifiques afin d'ôter l'effet de surface, fait apparaître les grands contrastes à l'échelle du territoire (Figure 3). La coexistence de régions plus ou moins productives s'explique notamment par les différences du régime des pluies, l'hydrogéologie, le relief, l'occupation des sols, etc. Les fortes valeurs sur le massif des Cévennes, l'ouest des Pyrénées, la Côte d'Azur et le sud des Vosges peuvent ainsi s'expliquer par la pluviométrie. En revanche la nature du sol peut expliquer les différences entre la Bretagne (aquifères discontinus, massifs cristallins imperméables) et le Bassin Parisien (grands aquifères de la Seine et de la Somme, couverture perméable).

La Figure 4 propose une répartition de la durée caractéristique ramenée à une unité de surface. Le choix de diviser par $S^{0.27}$ est expliqué plus loin (IV.2). Il permet de s'affranchir de l'effet taille du bassin versant dans l'identification des facteurs de variabilité. On note une dynamique de crue plus vive sur le pourtour méditerranéen, les Alpes du Sud, la Corse et les Pyrénées. A l'inverse les régions du Bassin Aquitain, du Massif Armoricaire, du Massif Central et des Vosges sont nettement moins réactives. Ces écarts de comportement observés entre les Vosges et les Alpes du Sud, malgré une morphologie commune, pourraient s'expliquer par la nature du sol et l'effet plus prégnant de la neige dans les Alpes. La Figure 5 renseigne la dispersion de la durée

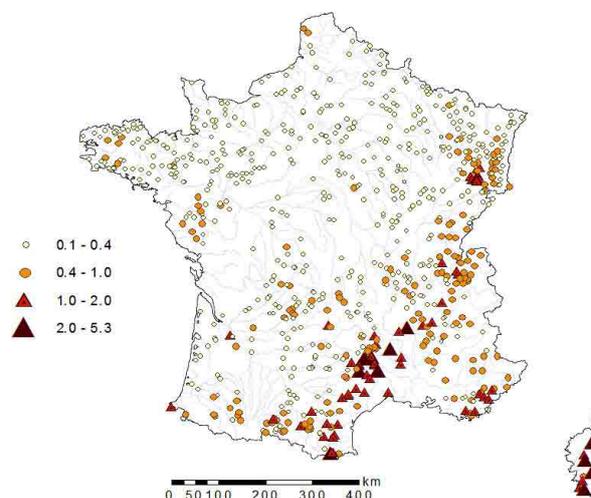


Figure 3 : Valeurs de QIXA10 spécifique aux stations de référence (en m³/s/km²)

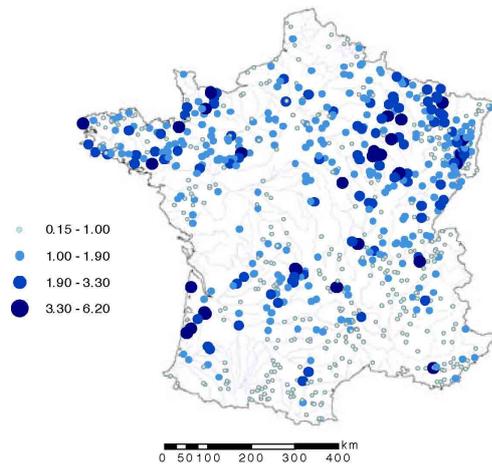


Figure 4 : Valeurs de $\theta_R / (7,45S^{0,27})$ aux stations de référence (θ_R en h, S en km^2)

caractéristique. Bien souvent σ/θ_R dépasse 1 (170 stations sur les 788 de l'échantillon), voire 2 (46 stations sur 788). Par conséquent la représentativité des valeurs θ_R peut être sujette à caution dans certains secteurs (en particulier dans le sud-est de la France).

V. CONSTRUCTION DE FORMULES EMPIRIQUES

Les formules ont été établies à partir d'un ensemble de variables candidates sélectionnées pour leur lien avec les débits et le régime des hautes eaux. La surface drainée S est une variable incontournable. En complément, il est nécessaire d'inclure une variable descriptive liée aux pluies extrêmes. Les données dont nous disposons sont les pluies de 24 heures décennales $P24hXA10$ interpolées à l'aide de la méthode AURELHY sur une grille à résolution kilométrique [Benichou et Lebreton, 1987]. Nous avons déduit de $P24hXA10$ une variable représentative du bassin donnée par la moyenne spatiale des $P24hXA10$ le long du réseau hydrographique via un plan de drainage kilométrique [Sauquet, 2006]. Soulignons que les valeurs obtenues $P24hXA10$ (en mm) ne fournissent pas de quantiles de la pluie de bassin et qu'il n'y a pas forcément recouvrement exact entre la saison des pluies fortes et la période des crues. En outre, rappelons que la pluie journalière décennale ne correspond pas à un débit de pointe de crue décennal. Cet indicateur élaboré à partir d'un quantile décennal doit donc être pris comme un descripteur des pluies fortes. Enfin, nous avons considéré deux autres descripteurs : l'un renseignant la géologie, *i.e.* la densité de drainage DDr et l'autre les processus liés à la neige, *i.e.* l'altitude moyenne du bassin versant $Zmoy$.

Les formules définies sont de type puissance, obtenues par régression linéaire des variables transformées par le logarithme népérien et ajustées selon le critère des moindres carrés. Dans ces formules, les débits sont exprimés en m^3/s et les surfaces (S) en km^2 . Les coefficients de détermination (r^2) et les écarts-types de l'erreur d'estimation (σ_{in}) sont calculés sur les variables transformées par le logarithme népérien. Nous présenterons les intervalles de confiance à 70%, c'est-à-dire qu'il y a une probabilité de 70% pour que la « vraie » valeur soit contenue dans cet intervalle ; les bornes de ces intervalles sous hypothèse de normalité des résidus sont des fonctions de σ_{in} et de l'estimation \hat{Q} issue de

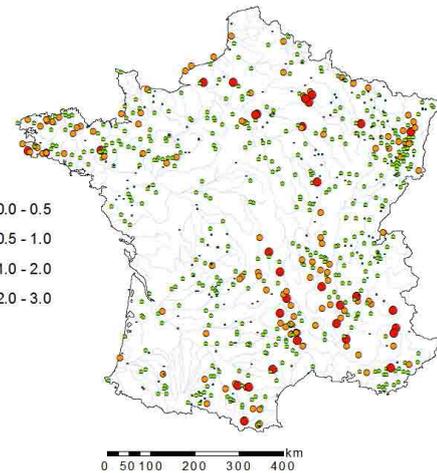


Figure 5 : Valeurs du coefficient σ/θ_R aux stations de référence

la formulation empirique. L'efficacité de Nash quant à elle est indiquée sur les variables non transformées. Ainsi, plus une méthode sera performante, plus le critère de Nash et le coefficient de détermination s'approcheront de 1, les écarts-types de zéro, et plus l'intervalle de confiance sera centré resserré autour de \hat{Q} .

Les formules ont été testées en validation croisée (VC), seule méthode objective permettant de juger de la performance de ces dernières. Pour cela nous éliminons tour à tour une station du jeu de données, nous construisons la régression sans elle à l'aide des variables explicatives, nous estimons la valeur en cette station et nous la comparons à l'observation.

V.1. Détermination du débit instantané de fréquence décennale QIXA10

Une première formule empirique sur l'ensemble du territoire a été établie. Les résultats obtenus par cette méthode pourront servir de base de comparaison :

$$\hat{Q} = 1,10 \cdot 10^{-4} \cdot S^{0,803} \cdot \overline{P24hXA10}^{2,097} \quad (1)$$

L'examen de (1) montre de fortes similitudes avec la formule de CRUPEDIX. Cette dernière estime le $QIXA10$ à l'aide de $PJXA10$, moyenne sur le bassin versant des pluies ponctuelles journalières de fréquence décennale, et d'un coefficient régional (R) ajusté sur un ensemble de 630 bassins versants :

$$\hat{Q} = S^{0,8} \cdot \left(\frac{PJXA10}{80} \right)^2 \cdot R \quad (2)$$

Les exposants de (2) sont proche de (1) même si les variables de pluie ne sont pas strictement les mêmes ($P24hXA10 \geq PJXA10$). A l'échelle d'un grand territoire hétérogène d'un point de vue climatique et géologique, comme la France, la formulation nationale (1) ne semble pas adaptée à tout contexte comme en témoigne la dispersion accrue du nuage sur la gamme de débit inférieur à $100 \text{ m}^3/\text{s}$ (Figure 6). Nous avons établi des relations empiriques à l'échelle d'entités géographiques en testant un découpage par grands secteurs hydrographiques (ZH) définis par la banque HYDRO et par hydro-écorégions de niveau 1 (HER). Ces HERs ont été construites par identification de grandes

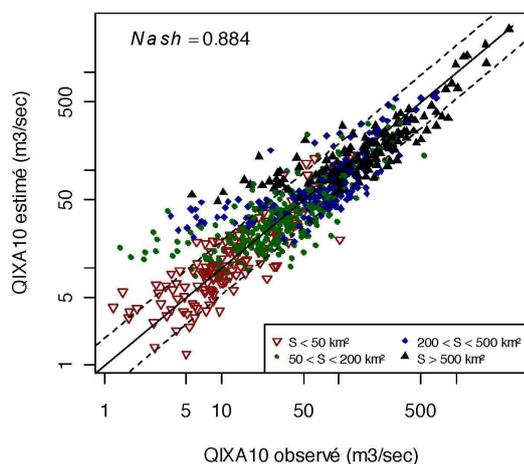


Figure 6 : Débits observés et estimés selon la formule nationale

structures délimitées par un changement important, d'ordre géophysique ou climatique [Wasson *et al.*, 2002]. Nous avons été amenés à en modifier légèrement les contours afin d'assurer un nombre minimum de stations au sein de chaque HER (Figure 7).

Les résultats obtenus sur les ZHs et les HERs ont montré que les HERs conduisent à de meilleures estimations en validation croisée. C'est ce découpage qui est conservé et les résultats associés sont présentés dans le Tableau 1. Ce dernier reprend les formules de plusieurs secteurs représentatifs de la qualité des estimations. Les résultats les moins satisfaisants sont enregistrés sur des régions peu instrumentées (Figure 8 : HER 16 « Corse », HER 18 « Grands Causses », et HER 19 « Dépôts Argilo-Sableux ») ou certainement hétérogènes car trop étendues (cf. HER 09 « Tables Calcaires »). Les coefficients associés à la surface varient peu d'une région à l'autre, à la différence de ceux associés à $P_{24hXA10}$. Le coefficient très élevé (5,03) amplifie les faibles contrastes pluviométriques observables sur la HER 09 « Tables Calcaires ». La Figure 8 permet de noter le gain de précision sur la dispersion générale ainsi que sur l'estimation des faibles valeurs de débits par rapport à la Figure 6.

Nous avons ensuite testé le pouvoir explicatif des deux autres descripteurs DDr (en m/km^2) et $Zmoy$ (en m). Le choix d'inclure définitivement une variable ou les deux est conditionné par l'amélioration du critère de Nash en validation croisée (VC). De nouvelles formulations incluant ces descripteurs émergent sur des régions de fortes variabilités climatique, géologique et morphologique (Tableau 2).

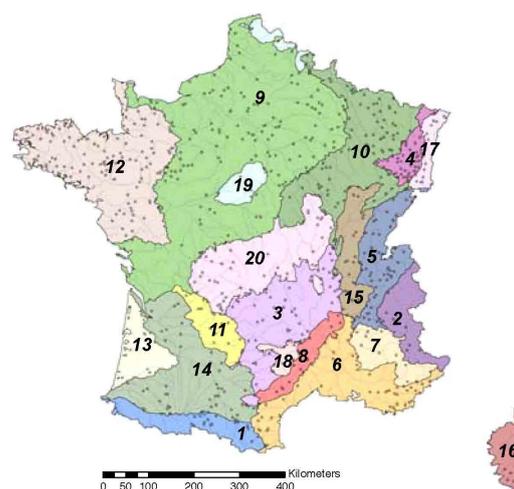


Figure 7 : Carte des hydro-écorégions de niveau 1 d'après [Wasson *et al.*, 2002], modifiées afin de limiter les régions contenant peu de stations

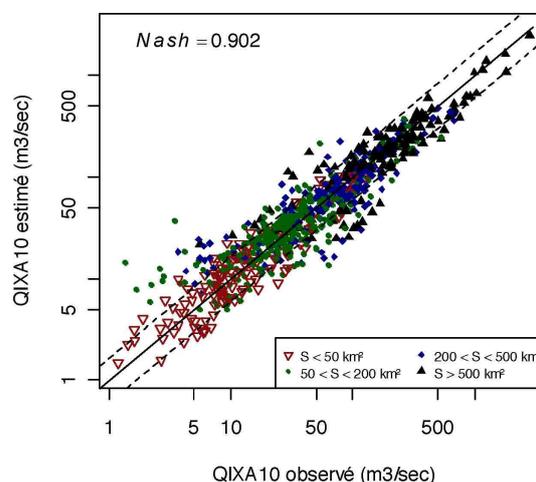


Figure 8 : Débits observés et estimés selon HER ($S, P_{24hXA10}$)

Aucune plus-value n'est en revanche décelable sur le critère de Nash sur les secteurs méditerranéen et armoricain.

Globalement retenons, sur la base du Tableau 1 et du Tableau 2, que les intervalles de confiance sont du même ordre de grandeur que ceux obtenus avec la formule CRUPEDIX et qu'une légère amélioration de la précision est obtenue en introduisant quatre variables explicatives

Tableau 1 : Formules empiriques ajustées sur les HERs de niveau 1 avec deux descripteurs et statistiques de performance associées (n : nombre de stations)

Hydro-écorégion (HER)	n	Formule HER ($S, \overline{P_{24hXA10}}$)	r^2 (VC)	Nash (VC)	I.C. 70%
01 – « Pyrénées »	32	$5,47 \cdot 10^{-2} \cdot S^{0,72} \cdot \overline{P_{24hXA10}}^{-0,86}$	0,73	0,60	$[0,64 \hat{Q} ; 1,56 \hat{Q}]$
02 – « Alpes internes »	9	$7,76 \cdot 10^{-3} \cdot S^{0,75} \cdot \overline{P_{24hXA10}}^{-1,17}$	0,85	0,82	$[0,68 \hat{Q} ; 1,47 \hat{Q}]$
06 – « Méditerranée »	30	$7,17 \cdot 10^{-6} \cdot S^{0,72} \cdot \overline{P_{24hXA10}}^{-2,72}$	0,70	0,72	$[0,58 \hat{Q} ; 1,73 \hat{Q}]$
09 – « Tables calcaires »	126	$1,28 \cdot 10^{-9} \cdot S^{0,65} \cdot \overline{P_{24hXA10}}^{-5,03}$	0,45	0,36	$[0,44 \hat{Q} ; 2,28 \hat{Q}]$
12 – « Massif Armoricain »	123	$1,64 \cdot 10^{-3} \cdot S^{0,89} \cdot \overline{P_{24hXA10}}^{-1,38}$	0,84	0,80	$[0,65 \hat{Q} ; 1,55 \hat{Q}]$
20 – « Massif Central Nord »	54	$2,93 \cdot 10^{-4} \cdot S^{0,94} \cdot \overline{P_{24hXA10}}^{-1,71}$	0,89	0,89	$[0,71 \hat{Q} ; 1,40 \hat{Q}]$

Tableau 2 : Formules empiriques ajustées sur les HERs de niveau 1 avec au maximum quatre descripteurs et statistiques de performance associées

HER	Formule HER (4 Var)	r^2 (VC)	Nash (VC)	I.C. 70%
01	$3,2 \cdot S^{0,84} \cdot \overline{P24hXA10}^{0,84} \cdot Zmoy^{-0,68}$	0,83	0,67	[0,71 \hat{Q} ; 1,41 \hat{Q}]
02	$6,8 \cdot 10^{-6} \cdot S^{0,69} \cdot \overline{P24hXA10}^{0,56} \cdot Zmoy^{1,31}$	0,93	0,82	[0,84 \hat{Q} ; 1,19 \hat{Q}]
06	$7,2 \cdot 10^{-6} \cdot S^{0,72} \cdot \overline{P24hXA10}^{2,72}$	0,70	0,72	[0,58 \hat{Q} ; 1,73 \hat{Q}]
09	$1,4 \cdot 10^{-11} \cdot S^{0,66} \cdot \overline{P24hXA10}^{4,41} \cdot DDr^{0,92} \cdot Zmoy^{0,31}$	0,75	0,66	[0,58 \hat{Q} ; 1,73 \hat{Q}]
12	$1,64 \cdot 10^{-3} \cdot S^{0,89} \cdot \overline{P24hXA10}^{1,38}$	0,84	0,80	[0,65 \hat{Q} ; 1,55 \hat{Q}]
20	$1,4 \cdot 10^{-4} \cdot S^{0,88} \cdot \overline{P24hXA10}^{1,76} \cdot DDr^{0,57} \cdot Zmoy^{-0,50}$	0,92	0,92	[0,76 \hat{Q} ; 1,32 \hat{Q}]

Si dans l'ensemble les performances sont supérieures (Figure 9), l'inconvénient de cette approche optimisée avec quatre variables est la totale liberté d'ajustement des coefficients pour la construction de la régression. Or, étant déterminés de manière à obtenir les meilleures performances statistiques, ils peuvent parfois paraître physiquement aberrants. Ainsi dans les Pyrénées ou le Massif Central, à superficie et pluie égales, notre formule abaisse la productivité des bassins d'altitude car le paramètre d'altitude moyenne ($Zmoy$) a un exposant négatif. Nous observons la situation inverse au niveau des Alpes. En effet dans ces deux cas, l'élément $Zmoy$ peut être porteur de significations différentes, en lien avec le régime hydrologique pour le premier et la pente pour le second. Une information intéressante pour vérifier ces hypothèses se base sur la saisonnalité des crues. D'après nos sites de mesure dans les Pyrénées, les plus fortes crues se produisent en majorité sur une période allant de janvier à avril. En cette saison, les bassins d'altitude élevée sont moins productifs que ceux situés en contrebas, vu que les précipitations peuvent encore être stockées sous forme de neige. Mais dans les Alpes, la saisonnalité s'étale de juin à septembre. Une saison où les précipitations ne sont pas sous forme neigeuse, et peuvent de plus ruisseler sur une morphologie plus acérée en altitude (*i.e.* sur de fortes pentes). L'altitude moyenne est sans doute un paramètre au rôle ambigu. Le fait d'avoir mêlé dans le traitement statistique crues de fonte nivale et crues pluviales peut en être à l'origine. Une solution aurait pu être de saisonnaliser les échantillons et de nous concentrer que sur les crues pluviales.

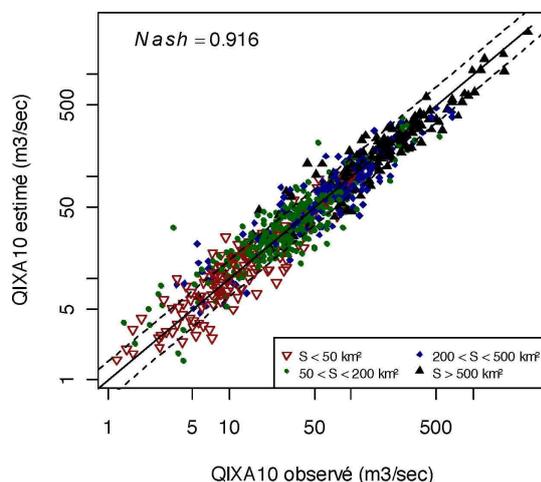


Figure 9 : Débits observés et estimés selon HER (4 Var)

Par ailleurs d'autres variables comme la température de l'air ont été testées. Mais elles n'ont pas autant renforcé les formulations que celles constituées par la densité de drainage ou l'altitude moyenne.

V.2. Détermination de la durée caractéristique θ_R

Une démarche identique à celle menée sur le $QIXA10$ a été appliquée sur la durée caractéristique q_R . Elle a conduit à une première formule basée sur la surface :

$$\hat{Q}_R = 7,45 \cdot S^{0,27} \quad (3)$$

La formule obtenue est cohérente avec les intuitions hydrologiques : le temps de concentration croît avec la taille du support du fait de contributions différées reçues le long du réseau. Nous retrouvons un exposant proche de 0,32 valeur fixée dans la méthode SOCOSE et donc une certaine cohérence avec cette étude référence malgré des définitions différentes de la durée caractéristiques. La surface n'explique que 20% de la variance ; la formule est trop incertaine pour être utilisée en sites non jaugés. L'introduction de $\overline{P24hXA10}$ dans une formulation nationale ne permet pas de gagner significativement en performance (Figure 10, $r^2 = 0,32$) :

$$\hat{Q}_R = 182,66 \cdot S^{0,25} \cdot \overline{P24hXA10}^{-0,73} \quad (4)$$

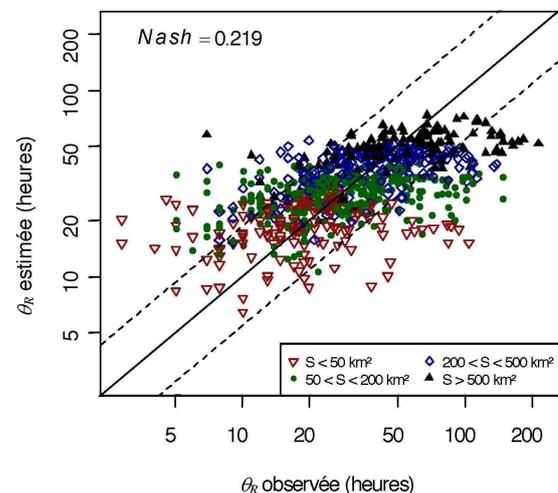


Figure 10 : Durées θ_R observées et estimées selon la formule nationale

En effet, la variable de pluie $\overline{P24hXA10}$ n'a pas de lien direct avec la dynamique des événements. Des formules régionales ont donc été recherchées intégrant toutes les variables (Tableau 3).

Bien que certaines HERs présentent une corrélation avérée comme le Jura, les Cévennes, l'Alsace, les Coteaux Aquitains (non présentées ici), les Pyrénées, le Massif Armoricain et le Massif Central Nord, les résultats des autres régions chutent en validation croisée. La comparaison des Figures 10 et 11 montre que malgré l'amélioration du critère de Nash (0,542 contre 0.219 pour la relation (4)) avec un découpage en régions, la dispersion du nuage reste grande et les formules empiriques demeurent peu efficaces.

VI. SYNTHÈSE

La section précédente a été consacrée à l'élaboration des formules empiriques. D'autres techniques exploitant la proximité géographique, mesurée par la distance entre les centres de gravité, ont été examinées en validation croisée (Tableau 4) : régressions exploitant S et $\overline{P24hXA10}$ établies sur un « Voisinage glissant » (dans un rayon de 55 km autour du site cible) et régressions couplées au krigage des résidus sur la base d'une distance entre centres de gravité (« HER (4 Var) & Krigage ») pour spatialiser

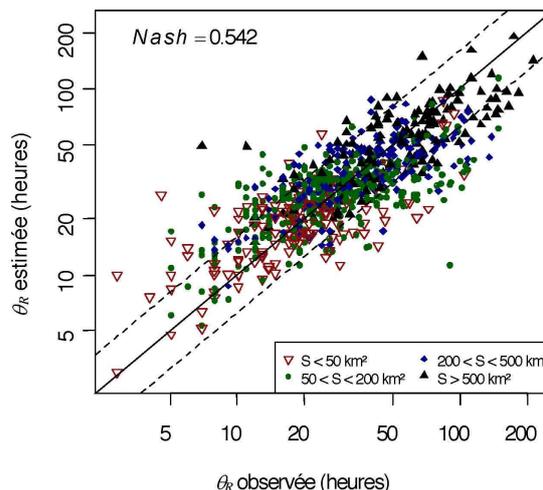


Figure 11 : Durées θ_R observées et estimées selon HER (4 Var)

un coefficient correcteur (la part non expliquée par les formules) à l'image du R de la formule CRUPEDIX. Les résultats en validation croisée ont été privilégiés ; de ce fait, certaines cellules du Tableau 4 n'ont pas été complétées (qui plus est le krigage est un interpolateur exact ;

Tableau 3 : Formules empiriques ajustées sur les HERs de niveau 1 avec au maximum quatre descripteurs et statistiques de performance associées

HER	Formule empirique HER (4 Var)	r^2 (VC)	Nash (VC)	I.C. 70%
01	$2,0 \cdot S^{0,14} \cdot \overline{P24hXA10}^{-0,22} \cdot Zmoy^{0,37}$	0,42	0,37	$[0,79 \hat{\theta}_R ; 1,27 \hat{\theta}_R]$
02	$1,7 \cdot 10^1 \cdot S^{0,23} \cdot \overline{P24hXA10}^{-0,17}$	0,04	-0,60	$[0,73 \hat{\theta}_R ; 1,38 \hat{\theta}_R]$
06	$2,8 \cdot S^{0,24} \cdot \overline{P24hXA10}^{-0,08}$	0,00	-0,10	$[0,50 \hat{\theta}_R ; 1,99 \hat{\theta}_R]$
09	$2,3 \cdot S^{0,40} \cdot \overline{P24hXA10}^{-0,42} \cdot DDr^{0,34}$	0,38	0,34	$[0,56 \hat{\theta}_R ; 1,77 \hat{\theta}_R]$
12	$5,5 \cdot S^{0,35} \cdot \overline{P24hXA10}^{-0,44} \cdot DDr^{0,36} \cdot Zmoy^{-0,12}$	0,40	0,25	$[0,61 \hat{\theta}_R ; 1,63 \hat{\theta}_R]$
20	$1,7 \cdot 10^{-4} \cdot S^{0,22} \cdot \overline{P24hXA10}^{-0,58} \cdot DDr^{1,09} \cdot Zmoy^{0,98}$	0,29	0,36	$[0,64 ; 1,57]$

Tableau 4 : Synthèse des statistiques de performance des procédures d'interpolation – les formulations nationales sont construites à partir des variables S et $\overline{P24hXA10}$

Détermination $QIXA10$	r^2	r^2 (VC)	Nash	Nash (VC)	σ_{ln}	I.C. 70%
Formulation nationale	0,775	0,773	0,884	0,883	0,611	$[0,53 \hat{Q} ; 1,89 \hat{Q}]$
CRUPEDIX	0,833	—	0,818	—	0,528	$[0,58 \hat{Q} ; 1,73 \hat{Q}]$
HER ($S, \overline{P24hXA10}$)	0,858	0,836	0,902	0,867	0,485	$[0,60 \hat{Q} ; 1,66 \hat{Q}]$
HER (4 Var)	0,903	0,874	0,916	0,882	0,400	$[0,66 \hat{Q} ; 1,52 \hat{Q}]$
Voisinage glissant	—	0,812	—	0,870	0,571	$[0,56 \hat{Q} ; 1,80 \hat{Q}]$
HER (4 Var) & Krigage	—	0,916	—	0,905	0,375	$[0,66 \hat{Q} ; 1,51 \hat{Q}]$
Détermination θ_R	r^2	r^2 (VC)	Nash	Nash (VC)	σ_{ln}	I.C. 70%
Formulation nationale	0,320	0,315	0,219	0,216	0,598	$[0,54 \hat{\theta}_R ; 1,86 \hat{\theta}_R]$
HER (4 Var)	0,602	0,513	0,542	0,474	0,458	$[0,62 \hat{\theta}_R ; 1,61 \hat{\theta}_R]$
Voisinage glissant	—	0,365	—	0,270	0,578	$[0,55 \hat{\theta}_R ; 1,82 \hat{\theta}_R]$
HER (4 Var) & Krigage	—	0,653	—	0,653	0,431	$[0,62 \hat{\theta}_R ; 1,60 \hat{\theta}_R]$

de ce fait l'analyse de l'application de cette technique hors VC n'apporte rien).

La qualité des estimations du $QIXA10$ fournies par ces modèles est comparable. Les formulations nationales constituent un cas particulier : la régression nationale est un cas de régionalisation triviale, et la formule CRUPEDIX ajoute un coefficient de correction régional R qui assure une bonne performance du modèle malgré une variable explicative légèrement différente de celle exploitée originellement. Les régressions établies par HER présentent un peu plus de degrés de liberté par rapport à la méthode CRUPEDIX (20 HER sont identifiées contre 10 zones homogènes vis à vis du coefficient R), ce qui lui permet de mieux s'adapter aux conditions locales. Un autre point traité concernait l'information pluie. Le voisinage glissant donne des résultats acceptables cependant tributaires de l'homogénéité de la région et de la qualité de la régression établie. Le couplage de « HER (4 Var) & Krigeage » fournit les meilleures estimations. La plus value est faible du fait de formules empiriques très efficaces. Concernant l'estimation de θ_R , nous avons noté la faible performance de la formule empirique calée à l'échelle nationale et une meilleure performance pour les modèles ajustés sur les HERs. La difficulté de spatialisation est réelle par comparaison aux résultats sur le $QIXA10$. L'apport d'un coefficient correcteur obtenu par krigeage a été mesuré. Cette correction s'avère pertinente car elle permet d'atteindre un coefficient de Nash de 0,65 en validation croisée. Ici, l'absence de variables explicatives est palliée par une interpolation d'un coefficient correcteur intégrant implicitement ces variables cachées structurées dans l'espace. Le krigeage des résidus compense l'inefficacité des régressions et son efficacité repose sur l'existence d'une structure spatiale de la variable interpolée. *A contrario* son inefficacité est révélatrice d'effets très locaux pas nécessairement capturés par le réseau de mesures.

D'autres sources de données pluviométriques sont désormais disponibles sous la forme de réanalyses qui reconstituent des séries climatiques complètes sur tout ou partie du territoire : Safran [Vidal *et al.*, 2010] et SPAZM [Gottardi *et al.*, 2012] ou de cartes de quantiles pluviométriques selon l'approche SHYPRE [Arnaud *et al.*, 2008]. Les réanalyses offrent les pistes certainement les plus prometteuses. Elles permettent le calcul de statistiques à l'échelle des bassins plus adaptées à l'analyse croisée débit-pluie sur les épisodes de crue que les sorties de générateurs stochastiques qui de fait ne respectent pas l'historique des événements. SPAZM fournit des valeurs à résolution kilométrique mais uniquement sur le sud de la France. Safran propose des données infra-journalières en France cependant sur une grille régulière plus grossière 8 km \times 8 km. Du fait de son extension spatiale adaptée pour cette étude, Safran est la seule réanalyse exploitée ici. Le pouvoir explicatif de plusieurs variables (cumul pluviométrique de bassin décennal sur 24 heures ou sur la durée caractéristique q_R , sur la saison des crues ou non...) dans les formules empiriques a pu être examiné. Les résultats ont montré l'absence de réelle plus-value d'une information temporelle plus cohérente avec la dynamique et l'occurrence des crues (augmentation du critère de Nash inférieure à 0,1 en validation croisée par rapport aux résultats obtenus avec le modèle HER (S , $P24hXA10$) calé sur les pluies AURELHY). L'usage de Safran est pénalisant en particulier sur les petits bassins versants du fait d'une résolution spatiale inadaptée. C'est bien la résolution spatiale qui semble faire défaut et la connaissance des contrastes spatiaux à fine échelle est donc

à privilégier dans une perspective d'estimation de quantiles de crue.

VII. CONCLUSION

Cet article décrit une partie des travaux de régionalisation des descripteurs du régime des crues opérée à l'échelle nationale à partir d'un jeu étendu de stations [Cipriani *et al.*, 2012] dans le but d'identifier les grands contrastes hydrologiques. Il propose une mise à jour des formules sommaires exploitant principalement la surface et la moyenne de bassin $\overline{P24hXA10}$ des pluies locales de 24 heures décennales. La variable $\overline{P24hXA10}$, il convient de le souligner, diffère de la moyenne des pluies journalières décennales privilégiée dans la formule CRUPEDIX publiée. Cependant, rappelons que l'application des formules dans une étude de risque ne peut se substituer au traitement statistique local si la donnée existe et même si l'enregistrement est court [Ribatet *et al.*, 2009] et nécessite un minimum de précaution (examen des I.C. à 70%, validité non garantie en dehors de la plage de superficie étudiée, incertitude forte dans les secteurs peu instrumentés, comparaison des estimations en sites jaugés voisins).

Même si les intervalles de confiance restent conséquents – en moyenne autour de l'intervalle $[0,6 \hat{Q} ; 1,7 \hat{Q}]$ – les formules empiriques permettent d'obtenir des estimations fortement corrélées aux $QIXA10$ estimés localement. Nous avons montré l'intérêt d'une partition en secteurs géographiques pour les construire. Le découpage en hydro-écorégions s'est avéré le plus satisfaisant. Deux variables explicatives optionnelles (DDr et $Zmoy$) améliorent la qualité de régressions établies sur la superficie et un indicateur de pluie extrême. Les cartes associées aux méthodes les plus efficaces sont données en Figure 12 et Figure 13. Les formules empiriques établies pour estimer la durée caractéristique θ_R en revanche fournissent des résultats très incertains. La marge de progression est grande pour atteindre une efficacité équivalente à celle obtenue sur le $QIXA10$, quelle que soit la méthode. C'est donc sur la définition de variables explicatives et adaptées aux processus hydrologiques en crue qu'il faut progresser. Au final la procédure associant krigeage et formules empiriques semble à ce jour la plus indiquée pour décrire les deux variables $QIXA10$ et θ_R .

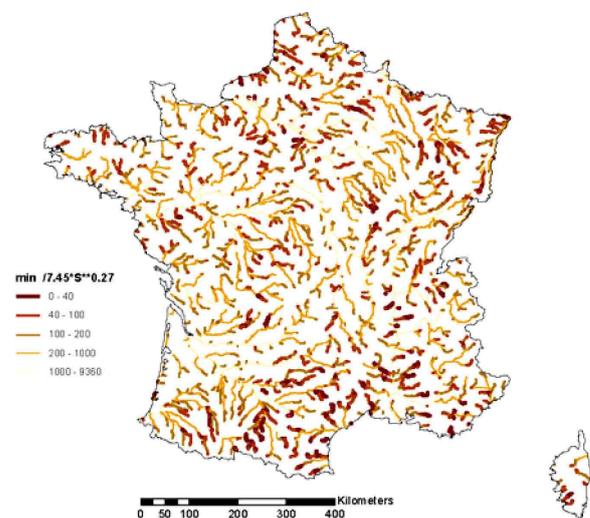


Figure 12 : Carte de $\theta_R / (7,45 S^{0,27})$ selon HER (4 Var) & Krigeage

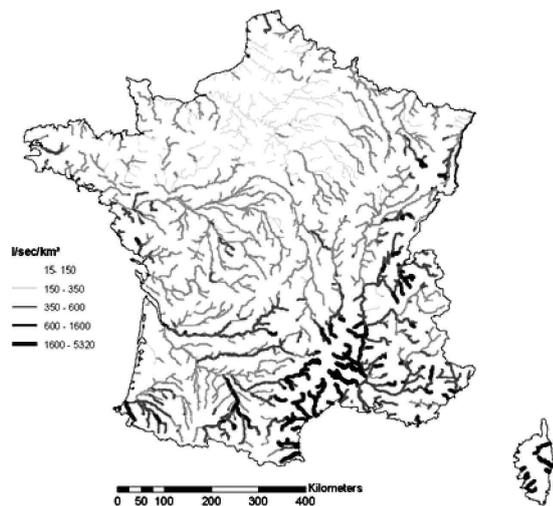


Figure 13 : Carte du QIXA10 selon HER (4 Var) & Krigeage

VIII. REMERCIEMENTS

Cette étude a été réalisée grâce au soutien financier du programme RiskNat de l'ANR dans le cadre du projet Extraflo (convention ANR-08-Risk-03-01) (<https://extraflo.cemagref.fr/>) et de l'Onema. Merci à Michel Lang, Benjamin Renard et Emmanuel Paquet pour leur relecture critique de cet article. Les auteurs remercient les fournisseurs de données hydrologiques et météorologiques (EDF DTG, SCHAPI, Météo France).

IX. RÉFÉRENCES

- [1] ARNAUD P., LAVABRE J., SOL B., DESOUCHES C. (2008) — Régionalisation d'un générateur de pluies horaires sur la France métropolitaine pour la connaissance de l'aléa pluviographique. *Journal des Sciences Hydrologiques*. **53(1)** : 34-47
- [2] BENICHO P., LE BRETON O. (1987) — Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie*. **7(19)** : 23-34
- [3] CIPRIANI T., TOILLIEZ T., SAUQUET E. (2012) *Caractérisation du régime des crues en France Métropolitaine. Rapport d'étude Irstea-Onema*.
- [4] CUNNANE C. (1979) — A note on the Poisson assumption in partial duration series models. *Water Resources Research*. **15(2)** : 489-494
- [5] FAURE D., GALÉA G., MATHYS N., OBERLIN G., EDOUARD J.L., VIVIAN H. (1991) — Synthèse cartographique des connaissances en crues observables dans le Sud-Est de la France. *La Houille Blanche*. **5** : 349-364
- [6] GOTTARDI F., OBLED C., GAILHARD J., PAQUET E. (2012) — Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. *Journal of Hydrology*. **432-433** : 154-167
- [7] HOSKING J.R.M., WALLIS J. (1997) — *Regional Frequency Analysis*. Cambridge University Press.
- [8] LANG M., LAVABRE J. (2008) — *Estimation de la crue centennale pour les plans de prévention des risques d'inondations*. Editions Quae.
- [9] LANG M., OUARDA T.B.M.J., BOBÉE B. (1999) — Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*. **225(3-4)** : 103-117
- [10] RIBATET M., SAUQUET E., GRÉSILLON J.M. et OUARDA T.B.J.M. (2007) — A Regional Bayesian POT Model for Flood Frequency Analysis. doi:10.1007/s00477-006-0068-z. *Stochastic Environmental Research and Risk Assessment*. **21(4)** : 327-339
- [11] Ministère de l'Agriculture, Services Régionaux de l'Aménagement des Eaux (S.R.A.E.), Direction de l'Aménagement Service de l'Hydraulique (D.A.S.H.), C.T.G.R.E.F. (1980) — *Synthèse nationale sur les crues des petits bassins versants*. Fascicule 3 : la méthode CRUPEDIX (juillet 1980) ; Fascicule 2 : la méthode SOCOSE (janvier 1980).
- [12] PETEUIL C., CARLADOUS S., MATHYS N. (2010) — La méthode ANETO : un outil pour la prédétermination des débits de crue des bassins versants torrentiels des Pyrénées françaises. *Sciences Eaux & Territoires*. **2** : 116-127
- [13] SAUQUET E. (2006) — Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies. *Journal of Hydrology*. **331(1-2)** : 300-314
- [14] SAUQUET E., RAMOS M.H., CHAPEL L., BERNARDARA P. (2008) — Streamflow scaling properties: investigating characteristic scales from different statistical approaches. DOI: 10.1002/hyp.6952 & erratum DOI: 10.1002/hyp.7192. *Hydrol. Processes*. **22(17)** : 3462-3475
- [15] TALLAKSEN L., VAN LANEN H. (2004) — Hydrological Drought : Processes and Estimation Methods for Streamflow and Groundwater. *Developments in Water Science*. Elsevier Science B.V. **48**
- [16] VIDAL J.P., MARTIN E., FRANCHISTÉGUY L., BAILLON M., SOUBEYROUX J.M. (2010) — A 50-year high-resolution atmospheric reanalysis over France with the Safran system. DOI: 10.1002/joc.2003. *International Journal of Climatology*. **30(11)** : 1627-1644
- [17] WASSON J.G., CHANDESRIS A., PELLA H., BLANC L. (2002) — *Définition des hydro-écorégions françaises métropolitaines. Rapport final Cemagref-MEDD*.