



HAL
open science

Les annuaires hydrologiques de la Société hydrotechnique de France : une source d'information patrimoniale pour la connaissance de l'hydrologie en France

C. Le Gros, Eric Sauquet, M. Lang, A.L. Achard, E. Leblois, B. Biton

► **To cite this version:**

C. Le Gros, Eric Sauquet, M. Lang, A.L. Achard, E. Leblois, et al.. Les annuaires hydrologiques de la Société hydrotechnique de France : une source d'information patrimoniale pour la connaissance de l'hydrologie en France. *La Houille Blanche - Revue internationale de l'eau*, 2015, 4 (4), pp.66-77. 10.1051/lhb/20150048 . hal-01249984

HAL Id: hal-01249984

<https://hal.science/hal-01249984>

Submitted on 4 Jan 2016

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.

Les annuaires hydrologiques de la Société hydrotechnique de France : une source d'information patrimoniale pour la connaissance de l'hydrologie en France

Chloé LE GROS¹, Eric SAUQUET¹, Michel LANG¹, Anne-Laure ACHARD²,
Etienne LEBLOIS¹, Brigitte BITON³

¹ Irstea, HHLX, 5 rue de la Doua CS70077, 69626 Villeurbanne, France – eric.sauquet@irstea.fr

² Irstea, IST, 5 rue de la Doua CS70077, 69626 Villeurbanne

³ SHF, 25 Rue des Favorites, 75015 Paris, France

RÉSUMÉ. – La Société Hydrotechnique de France a publié entre 1939 et 1969 une série d'annuaires hydrologiques incluant, entre autres, des observations de débits sur un ensemble de stations hydrométriques françaises et des contributions techniques ouvertes sur des sujets relatifs à l'hydrologie. Longtemps disponible sur support papier uniquement, le contenu de ces annuaires a été numérisé afin d'être porté à la connaissance de la communauté hydrologique. Les objectifs de cet article sont de préciser le contenu de ces annuaires, de partager l'expérience d'extraction des séries numériques et d'identifier les données complémentaires qui pourraient venir enrichir les connaissances apportées par la banque HYDRO.

Mots-clés : numérisation d'archives, sauvetage de données, données anciennes, débit

The hydrological yearbooks published by the Société Hydrotechnique de France: a valuable source of information on hydrology in France

ABSTRACT. – The Société Hydrotechnique de France has published a series of hydrological yearbooks between 1939 and 1969 including daily flows observed at a subset of French gauging stations and technical contributions on several topics related to hydrology. The exact contents are not well-known and data available in the tables have not been digitized. The objectives of this article are to present the content of the hydrological yearbooks and the way time series were extracted and to identify additional historical data that may enrich the HYDRO database.

Key-words: digital processing, data rescue, historic data, river flow

I. Introduction

Aujourd'hui, la connaissance des débits en France s'appuie principalement sur le réseau hydrométrique qui alimente la banque de données HYDRO (<http://hydro.eaufrance.fr>). La majorité des séries hydrologiques bancarisées (près de 85 %) débutent dans les années 1960 et un nombre réduit présente des enregistrements antérieurs à 1900. Or de longues séries sont requises notamment pour un calcul robuste des valeurs extrêmes (Lang *et al.*, 2014), pour l'examen de la stationnarité des régimes hydrologiques (Giuntoli *et al.*, 2013)...

Une manière de parfaire nos connaissances est de rechercher et d'exhumer des données anciennes renseignées dans des rapports techniques des services hydrométriques ou des retours d'expérience post-événements. Les Annuaires Hydrologiques de la Société Hydrotechnique de France font partie de ces sources archivées, peu accessibles par la communauté hydrologique. Edités de 1939 à 1969 par la Société Hydrotechnique de France (SHF), ils fournissent entre autres une synthèse de l'année écoulée au travers notamment de tableaux de débits journaliers et d'hydrographes annuels observés ou reconstitués sur un jeu de stations hydrométriques en fonctionnement. Ces annuaires n'étaient pas jusqu'à ce jour sur support informatique. Irstea

(centre Lyon-Villeurbanne) et la SHF ont conjointement souhaité numériser leur contenu, pour porter ces documents à la connaissance de la communauté hydrologique, et, à terme, mettre à disposition les séries numériques qu'ils contiennent.

L'objectif de cet article est de décrire le contenu historique et patrimonial des annuaires, de partager notre expérience en matière d'extraction et de contrôle de données hydrologiques anciennes numérisées et de faire état des stations « inédites », c'est-à-dire absentes à ce jour de la base de données HYDRO.

II. Présentation des Annuaire Hydrologique de la France Ancienne

Si des premières mesures de cotes sur un nombre limité de cours d'eau existent de façon discontinue dans les années antérieures, c'est au cours de la seconde moitié du XIX^e siècle que se met en place en France un réseau systématique d'observations hydrométriques avec la création des services d'annonce des crues, sous la direction des ingénieurs des Ponts et Chaussées. On commence à disposer sur un nombre significatif de cours d'eau d'informations

sur la topographie des rivières (Nivellement Général de la France), sur les niveaux d'écoulement (relevés des services d'annonce des crues ; cotes de crues historiques), sur des mesures (premiers jaugeages) ou des estimations de débit (notes techniques). L'arrêté du ministre de l'Agriculture du 25 mars 1903 porte sur la création du Service des Grandes Forces Hydrauliques, pour l'évaluation du potentiel énergétique des cours d'eau montagnards français. La mission est officiellement organisée sur le secteur des Alpes, puis dix ans plus tard sur les Pyrénées et le Massif Central, sous la direction des ingénieurs du Génie Rural. Une série d'annuaires hydrologiques (Côte, 1908) va alors être publiée de 1911 à 1924 pour environ 200 stations de jaugeage. On trouve dans ces annuaires la carte de localisation des stations hydrométriques, le nivellement de cours d'eau sous forme de profils en long et le zéro altimétrique des échelles limnimétriques, le tableau des cotes et débits moyens journaliers, les jaugeages, le barème de tarage et la courbe de tarage, ainsi qu'un bilan des débits moyens mensuels et annuels.

La première guerre mondiale met un terme à la forte activité hydrométrique du début du XXe siècle. A l'issue de ce premier conflit, la France retrouve ses ressources minières du Nord et elle compte profiter de celles de l'Allemagne. Elle est trop profondément éprouvée pour maintenir des réseaux hydrométriques d'intérêt général. Ce n'est qu'en 1942 que l'Etat décide de reprendre la publication systématique des données hydrologiques, sous la forme d'annuaires.

Sous l'égide du Secrétariat d'Etat à la Production Industrielle (Direction de l'Electricité) et de la Société Hydrotechnique de France, ces annuaires seront édités pour les années 1939 à 1969, avec un décalage de 1 à 3 ans, le temps de collecter et critiquer les données.

Les annuaires vont comporter un texte de synthèse sur l'année hydrologique (faits marquants en termes d'abondance ou de déficit, mention de crues). Le texte de synthèse sera particulièrement soigné sur la période 1947-1952, grâce à la contribution du géographe Charles Péguy avec une analyse météorologique et une interprétation de la variabilité spatio-temporelle des débits. A partir de 1953 le texte sera moins détaillé et centré sur l'hydraulicité des cours d'eau des Alpes, du Massif Central et des Pyrénées. A noter enfin que l'on peut trouver pendant la première décennie (1939-1949 et 1951) un ou deux articles scientifiques par annuaire, dont certains particulièrement intéressants. Le Tableau 2 rappelle les thèmes qui ont pu être traités. Les annuaires fournissent la liste des stations en fonctionnement (de l'ordre de 1000 stations, avec des variations sur la période 1939-1969) et les données journalières-mensuelles-annuelle accompagnées de graphiques sur une sélection de stations. Ces stations dites « fondamentales » (respectivement 68, 82 et 85 en 1939, 1957 et 1969) ont été choisies à partir de critères de qualité, de représentativité spatiale, de longueur de série et d'intérêt opérationnel. Un extrait de l'annuaire 1939 est proposé en Figure 1. En dehors des débits de l'année, d'autres

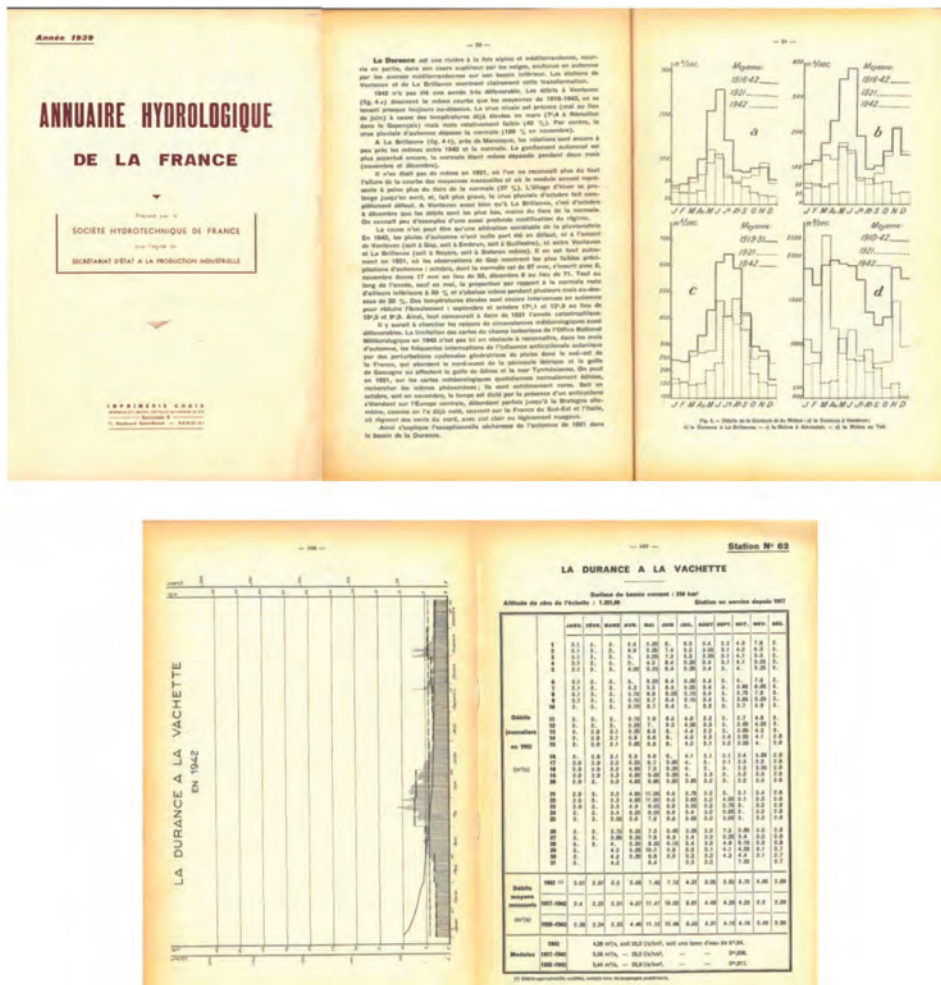


Figure 1 : Extrait de l'annuaire 1939.

Tableau 1 : Données quantitatives associées aux stations décrites dans les annuaires par période.

1939-1947	1948-1959	1960-1962	1963	1964-1969
Nom et numéro de station				
Surface du bassin versant et altitude (sauf en 1939)				
			Latitude et longitude	
Statut de la station (station usine ou non)				
Débits journaliers pour les 365/366 jours de l'année				
Débits moyens mensuels bruts de l'année		Débits moyens mensuels naturalisés de l'année		
Débits moyens mensuels corrigés de l'année (lorsqu'ils sont disponibles)			Moyennes mensuelles mesurées de l'année	
Débits moyens mensuels depuis la mise en service de la station et jusqu'à l'année considérée		Moyennes mensuelles interannuelles sur la plus longue période disponible		
Débits moyens mensuels depuis 1920 et jusqu'à l'année considérée		Moyennes mensuelles de 1921 à 1960		
Pluviométries mensuelles				
Maxima et minima journaliers par mois (valeur en m ³ /s et année correspondante)				
			Valeurs instantanées de débit maximum par mois de l'année (valeur en m ³ /s et jour correspondant) (sauf en 1960)	
Extrêmes annuels (valeur en m ³ /s et année correspondante)				

informations hydrologiques viennent compléter les tableaux des 365 valeurs journalières. La nature de ces informations a évolué au fil des années (Tableau 1). Chaque annuaire est précédé d'un erratum qui porte sur les annuaires précédents. Il s'agit de corrections à apporter aux métadonnées sur les stations hydrométriques (surface de bassin versant, altitude du zéro de l'échelle limnimétrique, période d'observation) et aussi sur les débits qui ont déjà été publiés (le plus souvent le débit moyen mensuel et parfois le débit moyen journalier).

Deux numéros spéciaux (1950 Alpes et 1951 Massif Central) donnent des informations complémentaires pour combler la lacune sur la période 1925-1938 (hors périodes visées par les Grandes Forces Hydraulique (1911-1924) et par les annuaires de la SHF publiés à partir de 1939). On dispose d'informations en termes de débit moyen mensuel et annuel sur respectivement 23 et 20 stations (Alpes, Massif central) depuis le début du XXe siècle jusque vers 1948 ou 1949. Un numéro spécial analogue sur les Pyrénées était annoncé mais n'a pas été retrouvé dans les différents annuaires disponibles à la Société Hydrotechnique de France.

III. MéTHod o Log IE Pour L'EXTr Ac Tion d'ES SEr IES d'ES Annu AIr ES

La numérisation de documents anciens est un enjeu important pour les services d'archives, bibliothèques, historiens, chercheurs... (Coüasnon *et al.* 2003). De nombreuses techniques de traitement et d'analyse d'images ont été développées ces dernières années, afin de sauvegarder et d'exploiter les informations consignées sous forme papier. Les météorologues (Dandin *et al.*, 2012) se sont engagés dans le sauvetage d'archives et l'extraction de données anciennes. Un exemple est donné par le projet initié par Météo France, les Archives Nationales et la Fondation BNP Paribas et baptisé AAA « Accès aux Archives du climat

en dépit de l'Amiante » (<http://www.bnpparibas.com/fondation-bnp-paribas/recherche-changement-climatique>) qui vise à exhumer les observations des fonds documentaires du service météorologique français, stockés actuellement dans les locaux de Fontainebleau des Archives Nationales.

Le recours à un logiciel d'OCR (Optical Character Recognition ou encore Reconnaissance Optique de Caractères) est incontournable afin d'automatiser au maximum l'extraction. L'OCR est un procédé qui va transformer un fichier image (feuille imprimée, feuillet dactylographié) en un fichier informatique texte exploitable pour des traitements ultérieurs. De nombreux logiciels existent sur le marché. Ils éprouvent des difficultés à extraire des données manuscrites ou issues de supports anciens partiellement endommagés mais sont généralement bien adaptés pour les documents imprimés assez récents (XXe siècle). Nous avons donc choisi d'en utiliser un parmi une sélection de logiciels commerciaux disponibles après des premiers tests effectués sur des feuillets extraits des annuaires. Le travail accompli ici, à la charnière de deux métiers (Gestion électronique des documents et hydrologie) a bénéficié de l'expérience antérieure acquise dans un contexte très différent d'un des coauteurs (Leblois, 2008).

L'obtention des séries présentes dans les annuaires s'est effectuée en trois étapes principales (Figure 2) : (1) la numérisation des fascicules originaux, (2) le traitement des fichiers numérisés par le logiciel d'OCR et (3) la correction et la vérification des données obtenues. Ces trois étapes sont décrites dans les sections suivantes. Notons que nous avons modifié les données publiées lorsqu'un erratum les concernant a été édité.

Afin de ne pas engager inutilement et de manière exhaustive (i.e. sur les 31 années) les étapes (2) et (3) (les plus fastidieuses), nous avons cherché à identifier quelles stations présentes dans les annuaires déjà étaient disponibles dans la banque HYDRO, sur la base des métadonnées disponibles

Tableau 2 : Liste des 14 articles scientifiques publiés dans les annuaires

Annuaire	Auteur	Titre	Commentaire
1939	Pardé M.	L'hydrologie fluviale française	notions générales sur le régime des cours d'eau (bilan annuel, variations saisonnières, étiages, crues, transport solide)
1940	Massé P.	Situation, perspectives et applications de l'hydrologie statistique	Etat des lieux encore embryonnaire sur l'hydrologie statistique (lois de probabilité, oscillations climatiques, corrélations temporelle et spatiale, prévision à court et long terme)
1940	Pardé M.	Les crues de 1940	Source d'information intéressante sur les crues du Cher (mai) Garonne (mai et déc.), Isère (sept.) et surtout Pyrénées orientales (oct.)
1941	Halphen E.	Etude statistique des débits du Rhin à Bâle	Analyse statistique (débits moyens annuel et mensuel) d'une longue série de débits sur le Rhin à Bâle (1808-1925)
1942	de Martonne E. et Bagnens S.	Deux années sèches : 1921 et 1942	Comparaison des débits mensuels avec les débits moyens, liens avec le régime pluviométrique, variabilité spatiale
1943	Waeber H.	Observations sur le régime des torrents alpestres en haute altitude et spécialement d'un torrent glaciaire	Analyse des premières mesures (1940-1943) sur un petit bassin versant (4 km ²) juste en aval du glacier de Tré-la-Tête
1944	Tissier J. et Zaccagnino E.	Monographie hydrologique de la Haute Dordogne	Caractéristiques du bassin en termes de relief, géologie, température, pluie, débit, crues (8 crues historiques de 1783 à 1912)
1944	Léo B.	Mesure et estimation des débits	Recommandations pour l'hydrométrie (jaugeage, transfert spatial, estimation à partir des pluies)
1945	Halphen E.	Un exemple d'application des méthodes statistiques : le problème du plan de développement de la production d'énergie électrique	Exposé pédagogique sur les statistiques appliquées à l'hydrologie (notions de régions hydrologiques, simulation de séries, optimisation du productible électrique)
1946	Halphen E., Morlat G. et Le Cam L.	Sur la valeur industrielle d'une chute d'eau	Optimisation basée sur une fonction de plusieurs variables (coût de l'aménagement, bénéfice pour le soutien d'étiage, économie de charbon)
1947	Serra L.	Evolution de l'hydrologie : de la statistique à la physique	Plaidoyer pour introduire des notions physiques en hydrologie. Analyse de la différence de comportement de 3 bassins versants sur l'Aubrac en fonction de leurs caractéristiques géologiques
1948	Coutagne A.	Quelques considérations générales sur la nature, les possibilités et les modalités des prévisions en hydrologie fluviale	Possibilités de prévisions hydrologiques à différentes échéances (régime interannuel, causalité physique – régime nival, autocorrélation)
1949	Reménieras G.	L'hydraulique des stations limnimétriques pour la mesure du débit des cours d'eau	Article de référence sur les relations hauteur-débit pour l'hydrométrie (écoulement uniforme, non uniforme permanent, non permanent). Cas des stations à deux échelles, courbe d'hystérésis
1951	Gaspard B.	Note sur la détermination des débits de la Seine à Paris	Modalités d'estimation du débit de la Seine à Paris et graphiques de débit (années 1944 à 1951). Différences notables avec les débits de la banque HYDRO

(confrontation des noms, coordonnées, surfaces). Lorsque nous soupçonnions avoir identifié un équivalent dans la banque HYDRO, l'examen des séries a permis de conclure : — lorsque des observations étaient disponibles sur une ou plusieurs année(s) commune(s), nous avons calculé les valeurs absolues des écarts entre les débits journaliers de la station de l'annuaire sur quelques années et ceux de la station de la banque HYDRO qui semblait correspondre.

Nous avons classé les valeurs et considéré les deux stations identiques lorsque le dernier décile est égal à zéro. Il en est de même lorsque le dernier décile est supérieur à zéro mais que la moyenne des différences relatives est inférieure à 5 %. Lorsque la moyenne des écarts est supérieure à 5 % mais reste inférieure à 10 %, nous avons multiplié les années de comparaison et, finalement, une expertise hydrologique permet de statuer ;

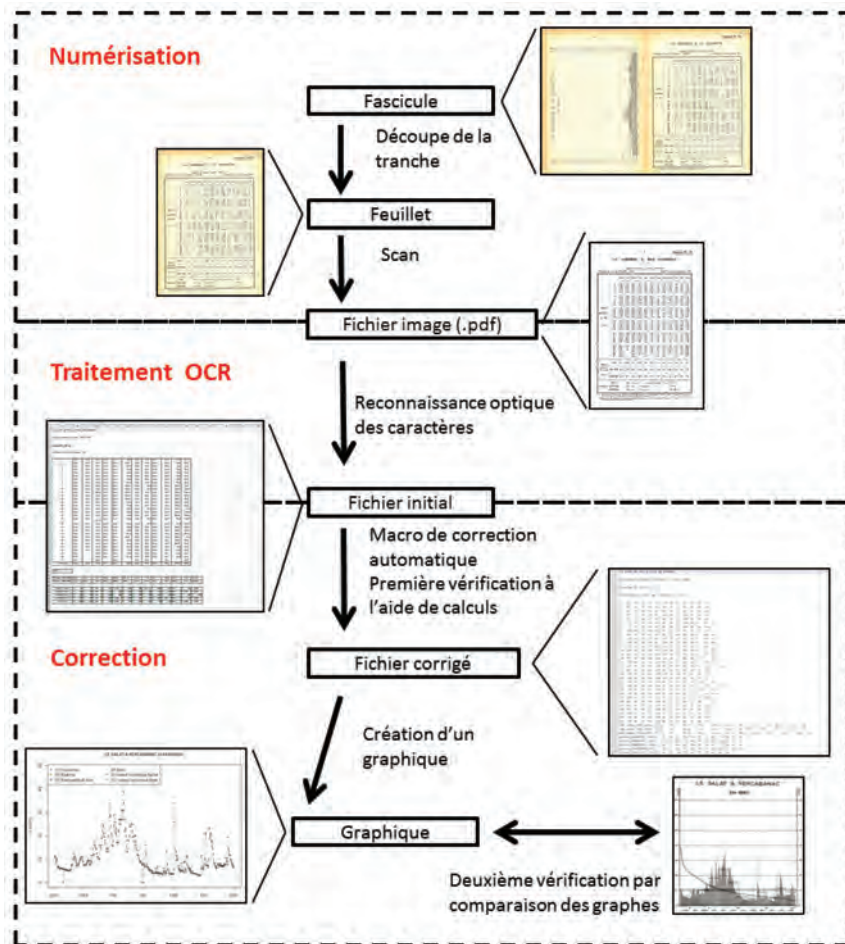


Figure 2 : Schéma récapitulatif des principales étapes du protocole d'extraction des données.

— lorsque la comparaison directe n'était pas possible (pas de chevauchement entre les chroniques des deux stations), des tests de stationnarité (Pettitt, Buishand, Lee et Heghinian, Hubert) ont été appliqués pour nous éclairer sur l'homogénéité de la série concaténée.

III.1. numérisation des fascicules

La première phase de la numérisation consiste à scanner les documents, après massicotage des originaux (découpe de la tranche des annuaires). Le choix de la résolution du scan est un critère technique important à définir, déterminant pour la suite des opérations. Compte tenu des documents (état physique des annuaires, polices d'écriture anciennes), la résolution du scan a été fixée à 600 dpi, avec détection automatique de couleurs puis binarisation en noir et blanc des pages. Le massicotage et le scan des feuillets ont été externalisés à une société spécialiste de la numérisation. Plus de 9000 pages A4 ont ainsi été scannées et converties au format pdf. Une version dégradée de ces fichiers (300 dpi), plus légère, a été mise en ligne pour consultation et téléchargement sur le site de la SHF (http://www.shf-hydro.org/annuaires_hydrologiques-73.html).

III.2. Traitement par un logiciel d'ocr

Une fois les pages numérisées, les images scannées ont été converties en données numériques exploitables.

Si l'ensemble des pages des 31 annuaires hydrologiques a été scanné, seuls les tableaux présentant des données de débits journaliers ont été traités par OCR pour associer à chaque tableau de données (en format pdf) une feuille de calcul de tableau contenant les données numérisées.

L'utilisation d'un modèle de zone permettant d'appliquer à toutes les pages la même structure d'éléments à reconnaître (textes, tableaux) a facilité le traitement. Cependant un ajustement manuel a été toutefois nécessaire à la marge sur toutes les pages afin que les zones soient bien positionnées (ex. décalage léger entre chaque page, nombre de lignes des tableaux non uniforme...). Il a fallu également corriger les erreurs de structuration des tableaux (lignes à ajouter ou supprimer, cellules fusionnées à séparer...). Nous avons mentionné précédemment l'introduction progressive de nouvelles données dans les tableaux synthétiques (Tableau 1) et les formats des tableaux et les polices d'écriture utilisées ont évolué (Figure 3). Ces transformations ont compliqué l'étape de traitement par OCR, car il a fallu régulièrement réadapter les paramètres du logiciel utilisés (en particulier le modèle de zone).

III.3. contrôle des fichiers de sortie

L'agencement des pages, la présence de tableaux plus ou moins bien délimités, l'alternance de texte et de chiffres, les polices anciennes, rendent difficile le travail de reconnaissance de caractères : les erreurs « d'interprétation »

Station N° 9

LE TAURION A PONTARION

Surface du bassin versant : 389 km² Station en service depuis 1939

	JANV.	FÉVR.	MARS	AVR.	MAI	JUIN	JUIL.	AOÛT	SEPT.	OCT.	NOV.	DÉC.	
Débites journaliers en 1939	1	23.7	15.6	10.9	10.	4.9	3.9	4.2	3.9	1.9	3.9	7.4	10.5
Débites mensuels (m³/s)	16	15.2	8.3	12.8	6.4	16.9	4.2	3.3	1.9	2.6	7.8	19.6	10.
Débites moyens mensuels (m³/s)	1939	22.7	10.8	11.5	7.	8.9	5.2	3.3	2.5	2.1	6.8	16.9	11.
Modèles	1919-1939	9,1 m ³ /s, soit 23,4 l/s/km ² , soit une lame d'eau de 0 ^m ,735.											
	1919-1939	5,1 m ³ /s, — 21 l/s/km ² , — — — 0 ^m ,656.											
	1920-1939	8 m ³ /s, — 20,1 l/s/km ² , — — — 0 ^m ,646.											

— 117 —

STATION EN SERVICE DEPUIS 1932

SURFACE DU BASSIN VERSANT : 389 km² ALTITUDE DU ZÉRO DE L'ÉCHELLE : 439 m

	JANV.	FÉV.	MARS	AVRIL	MAI	JUIN	JUIL.	AOÛT	SEPT.	OCT.	NOV.	DÉC.	
ÉCOULEMENT	1	27.0	8.4	8.8	6.1	5.9	1.93	2.1	2.6	3.3	4.4	8.1	6.1
Moyennes	16	15.2	8.3	12.8	6.4	16.9	4.2	3.3	1.9	2.6	7.8	19.6	10.
Précipitation	163	128	126	31	88	170	146	253	165	435	132	183	1 000
ÉCOULEMENT	1919-1939	9,1 m ³ /s, soit 23,4 l/s/km ² , soit une lame d'eau de 0 ^m ,735.											
Modèles	1919-1939	5,1 m ³ /s, — 21 l/s/km ² , — — — 0 ^m ,656.											
	1920-1939	8 m ³ /s, — 20,1 l/s/km ² , — — — 0 ^m ,646.											

Figure 3 : Exemple de deux formats de tableaux pour les données d'une même station en 1939 (sur la gauche) et en 1960 (sur la droite).

sont fréquentes et les données obtenues ne sont pas assez fiables pour être directement exploitées. Un contrôle des fichiers de sortie est donc essentiel pour s'assurer de la qualité des données numérisées, et le cas échéant les erreurs commises par le logiciel doivent être corrigées. Ce contrôle qualité a été réalisé avec un tableur et a nécessité le codage de différentes macros.

Nous avons rapidement pu identifier des erreurs récurrentes commises par le logiciel d'OCR, ce qui nous a permis de réaliser une procédure semi-automatique d'identification et de correction des erreurs. Parmi ces erreurs types, citons : — l'ajout de caractères « i », « / », « | », « ! », « : », ... sur des lignes des tableaux lorsque le trait de bordure de la colonne était mal identifié, — le remplacement d'un « 0 » par un « O », de « 11 » par un « U », de « 6 » par un « b »...

Ces premières corrections ont été complétées par deux autres phases de vérifications successives : l'une à l'aide de calculs et l'autre à l'aide des graphiques. Les annuaires présentent à la fois les données journalières de débits et les moyennes mensuelles correspondantes. Il est possible de recalculer ces moyennes mensuelles à l'aide des données journalières, et vérifier qu'elles sont similaires à la moyenne mensuelle indiquée dans le document source. Si ce n'est pas le cas, nous pouvons soupçonner une erreur d'identification de la part du logiciel sur des valeurs journalières intervenant dans le calcul en question.

Cette vérification par le calcul présente l'intérêt d'identifier rapidement des erreurs d'OCR liées au remplacement d'un chiffre par une lettre (le calcul de moyenne ne se fait pas). Toutefois, cette vérification n'est pas toujours très efficace en cas de confusion d'un chiffre avec un autre. Une seconde étape de vérification nous a donc paru nécessaire.

En complément, nous avons donc exploité les informations contenues dans les graphiques adjacents aux tableaux de données annuels. Des hydrogrammes ont été élaborés à partir des séries issues de l'étape précédente, ce qui nous a permis de réaliser la seconde étape de vérification de l'OCR. La comparaison visuelle des deux graphiques permet de déceler de nouvelles erreurs.

III.4. r é s u l t a t s

Nous avons identifié près de 2300 tableaux synthétiques contenant les 365 ou 366 débits journaliers observés sur un sous-ensemble de stations dans les 31 annuaires hydrologiques. L'examen des redondances « évidentes » avec les données de la banque HYDRO (interrogation en mai 2015) a réduit le nombre de tableaux à environ 1600, répartis sur une centaine de stations hydrométriques.

Ainsi, sur les 120 stations décrites dans les annuaires, trois d'entre elles sont implantées en Suisse (Rhin à Rheinfelden, Doubs à Ocourt, Arve au Pont de Carouge). Les données associées sont mises à disposition pour la période 1993-2015 sur le site de l'Office fédéral de l'environnement OFEV (<http://www.hydrodaten.admin.ch/fr>). Les données du Rhône à Penay et à Chancy (Aux Ripes) sont bancarisées en France et en Suisse. La Figure 4 fait apparaître l'ensemble des stations pour lesquelles les débits sont disponibles dans les annuaires. Sans surprise elles sont situées majoritairement dans les secteurs à enjeu énergétique. Finalement, 15 points de mesure semblent correspondre à des stations inédites, non présentes dans la banque HYDRO (Tableau 3 ; Figure 4), et 41 séries apportent des données complémentaires à des stations de la banque HYDRO (Tableau 4 ; Figure 4). Un total

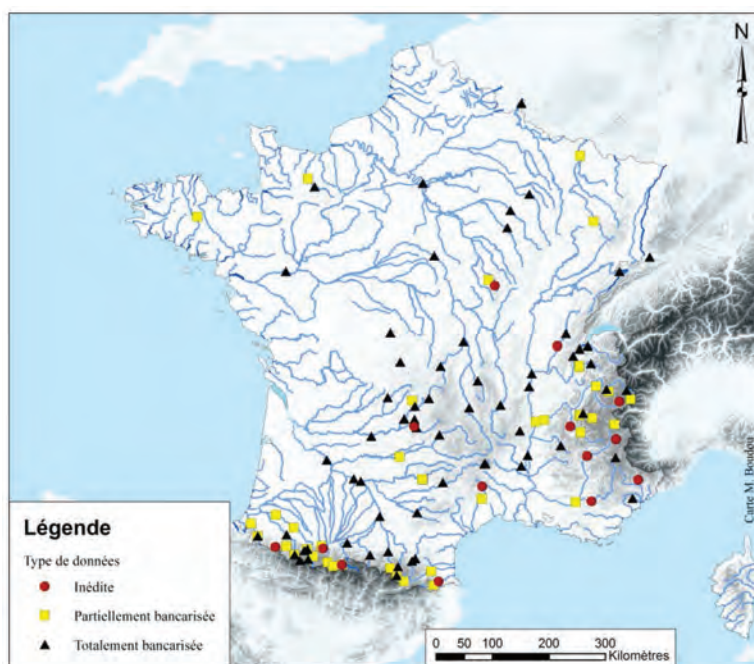


Figure 4 : Localisation et statut des stations présentes dans les annuaires.

Tableau 3 : Descriptif des stations inédites.

n° (jusqu'en 1956)	n° (à partir de 1957)	nom (annuaire)	Surface (km ²)	date de mise en service	chronique disponible
3		LA CURE AUX SETTONS	50	1926	1939
16		LA CERE A MONTVERT (Usine de Lamativie)	764	1909	1939-1948
24		LA PIQUE A LA PIQUE SUPÉRIEURE	50.5/58.6	1931	1939-1947
33		L'ADOUR A ARTHEIL	160.6	1912	1939-1941
36bis	36-1	LE GAVE DE SAINTE ENGRACE A LICQ-ATHEREY	61		1956-1969
41		LE TECH A CANTAÏRÉ	393	1920	1939-1940
	42-3	L'HERAULT A LA VALETTE	109		1957-1958
44	44	L'AIN A CIZE-BOLOZON	2560	1923	1939-1969
56		L'ARC A TERMIGNON	359	1924	1939
58		LE DRAC A AVIGNONET	1968/2006.7	1904	1939-1954
63	63	LA DURANCE A VENTAVON	4215	1912	1939-1969
65		LE GUIL A PONT-LA-PIERRE	475	1914	1939-1948 ; 1952
65bis	65-1	LE GUIL A LA MAISON DU ROY	580		1953-1958 ; 1960-1969
67	67	LE VERDON A QUINSON	1661	1906	1939-1969
68	68	LA TINÉE A BANCAIRON	532/450	1929	1939-1969

de 577 années stations inédites ont été exhumées (209 sont associées aux stations inédites).

III.5. Problèmes rencontrés

Le travail d'extraction des données des annuaires hydrologiques publiés par la SHF a soulevé un certain nombre de difficultés, malgré l'existence d'outils d'OCR performants.

Certaines d'entre elles sont décrites ici, afin de donner une idée aux lecteurs des complications qu'ils pourraient rencontrer lors de l'extraction de données anciennes. La qualité de l'impression est un paramètre déterminant pour une bonne reconnaissance par OCR. Certains tableaux, rédigés à la machine à écrire, ne présentaient pas toujours une qualité suffisante pour nous permettre d'être sûrs de ne pas faire d'erreur de lecture : décalage de lignes, caractères effacés...

Tableau 4 : Descriptifs des stations bancarisées partiellement.

n° (jusqu'en 1956)	n° (à partir de 1957)	nom (annuaire)	Surface (km²)	date mise en service	c hronique disponible	r éférence Hydro	nom (Hydro)	Surface Hydro (km²)	c hronique disponible Hydro
2		LA MOSELLE A ÉPINAL	1250	1891	1939-1941	A4250640	LA MOSELLE A ÉPINAL	1217	1960-2015
2bis	2-1		1215	1952	1954-1969				
3	2-2	LA MOSELLE A HAUCONCOURT	9400	1956	1957-1969	A7930610	LA MOSELLE A HAUCONCOURT	9387	1960-2013
4	3	LE CHALAUX A CHAUMEÇON	98/100	1935	1940-1969	H2133010	LE CHALAUX A SAINT-MARTIN-DU-PUY	98	1946-1981
	4	LE BLAVET A GUERLÉDAN	620	1912	1939-1969	J5412110	LE BLAVET A MUR-DE-BRETAGNE [GUERLEDAN]	620	1948-2011
	4-1	L'ORNE A COSSESSEVILLE	2040	1955	1957-1963	I3501010	L'ORNE A CLECY	2060	1977-1979
14bis	14-1	LA LUZEGE A PONT DES BOUYGES	252	1951	1951-1969	P1154010	LA LUZEGE A LAMAZIERE-BASSE [PONT DE BOUYGES]	252	1961-2011
	21-2	L'ARIÈGE A L'HOSPITALET AMONT	45	1961	1961-1969	O1002510	L'ARIÈGE A L' HOSPITALET-PRES-L'ANDORRE [AMONT]	45	1967-1973
23		L'ASTON A CHATEAU-VERDUN	162.2	1911	1939-1941 ; 1946-1950	O1074320	L'ASTON A CHÂTEAU-VERDUN	52.8	1948-1983
24bis	24-1	LA PIQUE A RAVI	81	1921	1953-1969	O0024010	LA PIQUE A BAGNERES-DE-LUCHON [RAVI, USINE]	81	1960 - 1974
25bis		LA NESTE DE RIOUMAJOU A SAINT-LARY	63.7	1942	1943-1947	O0126210	LA NESTE DE RIOUMAJOU A TRAMEZAIGUES [MAISON BLANCHE]	63.7	1948-2011
						O0126220	LA NESTE DE RIOUMAJOU A SAINT-LARY-SOULAN [I]	67	1939-1941
26	26	LA NESTE DE CLARABIDE A LASSOULA	16	1933	1939-1969	O0155010	LA NESTE DE CLARABIDE A LOUDENVIELLE [RIOUMAJOU]	16	1960-1974
29	29	LE VIAUR A THURIES / THURIES-ROC-MIQUEL	1050	1921	1939-1969	O5482920	LE VIAUR A SAINT-JUST-SUR-VIAUR [ROC MIQUEL]	1008	1962-2008 ; 2010-2011
						O5482925	LE VIAUR A SAINT-JUST-SUR-VIAUR [LE CAMBON ANCIEN]	1010	1930-1940
						O5482930	LE VIAUR A SAINT-JUST-SUR-VIAUR [LE CAMBON]	1010	1930-1940 ; 1991- 2015
31	31	LE LOT A CAJARC	7033	1913	1939-1945 ; 1949-1969	O7981510	LE LOT A CAJARC [PONT]	7013	1913-1942
						O7991510	LE LOT A CAJARC	7033	1960-1983
33bis	33-1	L'ADOUR A ASTÉ	272	1911	1942-1969	Q0100025	L'ADOUR [ADOUR+CANAL] A ASTE [PONT D ASTE]	272	1950-2015

Tableau 4 (suite)

n° (jusqu'en 1956)	n° (à partir de 1957)	nom (annuaire)	Surface (km²)	date mise en service	c hronique disponible	r éférence Hydro	nom (Hydro)	Surface Hydro (km²)	c hronique disponible Hydro
34	34	LE GAVE DE BROUSSET AUX ALLIAS	54/61/63	1912	1939-1969	Q6014010	LE GAVE DE BROUSSET A LARUNS [ALLIAS]	63	1960-1974
35	35	LE GAVE D'ASPE AU PONT D'ESCOT	425	1911	1939-1969	Q6332510	LE GAVE D'ASPE A BEDOUS [PONT D'ESCOT]	425	1948-2011
37	37	LE GAVE DE GAVARNIE A LUZ	236	1927	1939-1959	Q4231010	LE GAVE DE PAU A LUZ-SAINTE-SAUVEUR	274	1909-1926 ; 1929-1939 ; 1941-1959
38		LE GAVE DE PAU A ARTIGUELOUVE	1905	1920	1939-1941	Q5231010	LE GAVE DE PAU A ARTIGUELOUVE [PONT DE L'ESCAR]		2012-2015
38bis	38-1	LE GAVE DE PAU A PONT-DE-BÉRENX	2575	1924	1944-1969	Q5501010	LE GAVE DE PAU A BERENX [PONT DE BERENX]	2575	1923-1945 ; 1948-2015
39 ou 39ter	39-2	LE GAVE D'AZUN A ARRAS A LAU-BALAGNAS	191	1920	1939-1947 1951-1959	Q4674010	LE GAVE D AZUN A ARRAS-EN-LAVEDAN [ARRAS-NOUAUX]	191	1960-1986 ; 1989-2011
40		LA NIVE A CAMBO	839,5	1904	1939-1941	Q9302510 Q9312510	LA NIVE A ITXASSOU LA NIVE A CAMBO-LES-BAINS	850 870	1967-1973 ; 1975-1999 1999-2000 ; 2004-2015
40bis	40-1	LA NIVE DE BÉHÉROBIE A SAINT-JEAN-PIED-DE-PORT	115	1912	1942-1969	Q9034610	LA NIVE DE BEHEROBIE A SAINT-JEAN-PIED-DE-PORT	115	1918-1951 ; 1953-1968 ; 1970-1991 ; 1994-2000
41bis		LA TÊT A VINÇA	900	1936	1941	Y0464040	LA TET [TOTALE] A VINÇA	942	1968-1974
41ter	41-2	LE TECH A PAS-DU-LOUP	236	1935	1942-1969	Y0224011	LE TECH A MONTFERRER [PAS DU LOUP]	250	2005-2015
42ter	42-2	L'ANGOUSTRINE A ANGOUSTRINE	46	1913	1956-1969	Y0025010	L'ANGOUSTRINE A ANGOUSTRINE-VILLENEUVE-DES-ESCALDES	45.9	1970-1987 ; 2005-2015
	42-4	L'HERAULT AU MOULIN DE BERTRAND	1090	1928	1959-1969	Y2132010	L'HERAULT A CAUSSE-DE-LA-SELLE [MOULIN DE BERTRAND]	1090	1967-1981

Tableau 4 (suite)

n° (jusqu'en 1956)	n° (à partir de 1957)	nom (annuaire)	Surface (km²)	date mise en service	chronique disponible	référence Hydro	nom (Hydro)	Surface Hydro (km²)	chronique disponible Hydro
52		LE FIER A MOTZ - VAL DE FIER	1370	1911	1939-1942	V1264010	LE FIER A VALLIERES	1350	1948-2011
52bis	52-1	LE FIER A VALLIÈRES	1350	1943	1943-1969				
54		L'ISÈRE A BEAUMONT-MONTEUX	11500	(1918)-1921	1939-1951	W3540010	L'ISERE A BEAUMONT-MONTEUX	11800	1956-2013
54bis	54-1	L'ISERE A PIZANÇON	11320	1932	1952-1969				
55	55	LE DORON DE BOZEL A LA PERRIERE	330/301	1930	1939-1969	W0224010	LE DORON DE BOZEL A LA PERRIERE [VIGNOTAN]	330	1948-2003
56bis		L'ARC A AIGUEBELLE	1928	1930	1941-1944	W1074030	L'ARC A AIGUEBELLE	1946	2011-2015
56ter	56-2	L'AVEROLE A AVEROLE	46	1948	1952-1969	W1006010	L'AVEROLE A BESSANS [AVEROLE]	45.4	1969-1995 ; 1998-2011
57	57	LE DRAC AU SAUTET	990/1020	1904	1939-1969	W2222010	LE DRAC A CORPS [LE SAUTET]	990	1969-1999
59	59	LA ROMANCHE AU CHAMBON	220	1936	1939-1969	W2714010	LA ROMANCHE A MIZOËN [CHAMBON AMONT]	220	1948-2011
60		LA ROMANCHE A GAVET	1047	1918-1906	1939-1943 ; 1945-1946	W2764010	LA ROMANCHE AU BOURG-D OISANS [CHAMPEAU]	1000	1951-2011
60bis	60-1	LA ROMANCHE A RIOUPÉROUX	1025	1944 (à partir de 1947 : 1906)	1944 ; 1947-1957				
62	62	LA DURANCE A LA VACHETTE	210	1917	1939-1969	X0010010	LA DURANCE A VAL-DES-PRES [LES ALBERTS]	203	1975-2015
						X0010020	LA DURANCE A VAL-DES-PRES [LA VACHETTE]	210	1950 ; 1952 ; 1954-1956 ; 1958-1973 ; 1975
64	64	LA DURANCE A MIRABEAU	11920	1904	1939-1959	X3000010	LA DURANCE A SAINT-PAUL-LES-DURANCE [JOUQUES-CADARACHE]	11700	1918-1958 ; 1969-2011
	64-1	LA DURANCE A JOUQUES-CADARACHE	11920	1904	1960-1969				

L'ancienneté des documents que l'on souhaite récupérer peut, en elle-même, poser problème en raison de la difficulté de vérifier des informations datant de plusieurs dizaines d'années. Il faut accepter qu'il existe une part de doute dans les données récupérées, en raison de l'impossibilité de vérifier d'éventuelles erreurs. Nous avons pu constater des incohérences entre les tableaux de données et les graphiques correspondant, et entre les moyennes indiquées dans les tableaux et les moyennes recalculées. Il est impossible de connaître la raison de ces incohérences (erreur de calcul, faute de frappe, correction ultérieure d'une donnée...). Quand les erreurs étaient évidentes, elles ont été corrigées. Le temps nécessaire pour la récupération des données par annuaire a été très variable : il a évolué entre 4 heures pour la décennie 1940 et 10 heures pour la décennie 1960. Cette augmentation est pour partie liée au nombre de données à extraire, et donc au nombre de données à vérifier. Par exemple un tableau présentant les valeurs de débit mensuel maximum et minimum sur une période longue requiert plus de temps pour la vérification, car il contient un grand nombre de données non vérifiables par le calcul, et des chiffres à saisir manuellement (les années, inscrites ici sur un fond grisé, n'ont pas été bien reconnues par l'OCR). La complexité croissante des tableaux a également allongé considérablement le temps d'ajustement des modèles de zone.

Les annuaires présentent des données de débits journaliers sur 120 stations hydrologiques, pour la période 1939-1969. Durant cette période, les stations hydrologiques ont évolué : certaines stations ont été déplacées, remplacées par d'autres, ont changé de nom car la dénomination n'était plus appropriée... Les annuaires ne sont pas toujours très explicites sur ces évolutions, et des erreurs se glissent régulièrement au milieu d'informations exactes. Il y a ainsi beaucoup d'incohérences temporelles sur les numéros et noms de station, les valeurs d'altitude, de surface du bassin versant, les dates de mise en service, les commentaires... Il est donc parfois difficile de savoir si des chroniques de débits journaliers mesurés deux années consécutives l'ont été à la même station, ou si la station était différente la seconde année. Le problème de fiabilité des informations prend une autre dimension quand il s'agit de confronter deux sources de données. Nous avons cherché à identifier des stations identiques entre les annuaires et la banque HYDRO. Outre l'évolution des dénominations des stations et les incohérences des informations présentées dans les annuaires, des corrections réalisées ultérieurement sur les données de la banque HYDRO (modification des courbes de tarage par exemple), ont pu creuser l'écart entre les valeurs extraites des deux sources, pourtant initialement identiques. Pour ces deux raisons, il persiste un doute sur la correspondance réelle entre certaines stations décrites dans les annuaires et les stations décrites dans la banque HYDRO.

IV. conc Lu Slon

Les séries hydro-météorologiques anciennes ont une valeur manifeste pour la connaissance générale du cycle de l'eau et de manière pragmatique pour la mise en place de moyens de prévention contre les événements extrêmes (peu échantillonnés dans les séries courtes). Elles permettent de mieux apprécier la variabilité temporelle des débits dans une perspective de changement climatique. Elles sont le support de validation direct ou indirect de travaux de reconstitution de séries climatiques en France (Caillouet *et al.*,

2015 ; Kuentz *et al.*, 2015 ; Minvielle *et al.*, 2015). La prise en compte de ces séries anciennes offre enfin une source d'information non négligeable pour parfaire la connaissance d'événements exceptionnels passés, souvent peu connus et dont l'occurrence future demeure un enjeu de taille pour les gestionnaires du risque (Boudou, 2015).

La communauté des météorologues s'est déjà mobilisée pour identifier des sources potentielles de données, numériser les archives et sauvegarder les observations sur un support numérique. Nous avons souhaité engager une démarche similaire en se focalisant sur les annuaires hydrologiques publiés par la Société Hydrotechnique de France, qui avaient l'intérêt de centraliser une information multi-sites sur une période suffisamment ancienne pour espérer identifier des données inédites.

Les annuaires de 1939 à 1969 ont été numérisés et sont désormais accessibles (http://www.shf-hydro.org/annuaires_hydrologiques-73.html). Le contenu de ces documents a été détaillé ici. Outre les données pour chaque année, des contributions libres de différents auteurs apportent des éléments d'expertise utile pour les hydrologues. La récupération des séries via un procédé d'OCR est rappelée. Il requiert une phase de contrôle qui peut être longue et fastidieuse selon la complexité du support de base numérisé. Elle a porté ses fruits : nous avons pu exhumer un ensemble de 56 stations vraisemblablement inédites ou apportant des compléments possibles de chroniques pour des stations partiellement bancarisées (un équivalent d'environ 58 années stations), soit près de 50 % des stations mentionnées dans les annuaires.

Il reste à définir une stratégie de mise à disposition de ces données. Elles ne sont pas forcément directement exploitables. Un travail de qualification des données pour une exploitation optimale devra être engagée en s'appuyant sur les commentaires relevés sur les annuaires (sauvegardés également) et au regard des techniques de mesure employées en s'inspirant des travaux de Kuentz (2013). Ce travail de sauvegarde permettra dans un premier temps de valider un modèle hydrologique élaboré sur le Rhône (projet « Modélisation Hydrologique Distribuée du Rhône ») sous réserve de réajustement climatique sur des stations « inédites » ou qui présentent un régime influencé sur la période récente.

V. r EMERc IEMEn TS

Cette action d'extraction des séries a été réalisée dans le cadre du projet de recherche « Modélisation Hydrologique Distribuée du Rhône », qui vise à déployer un modèle distribué à base physique sur le Rhône et ses affluents majeurs (partie française). Ce projet est soutenu financièrement par l'Agence de l'Eau Rhône-Méditerranée-Corse, la Compagnie Nationale du Rhône et le fond européen FEDER (action Présage 45537).

VI. r EFER Enc ES

BOUDOU M. (2015) — Caractérisation de neuf événements remarquables d'inondations en France au XX^e siècle : contribution à la valorisation de l'information historique pour la gestion prospective des risques. Thèse, Montpellier 3 *en cours*.

CAILLOUET L., VIDAL J.-P., SAUQUET E., GRAFF B. (2015) — Probabilistic precipitation and temperature downscaling of

- the Twentieth Century Reanalysis over France. *Climate of the Past*, submitted
- COTE E.-F. (1908) — Le service d'étude des grandes forces hydrauliques. *La Houille Blanche*. **12** : 284-287
- COÛASNON B., DALBERA J.-P., EMPTOZ H. (2003) — Numérisation et patrimoine. *Document numérique*. **7(3-4)** : 188 pages
- DANDIN P., ARESSY P., DEAUX N., DUBUISSON B., FLEUTER G., GIBELIN A.-L., JOURDAIN S., LAVAL L., MENASSERE S., ROUCAUTE E. & WIECZOREK A.-M. (2012) — Data rescue: a necessary look at climate. In *Climate ExChange, WMO*, (<http://digital.onwindows.com/climate-exchange/>). 268-271
- GIUNTOLI I., RENARD B., VIDAL J.-P. BARD A. (2013) — Low flows in France and their relationship to large-scale climate indices. *Journal of Hydrology*. **482** : 105-118
- KUENTZ A. (2013) — *100 ans de variabilité climatique observée et reconstituée sur le bassin de la Durance*. Thèse, AgroParisTech, Paris. 373 pages
- KUENTZ A., MATHEVET T., GAILHARD J., HINGRAY B. (2015) — Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM method. *Hydrol. Earth Syst. Sci. Discuss.* **12** : 311-361
- LANG M., ARNAUD P., CARREAU J., DEAUX N., DEZILEAU L., GARAVAGLIA F., LATAPIE A., NEPPEL L., PAQUET E., RENARD B., SOUBEYROUX J.M., TERRIER B., VEYSSEIRE J.M., AUBERT Y., AUFFRAY A., BORCHI F., BERNARDARA P., CARRE J.C., CHAMBON D., CIPRIANI T., DELGADO J.L., DOUMENC H., FANTIN R., JOURDAIN S., KOCHANEK K., PAQUIER A., SAUQUET E., TRAMBLAY Y. (2014) — Résultats du projet ExtraFlo (ANR 2009-2013) sur l'estimation des pluies et crues extrêmes. *La Houille Blanche*. **2(5)** : 13
- LEBLOIS E. (2008) — Extraction de lexique bilingue à partir d'un corpus de traduction : une stratégie par écrémage. *Actes des 9^{èmes} Journées internationales d'Analyse statistique des Données Textuelles, 12-14 mars 2008, Lyon, France* (<http://lexicometrica.univ-paris3.fr/jadt/jadt2008/tocJADT2008.htm>)
- MINVIELLE M., PAGE C., CERON J.-P., BESSON F. (2015) — Extension of the SIM Reanalysis by Combination of Observations and Statistical Downscaling. *Engineering Geology for Society and Territory*, doi10.1007/978-3-319-09300-0_36, 2015. **1** : 189-192