

Projet ANR-08-RISK-03-01

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

Programme RISKNAT 2008

Tâche VI : Transfert en milieu opérationnel

Rapport VI.2 « *Digitalisation et valorisation de longues séries de précipitations* »

Date : Avril 2013

Rapport réalisé par :

⁽¹⁾ Météo-France, Direction de la Climatologie

⁽²⁾ Météo-France, Direction Interrégionale Centre Est

Auteurs :

N Deaux¹

J.M. Soubeyroux¹

F Borchì¹

J-M Veysseire¹

S Jourdain¹

A Auffray²

SOMMAIRE

<i>I La digitalisation des pluviogrammes</i>	4
I.1. Appareils de mesure et supports d'enregistrement	4
I.2. L'outil de digitalisation : le logiciel DIGITALISE	5
I.3. Comparaison des techniques de numérisation	6
<i>II Sélection et validation des données</i>	7
II.1. Sélection des pluviogrammes	7
II.2. Précision de la méthode	8
II.3. Contrôles des données	8
a) Contrôles élémentaires	8
b) Evaluation de la qualité des données	9
<i>III Application aux valeurs extrêmes</i>	10
III.1. Evaluation au pas de temps quotidien	10
III.2. Apport aux pas de temps infra-quotidiens	12
<i>IV Conclusions</i>	14
<i>V Références</i>	15

I La digitalisation des pluviogrammes

La disponibilité de longues séries de données est particulièrement importante en climatologie, tant pour l'analyse des évolutions du climat que pour la connaissance des valeurs extrêmes. Météo France, dans le cadre du Data Rescue (DARE) a recherché et inventorié des milliers d'années stations d'enregistrements papier de données pluviographiques anciennes et souhaite les exploiter.

Ce rapport explique les différents traitements et valorisations des données digitalisées pour la connaissance des précipitations extrêmes menés dans le cadre du projet EXTRAFLO.

1.1. Appareils de mesure et supports d'enregistrement

La connaissance des métadonnées météorologiques, notamment du lieu de mesure, du type d'appareil et de support d'enregistrement, est un élément essentiel du dépouillement des données pluviographiques. Les pluviographes sont de forme, de dimension et de conception très variées [1]. On peut regrouper les différents instruments selon les trois principes météorologiques utilisés : les pluviographes à flotteur ou siphon, les pluviographes à balance et les pluviographes à augets.

Au sein des réseaux météorologiques nationaux, les premiers pluviographes ont été de type pluviographe à flotteur ou siphon, basé sur la mesure d'une hauteur d'eau à partir d'un flotteur, exploité dès la fin du XIX^{ème} siècle selon le modèle Helmann-Fuess (figure 1). Ces pluviographes sont encore parfois en service chez nos voisins, notamment en Belgique [2].



Figure 1 : Pluviographe à siphon de type Helmann-Fuess exploité à Uccle (Belgique) depuis 1898

Pour le poste de Lyon Bron, dont la série est étudiée plus spécifiquement dans le cadre du projet Extraflo, ce type d'appareil a été installé au début des années 1920 et a fait l'objet d'une exploitation régulière jusqu'au début des années 1960, avec plusieurs versions successives d'enregistrement, hebdomadaire puis quotidien.

Ces appareils à flotteur ou siphon, très difficiles à régler [3], ont été progressivement remplacés en France, après la seconde guerre mondiale, par des pluviographes à augets, toujours utilisés aujourd'hui en opérationnel dans des versions modernes.

Le support d'enregistrement, nommé pluviogramme, est un élément essentiel de la mémoire de la mesure. Le pluviogramme (figure 2) est une bande papier de format variable où les épisodes secs sont représentés par des lignes horizontales et les épisodes pluvieux par des lignes obliques, plus ou moins verticales selon l'intensité de la pluie. On peut repérer ainsi le début et la fin de pluie, ainsi que la hauteur de précipitation, et en déduire son intensité. Une difficulté du dépouillement des données est l'existence, au cours du dernier siècle, de toutes sortes de pluviogrammes, selon la fréquence du relevé (quotidien, hebdomadaire, mensuel), le type et le modèle du pluviographe et l'imprimé utilisé.

La qualité et la rigueur des conditions passées d'exploitation de ces instruments, tant pour le réglage des appareils (vidange, étalonnage, débouchage) et de leur système d'enregistrement (stylet, encre), que pour la conservation des relevés (documentation sur les problèmes d'exploitation, classement), sont les déterminants d'une possible valorisation de nos jours.

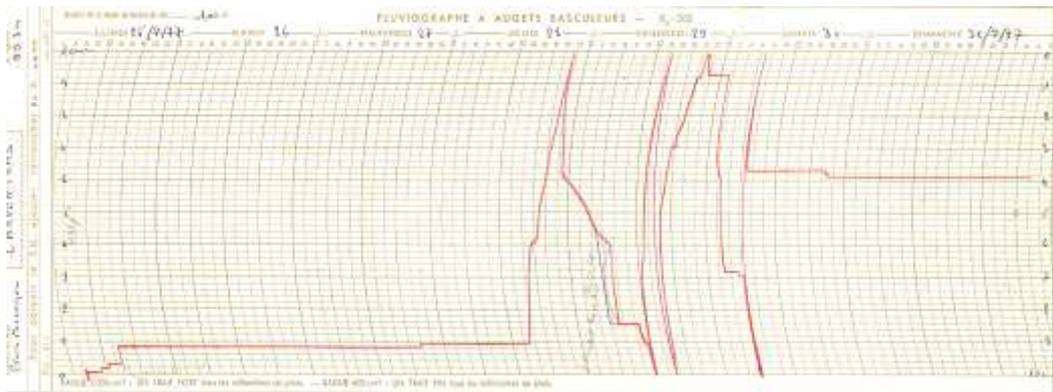


Figure 2 : *Pluviogramme hebdomadaire de Carpentras : semaine du 25 au 31 juillet 1977*

1.2. L'outil de digitalisation : le logiciel DIGITALISE

Le dépouillement des pluviogrammes peut s'avérer être une opération longue et fastidieuse sans l'aide d'un outil adapté. Le logiciel DIGITALISE, développé par Météo France au début des années 2000, permet, grâce à la connexion d'un PC à une table à digitaliser, de saisir les données de précipitations. Le principe consiste à décomposer le tracé en une succession de paliers de durée variable. On obtient ainsi une ligne brisée dont les sommets, appelés points caractéristiques, sont choisis par l'opérateur suivant une technique équivalente aux traitements manuels des années 1980. Lorsque le pluviogramme a été digitalisé, le logiciel affiche les points saisis à l'aide de la souris de la table à digitaliser, effectue les calculs de durées et de quantités de précipitation puis indique une valeur quotidienne (en mm) calculée entre 6h TU (jour J-1) et 6h TU (jour J). Il affiche, en correspondance du jour, la valeur (en mm), quand elle existe, issue de la base de données, ce qui permet de comparer et de contrôler les résultats de la digitalisation. Enfin, le logiciel permet de visualiser le pluviogramme digitalisé pour contrôler l'allure générale du tracé (figure 3). Plusieurs formats de fichier peuvent être obtenus en sortie, sur des pas de temps fixes (quotidien, horaire) ou variables (PRECIP).

Parmi les fonctionnalités les plus importantes, il convient de noter la comparaison en temps réel avec les données externes disponibles (poste pluviométrique de référence), permettant d'identifier les erreurs potentielles liées aux conditions de fonctionnement de l'appareil (pluviomètre bouché, effets de la neige, etc).

Le programme de digitalisation à Météo France a permis à ce jour d'allonger les longues séries de précipitations de plusieurs stations : Cambrai (11 ans), Dijon (15 ans), Châtillon sur Seine (10 ans), Rennes (14 ans), Dinard (11 ans) et Limoges (9 ans).

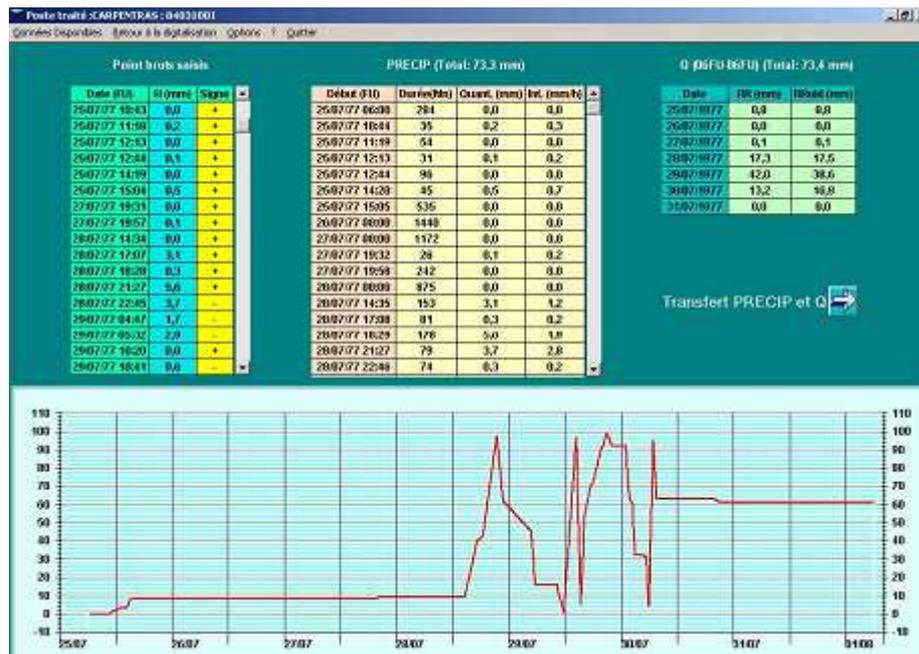


Figure 3 : Interface du logiciel Digitalise - exemple du pluviogramme digitalisé pour Carpentras

1.3. Comparaison des techniques de numérisation

DIGITALISE n'est pas le seul outil permettant la numérisation de pluviogramme. Un logiciel a été développé par le CETE Méditerranée / service hydraulique pour la numérisation de limnigrammes et de marégrammes papiers. Il s'agit de NUNIEAU (NUMérisation des enregistrements graphiques de Niveaux d'EAU). L'outil se base sur la reconnaissance des couleurs du signal et sur le calage de la planche. Le principe est de scanner les feuilles papier (images en format .tif de préférence, 300 DPI au moins de résolution), créer un fichier comportant les informations de calage, caler l'image, la numériser (extraire le signal sans rien oublier ni ajouter) et la vérifier. Un rapport de digitalisation est ensuite généré. Une version spécifique « NUNIEAU pluie » a été développée pour numériser les pluviogrammes exclusivement.

En dehors du programme EXTRAFLO, une comparaison élémentaire a été effectuée entre Digitalise et NUNIEAU 'pluie' en 2011 sur des pluviogrammes de Lyon et Carpentras, en liaison avec le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision aux Inondations) qui a initié en 2009 un inventaire des données pluviométriques de la DREAL Midi-Pyrénées pour exploiter les données pluviographiques anciennes (plus de mille années stations d'enregistrement identifiées sur une trentaine d'années).

Il en ressort que :

- Les résultats entre les deux méthodes sont très proches,
- Les différences entre les méthodes sont surtout visibles sur les pas de temps les plus fins (6 ou 15 minutes),
- DIGITALISE peut avoir tendance à minorer les pics ou au contraire NUNIEAU peut avoir tendance à les majorer,
- En terme d'utilisation, DIGITALISE, plus ciblé pour l'analyse des précipitations est plus rapide à prendre en main mais nécessite d'être utilisé par un public averti, ayant une bonne expertise de la mesure de la pluie,

- NUNIEAU, plus complexe dans ses réglages, est aussi plus générique (tout type d'enregistrement pris en charge) et plus facilement déployable (pas de table à digitaliser), notamment pour des actions de sous-traitance (archivage numérique).

Des investigations plus poussées mériteraient d'être menées à l'avenir, pour préciser les performances relatives et les domaines d'application spécifiques de ces deux outils a priori complémentaires.

II Sélection et validation des données

II.1. Sélection des pluviogrammes

Dans le cadre du projet Extraflo, une opération spécifique de digitalisation a été engagée, afin d'évaluer son intérêt dans l'étude des précipitations extrêmes. Les données digitalisées ont été sélectionnées à partir de l'inventaire effectué à la demande de l'OHMCV, complété par la station professionnelle de Lyon Bron. Cette station a permis d'allonger la série pluviographique numérisée de 1921 à 1960, en complément de données quotidiennes déjà disponibles pour les contrôles. Six autres séries du Sud-Est de la France ont été complétées ou créées dans le cadre de cette opération: La Grand-Combe (30), Lussan-Prat (30), Pompignan (30), Saint Christol les Alès (30), Montpellier (34) et Saint Martin de Londres (34).

dépt	station	num_poste	type station	date début pluviogramme	date de fin pluviogramme	type pluviogramme
69	Lyon Bron	69029001	0	08/1920	1964	hebdomadaire et quotidien
30	La grande Combe	30132003	2	1977	1992	hebdomadaire
30	Lussan	30151002	4	06/1978	2006	hebdomadaire
30	Pompignan	30200002	4	10/1970	04/2006	hebdomadaire
30	St Christol les Ales	30243001	4	01/11/1978	10/2006	hebdomadaire
34	Montpellier	34154001	0	15/03/1956	28/02/1957	quotidien
34	St Martin de Londres	34274001	2	03/1981	1995	hebdomadaire

Cette action de digitalisation a nécessité six mois de travail pour plus de 150 années stations numérisées (rythme mensuel moyen de 25 années pour 1 station). D'autres données précédemment digitalisées (Carpentras 1977-1990) ont été utilisées pour consolider les résultats des analyses.

La phase de sélection des séries à digitaliser s'est avérée une étape particulièrement importante pour la réussite de l'opération. La disponibilité de l'ensemble des archives visées est vérifiée (examen du contenu des différents cartons et classeurs) ainsi que la qualité des données (échantillonnage sur quelques journées test). Ces tests ont été menés pour les événements pluvieux les plus importants, en comparant les données digitalisées avec les données quotidiennes des postes climatologiques de référence.

II.2. Précision de la méthode

Les comptages des basculements d'augets, la précision temporelle et les erreurs de lecture sont surtout responsables de différentes sources d'incertitudes de la méthode de digitalisation [6]. Pour les fortes intensités (supérieures à 50mm/h), les risques d'erreur de lecture prédominent. En effet, l'œil humain est incapable de distinguer des variations de quelques degrés sur une droite oblique. Cette source d'erreur peut atteindre 25 % pour les pas de temps les plus courts, lisibles sur les différents enregistrements.

Une vérification de ces résultats avec les sorties du logiciel DIGITALISE a été menée avec un double dépouillement des archives pluviographiques sur cinq années de la série de Lyon Bron, pour la période entre 1960 et 1964. Les tests ont concerné les événements de plus forte intensité aux pas de temps courts inférieurs à 2 heures, détectables à partir de pluviogramme quotidien (environ vingt épisodes par pas de temps). La figure 4 montre que la dispersion (rapport de l'écart type sur la valeur moyenne) diminue fortement avec le pas de temps, et se stabilise à partir du pas de temps horaire à des valeurs comprises entre 2 et 3 %, inférieures aux incertitudes de mesure.

Ce résultat confirme la possibilité d'utilisation des données issues de digitalisation de pluviogramme quotidien pour les intensités les plus fortes à partir du pas de temps horaire.

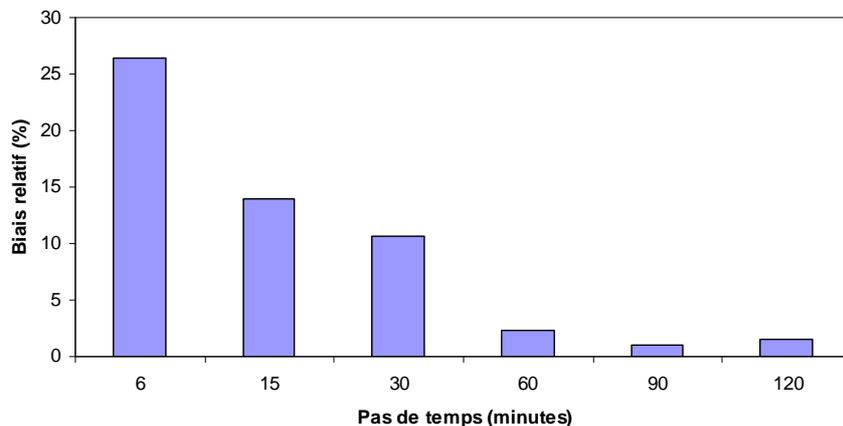


Figure 4 : Dispersion suivant l'opérateur des données digitalisées pour des événements pluvieux intenses à Lyon Bron entre 1960 et 1964.

II.3. Contrôles des données

a) Contrôles élémentaires

Différents types de contrôle ont été testés sur les séries digitalisées et notamment les séries de Lyon Bron (1921-1964) et Carpentras (1963-1978), pour laquelle des données quotidiennes de référence sont disponibles dans la base de données climatologiques.

La première étape de contrôle visuel de la qualité des archives a conduit à éliminer les deux premières années de la série de Lyon Bron (1920 et 1921). En effet, certains tracés se superposaient et l'encre était partiellement effacée. De même, un nombre important d'archives manquantes ont été identifiées entre 1935 et 1947, rappelant le poids de l'histoire dans les séries climatologiques.

Au cours de la digitalisation, la comparaison avec les données quotidiennes de la base de données a permis d'identifier des journées ayant connu des problèmes d'exploitation du pluviographe (panne du capteur, pluviomètre bouché, neige). Ces journées représentent environ 2% des chroniques totales pour la série de Lyon moins de 1% pour la série de Carpentras.

b) Evaluation de la qualité des données

Ces contrôles élémentaires, portant sur la présence des données et leur vraisemblance, restent cependant insuffisants pour évaluer correctement la qualité ces séries. Plusieurs tests ont complété notre diagnostic.

En premier lieu, nous avons comparé systématiquement les cumuls annuels des séries digitalisées aux séries quotidiennes de référence disponibles, issues de mesures réalisées par un pluviomètre totalisateur.

Ce test a mis en évidence la qualité irrégulière des données de la série de Lyon Bron (figure 5), indépendamment du taux de valeurs manquantes. Certaines années présentent des ratios de l'ordre de 0,6 non explicables par un taux significatif de valeurs manquantes.

Les évaluations des données de digitalisation ont été menées également pour les différentes gammes d'intensité des événements pluvieux (en coïncidence de date pour toutes précipitations supérieures à 2 mm). La figure 6, toujours à Lyon Bron, montre ainsi un biais négatif important de la série digitalisée par rapport aux données quotidiennes (médiane des écarts de -7,5%). Ce biais se retrouve dans toutes les gammes d'intensité. Par contre, ce biais se réduit à 3.5% pour la période de fonctionnement du pluviographe à augets. On note toutefois que la précision du pluviographe à augets diminue légèrement avec l'intensité, conformément aux résultats d'autres études [4]. La série de Carpentras présente de meilleures performances : biais médian inférieur à 2%, restant inférieur à 3% pour les plus fortes intensités.

Ces différents résultats confirment l'importance de la connaissance des métadonnées des séries anciennes, lors des actions digitalisation.

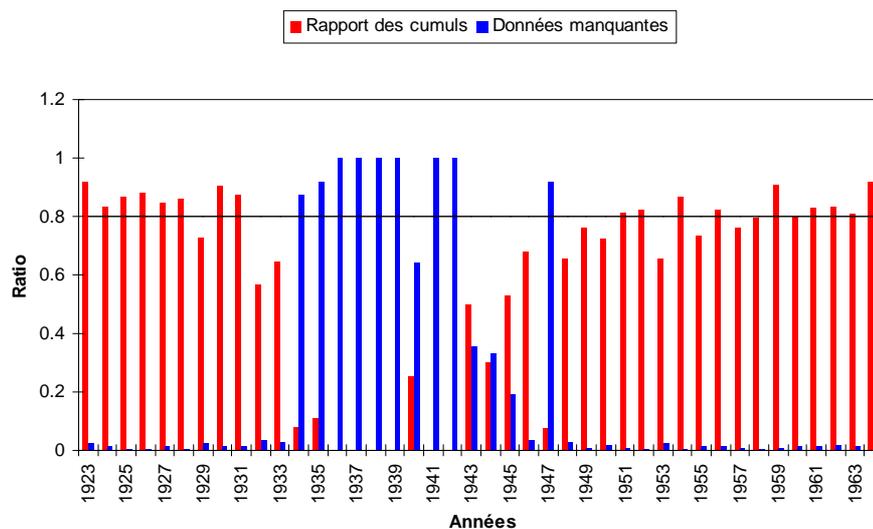


Figure 5 : Taux annuel de données de digitalisation manquantes (barre bleue) et ratio des cumuls annuels entre digitalisation et données quotidiennes à Lyon Bron (1923-1964)

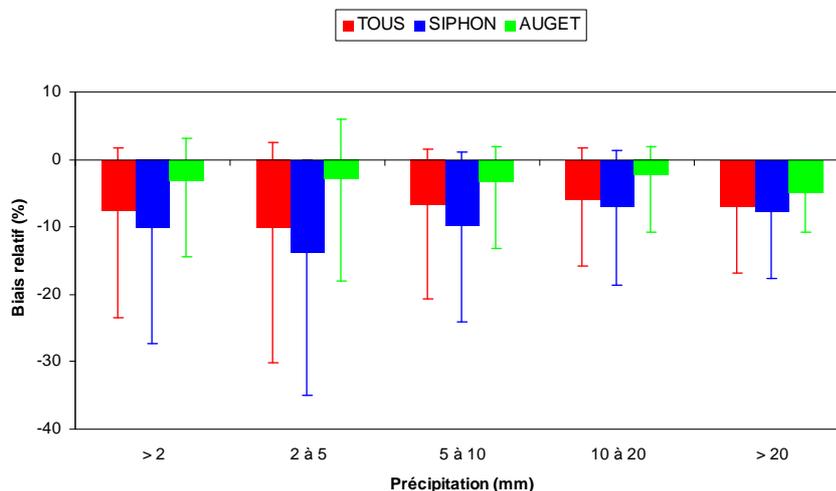


Figure 6 : Distribution de l'écart relatif (médian, 1^{er} et 4^{eme} quintile) entre les données digitalisées et celles quotidiennes à Lyon entre 1923 et 1964, selon le type de pluviographe

III Application aux valeurs extrêmes

Les données digitalisées présentent un intérêt potentiel très fort pour l'étude des précipitations extrêmes à des pas de temps quotidiens et surtout infra-quotidiens où les séries actuelles disponibles restent assez courtes. L'apport de la digitalisation peut ainsi permettre la création de nouvelles séries densifiant la couverture spatiale des longues séries mais aussi le prolongement de séries actuelles pour lesquelles les inventaires ont mis en évidence des archives anciennes.

III.1. Evaluation au pas de temps quotidien

Nous avons comparé l'estimation des valeurs extrêmes issue des séries digitalisées aux séries quotidiennes de référence de la base de données climatologiques (Bdclim) à Lyon et Carpentras.

Sur les maxima annuels (figure 7), on retrouve l'effet de sous-estimation des valeurs maximales, ainsi que l'impact des années manquantes. Ainsi, le record historique de pluie quotidienne à Lyon a été observé en 1935 (97 mm en octobre) sur une période où les archives pluviographiques ne sont pas disponibles. Le biais relatif moyen entre les maxima annuels de la série digitalisée et celle de la Bdclim est de 9% pour les événements en coïncidence de dates. Les résultats s'avèrent légèrement moins sensibles pour les dépassements de seuil (événements quotidiens supérieurs à 34mm), que pour les maxima annuels, avec un biais relatif moyen en coïncidence de date de 6,5%. Mais le nombre total d'événements obtenus avec les données digitalisées est fortement réduit par rapport à la série complète de la Bdclim (39 événements au lieu de 110).

Pour l'estimation des quantiles, nous avons comparé les résultats des deux méthodes statistiques couramment utilisées pour l'extrapolation des valeurs extrêmes [5]: la Generalized Extreme Values (GEV) s'appuyant sur les maxima annuels, et la Generalized Pareto Distribution (GPD) à partir des dépassements de seuil. L'application de ces deux méthodes a nécessité, d'abord de définir les règles de sélection des années à prendre en compte en contexte de données manquantes et de biais instrumental.

Trois règles ont été successivement envisagées à partir des années disposant de moins de 10% de valeurs manquantes (34 années), possédant un rapport des cumuls annuels, entre données digitalisées et celles quotidiennes de la Bdclim, supérieur à 80% (13 années disponibles), ou celles correspondant aux mesures de meilleure qualité, obtenues à partir de pluviographes à

augets. Dans le dernier cas, la période disponible (10 années entre 1955 à 1964) s'avère trop courte pour l'utilisation des lois extrêmes.

Au final, pour les lois GPD et GEV, ajustées selon la méthode du maximum de vraisemblance, le maintien d'un nombre important d'événements dans la série s'avère plus important que la représentativité individuelle de chacun. Pour la série de 34 années, Le biais moyen de l'estimation des quantiles est du même ordre de grandeur que celui obtenue à partir des dépassements de seuil, soit 6% pour la loi GPD (figure 8). Les intervalles de confiance à 70% s'élargissent avec la réduction de la longueur de la série, mais conservent dans tous les cas une plage de valeurs communes. La série de 13 années s'avère pénalisée par des effets d'échantillonnage la privant des valeurs observées les plus fortes.

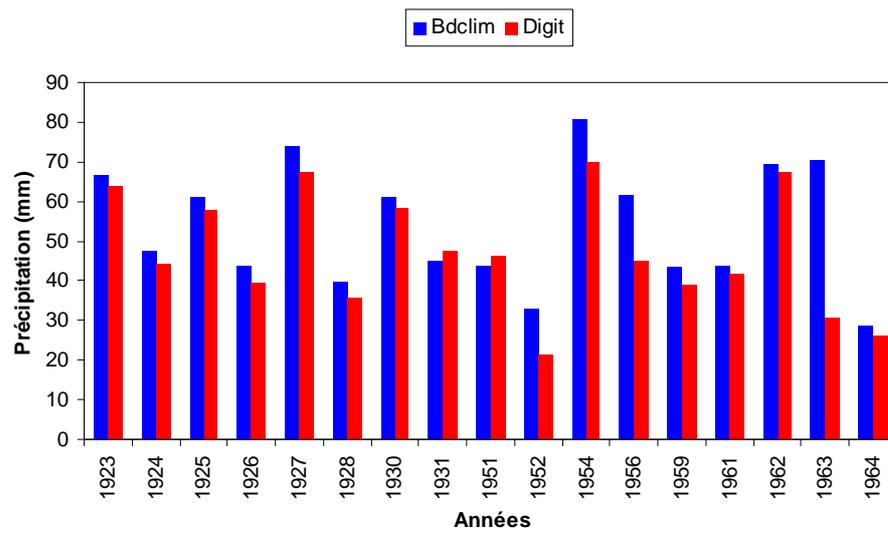


Figure 7 : Comparaison des maxima quotidiens annuels des séries digitalisées avec celles de la Bdclim, à Lyon Bron (1923-1964)

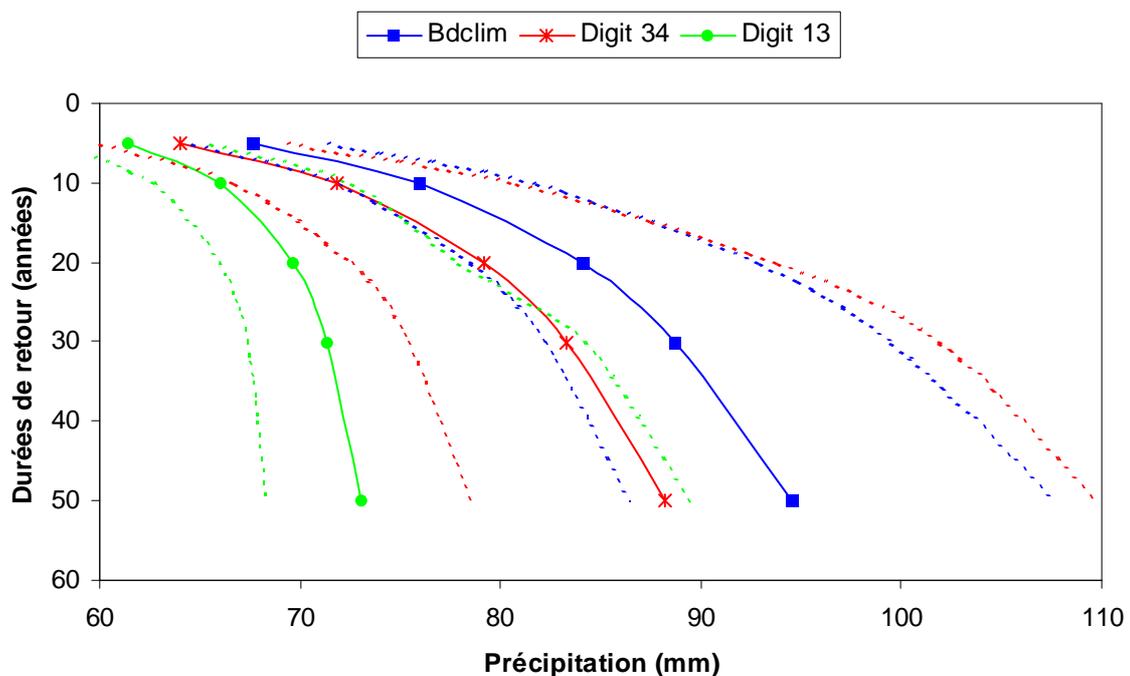


Figure 8 : Comparaison des ajustements à Lyon Bron (1923-1964) à partir de la loi GPD (hauteur moyenne en trait plein et intervalle de confiance à 70% en pointillé) entre les séries digitalisées (34 ou 13 années) et celle de la Bdclim

III.2. Apport aux pas de temps infra-quotidiens

En l'absence de série de référence indépendante, l'apport des valeurs digitalisées au pas horaires a été évalué par l'effet de la prolongation d'une série existante. L'exemple développé ici est la série horaire de Carpentras, d'une durée initiale de 30 ans dans la Bdclim (1978-2009), étendue par 16 années supplémentaires issues des données digitalisées. On appelle MIX la nouvelle série créée.

Le complément par digitalisation des pluies horaires maximales depuis 1963 (figure 9) permet d'améliorer le record de la série Bdclim avec 96 mm le 22/08/1965. Par contre, les maxima annuels entre 1963 et 1978, présentent une moyenne inférieure de près de 3 mm à ceux de la période de 1978 à nos jours, sans que l'on puisse discriminer les biais éventuels instrumentaux, les effets d'échantillonnages ou les prémices de tendance climatique sur les précipitations extrêmes. A partir de la loi GPD (figure 10), l'estimation des quantiles, inférieurs à 50 ans, de la nouvelle série MIX est légèrement inférieure à celle de la série Bdclim (de 1 à 5%). Par contre, l'allongement de la série MIX permet de réduire de plus de 20% les intervalles de confiance associés à la méthode.

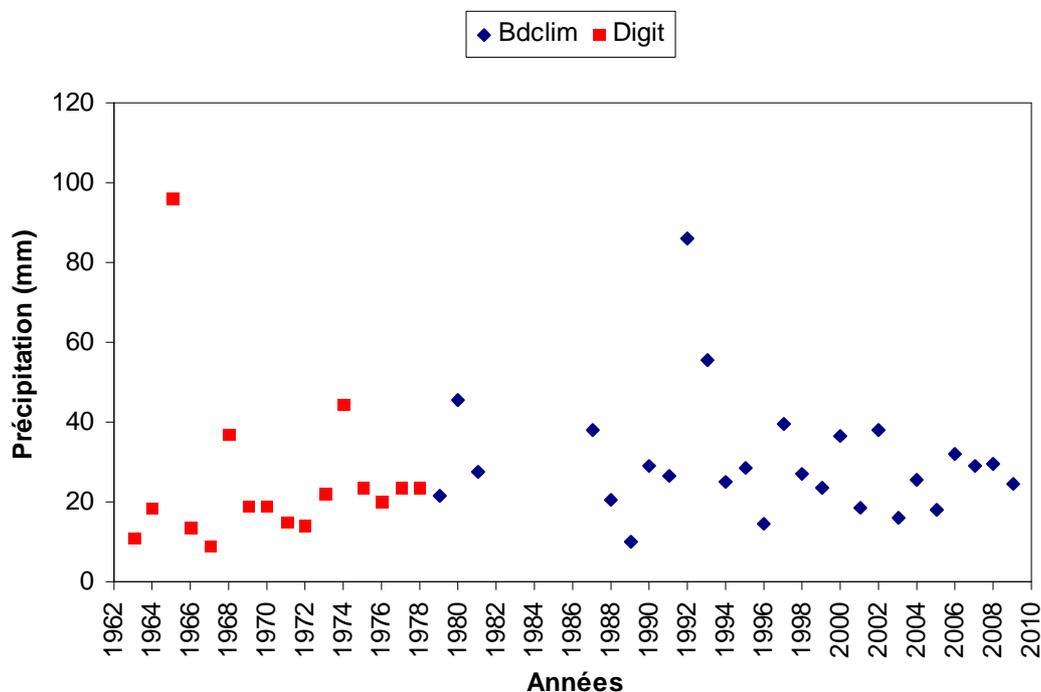


Figure 9 : Maxima de pluie horaire à Carpentras issue de la nouvelle série digitalisée (1963-1978) et de celle de la Bdclim (1978-2009)

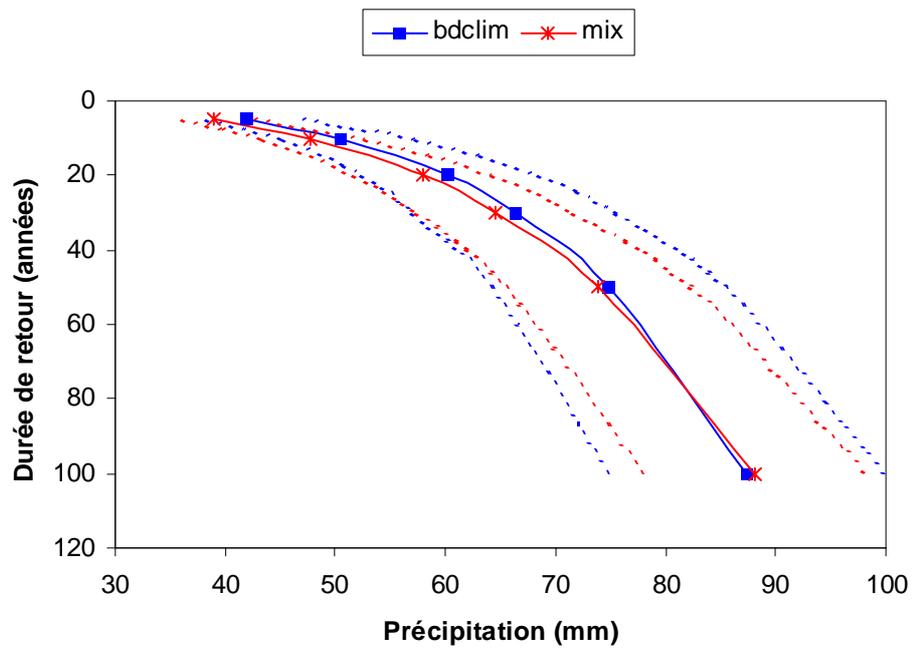


Figure 10 : Comparaison des ajustements GPD pour la série horaire à Carpentras de la Bdclim (1978-2009) et de la nouvelle série complétée MIX (1963-2009)

IV Conclusions

L'inventaire des données hydro-climatologiques anciennes a remis à la lumière le gisement important d'enregistrements pluviographiques, peu valorisés jusqu'à ce jour. Ces données présentent un grand intérêt potentiel pour les applications sur les extrêmes hydrologiques du fait des limitations des séries numériques actuelles tant en nombre qu'en profondeur temporelle.

La numérisation des pluviogrammes est aujourd'hui facilitée grâce au nouveau logiciel DIGITALISE, éprouvé dans l'environnement climatologique opérationnel. Les nouvelles séries digitalisées permettent de prolonger les séries existantes à pas de temps fin, généralement courtes (une vingtaine d'année au mieux) et d'en créer de nouvelles, dans une perspective d'amélioration de la connaissance à haute résolution spatiale des intensités pluviométriques en France.

Les évaluations menées dans le cadre du projet EXTRAFLO ont permis de préciser les modalités générales d'utilisation de ces nouvelles données. En premier lieu, il a été montré que la digitalisation des pluviogrammes quotidiens ne doit pas s'appliquer à des pas de temps infra-horaires. Par ailleurs, la difficulté de réglage et de suivi de fonctionnement des anciens pluviographes impose une grande rigueur dans la vérification de la qualité des séries anciennes. L'analyse des données devra prendre en compte la connaissance des métadonnées (type d'appareils) et s'appliquer aux différentes gammes d'intensité pluviométriques. En fonction des résultats, l'estimation des quantiles pluviométriques extrêmes pourra être conduite avec confiance, en privilégiant le nombre d'années traitées permettant de limiter les effets liés au sous échantillonnage. Ainsi, la prolongation de séries existantes par des données digitalisées, peut permettre de réduire fortement les intervalles de confiances des ajustements : 20% de réduction du quantile centennal pour un prolongement de 50% de la longueur de la série.

Sur la base de ces résultats, un programme de valorisation de ces données anciennes a été lancé avec les services régionaux de Météo-France et pourra être étendu aux différents partenaires intéressés, avec la mise à disposition du logiciel DIGITALISE : des contacts ont été pris à ce jour avec le SCHAPI et EDF/DTG.

V Références

- [1] Harrang C., 1970, La mesure des précipitations, Monographie de la météorologie nationale n°78, Direction de la Météorologie Nationale
- [2] Brouyaux F. et Ticaux C., 2006, Etude en support au « Plan Pluies » pour la Région de Bruxelles-Capitale, Institut Bruxellois pour la Gestion de l'Environnement, 69 p
- [3] Laborde J-P, 2009, Eléments d'hydrologie de surface, Ecole Polytechnique de l'Université de Nice, 202 p
- [4] Duchon C,E et Biddle C,J, 2010, Undercatch of tipping-bucket gauges in high rain rate events. Advances in Geosciences, vol 25, pp 11-25
- [5] Coles, S., 2001, An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag, 224 p
- [6] Grosse J.Y., 1980, Un mode d'enregistrement des intensités de précipitations : le fichier à pas de temps variable, La Météorologie n°20 VI série, numéro spécial précipitations et hydrologie, 45-50