



**Recueil des données des annuaires hydrologiques
1939-1969 de la Société Hydrotechnique de France -
Contribution au projet “ Modélisation hydrologique
distribuée du Rhône ”**

C. Le Gros, Eric Sauquet, Etienne Leblois, A.L. Achard

► **To cite this version:**

C. Le Gros, Eric Sauquet, Etienne Leblois, A.L. Achard. Recueil des données des annuaires hydrologiques 1939-1969 de la Société Hydrotechnique de France - Contribution au projet “ Modélisation hydrologique distribuée du Rhône ”. [Rapport de recherche] irstea. 2015, pp.96. hal-02601545

HAL Id: hal-02601545

<https://hal.inrae.fr/hal-02601545>

Submitted on 16 May 2020

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.

Recueil des données des annuaires hydrologiques 1939-1969 de la Société Hydrotechnique de France

CONTRIBUTION AU PROJET « MODELISATION HYDROLOGIQUE
DISTRIBUEE DU RHONE »

Janvier 2015

Chloé LE GROS

Eric SAUQUET

Etienne LEBLOIS

Irstea, UR HHLY

5 rue de la Doua CS70077

69626 VILLEURBANNE Cedex

Anne-Laure ACHARD

Irstea, IST

5 rue de la Doua CS70077

69626 VILLEURBANNE Cedex



RESUME

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 ce rapport sont 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.

SOMMAIRE

1.	Introduction.....	9
2.	Méthodologie.....	10
2.1.	Données disponibles.....	10
2.2.	Protocole utilisé	14
2.2.1.	Numérisation des archives	15
2.2.2.	Définition des paramètres du logiciel.....	15
2.2.3.	Retouche des images.....	15
2.2.4.	Lecture des images	15
2.2.5.	Enregistrement des résultats de l'OCR.....	17
2.2.6.	Modifications de la mise en forme des fichiers de sortie.....	18
2.2.7.	Corrections automatiques et vérifications visuelles.....	19
2.2.8.	Schéma récapitulatif : les étapes du protocole	24
2.2.9.	Codes utilisés dans les chroniques de débits journaliers	25
2.2.10.	Evolution du protocole	25
2.3.	Evaluation du temps nécessaire au traitement des données	27
3.	Problèmes rencontrés	35
3.1.	Manque de clarté des informations disponibles	35
3.2.	Pertinence des vérifications effectuées.....	38
3.2.1.	Vérification par le calcul de moyennes.....	38
3.2.2.	Vérification des débits par visualisation des graphes.....	39
3.3.	Continuité des mesures sur les stations des annuaires	41
3.4.	Correspondance avec les stations de la banque HYDRO	41
3.5.	Analyse par station	42
3.5.1.	Station 2/2bis - La Moselle à Epinal.....	42
3.5.2.	Station 3 - La Cure aux Settons/Le Chaux à Chaumeçon	43
3.5.3.	Station 5ter – La Loire à Montjean	44
3.5.4.	Station 10 - La Creuse à Eguzon.....	44
3.5.5.	Station 12 – La Dordogne à Argentat	45
3.5.6.	Station 13 – La Dordogne à Cénac (Domme).....	45
3.5.7.	Station 15-1 – La Maronne à Basteyroux	45
3.5.8.	Station 17 – La Vézère à Uzerche	46
3.5.9.	Station 18 – La Garonne à Valentine	46
3.5.10.	Station 18bis – La Garonne à Portet.....	47
3.5.11.	Station 18-2/18-3 - La Garonne à Malause-Lamagistère.....	47

3.5.12.	Station 19 – La Garonne au Mas d’Agenais	48
3.5.13.	Station 20 – Le Salat à Kercabanac	49
3.5.14.	Station 22 – L’Ariège à Foix	49
3.5.15.	Stations 25/25bis/25ter - La Neste de Rioumajou.....	49
3.5.16.	Station 27/27bis - Le Tarn au Pont de Montvert/Le Tarn à Fonchalettes	52
3.5.17.	Station 29 - Le Viaur à Thuriès / Thuriès-Roc-Miquel.....	53
3.5.18.	Station 30 – L’Agout à Clot	54
3.5.19.	Station 31 – Le Lot à Cajarc.....	55
3.5.20.	Station 33-1 - L’Adour à Asté.....	55
3.5.21.	Station 34 : Le Gave de Brousset à Miégebat/aux Allias	56
3.5.22.	Station 35 – Le Gave d’Aspe au Pont d’Escot	56
3.5.23.	Station 36 – Le Gave d’Oloron à Oloron	57
3.5.24.	Station 37 - Le Gave de Gavarnie à Luz.....	57
3.5.25.	Station 38 – Gave de Pau à Artiguelouve	58
3.5.26.	Station 38bis- Gave de Pau à Pont de Berenx	58
3.5.27.	Station 39/39-2/39-ter/39-3 - Le Gave d’Azun à Arras / Lau-Balagnas / Arras-Nouaux.....	59
3.5.28.	Station 40 – Nive à Cambo	60
3.5.29.	Station 40bis - La Nive à St-Jean-Pied-de-Port/La Nive de Béhérobie à St-Jean-Pied-de-Port.....	60
3.5.30.	Station 41bis – Têt à Vinca	61
3.5.31.	Station 41-2 – Le Tech à Pas du Loup	61
3.5.32.	Station 42 – L’Aude à Belvianes.....	62
3.5.33.	Station 42-2 – L’Angoustrine à Angoustrine.....	62
3.5.34.	Station 43 - L’Ain à la Chartreuse de Vaucluse/Ain à Chartr. Voulgans.....	63
3.5.35.	Station 47 - Le Rhône à Génissiat	64
3.5.36.	Station 47-1 - Le Rhône à Lyon Poincaré/ Saint Clair	64
3.5.37.	Station 47-2 - Le Rhône à la Mulatière / Mulatière-Givors	64
3.5.38.	Station 49 - Le Rhône au Teil/ Teil (Pouzin).....	65
3.5.39.	Station 52 – Le Fier à Motz – Val de Fier	66
3.5.40.	Station 54 – L’Isère à Beaumont - Monteux	66
3.5.41.	Station 53bis - L’Isère au Pont du Laisinant/ à Val d’Isère - Laisinant	67
3.5.42.	Station 56bis – L’Arc à Aiguebelle.....	67
3.5.43.	Station 56ter – L’Averole à Avrole	68
3.5.44.	Station 57 – Le Drac au Sautet.....	68
3.5.45.	Station 64/64-1 – La Durance à Mirabeau / La Durance à Jouques-Cadarache	69
4.	Synthèse	70
5.	Conclusion	74
1.	Annexe 1 : Fonctionnement du logiciel ABBYY Fine Reader	76

1.1.	Etape 1 : Numérisation du document original	77
1.2.	Etape 2 : Définition des paramètres du logiciel	77
1.3.	Etape 3 : Retouches de l'image.....	77
1.4.	Etape 4 : Lecture de l'image.....	78
1.5.	Etape 5 : Vérification des caractères incertains.....	79
1.6.	Etape 6 : Enregistrement du résultat	80
1.7.	Outils de correction et vérification	80
1.8.	Difficultés et solutions	81
1.8.1.	Vérifier la qualité des documents disponibles.....	81
1.8.2.	Structure de page se répétant : utiliser un modèle de zone	81
1.8.3.	Les zones présentes sur une même page ont des caractéristiques différentes	82
1.8.4.	Police de caractère inconnue : lecture avec apprentissage	83
1.8.5.	Modification de l'ordre des zones dans le fichier de sortie.....	84
1.8.6.	Autres problèmes	85
2.	Annexe 2 : Liste détaillée des étapes du protocole (valable pour les années 1964-1969)	88
3.	Annexe 3 : Offre commerciale de la société IMSO	91

1. Introduction

Le projet « Modélisation Hydrologique du Rhône »¹¹ vise à déployer un modèle distribué à base physique sur le Rhône et ses affluents majeurs (partie française), et d’y intégrer des modules liés à la gestion de l’eau, afin d’obtenir un outil de modélisation pouvant être utilisé dans des exercices de prospective (études d’impact du changement global incluant des modifications du climat et des changements de pratiques liées à la gestion de l’eau). Il est soutenu financièrement par l’Agence de l’Eau Rhône-Méditerranée-Corse, la Compagnie Nationale du Rhône et les fonds européens FEDER. Ce projet nécessite donc de modéliser le fonctionnement hydrologique des cours d’eau dans le bassin versant du Rhône. Or, de nombreux cours d’eau sont aujourd’hui influencés par les actions humaines, principalement en raison de la présence d’aménagements pour la production d’hydro-électricité. Il est donc intéressant d’avoir accès à des données anciennes, car en faisant l’hypothèse que les cours d’eau étaient moins influencés par le passé qu’aujourd’hui, on peut avoir une connaissance approximative du fonctionnement naturel des cours d’eau et ainsi adapter ou contrôler de manière plus juste le modèle hydrologique.

Des données anciennes pouvant contribuer à améliorer la calibration du modèle dans le cadre de ce projet ont été identifiées dans des archives de la Société Hydrotechnique de France (SHF) : il s’agit d’annuaires hydrologiques contenant notamment des débits journaliers pour un ensemble de stations hydrométriques mesurées sur la période 1939-1969. Ces données pourront également être utilisées ultérieurement dans le cadre d’autres projets.

La numérisation de documents anciens est un enjeu important pour les services d’archives, les bibliothèques, les historiens et les chercheurs en sciences littéraires². De nombreuses techniques de traitement et d’analyses d’images ont été développées ces dernières années, afin de sauvegarder et d’exploiter des documents patrimoniaux. Ces techniques sont aussi nombreuses que complexes, et nécessitent de savoir évaluer quelle sera la méthode la plus adaptée selon les caractéristiques du document à traiter. Parmi ces techniques, les logiciels d’OCR (Optical Character Recognition) sont bien adaptés pour les documents imprimés assez récents (fin du XXe siècle). La reconnaissance optique de caractères est un procédé informatique qui permet de transformer un fichier image (feuille imprimée, feuillet dactylographié) en un fichier texte, exploitable ensuite via les méthodes informatiques modernes. La numérisation de ces documents anciens a donc été réalisée grâce à l’utilisation d’un des logiciels OCR. Cette reconnaissance n’est pas exempte d’erreurs. Une étape de vérification des données extraites a donc été rendue nécessaire. La banque HYDRO (www.hydro.eaufrance.fr) centralise les données hydrométriques produites en France depuis la fin du XIXe siècle. La dernière étape du travail a consisté à isoler les stations extraites des annuaires non présentes dans la base de données Hydro.

¹ Tilmant, F., Cipriani, T., Le Gros, C., Barachet, C., Branger, F., Sauquet, E., Braud, I., Leblois, E. and Gouttevin, I. (2014). Modélisation Hydrologique Distribuée du Rhône : rapport d’avancement, Irstea Lyon.

² Couasnon, B., Dalbera, J.-P., Emptoz, H. (2003). Numérisation et patrimoine. Document numérique, 7(3-4), 188 pages.

La numérisation des documents a été externalisée sous la coordination de A.-L. Achard. Le choix du logiciel d'OCR a été effectué après expertise des logiciels sous licence disponibles sur le marché par E. Leblois. L'analyse hydrologique des données a été effectuée avec le soutien de E. Sauquet.

Le présent document présente la méthodologie utilisée pour la récupération de ces données et la confrontation avec les stations de la banque HYDRO. Les difficultés qui ont été rencontrées y sont détaillées. Ce rapport a également pour but de donner un retour d'expérience pouvant être utile à toute personne débutant une tâche de numérisation de données par OCR.

2. Méthodologie

2.1. Données disponibles

Les annuaires hydrologiques de la SHF présentent des données sur la période 1939-1969. Ils contiennent entre autres : une description de l'hydrologie générale de l'année, des contributions libres de différents auteurs, et des séries de données (dont le contenu et la nature varient au cours du temps).

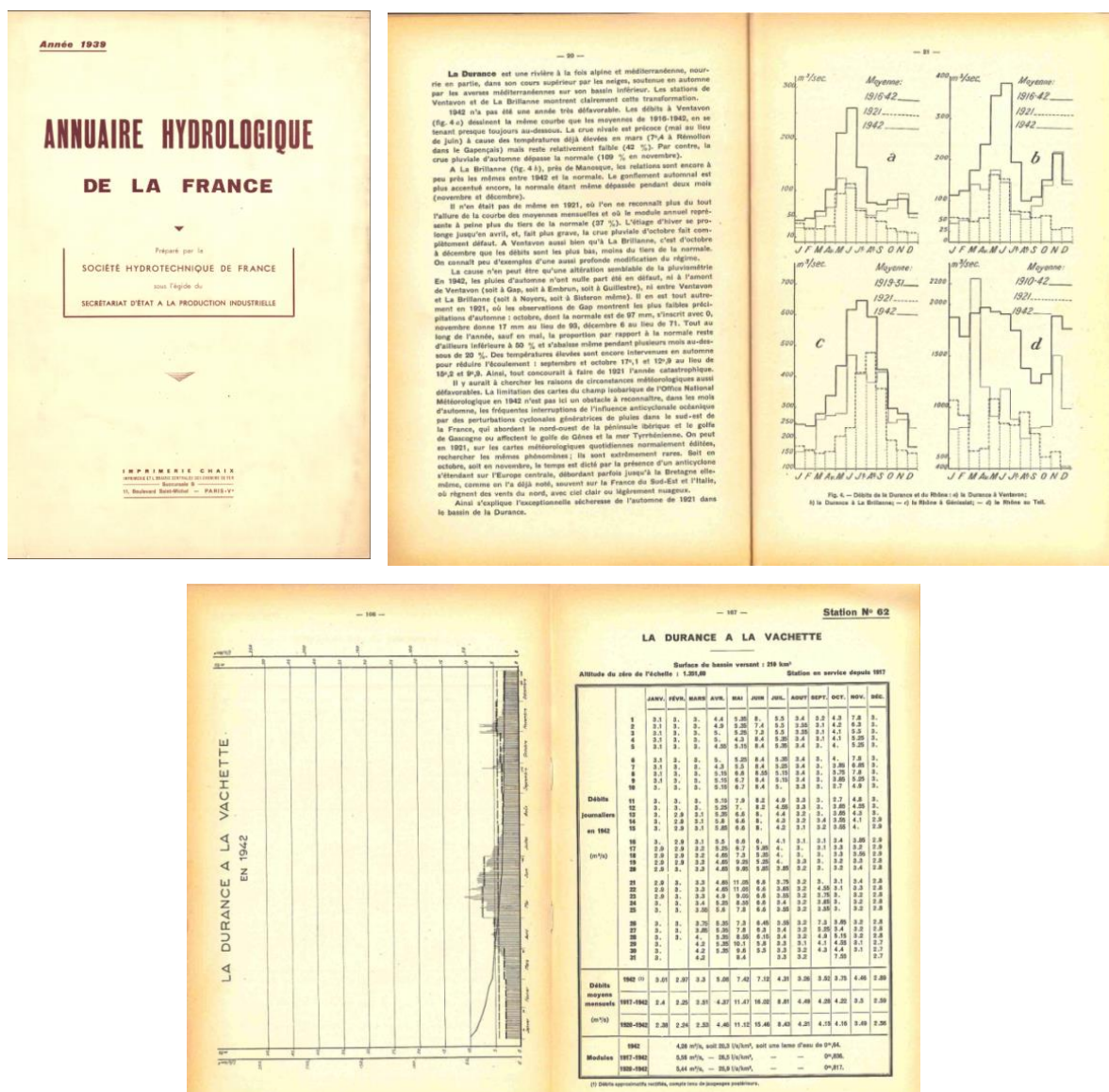


Tableau 1 : Données récupérées selon les années.

1939-1947	1948-1959	1960-1962	1963	1964-1969
Nom et numéro de station	Nom et numéro de station	Nom et numéro de station	Nom et numéro de station	Nom et numéro de station
Surface du bassin versant et altitude (sauf en 1939)	Surface du bassin versant et altitude	Surface du bassin versant et altitude	Surface du bassin versant et altitude	Surface du bassin versant et altitude
			Latitude et longitude	Latitude et longitude
Station usine ou non	Station usine ou non	Station usine ou non		
Débits journaliers pour les 365/366 jours de l'année	Débits journaliers pour les 365/366 jours de l'année	Moyennes journalières pour les 365/366 jours de l'année	Moyennes journalières pour les 365/366 jours de l'année	Moyennes journalières pour les 365/366 jours de l'année
Débits moyens mensuels bruts de l'année	Débits moyens mensuels bruts de l'année	Moyennes mensuelles naturelles de l'année	Moyennes mensuelles naturelles de l'année	Moyennes mensuelles naturelles de l'année
Débits moyens mensuels corrigés de l'année (lorsqu'ils sont disponibles)	Débits moyens mensuels corrigés de l'année (lorsqu'ils sont disponibles)	Moyennes mensuelles mesurées de l'année (lorsqu'elles sont disponibles)	Moyennes mensuelles mesurées de l'année	Moyennes mensuelles mesurées de l'année
Débits moyens mensuels depuis la mise en service de la station et jusqu'à l'année considérée	Débits moyens mensuels depuis la mise en service de la station et jusqu'à l'année considérée	Moyennes mensuelles sur la plus longue période disponible	Moyennes mensuelles sur la plus longue période disponible	Moyennes mensuelles sur la plus longue période disponible
Débits moyens mensuels depuis 1920 et jusqu'à l'année considérée	Débits moyens mensuels depuis 1920 et jusqu'à l'année considérée	Moyennes mensuelles de 1921 à 1960	Moyennes mensuelles de 1921 à 1960	Moyennes mensuelles de 1921 à 1960

1939-1947	1948-1959	1960-1962	1963	1964-1969
	Pluviométries mensuelles	Pluviométries mensuelles	Pluviométries mensuelles	Pluviométries mensuelles
		Maxima et minima journaliers par mois (valeur en m3/s et année correspondante)	Maxima et minima journaliers par mois (valeur en m3/s et année correspondante)	Maxima et minima journaliers par mois (valeur en m3/s et année correspondante)
		Valeurs instantanées de débit maximum par mois de l'année (valeur en m3/s et jour correspondant) (sauf en 1960)	Valeurs instantanées de débit maximum par mois de l'année (valeur en m3/s et jour correspondant) (sauf en 1960)	
		Extrêmes annuels (valeur en m3/s et année correspondante)	Extrêmes annuels (valeur en m3/s et année correspondante)	Extrêmes annuels (valeur en m3/s et année correspondante)

Cela représente au total 2293 tableaux de données journalières, qui correspondent aux données d'intérêt dans le cadre de la récupération des données anciennes pour le projet MDR. Cependant, certaines stations correspondent à des stations de la banque HYDRO, les données de certains tableaux étaient donc déjà présentes dans cette base de données. La priorité a été donnée aux stations qui semblaient absentes de la banque HYDRO. Tous les tableaux n'ont donc pas été récupérés. Nous avons finalement manipulé un total de 1638 tableaux de données. Ce nombre correspond à la somme des tableaux comprenant des données inédites par rapport à la banque HYDRO, et les tableaux que nous avons récupérés afin de les comparer à ceux présents dans la banque HYDRO.

La localisation des stations est représentée sur la Figure 2.



Figure 2 : Localisation des stations présentant des données journalières dans les Annuaire Hydrologiques de la SHF.

2.2. Protocole utilisé

En raison du grand nombre de fichiers à traiter, l'avancement du travail de numérisation et de traitement des données a été consigné au fur et à mesure dans un fichier d'avancement.

La récupération des données est basée sur l'utilisation d'un logiciel de reconnaissance optique de caractères (OCR). Un logiciel d'OCR fonctionne selon le principe général suivant : on donne au logiciel un fichier d'entrée en format image (scan de feuille imprimée, de feuillet dactylographié...) et on obtient un fichier de sortie en format texte. Cela permet donc d'obtenir des données numériques exploitables (textes, chiffres) à partir d'images scannées. Le processus comporte plusieurs étapes :

- Numérisation du document original (si elle n'a pas eu lieu en amont),
- Définition des paramètres du logiciel d'OCR,
- Retouches de l'image,
- Lecture de l'image,
- Enregistrement du résultat.

Le logiciel utilisé ici est ABBYY FineReader 11 Professional Edition après expertise de logiciels disponibles sur le marché. Une description précise du fonctionnement de ce logiciel étape par étape est donnée en Annexe 1. Cette annexe peut permettre de se familiariser avec le fonctionnement d'un logiciel d'OCR. Ici, seuls les paramètres utilisés pour la récupération des annuaires hydrologiques sont indiqués.

La reconnaissance optique des caractères a été réalisée annuaire par annuaire, les données ont par la suite été regroupées par stations.

Un schéma récapitulatif des différentes étapes du protocole est donné dans le paragraphe 2.2.8. (Figure 10). Un exemple de liste précise des opérations à effectuer est présenté en Annexe 2.

2.2.1. Numérisation des archives

Les annuaires hydrologiques sont des fascicules qui nous ont été transmis par la Société Hydrotechnique de France en format papier. Ils ont été « sacrifiés » afin de pouvoir les scanner : la tranche a été découpée afin d'obtenir des feuillets séparés.

Le massicotage et le scan des feuillets ont été réalisés par la société IMSO. La numérisation a été faite avec une résolution de 600 dpi. Le détail des prestations réalisées par IMSO est présenté en Annexe 3. Les données initiales correspondent donc à 31 fichiers image en format PDF (un fichier par annuaire). Cette première phase du projet a été coordonnée par A.-L. Achard (IST Lyon).

2.2.2. Définition des paramètres du logiciel

Parmi les nombreux paramètres à régler dans les options du logiciel, on peut préciser qu'on a utilisé :

- Langue du document : Nombres ; Français,
- Ne pas lire et ne pas analyser les images de pages acquises automatiquement,
- Lecture approfondie,
- Convertir les valeurs numériques en nombres.

2.2.3. Retouche des images

Aucune retouche d'image n'a été effectuée, car la qualité des fichiers fournis par IMSO était bonne.

2.2.4. Lecture des images

La lecture s'effectue en deux étapes : dans un premier temps le logiciel analyse la structure des pages (position des zones de textes, de tableaux, d'images, nombre de lignes et de colonnes dans un tableau...), et dans un second temps il reconnaît les caractères au sein de chaque zone identifiée par l'analyse.

L'analyse automatique des pages s'est vite révélée inappropriée : la structure des tableaux était systématiquement mal reconnue (en raison de l'absence de lignes de démarcation assez explicites), et de nombreuses informations présentes sur les pages n'étaient pas intéressantes à récupérer (numéro en haut de page, données mensuelles de lames d'eau, commentaires...). Une rectification manuelle est possible mais aurait demandé un temps bien trop important. On s'est donc aidé des modèles de zone pour faire l'analyse des pages (Figure 3). Cela a été possible car toutes les pages d'un annuaire présentent la même structure, à peu de choses près (on ne considère que les pages contenant les données de débit journalières).

L'utilisation d'un modèle de zone permet d'appliquer à toutes les pages la même structure d'éléments à reconnaître (textes, tableaux). Un ajustement manuel est toutefois nécessaire 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 faut également corriger les erreurs de

structuration des tableaux (lignes à ajouter/supprimer, cellules fusionnées à séparer...). Cette étape s'est révélée assez longue, mais était essentielle pour une bonne reconnaissance des caractères.

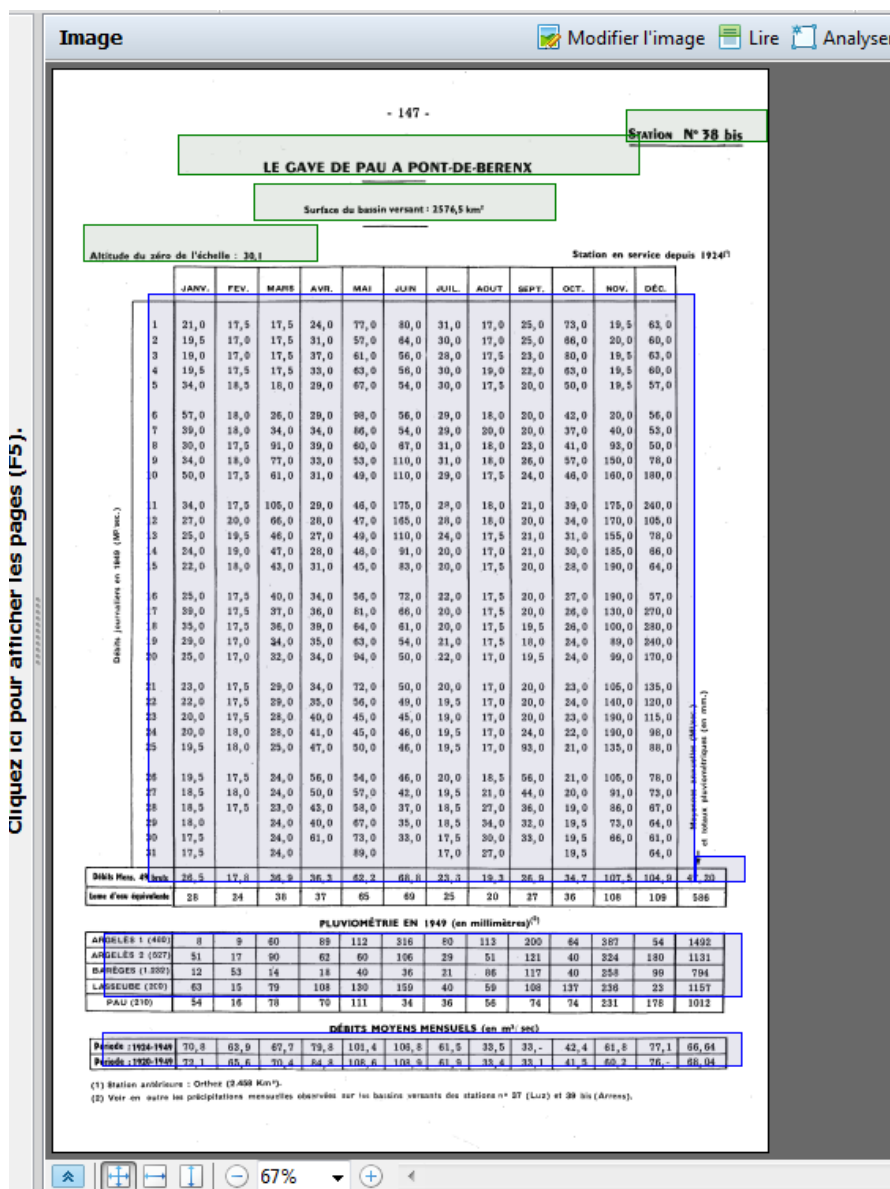


Figure 3 : Le modèle de zone permet de définir la position de zones de texte (en vert) et de zones de tableaux (en bleu) sur toutes les pages du document. On constate ici qu'un ajustement manuel est nécessaire pour caler les zones correctement sur la page.

Lorsque l'analyse des pages était satisfaisante, on a relancé l'étape de lecture (seconde partie du processus uniquement : la reconnaissance des caractères).

En cas de difficulté de reconnaissance des caractères (liée notamment à une police particulière ou à une mauvaise qualité d'impression), on a pu avoir recours à un gabarit personnalisé. Cela consiste en une lecture pendant laquelle le logiciel s'arrête sur chaque caractère qu'il ne reconnaît pas, et demande à l'utilisateur de lui indiquer de quel caractère il s'agit (Figure 4). Le logiciel garde ensuite en mémoire chaque couple image/caractère dans un « gabarit utilisateur ». Il utilise ensuite ce gabarit pour améliorer le reste de la lecture. L'apprentissage par le logiciel peut être long, mais permet ensuite une meilleure reconnaissance des caractères.

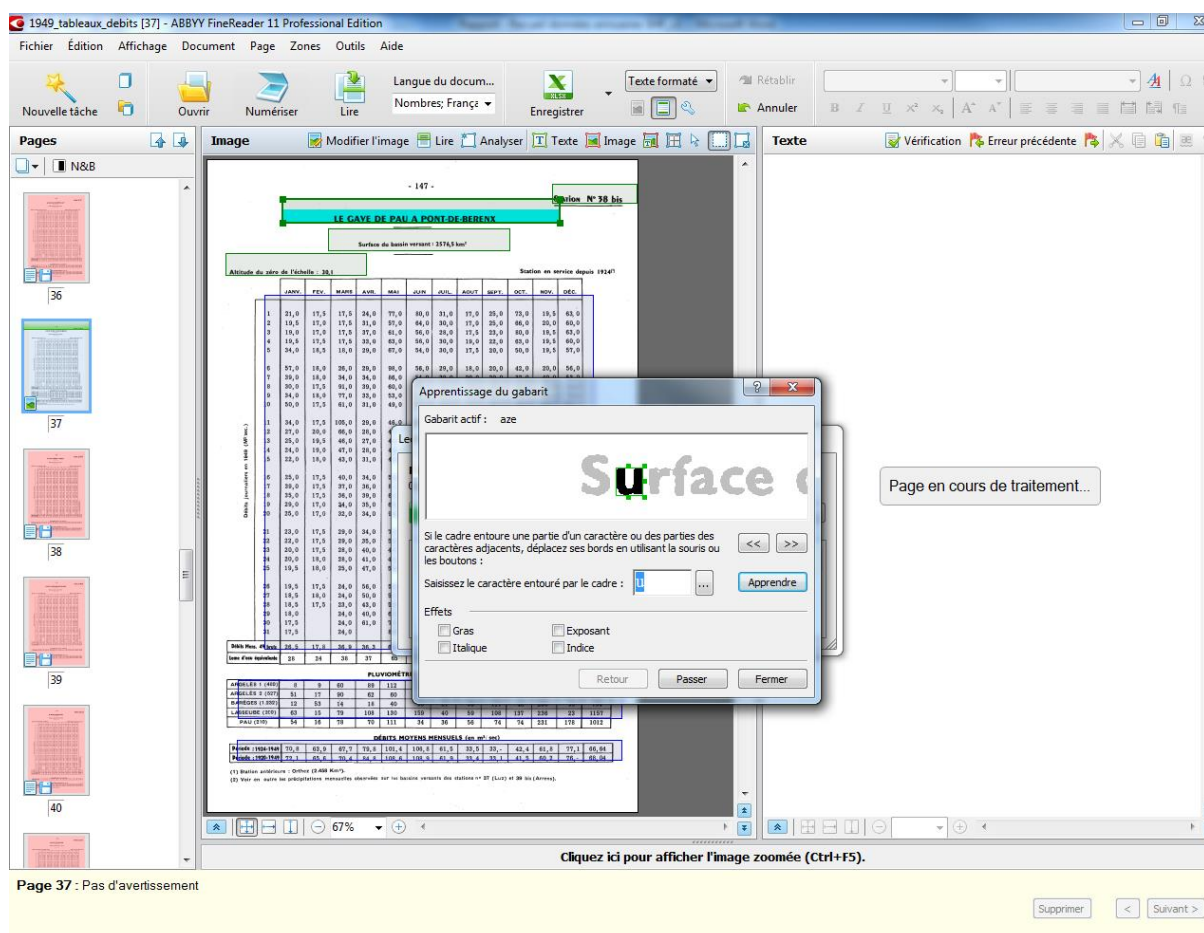


Figure 4 : Aperçu de l'étape d'apprentissage du gabarit. Le logiciel propose pour chacun des caractères une interprétation, que l'utilisateur valide ou modifie.

2.2.5. Enregistrement des résultats de l'OCR

Les résultats ont été générés en format Excel, avec un fichier séparé par page. Les données de sortie issues de l'analyse par OCR se présentent donc sous la forme de N fichiers Excel (N étant le nombre de tableaux de données présentés l'année en question). Chaque fichier Excel contient les caractéristiques de la station (nom de station, numéro, surface du bassin versant, altitude...) et le tableau de données chiffrées (données pouvant varier selon les années). Un exemple est donné en Figure 5.

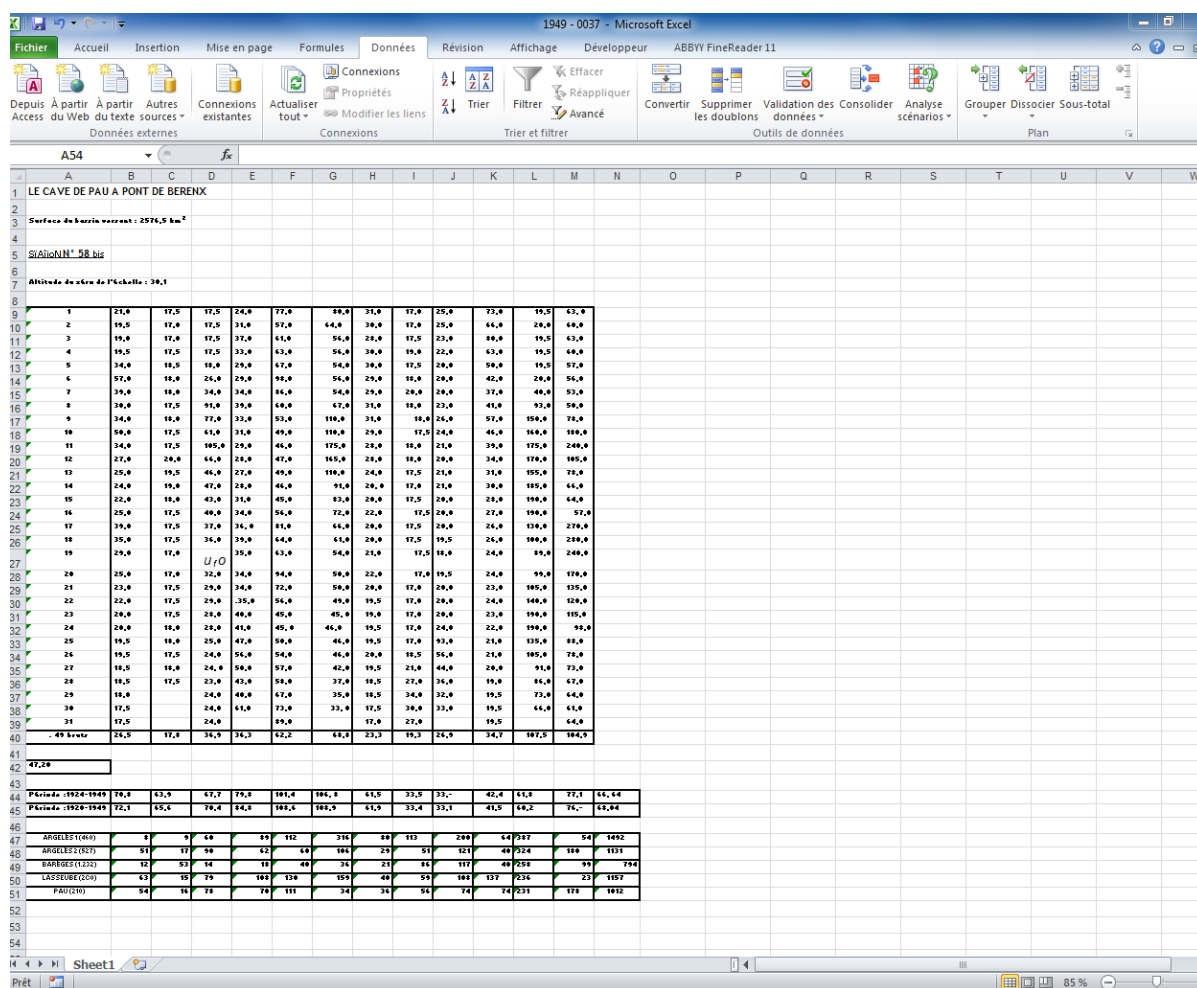


Figure 5 : Exemple d'un fichier de sortie en format Excel, après traitement par le logiciel ABBYY.

2.2.6. Modifications de la mise en forme des fichiers de sortie

L'objectif principal de ce travail de récupération était d'obtenir des chroniques de débits journaliers exploitables. Il fallait donc obtenir des fichiers se présentant au format QJ actuel de la banque HYDRO (avec une valeur de débit journalier par ligne). Or, les annuaires hydrologiques présentent des tableaux de données sous la forme de douze colonnes de données, une colonne contenant un mois de données journalières. Une modification de la mise en forme des fichiers de sortie était donc nécessaire.

Les fichiers de sorties ont été générés en format Excel, afin de pouvoir corriger automatiquement certaines erreurs commises par le logiciel d'OCR (Paragraphe 2.2.7). Ces fichiers ont ensuite été enregistrés sous format texte, sans modification de la mise en forme. Ces fichiers texte présentent les données de débits journaliers en colonne, mais également d'autres données (débits mensuels naturels/corrigés ou mesurés/naturels, débits sur des périodes longues, données de précipitations...) lorsque ces données sont disponibles. **Ils contiennent donc toutes les données récupérées depuis les tableaux initiaux.**

Un traitement avec le logiciel R a ensuite permis de convertir cette mise en forme initiale pour obtenir le format souhaité (similaire au format HYDRO2 de la banque HYDRO, Figure 6). **Ces fichiers ne contiennent pas de données autres que les débits journaliers de l'année (et les métadonnées).**

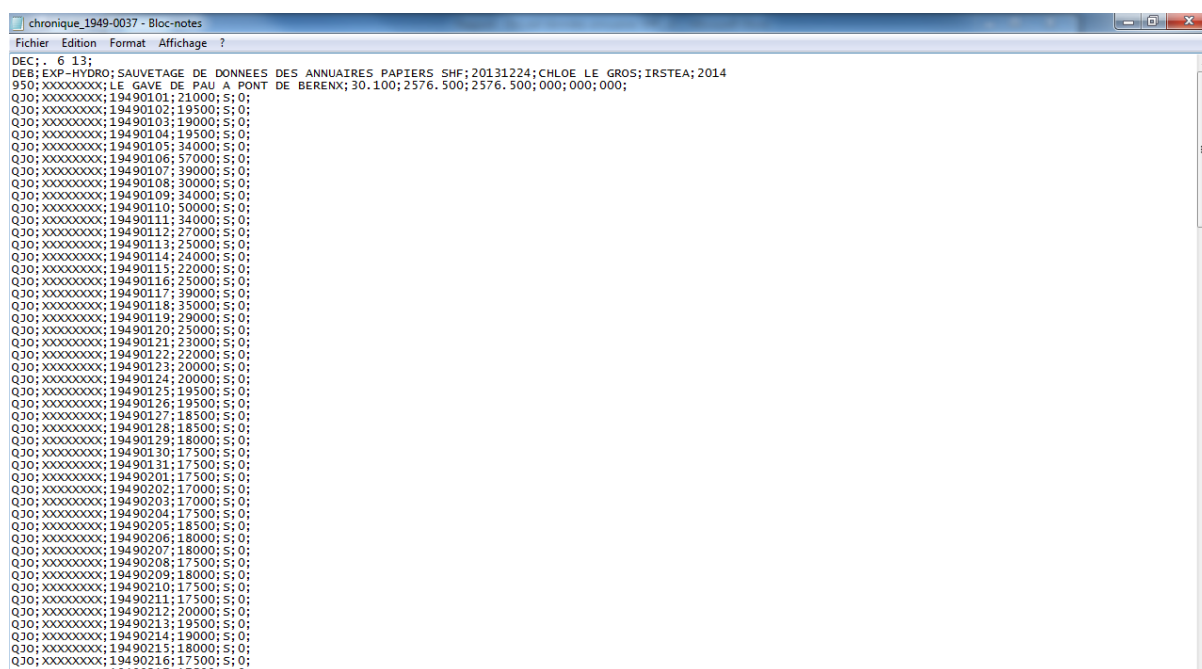


Figure 6 : Exemple d'un fichier en format QJ actuel de la banque HYDRO.

2.2.7. Corrections automatiques et vérifications visuelles

Le logiciel d'OCR commet des erreurs de lecture, et c'est d'autant plus vrai que la page contient à la fois des données chiffrées et du texte. Une vérification visuelle des fichiers de sortie est donc essentielle pour contrôler la qualité des données numérisées, et le cas échéant corriger les erreurs commises par l'OCR.

2.2.7.1. Corrections automatiques sous Excel

On a pu identifier rapidement des erreurs récurrentes. Ces erreurs ont été éliminées automatiquement à l'aide d'une macro Excel, formulée en Visual Basic. Il s'agit notamment de la suppression des caractères « i », « l », « / », « | », « ! », « : », « ; », ... régulièrement ajoutés par ABBYY lorsque le trait de bordure de la colonne était mal identifié. Cette macro permet également de remplacer certaines lettres par des chiffres (l'OCR commet souvent les erreurs suivantes : remplacement d'un « 0 » par un « O », de « 11 » par un « U »...). Chaque modification réalisée automatiquement est visible grâce à une coloration des cellules modifiées (voir Figure 7). Une vérification visuelle est ensuite réalisée pour chaque cellule apparaissant en couleur, par une comparaison avec le tableau originel, afin de s'assurer que la correction automatique est pertinente (on peut, par exemple, en correction automatique remplacer les « b » par des « 6 », alors que parfois l'OCR prend aussi les « 5 » pour des « b »).

Cette correction automatique s'applique à tous les fichiers. Il faut ensuite vérifier fichier par fichier les autres erreurs éventuellement présentes. Les données contenues dans les annuaires hydrologiques nous ont permis de réaliser deux vérifications successives : la vérification à l'aide de calculs et la vérification à l'aide des graphiques.

2.2.7.2. Vérification à l'aide de calculs

Dans un premier temps, une vérification a été réalisée à base de calculs. Les annuaires présentent à la fois les données journalières de débits et les moyennes mensuelles correspondantes. On peut donc recalculer ces moyennes mensuelles à l'aide des données journalières, et vérifier qu'on trouve un résultat similaire à la moyenne mensuelle indiquée dans le fichier d'origine. Si ce n'est pas le cas, c'est peut-être que l'OCR a fait une erreur d'identification d'une des valeurs journalière intervenant dans le calcul en question. On a rapidement pu constater que les valeurs moyennes indiquées dans les tableaux s'écartaient parfois sensiblement de la valeur recalculée à l'aide des valeurs journalières, sans pour autant que l'OCR ait fait d'erreur. Ces écarts peuvent être dus à des accumulations d'arrondis, des erreurs de calcul, des fautes de frappe, des modifications ultérieures soit des valeurs journalières soit des valeurs mensuelles... On a donc calculé, pour chacune des 12 valeurs moyennes mensuelles, l'écart entre la valeur indiquée par l'annuaire et la valeur recalculée à l'aide des données journalières de ce mois. Lorsque la différence était supérieure ou égale à 10%, l'ensemble des données journalières du mois en question ont été vérifiées. Ce seuil a été ramené à 1% à partir de 1952, car les moyennes mensuelles indiquées dans les tableaux de données semblent plus précises à partir de cette date.

A partir de 1948, les annuaires présentent également la moyenne annuelle des débits, la moyenne des débits mensuels pour deux périodes de référence, des données mensuelles et le cumul annuel de précipitations. On a donc des données permettant de vérifier les valeurs mensuelles pour les débits sur les deux périodes et pour la pluviométrie. De la même manière, on regarde l'écart entre la valeur annuelle donnée par l'annuaire et la valeur recalculée.

Cette vérification par le calcul présente l'intérêt d'identifier rapidement des erreurs d'OCR liées au remplacement d'un chiffre par une lettre. Des confusions entre « l » et « 1 », « O » et « 0 », « S » et « 8 », ... sont en effet assez fréquentes. Sous Excel, la présence d'une lettre empêche le calcul (affichage de « !VALEUR »). On peut ainsi rapidement identifier la valeur qui a mal été reconnue par l'OCR et la corriger en vérifiant visuellement sur le tableau de l'annuaire. Toutefois, cette vérification n'est pas toujours très efficace en cas de confusion d'un chiffre avec un autre chiffre. En effet, selon la police de caractère utilisée, la confusion entre les 3, les 5 et les 7 peut être facile. De même, en cas de caractère un peu effacé, un "6" est facilement pris pour un "5", ou un "8" pour un "6"... Une seconde étape de vérification nous a donc paru nécessaire.

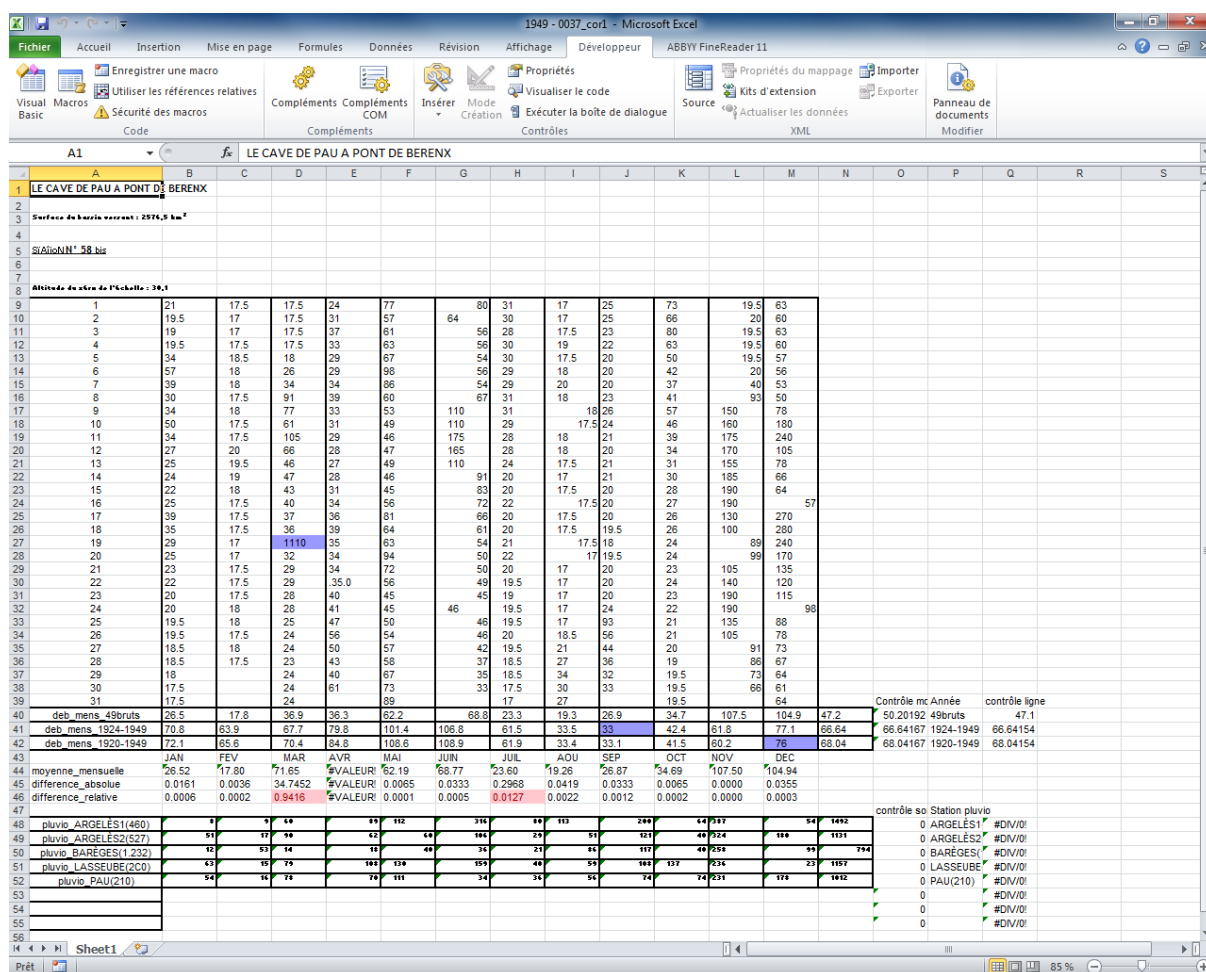


Figure 7 : Résultat après correction automatique de la macro (cases violettes) et calcul des moyennes. La correction automatique du 19 mars n'est pas pertinente, car la moyenne recalculée à l'aide des données journalières (71.65 m3/s) est bien plus élevée que la moyenne indiquée dans l'annuaire (36.9 m3/s). La présence d'un point en début de nombre le 22 avril empêche le calcul de la moyenne. L'écart entre les moyennes de juillet indique qu'une erreur de faible ampleur doit être présente (par exemple remplacement d'un 6 par un 8).

Une fois ces vérifications par calcul effectuées, on a converti le fichier Excel en un fichier texte, qui ne présente plus que les données brutes (on enlève toutes les lignes et les colonnes de calculs qui ont servi aux vérifications). Ce fichier texte est traité sous R afin de modifier son format (à l'origine : données mensuelles en colonnes) et obtenir le format HYDRO2 de la banque HYDRO (au final : une donnée journalière par ligne, voir Figure 6).

2.2.7.3. Vérification à l'aide des graphiques

On produit ensuite un graphique pour chacun de ces fichiers ascii, ce qui nous permet de réaliser la seconde étape de vérification de l'OCR. En effet, les annuaires hydrologiques présentent pour chaque station un tableau de données et le graphique correspondant (voir Figure 1). La comparaison visuelle de ces deux graphiques permet de déceler des erreurs qui n'auraient pas été identifiées à l'étape de vérification précédente (deux exemples sont présentés en Figures 8 et 9).

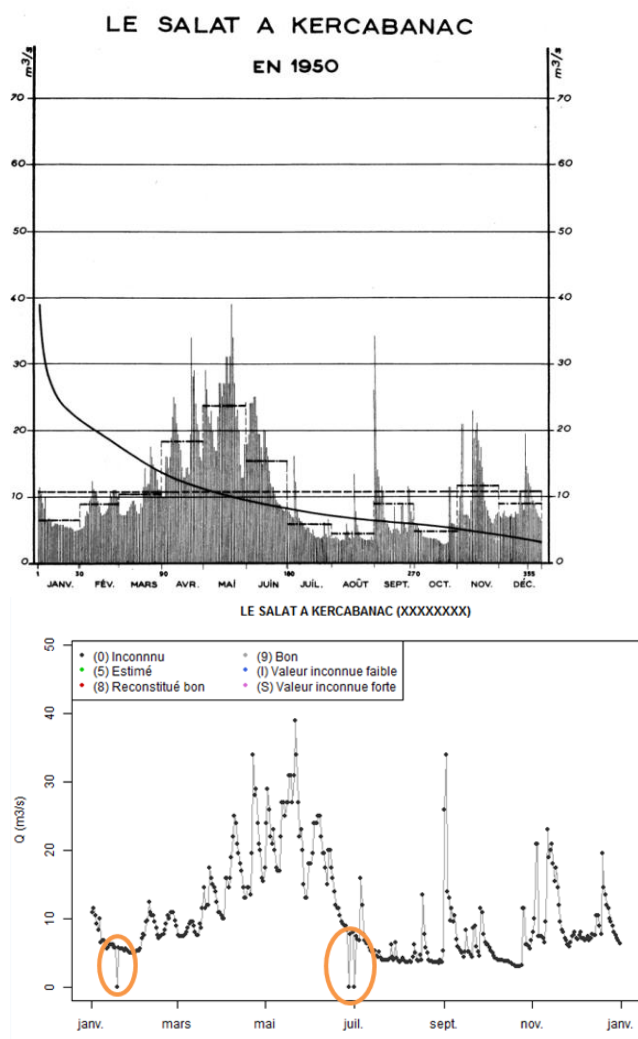


Figure 8 : A gauche : graphique présenté dans l'annuaire. A droite : graphique obtenu grâce aux données issues de l'OCR. On peut voir que l'OCR a mal identifié 3 nombres (un zéro a été reconnu pour ces trois cas).

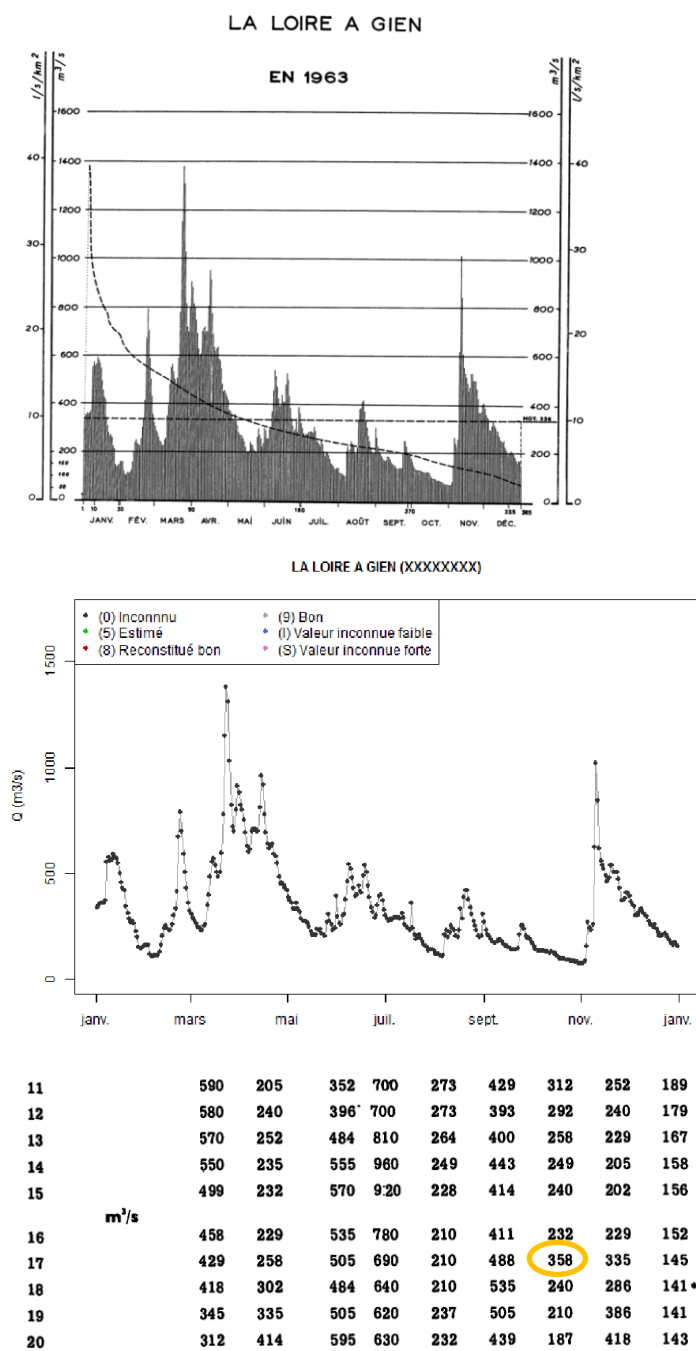


Figure 9 : A gauche : graphique présenté dans l'annuaire. A droite : graphique obtenu grâce aux données issues de l'OCR. En bas : tableau de données de l'annuaire. On peut voir une valeur plus élevée sur le graphique issu des données après traitement OCR. Cette valeur élevée est également présente dans le tableau de donnée de l'annuaire, il ne s'agit donc pas d'une erreur d'OCR. En corrigeant cette valeur par un nombre cohérent avec le graphique de l'annuaire (ici, on remplace 358 par 258), on diminue sensiblement l'écart entre la moyenne indiquée dans l'annuaire (231 m³/s) et la moyenne recalculée (on passe de 234 à 231 m³/s). Il s'agit donc probablement d'une faute de frappe dans le tableau de l'annuaire.

2.2.8. Schéma récapitulatif : les étapes du protocole

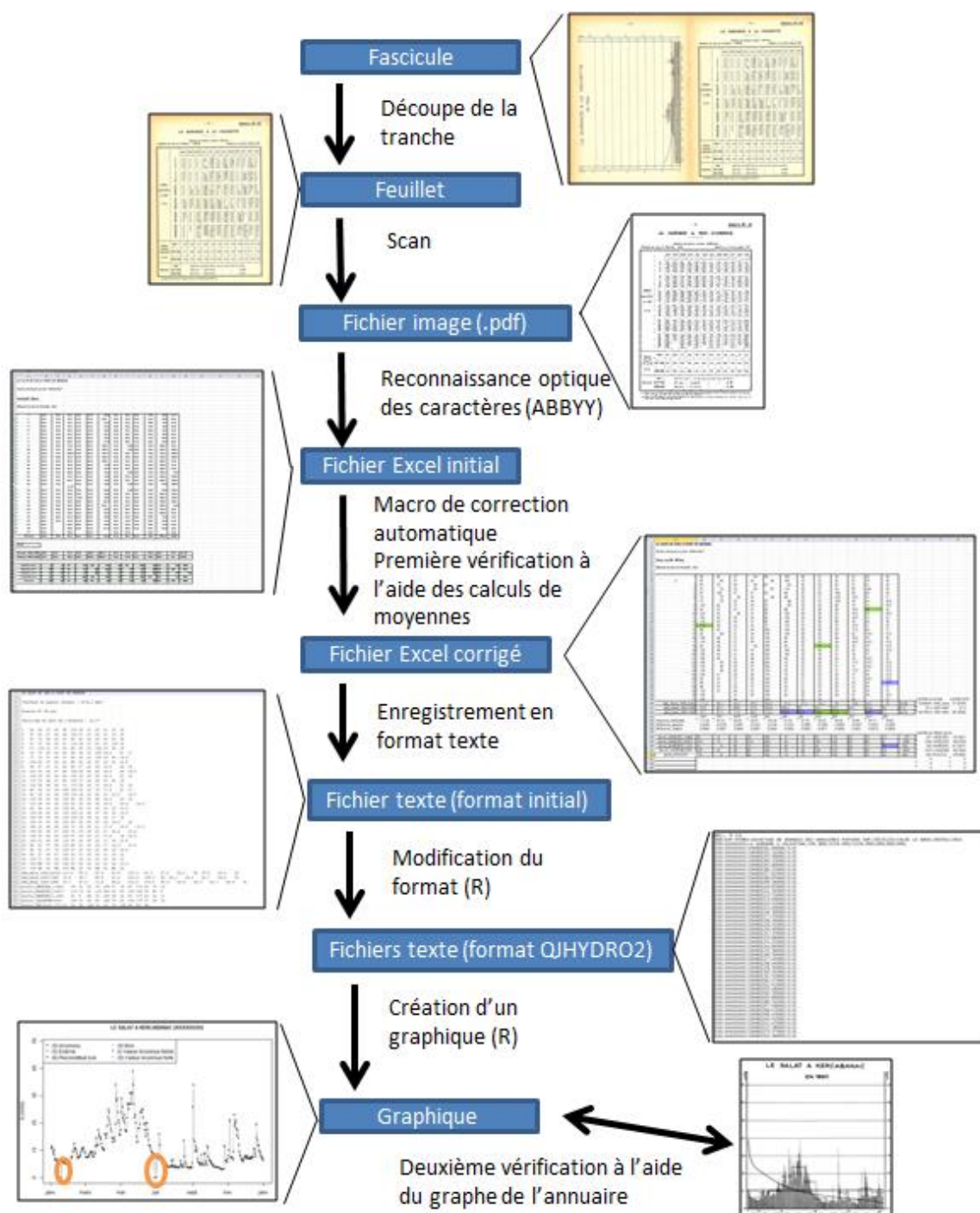


Figure 10 : Schéma récapitulatif des étapes du protocole.

2.2.9. Codes utilisés dans les chroniques de débits journaliers

Lors de la transformation en format HYDRO2, les colonnes renseignant l'origine et la validité des données ont été remplies. Pour l'origine, étant donné le travail de récupération par OCR, le code S a été appliqué à toutes les données (S : saisie manuelle). Pour la validité, un code spécifique a été appliqué. On a voulu traduire l'erreur relative entre la valeur de débit mensuelle calculée par le gestionnaire et la valeur de débit mensuelle recalculée avec les données journalières. En effet, les annuaires présentent parfois une valeur de débit mensuelle différente de celle que l'on peut recalculer, sans pour autant qu'aucune donnée journalière n'ait été mal lue par le logiciel d'OCR. Dans ce cas, on applique pour l'ensemble des données journalières du mois en question un code de validité qui traduit cette différence. Le code de validité peut ainsi prendre les valeurs suivantes :

- 0: de 0 à 10% de différence relative
- 1: de 10 à 20% de différence relative
- 2: de 20 à 30% de différence relative
- 3: de 30 à 40% de différence relative
- 4: de 40 à 50% de différence relative
- 5: de 50 à 60% de différence relative
- 6: de 60 à 70% de différence relative
- 7: de 70 à 80% de différence relative
- 8: de 80 à 90% de différence relative
- 9: plus de 90% de différence relative

Ce code de validité est donc le même pour tous les jours d'un même mois.

On a de plus ajouté le code A : débits journaliers approximatifs selon le gestionnaire. En effet, les commentaires précisent parfois que certaines données sont douteuses ou approximatives (débits reconstitués, déduits d'une corrélation avec une autre station...). Dans ce cas, le code « A » a été appliqué aux données journalières concernées.

Ces codes sont présents uniquement dans les fichiers au format de la banque HYDRO (fichiers contenant uniquement les données de débits journaliers). Ces codes sont différents de ceux présents dans les fichiers extraits de la banque HYDRO pour permettre d'identifier a posteriori les valeurs issues des numérisations.

2.2.10. Evolution du protocole

En raison de l'évolution au cours du temps des données disponibles dans les annuaires, le protocole utilisé a pu varier et être rectifié a posteriori :

- Un écart a souvent pu être observé entre les moyennes indiquées dans les annuaires et les moyennes recalculées (sans qu'une erreur d'OCR en soit l'origine), c'est pourquoi le protocole de vérification des erreurs de l'OCR a été modifié au cours du temps :
 - De 1949 à 1951, un seuil de 10% d'écart a été fixé. Pour une différence supérieure à ce seuil, l'ensemble des valeurs journalières du mois étaient vérifiées une à une.

- Ce seuil de vérification a été abaissé à 1% en 1952, et couplé avec la vérification que les moyennes des annuaires étaient cohérentes avec les moyennes recalculées à l'arrondi près.
 - A partir de 1954 et jusqu'en 1962, seuls les arrondis ont été vérifiés, en raison du grand nombre de dépassement du seuil de 1% d'écart.
 - A partir de 1963 et jusqu'en 1969, on est repassé à une vérification basée uniquement sur le seuil de 1% (modification des règles d'arrondis présentée en début d'annuaire 1963).
- Les modifications de mise en page des annuaires a conduit à modifier à plusieurs reprises les macros Excel et les scripts R.
 - Jusqu'en 1952, l'écriture par défaut des fichiers en format QJ de la banque HYDRO consiste à indiquer pour les fortes valeurs de débit leur mantisse et exposant. Cette écriture par défaut a été modifiée à partir de 1953.
 - Selon l'affichage sur une ligne ou deux lignes des dates des périodes longues, le « 19 » peut parfois être omis dans le nom « moy_mens_1921-61 ».
 - A partir de 1961, un « m » apparaît dans l'altitude des stations pluviométriques.
 - Les fautes de frappe évidentes ont été corrigées à partir de 1962 (beaucoup de fautes de frappe identifiées lors de la vérification des graphes en 1962 et 1963, un exemple est présenté en Figure 9).
 - La vérification des valeurs de surface du bassin versant et de l'altitude a été supprimée en 1966, car cette étape est aussi réalisée lors de la correction des fichiers Excel.
 - Il y a eu des évolutions dans le vocabulaire utilisé dans les annuaires (Tableau 2), qui ont été répercutées dans les fichiers de sortie :

Tableau 2 : Correspondance des termes utilisés pour les débits moyens mensuels.

1939-1959		1960-1969		Remarques
Débits mensuels (toujours disponibles)	moyens bruts	Moyennes mesurées disponibles	mensuelles (parfois en 60-62, puis toujours disponibles)	Débits qui peuvent être influencés ou non. Ils correspondent aux données journalières
Débits mensuels (parfois disponibles)	moyens corrigés	Moyennes naturelles disponibles)	mensuelles (toujours disponibles)	Débits qui peuvent être mesurés (dans ce cas, ils correspondent aux données journalières) ou reconstitués (correction des effets des ouvrages amont)

De 1939 à 1959, les moyennes que l'on recalcule à partir des données journalières correspondent toujours aux débits moyens mensuels bruts. A partir de 1960, les moyennes que l'on recalcule à partir des données journalières correspondent aux moyennes mensuelles naturelles quand ce sont les seuls disponibles, alors qu'elles correspondent aux moyennes mensuelles mesurées quand les deux types de données sont disponibles. A partir de 1963, les deux lignes sont toujours renseignées, et contiennent donc parfois les mêmes données. Les moyennes recalculées correspondent toujours aux moyennes mesurées.

2.3. Evaluation du temps nécessaire au traitement des données

Le temps nécessaire pour la récupération des données a été très variable selon les annuaires traités (Tableau 3), il est donc difficile d'évaluer le temps nécessaire pour ce travail. On peut cependant donner un ordre d'idée. La complexité des tableaux et le nombre de données à récupérer ayant varié selon les années, le temps nécessaire a lui aussi fortement varié. On donne un ordre de grandeur général pour le traitement dans les années 40, 50 et 60. Le détail est ensuite donné pour chacune des étapes du protocole uniquement pour les années 60 (le détail du temps nécessaire n'a pas été relevé avant 1960).

Tableau 3 : Evaluation du temps nécessaire pour le traitement des données selon les années (évolution de la complexité et du nombre de données).

Etapes	Années 40	Années 50	Années 60
Total	Environ 5 heures par année	Environ 4.5 heures par année	Environ 10 heures par année
Préparation (remplissage du fichier d'avancement, identification des différents types de fichiers disponibles)			30 min
Reprise des commentaires			45 min
Traitement préliminaire ABBYY (ouverture du fichier, suppression des pages inutiles, intercalage de pages blanches):			15 minutes
Traitement ABBYY (lecture, calage des modèles de zone) :			Entre 70 et 120 min selon la complexité
Traitement Excel (première vérification)			Entre 5 et 8 min par fichier en moyenne. Au total, de 4 à 5h
Modification suite aux errata			5 min
Traitement R (modification du format)			2 à 5 min
Traitement des lacunes			Quelques minutes par fichier concerné

Etapes	Années 40	Années 50	Années 60
Modification du code de validité			Quelques minutes par fichier concerné
Production des graphes			2 min
Vérification des graphes			1 à 2 min par fichier. Au total, 30min à 1h30
Vérification des surfaces de bassin versant et de l'altitude			environ 15 minute pour l'ensemble des 50 fichiers
Comparaison des stations avec celles de la banque HYDRO			environ 6 minutes par station

On indique ici certains éléments qui ont pu allonger considérablement le temps nécessaire au traitement des données :

- On peut préciser que le temps de chargement des fichiers dans ABBYY est assez long (fichiers de résolution 600 dpi : 3-4 minutes pour charger plus de 250 pages). De même, les étapes de lecture peuvent prendre plusieurs minutes (5-7 minutes).
- La mise en page des tableaux de débits journaliers a évolué au fil du temps (Figures 11 à 15). Le traitement a pris plus de temps à partir de 1960, en raison de l'augmentation du nombre de données à récupérer, et donc du nombre de données à vérifier. Par exemple, un tableau de données présentant les valeurs de débit mensuels maximum et minimum sur une période longue prenait plus de temps pour la vérification sous Excel, car ce tableau présentait un grand nombre de données non vérifiables par le calcul, et des données à retaper manuellement (les années, inscrites sur un fond grisé, ne sont pas bien reconnues par l'OCR). La complexité croissante des tableaux a également allongé considérablement le temps d'ajustement des modèles de zone.
- Au cours des années, les annuaires ne sont pas restés les mêmes. Il y a eu des changements dans les mises en page, dans les polices de caractère utilisées, dans les normes utilisées pour les séparateurs de millier/décimale... (Figures 11 à 15). Toutes ces évolutions ont conduit à modifier régulièrement les modèles de zone et les gabarits utilisateur du logiciel ABBYY, les macros Excel de correction automatique, les scripts R de lecture et de modification des fichiers...
- A partir de 1951, les noms de stations ne sont plus situés sur la page contenant le tableau des données de débits journaliers. Ceci a allongé le temps de traitement préliminaire à la lecture par OCR, car il a fallu dans ABBYY rajouter des pages blanches entre chaque couple de pages (seule manière d'obtenir pour chaque station un fichier unique comprenant les deux pages).

- La lecture avec apprentissage a été nécessaire pour faire un gabarit en 1948 car il y avait beaucoup trop d'erreur de reconnaissance de caractères suite au changement de police.
- En 1952, malgré une modification du gabarit, il y a eu de nombreuses erreurs à corriger en raison d'un gros problème d'identification des « 8 ».

LE TAURION A PONTARIONSurface du bassin versant : 389 km²

Station en service depuis 1919

		JANV.	FÉVR.	MARS	AVR.	MAI	JUIN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DÉC.
Débits journaliers en 1939 (m ³ /s)	1	23.7	18.6	10.9	10.	4.9	3.9	4.2	3.9	1.9	3.9	7.4	10.5
	2	24.2	16.2	8.6	9.	4.8	3.9	4.8	2.6	1.9	7.8	6.1	9.7
	3	20.2	14.5	7.1	9.	4.6	10.	3.5	2.3	1.9	4.	5.8	10.2
	4	17.9	13.2	6.6	9.3	4.4	5.8	3.	4.8	5.9	4.	21.3	16.2
	5	24.	12.3	6.4	12.1	4.4	7.6	2.6	3.5	3.3	6.8	16.7	24.2
	6	16.9	11.4	10.5	10.2	5.3	6.6	3.2	2.6	2.1	6.	11.6	14.8
	7	14.5	10.7	10.5	8.1	23.2	4.9	2.8	2.5	1.9	6.1	10.	15.
	8	14.3	10.2	14.	7.4	26.3	4.6	2.5	3.2	1.8	6.6	8.1	12.3
	9	12.3	10.	10.5	6.6	13.2	5.8	2.4	4.6	1.6	4.4	13.7	13.7
	10	11.6	10.	12.8	5.9	10.	5.3	2.4	3.	1.4	5.8	20.7	19.7
	11	11.4	9.5	10.	6.1	9.	4.9	2.3	2.5	1.3	4.9	14.7	18.9
	12	12.8	9.2	9.5	5.9	8.3	6.8	2.3	2.3	1.7	4.8	34.8	16.
	13	15.4	9.5	10.	5.8	10.5	7.1	2.1	2.1	1.9	7.4	19.1	13.5
	14	12.3	10.	9.	7.6	13.	7.6	2.4	2.	2.	8.8	15.	12.3
	15	10.2	8.6	12.	9.3	12.	4.9	2.4	1.9	2.5	15.	16.2	10.9
	16	15.2	8.3	12.8	6.4	16.9	4.2	3.3	1.9	2.6	7.8	19.6	10.
	17	30.	8.3	13.7	6.6	13.	3.9	4.6	1.9	2.3	6.6	32.4	10.
	18	24.7	8.3	16.2	6.1	12.5	3.5	4.4	1.8	4.	8.1	30.6	11.4
	19	22.2	8.3	12.3	4.9	11.1	3.3	4.6	1.8	3.3	10.2	27.4	11.4
	20	20.2	8.8	9.5	4.6	10.2	3.3	5.6	2.8	2.5	6.6	24.	10.5
	21	31.1	6.1	11.6	4.1	9.	3.9	5.1	3.2	2.1	5.8	19.1	9.
	22	45.9	5.8	12.	4.6	6.6	5.8	3.3	2.4	2.	5.3	16.4	8.1
	23	41.7	16.7	18.4	3.9	6.1	5.9	3.9	2.3	2.	4.9	14.7	7.6
	24	42.2	17.6	17.6	3.9	6.	4.6	5.6	2.1	1.9	4.8	13.5	7.1
	25	11.9	12.3	17.2	5.8	5.8	5.1	5.8	2.6	1.6	6.	12.3	6.4
	26	34.3	10.9	13.7	10.7	4.9	6.4	3.5	3.	1.6	6.6	11.1	6.1
	27	27.9	10.5	12.8	9.	4.8	5.3	3.2	2.3	1.6	9.	13.	5.8
	28	24.	8.6	11.4	7.1	4.6	3.9	2.5	1.9	1.6	6.	17.6	5.8
	29	22.2	10.	6.4	4.4	4.6	2.4	1.9	1.6	9.	17.4	5.5	
	30	25.8		9.5	6.	4.2	5.1	2.5	1.9	1.6	6.8	12.3	5.1
	31	22.4		9.		4.1		4.6	1.9		10.		5.1
Débits moyens mensuels (m ³ /s)	1939	22.7	10.8	11.5	7.	8.9	5.2	3.3	2.5	2.1	6.8	16.9	11.
	1919-1939	13.	10.8	12.2	11.6	8.9	6.	5.	2.8	2.8	4.6	8.7	10.7
	1920-1939	12.9	10.6	11.9	11.4	8.9	6.1	4.8	2.8	2.8	4.7	8.6	10.2
Modules	1939	9,1 m ³ /s, soit 23,4 l/s/km ² , soit une lame d'eau de 0 ^m ,738.											
	1919-1939	8,1 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 11 : Exemple de mise en page des tableaux de débits journaliers pour la période 1939-1947

LE TAURION A PONTARION

Surface du bassin versant : 389,4 km²

Altitude du zéro de l'échelle : 436,37

Station en service depuis 1919

		JANV.	FEV.	MARS	AVR.	MAI	JUIN	JUIL.	AOÛT	SEPT.	OCT.	NOV.	DÉC.	
Débits journaliers en 1948 (M ³ /sec.)	1	11,5	13,5	3,0	2,1	4,4	4,4	1,6	1,0	0,7	5,2	0,6	0,9	
	2	10	9,2	2,7	4,4	6,7	5,6	1,6	0,8	0,7	5,2	0,6	0,9	
	3	9,7	8,0	2,4	2,7	6,3	6,3	1,6	0,8	5,9	5,2	0,9	0,9	
	4	6,3	15,0	2,4	4,4	4,1	8,4	1,3	0,8	5,9	5,2	1,2	1,7	
	5	5,9	9,4	2,1	3,0	3,4	4,8	1,3	0,7	6,7	5,2	1,8	1,9	
	6	5,9	7,6	2,1	10,5	3,4	3,7	1,3	0,9	7,8	5,0	3,5	1,5	
	7	26	7,2	1,8	5,2	4,4	3,4	1,3	0,9	2,7	4,8	2,7	1,1	
	8	27	6,3	1,8	4,1	3,7	2,7	1,3	4,3	1,5	5,0	6,6	1	
	9	17	6,3	1,6	4,4	3,0	2,7	1,3	2,5	1,6	4,8	5,6	1,3	
	10	21	6,3	1,6	3,7	5,9	5,6	1,3	1,6	3,2	4,8	3,2	1,7	
	11	17	6,3	1,6	2,7	6,3	4,4	1,3	1,3	4,8	4,8	4,1	1,3	
	12	15	5,6	1,3	5,6	4,4	3,7	1,3	1,6	9,7	5,0	4,3	1,8	
	13	17	5,9	1,3	3,7	6,3	4,1	1,8	3,4	5,6	5,0	3,4	3,4	
	14	13	5,2	1,3	2,7	4,4	3,4	2,2	2,4	4,6	5,2	1,9	1,8	
	15	23	10	1,3	2,4	4,4	3,4	1,7	1,6	4,4	5,6	1,7	1,8	
	16	17,5	9,7	1,3	2,1	9,7	3,0	1,8	1,3	4,3	5,2	1,8	2,8	
	17	12	8,9	1,6	9,4	5,2	2,7	1,5	1,8	4,1	5,2	1,6	1,8	
	18	11	8,4	1,8	4,4	17,5	2,4	1,5	2,4	3,9	5,0	1,7	1,6	
	19	14,5	7,6	2,4	7,2	6,3	2,4	1,7	3,0	3,7	5,9	1,3	1,6	
	20	12,5	7,2	1,8	4,4	5,6	7,6	1,3	2,7	4,3	5,7	1,0	1,5	
	21	11,0	7,2	1,8	4,4	4,4	5,2	1,2	1,6	5,7	5,0	1,0	1,1	
	22	13	8,0	1,6	4,4	4,4	3,4	5,7	1,3	5,6	4,6	0,9	1,3	
	23	16,5	7,6	1,3	4,1	3,7	2,7	3,2	1,1	5,6	4,4	0,9	1,2	
	24	15,5	8,4	1,3	3,4	3,7	2,4	1,6	1,0	5,6	3,7	0,9	1,1	
	25	11,5	8,4	1,1	3,0	3,7	2,4	1,3	0,9	5,6	2,8	0,8	0,7	
	26	11,0	8,4	1,1	2,7	3,7	2,4	1,3	1,2	5,4	2,3	0,9	0,8	
	27	11,0	4,1	1,1	2,4	3,7	2,1	1,1	1,1	5,2	1,8	0,9	0,9	
	28	10,5	3,4	1,1	2,4	16	1,8	1,0	1,2	5,2	1,6	0,8	0,8	
	29	11,0	3,0	1,1	2,1	8,4	1,8	1,0	1,0	5,2	1,6	0,9	1,2	
	30	16,5		1,1	2,4	7,2	1,8	1,0	0,9	5,4	1,5	0,9	2,2	
	31	12,5		1,3		5,2		1,1	0,6		0,7		6,5	
Débits mensuels 1948	Bruts	13,9	7,7	1,7	4,0	5,8	3,7	1,6	1,5	4,7	4,3	1,9	1,6	4,36
	Corrigés ⁽¹⁾	17,9	6,7	2,5	5,1	6,7	4,4	2,0	1,5	1,2	0,3	1,9	2,1	4,36
Lame d'eau équivalente		123	43	17	34	46	29	14	10	8	2	13	14	353
Moyennes annuelles (M ³ /sec.) et totaux pluviométriques (en mm.) ↓														

Moyennes annuelles (M³/sec.)
et totaux pluviométriques (en mm.)

PLUVIOMÉTRIE EN 1948 (en millimètres)

	GENTIOUX (780)	115	16	101	87	105	25	41	72	31	89	68	91	841
PONTARION (448)		170	23	17	124	182	71	55	88	72	35	73	72	982

DÉBITS MOYENS MENSUELS (en m³ sec)

Période : 1919-1948	12,2	10,9	10,5	9,7	7,9	5,4	4,2	2,6	2,8	4,4	8,1	9,9	7,39
Période : 1920-1948	12,1	10,8	10,3	9,3	7,9	5,4	4,1	2,6	2,7	4,5	8,0	9,5	7,27

(1) Débits moyens mensuels corrigés du jeu du réservoir de la VAUGELADE.

Figure 12 : Exemple de mise en page des tableaux de débits journaliers pour la période 1948-1950.

- 77 -

Surface du Bassin Versant : 390 km²
 Altitude du zéro de l'échelle : 436 37

Station N° 9 en service depuis 1919

	JANV.	FEV.	MARS	AVRIL	MAI	JUIN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DEC.	
1	120	115	125	120	95	105	65	11	17	24	55	73	
2	21	110	120	150	91	99	55	12	19	49	43	66	
3	120	120	120	150	57	94	48	19	23	49	31	66	
4	64	130	115	115	62	75	41	36	24	46	24	66	
5	88	155	125	105	62	64	38	26	24	46	24	66	
6	88	150	145	105	57	62	36	23	48	46	19	91	
7	110	135	140	130	68	59	35	23	46	46	19	91	
8	100	155	155	160	68	57	31	24	46	46	20	97	
9	110	170	20	20	73	55	36	51	48	43	19	110	
10	96	135	165	32	58	39	35	49	49	43	43	150	
11	94	92	165	20	61	43	31	49	48	43	45	73	
12	20	79	170	160	55	39	31	35	46	39	26	62	
13	165	180	150	140	49	35	29	29	84	39	43	90	
14	130	170	20	130	46	29	29	26	190	43	70	93	
15	21	130	175	115	45	24	35	20	41	43	43	79	
16	20	100	155	100	46	26	31	18	29	49	33	75	
17	140	120	175	98	58	49	26	17	33	49	26	73	
18	165	26	26	120	56	92	24	16	33	48	29	72	
19	24	185	24	115	59	64	21	13	29	48	43	70	
20	160	190	28	130	59	51	21	24	51	46	39	70	
21	140	31	25	140	56	39	18	39	49	46	39	66	
22	125	28	195	120	53	43	18	23	48	46	88	48	
23	120	22	170	125	49	98	18	18	51	49	59	41	
24	115	165	185	91	46	140	53	17	57	46	62	36	
25	110	175	180	62	43	115	38	145	28	48	62	41	
26	110	150	140	66	51	81	28	145	66	49	62	110	
27	110	170	130	70	61	72	24	145	49	46	57	88	
28	88	150	125	64	59	72	23	145	36	46	70	92	
29	84	115	115	61	48	84	19	13	33	48	100	100	
30	81	130	130	59	49	98	19	17	28	53	98	125	
31	110	135	135	105	105	13	13	16	46	46	105	105	
DÉBITS MENSUELS EN 1951	1304	1615	1676	1262	574	674	315	233	458	456	468	786	813
Corrigés	1590	1670	1670	1250	580	670	310	230	340	120	550	630	797
Lame d'eau équivalente	90	100	115	84	39	45	22	16	30	31	31	54	657

PLUVIOMÉTRIE EN 1951 (en mm.)												
	JANV.	FEV.	MARS	AVRIL	MAI	JUN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DEC.
Gentoux (780)	202	234	174	160	121	145	49	139	120	21	97	91
Pontarion (448)	119	178	140	92	127	132	33	91	77	35	103	130

DÉBITS MOYENS MENSUELS (en M ³ /sec.)												
	JANV.	FEV.	MARS	AVRIL	MAI	JUN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DEC.
Période 1919-1951	1155	1164	1040	933	764	525	389	240	247	398	822	979
Période 1920-1951	1140	1152	1015	911	756	527	377	242	247	403	813	948

(1) Débits moyens mensuels corrigés du jeu du réservoir de la VAUGELADE.

Figure 13 : Exemple de mise en page des tableaux de débits journaliers pour la période 1951-1959.

SURFACE DU BASSIN VERSANT : 389 km²

- 117 -

N° 9

ALTITUDE DU ZÉRO DE L'ÉCHELLE : 436 m

STATION EN SERVICE DEPUIS 1919

				JANV.	FÉV.	MARS	AVRIL	MAI	JUIN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DÉC.			
1960	ÉCOULEMENT		1	17,0	8,4	9,8	6,1	5,6	1,85	2,5	2,8	3,2	4,9	20	6,1	K m 3 s		
			2	18,0	7,8	12,5	8,0	5,7	1,45	2,1	4,3	2,8	3,1	19,5	6,8			
			3	29	7,2	21,0	8,0	5,6	1,35	1,80	2,1	2,7	13,5	16,5	6,3			
			4	19,5	10,5	18,0	9,0	5,2	1,55	1,75	1,75	3,2	55	16,0	6,1			
			5	14,5	13,5	13,5	8,6	5,0	3,1	2,0	3,1	4,3	37	21	20			
			6	12,0	8,4	12,0	8,2	4,9	4,3	2,1	2,8	2,9	32	19,0	33			
			7	10,0	7,2	12,0	7,6	4,9	5,9	2,5	2,8	4,2	29	16,5	25			
			8	9,2	6,1	11,5	7,4	4,9	1,90	2,5	8,0	5,4	26	13,5	19,5			
			9	8,4	5,7	15,5	7,0	4,7	1,75	2,2	4,7	3,4	38	8,2	16,0			
			10	7,2	5,4	15,5	7,2	3,2	2,4	2,1	3,2	2,8	37	8,0	15,0			
		Moyennes	11	6,3	5,2	20	7,2	2,8	2,5	2,0	2,5	2,4	33	8,8	14,5			
			12	6,3	5,7	13,5	7,0	2,7	2,4	2,0	4,9	2,4	31	8,4	12,0			
			13	5,0	7,0	11,0	6,5	2,5	2,4	1,90	4,3	2,0	25	8,0	7,6			
			14	5,6	6,8	11,0	6,6	2,9	2,8	1,65	3,1	2,0	19,5	7,2	6,8			
			15	5,7	6,1	12,5	6,6	9,4	2,4	1,75	2,8	2,2	16,0	7,4	6,3			
			m ³ /s															
		journalières	16	5,2	8,4	13,0	7,2	8,6	2,1	1,65	2,4	4,3	11,5	14,5	5,7			
			17	5,4	12,5	10,5	8,0	5,6	3,0	1,55	2,2	3,2	9,6	13,0	5,4			
			18	5,2	25	9,4	7,0	4,7	2,0	1,65	4,3	2,5	8,2	12,0	5,0			
			19	5,9	21	8,4	6,6	6,5	1,75	1,90	7,6	2,8	7,0	13,5	5,6			
			20	21	14,0	7,6	6,6	3,8	1,75	1,65	4,0	5,7	8,2	12,0	15,5			
			21	23	12,0	7,0	6,5	3,5	1,65	1,65	2,8	5,2	12,5	11,0	11,5			
			22	12,0	12,5	7,6	6,3	4,5	1,55	1,55	2,2	3,1	8,6	22	8,4			
			23	9,2	11,5	9,2	6,1	3,5	2,5	1,55	2,2	2,7	12,5	16,0	8,2			
			24	8,8	11,0	8,8	5,9	4,0	3,4	2,1	2,1	4,2	11,0	12,5	6,5			
			25	11,0	9,6	8,8	5,9	3,8	2,5	2,2	2,0	4,0	14,0	9,4	6,5			
			26	12,5	8,4	8,0	5,9	3,5	13,5	1,90	7,4	2,1	25	8,8	7,4			
			27	11,0	7,6	8,2	5,7	2,9	11,5	1,75	13,0	2,1	19,0	7,4	16,0			
			28	22	6,6	8,8	5,6	2,7	5,2	1,55	11,5	3,2	20	9,2	10,0			
			29	12,5	6,3	8,6	5,4	2,5	3,5	1,45	5,9	3,7	33	9,6	14,5			
			30	10,0		5,6	5,4	2,4	2,7	1,90	4,2	4,3	38	7,0	16,5			
			31	9,0		5,0		2,2		2,5	3,7		27		18,0			
		Valeurs instantanées	maximum jour	m ³ /s														
		Moyennes	mesurées	m ³ /s	11,52	9,57	11,22	6,84	4,33	3,18	1,92	4,21	3,30	21,42	12,58	11,67	8,50	
			naturelles	m ³ /s	14,00	11,60	11,10	3,30	3,60	3,50	2,30	5,10	4,10	23,40	12,80	11,80	8,30	
				l/s/km ²	36	29,8	28,5	8,2	9,3	9,0	5,9	13,1	10,5	60,1	33,0	30,3	22,9	
				Mm ³	37	29	30	8	10	9	6	14	11	63	33	32	282	
mensuelles	mm	96	75	76	21	20	23	15	36	27	161	80	81	722				
1921-1960 48 ans	PRÉCIPITATIONS	Moyennes	mm													=		
		- Royère (850)	mm	145	128	136	31	88	170	116	253	166	436	152	183	2 003		
		- La Nouaille (880)	mm	98	100	85	13	67	161	73	157	82	378	96	122	1 430		
		- St-Sulp. les Charges (590)	mm	97	90	92	24	95	114	67	166	103	298	81	99	1 336		
		- Vallières (580)	mm	93	98	86	19	80	162	81	181	89	352	84	113	1 471		
		Moyennes 1921-1960	mm															
1919-60	ÉCOULEMENT	Déficit d'écoulement	mm															
		Moyennes	naturelles	mm	78	77	58	56	49	35	25	18	18	30	53	68	575	
				Mm ³	31	30	26	22	19	13	10	7	7	12	21	26	224	
				l/s/km ²	28,4	30,6	25,3	21,7	18,3	13,4	9,5	6,6	6,8	11,3	20,5	25,4	18,2	
				m ³ /s	11,44	11,89	9,84	8,44	7,13	5,20	3,69	2,56	2,65	4,40	7,99	9,87	7,08	
		mensuelles	m ³ /s	11,76	11,85	10,05	8,66	7,16	5,14	3,80	2,52	2,63	4,36	7,94	10,03	7,14		
1919-60	ÉCOULEMENT	Maxima journaliers	m ³ /s	61,0	52,0	63,0	57,0	56,5	62,0	74,0	34,0	24,0	55,0	47,0	56,0	10,8		
		Minima journaliers	m ³ /s	1,30	0,92	0,84	0,60	0,68	0,30	0,12	0,04	0,12	0,12	0,16	0,60	2,62		
AVATERS		STATION INFLUENCÉE depuis 1943 par le jeu de la réserve de LAVAUD-GELADE.																
		Autres références pages : 3, 17, 34, 52, 53, 252, 257.																

Figure 14 : Exemple de mise en page des tableaux de débits journaliers pour la période 1960-1963.

ALTITUDE DU ZÉRO DE L'ÉCHELLE : 438 m		LATITUDE : 46° 00' N												LONGITUDE : 1° 50' E	
		JAN.	FÉV.	MARS	AVRIL	MAI	JUIN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DÉC.		
1964	ÉCOULEMENT	1	9.95	13.5	6.20	6.85	10.2	1.22	.510	.420	.650	.830	4.44	Moyennes annuelles de pluies et débits	
		2	9.50	9.95	4.66	9.30	9.05	1.09	.460	.370	1.22	.960	4.44		
		3	9.30	7.95	3.80	7.50	8.20	1.09	.460	.370	.700	.830	4.22		
		4	8.40	7.30	3.60	7.10	11.3	.960	.460	.460	.830	1.09	6.40		
		5	7.95	6.40	3.60	9.70	13.5	.830	.420	.830	.650	2.60	6.00		
		6	7.95	5.55	3.60	12.8	10.8	1.09	.420	.960	.610	3.80	6.00		
		7	7.50	3.60	3.80	11.0	8.20	1.09	.460	.830	.650	4.22	6.05		
		8	7.50	3.60	3.80	9.40	7.30	.960	.560	.830	4.00	4.00	6.20		
		9	7.10	3.60	3.80	7.30	3.80	1.09	.560	.560	3.60	3.80	6.00		
		10	7.10	3.60	4.22	6.85	3.20	1.35	.560	.460	3.40	3.80	5.55		
Moyennes	11	7.75	5.55	5.10	6.20	6.85	3.00	1.22	.460	.420	3.80	3.60	5.30		
	12	7.50	5.55	5.55	6.60	6.20	1.35	.460	.560	3.80	3.80	5.10			
	13	7.30	5.30	5.10	6.20	5.75	1.35	.460	.460	3.80	3.80	5.10			
	14	6.20	5.55	4.22	7.10	3.80	3.60	.370	.460	1.87	4.00	4.66			
	15	5.75	6.65	6.20	6.00	5.55	4.00	.370	.510	1.87	4.44	4.44			
	16	5.30	6.85	6.00	5.55	6.40	4.00	.370	.560	3.40	4.44	4.44			
	17	5.30	6.40	4.44	6.40	3.80	3.80	.320	.510	3.40	4.22	4.44			
	18	5.10	7.75	4.00	21.0	9.50	3.80	.420	.460	2.60	4.44	4.44			
	19	5.10	7.10	5.10	48.0	9.05	4.00	.560	.510	4.00	4.00	4.44			
	20	4.88	7.10	9.05	30.7	4.66	3.80	.560	.510	1.61	3.80	4.22			
journalières	21	4.88	4.44	8.60	26.9	4.44	3.80	.460	.460	1.48	3.80	4.22			
	22	4.66	3.80	15.9	4.22	2.00	3.80	.830	.460	1.35	3.60	4.00			
	23	4.66	3.80	15.7	19.8	4.66	3.80	1.35	.420	1.35	3.80	4.00			
	24	4.44	3.80	12.1	28.9	7.75	3.80	.830	.420	1.74	3.80	4.00			
	25	4.88	4.00	18.1	21.6	8.20	3.60	.610	.460	1.48	3.60	2.80			
	26	4.88	4.88	14.7	16.4	6.85	3.60	.560	.460	1.22	3.60	3.00			
	27	4.66	4.44	11.3	14.2	9.05	3.80	.460	.610	1.09	3.80	3.00			
	28	4.66	4.44	9.70	12.6	8.60	3.80	.420	.560	.960	4.66	3.80			
	29	4.88	6.20	8.85	11.9	6.00	3.60	.460	.560	.830	4.88	2.60			
	30	5.10	7.75	12.1	4.88	1.35	3.00	.460	.560	.830	4.44	3.00			
Valeurs instantanées	31	8.60	7.30	7.30	4.66	.650	.650	.510							
	maximum jour														
	Moyennes	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	mesurées	4.05	5.10	8.35	16.1	8.05	1.28	.850	.560	2.10	2.25	3.06			
	naturelles	11.4	13.1	21.5	41.4	20.7	3.3	1.7	1.4	5.4	5.8	7.8			
		28	33	57	107	55	9	4	4	14	15	21			
	mensuelles														
PÉRIODES	Moyennes	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
		1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54		
Déficit d'écoulement	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
Moyennes	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
Déficit d'écoulement	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
Moyennes	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
Déficit d'écoulement	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45	5.80	7.25	18.0	7.45	2.54	.520	.675	1.76	3.55	4.54			
	1921-1960	6.45													

3. Problèmes rencontrés

Des difficultés ont été rencontrées lors du travail de numérisation des données contenues dans les annuaires de la SHF. Nous les avons listées ici afin d'en garder une trace et de présenter les solutions qui ont été apportées.

Les difficultés techniques liées spécifiquement à l'utilisation du logiciel d'OCR sont présentées en Annexe 1.

3.1. Manque de clarté des informations disponibles

Il était parfois difficile d'être sûr de la signification de certaines informations contenues dans les annuaires hydrologiques.

- La qualité de l'impression des annuaires n'était pas toujours suffisante pour être certain de ne pas faire d'erreur de lecture :

Ce problème s'est posé pour les données chiffrées (valeurs à moitié effacées) et pour les graphiques (à partir des années 1960) : certains pics de débits sont très mal passés à l'impression, et il est très difficile voire impossible de les distinguer.

- Les dates des données longues périodes n'étaient pas toujours très claires depuis la mise en page de 1960 (alignement non cohérent avec le reste du tableau).
- L'alignement du nom des stations hydrologiques était parfois source de confusion quant à la correspondance avec les données chiffrées (Figure 16).

Moyennes		<i>mm</i>								
— Le Thillot	(492 m)	<i>mm</i>	135	32	155	137	95	102	146	
— Rupt/s/Mos	(428 m)	<i>mm</i>	279	46	312	187	181	99	116	
— Remiremont	(380 m)	<i>mm</i>	206	43	216	156	153	125	86	
		<i>mm</i>	285	74	255	182	184	108		

Figure 16 : Les noms de stations pluviométriques ne sont pas très bien alignés avec les données chiffrées, et il manque un nom de station.

- Les valeurs d'extrêmes annuels étaient parfois incohérentes avec les valeurs du tableau.
- On a constaté des incohérences entre tableau de données et graphique :

Parfois, le graphique présenté dans l'annuaire n'est pas cohérent avec les données présentées dans le tableau correspondant. Cette incohérence est identifiée lors de la comparaison visuelle entre les deux graphiques (celui obtenu grâce aux données issus de l'OCR du tableau de données, et le graphique de l'annuaire). Un exemple est présenté en Figure 17 : le 1^{er} et 2 août, le tableau de l'annuaire indique une hausse du débit (64 et 57 m³/s au lieu des 5-10 m³/s en moyenne autour de ces deux dates sur le graphique). Cette hausse n'apparaît pas sur le graphe présenté dans l'annuaire. Elle apparaît cependant sur le graphe issu de l'analyse OCR, puisqu'elle se base sur les données du tableau. Il ne s'agit pas dans ce cas d'une erreur de l'OCR. Il s'agit soit d'une erreur de frappe dans le tableau, soit d'une correction ultérieure (mais on ne sait pas si la correction concerne le tableau ou le graphique). Ces incohérences ont été indiquées dans les commentaires du tableau d'avancement.

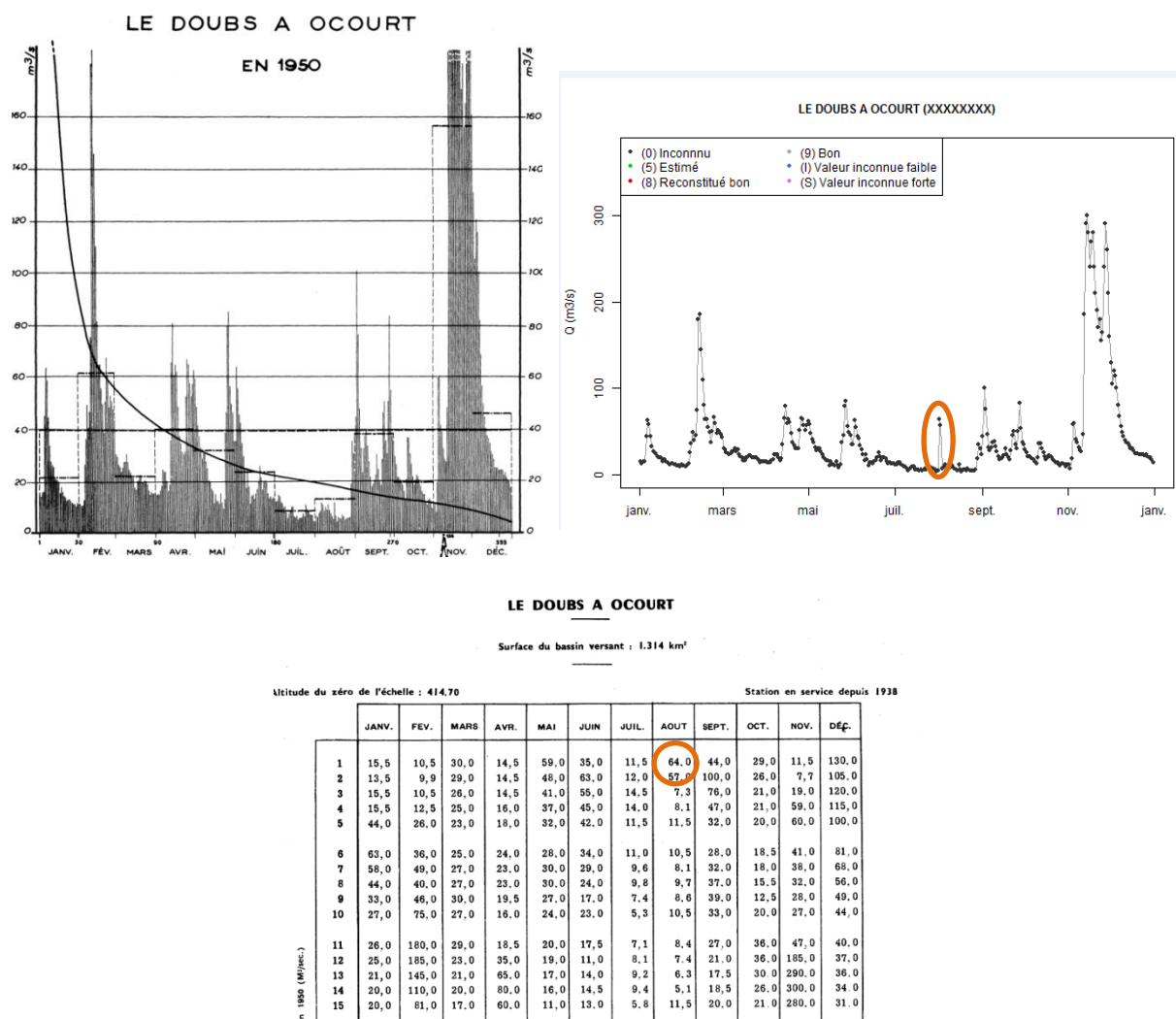


Figure 17 : Exemple d'incohérence dans l'annuaire identifiée grâce à la vérification visuelle des graphes. A gauche : graphique présenté dans l'annuaire. A droite : graphique obtenu grâce aux données issues de l'OCR. En bas : tableau présenté dans l'annuaire.

- Parfois, les séparateurs de décimales étaient mal indiqués.

Dans l'exemple ci-dessous (Figure 18), le logiciel d'OCR fait une erreur d'unité car la seule manière de comprendre que les sept nombres entourés possèdent une décimale est de voir l'alignement des nombres sur la colonne (pas de point ou de virgule représentés).

		JANV.	FÉV.	MARS	AVRIL	MAI
Débits journaliers en m³/s	1	100	150	22	105	115
	2	125	230	31	91	150
	3	160	210	44	73	135
	4	195	140	56	59	195
	5	250	96	56	30	185
	6	260	160	115	43	170
	7	290	190	125	54	115
	8	420	120	82	54	66
	9	370	125	87	66	55
	10	290	156	63	79	63
	11	150	110	62	135	84
	12	195	120	55	165	105
	13	220	135	55	110	105
	14	310	96	41	73	135
	15	390	84	30	135	175
	16	430	110	25	185	230
	17	350	59	22	210	125
	18	170	53	23	220	90
	19	150	48	190	100	170
	20	220	26	150	150	220
	21	280	150	100	140	210
	22	250	150	100	140	105
	23	190	140	100	60	150
	24	135	140	150	30	99
	25	61	195	180	25	68
	26	140	23	34	20	82
	27	260	24	59	49	68
	28	330	22	67	105	100
	29	340		43	170	67
	30	310		50	180	60
	31	185		86		60
	DÉBITS MENSUELS EN 1959	Bruts 242,8	Corrigés (1) 195,5	46,6 91,3	104,3 166,1	121,2 124,1
	Lame d'eau équivalente	119	33	55	97	71

Figure 18 : Illustration d'un cas où l'OCR commet des erreurs d'unités.

Dans de rares cas, un séparateur de millier était indiqué comme dans la Figure 19. Dans les tableaux de toutes les autres stations, le point représente le séparateur pour les décimales, mais ici c'est le séparateur des milliers.

LE RHONE AU TEIL

Surface du bassin versant : 67.332 km²

Station en service depuis 1910

		JANV.	FÉVR.	MARS	AVR.	MAI	JUIN	JUIL.	AOUT	SEPT.	OCT.	NOV.	DÉC.
Débits journaliers en 1939 (m³/sec.)	1	651	2.510	1.390	1.249	1.727	1.510	2.170	1.575	1.060	595	2.827	2.392
	2	815	2.358	1.302	1.302	1.693	1.510	2.635	1.470	1.074	595	3.040	2.302
	3	995	2.170	1.294	1.655	1.655	1.446	2.510	1.350	1.020	655	3.310	2.210
	4	1.478	1.955	1.235	1.874	1.575	1.486	2.180	1.193	995	655	3.340	2.060
	5	1.620	1.685	1.109	1.920	1.874	1.414	2.034	1.130	1.047	995	4.532	1.955
	6	1.655	1.510	1.095	2.170	1.655	1.510	1.655	1.318	1.109	1.270	4.732	2.034
	7	1.532	1.350	1.109	2.621	1.612	1.510	1.858	1.302	1.130	1.767	5.087	2.405
	8	1.470	1.172	1.655	3.385	1.612	1.575	1.735	1.235	1.034	1.920	4.860	2.180
	9	1.334	1.060	2.302	3.520	1.900	1.655	1.612	1.350	1.047	2.010	4.310	2.120
	10	1.235	1.027	2.051	3.115	2.080	1.655	1.560	1.532	995	1.858	3.895	2.018
	11	1.130	995	2.100	2.885	2.120	1.590	1.532	1.655	900	2.034	3.735	2.405
	12	1.350	943	2.018	2.840	1.990	1.774	1.470	1.430	900	2.405	3.780	3.190
	13	1.670	969	1.955	2.762	1.990	2.000	1.390	1.350	930	1.955	3.445	2.900
	14	1.670	930	1.920	2.696	1.942	2.250	1.350	1.310	900	1.928	3.040	2.560
	15	1.567	930	1.891	2.995	1.990	1.955	1.414	1.165	900	2.080	2.800	2.302
	16	1.655	918	1.753	3.010	1.942	1.693	1.575	1.060	930	2.260	2.560	2.220
	17	1.835	900	1.750	2.855	2.080	1.655	1.842	1.095	900	2.740	2.696	2.180
	18	2.336	894	1.767	2.560	2.370	1.575	1.842	1.060	900	2.980	4.362	2.060
	19	3.265	804	1.774	2.336	3.055	1.462	1.670	1.060	1.047	2.621	5.175	2.060

Figure 19 : Illustration d'un cas où le point représente le séparateur des milliers

Dans certains cas, sur un même tableau on pouvait avoir les deux types de séparateurs (point ou virgule) pour indiquer les décimales.

3.2. Pertinence des vérifications effectuées

3.2.1. Vérification par le calcul de moyennes

Parfois, la moyenne mensuelle indiquée dans l'annuaire ne correspond pas à la moyenne que l'on a recalculé avec les données journalières. Aucune erreur d'OCR n'explique cet écart : une vérification visuelle a été faite pour toutes les données journalières du mois en question. Il s'agit peut-être d'une erreur de calcul lorsque le rédacteur de l'annuaire a calculé la moyenne, ou bien d'une erreur de frappe (soit sur une donnée journalière, soit sur la donnée de moyenne), ou encore d'une correction ultérieure des moyennes mensuelles sans retouche des données journalières... Il est impossible de connaître la raison de cette incohérence. Dans le fichier de chronique, un code indique l'écart entre ces valeurs (Paragraphe 2.2.9.).

Ces écarts ont été souvent observés, et la vérification systématique de toutes les données journalière est assez chronophage. C'est pourquoi une certaine tolérance a été acceptée concernant ces écarts (Paragraphe 2.2.10.).

Au contraire, parfois une erreur d'OCR a été détectée alors que la différence entre les valeurs des moyennes mensuelles inscrite et calculée était faible. Ce critère d'écart entre les moyennes n'était donc pas toujours très discriminant. C'est pourquoi le seuil de tolérance a été abaissé en 1952 (de 10 à 1%), dès lors que les erreurs de calcul des annuaires (ou supposées telles) étaient moins fréquentes (Paragraphe 2.2.10.).

La vérification par le calcul n'était pas toujours possible : parfois certaines valeurs journalières étaient manquantes. Dans ce cas, il n'y avait pas de moyenne mensuelle permettant une vérification rapide des valeurs journalières. Elles ont donc dû être vérifiées une par une.

En 1962 et 1963, la comparaison avec les graphes a montré qu'à plusieurs reprises, les écarts qu'on avait pu observer entre les moyennes sans qu'il y ait d'erreur d'OCR peuvent s'expliquer par une erreur de frappe dans le tableau de l'annuaire : en corrigeant la valeur qui semble abhéroente à la vue du graphe, on retrouve par le calcul une valeur de moyenne cohérente avec celle indiquée dans l'annuaire (exemple en Figure 20). Ces erreurs ont été corrigées en remplaçant le nombre qui semble faux par un nombre de l'ordre de grandeur de celui qu'on peut lire sur le graphe, et en tenant compte des erreurs facilement faisables (remplacement d'un 5 par un 6, d'un 0 par un 9...).

33	24	13,0	6,4	5,9
31	25	13,0	6,2	5,6
30	24	13,0	6,1	5,7
21	27	12,5	5,8	5,5
22	26	11,5	6,5	5,4
21	38	94	6,5	5,2
21	36	11,5	6,0	5,9
27	30	11,5	5,9	5,3
34	29	11,5	5,8	5,2
35	29	11,0	5,4	5,1

Figure 20 : Exemple d'une faute de frappe manifeste : le 94 doit en fait être lu 9.4 pour retrouver la bonne moyenne mensuelle.

La vérification par le calcul n'est pas toujours suffisante, comme le montre l'exemple de la Figure 21 : si dans le même mois on cumule une surévaluation d'une valeur journalière et une sous-évaluation d'une autre valeur, on peut ne pas observer d'écart sur la moyenne mensuelle.

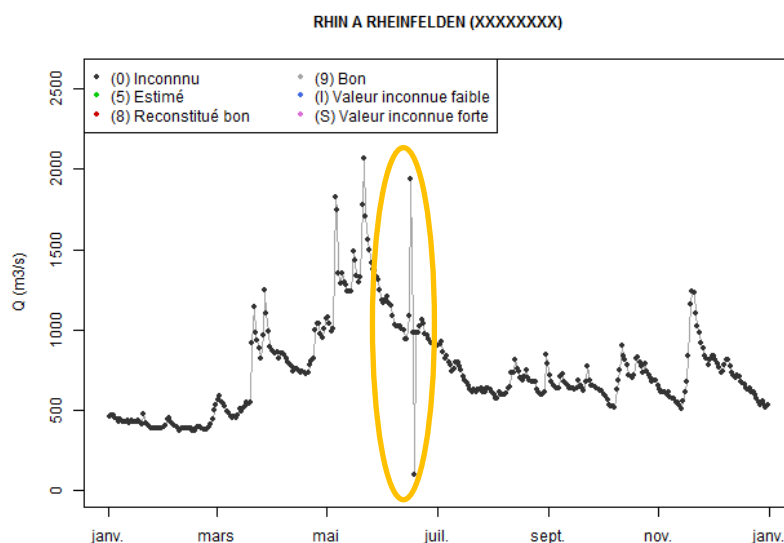


Figure 21 : Exemple pour lequel la vérification par le graphe a permis d'observer deux erreurs, non identifiables par la vérification par le calcul des moyennes mensuelles.

Ces différents éléments illustrent bien la pertinence d'utiliser un second moyen de vérification des erreurs ayant pu être commises par l'OCR.

3.2.2. Vérification des débits par visualisation des graphes

Souvent, la seconde vérification (à l'aide des graphiques) n'était pas possible pour les débits élevés, car ils n'étaient pas représentés sur le graphique de l'annuaire (exemple en Figure 22).

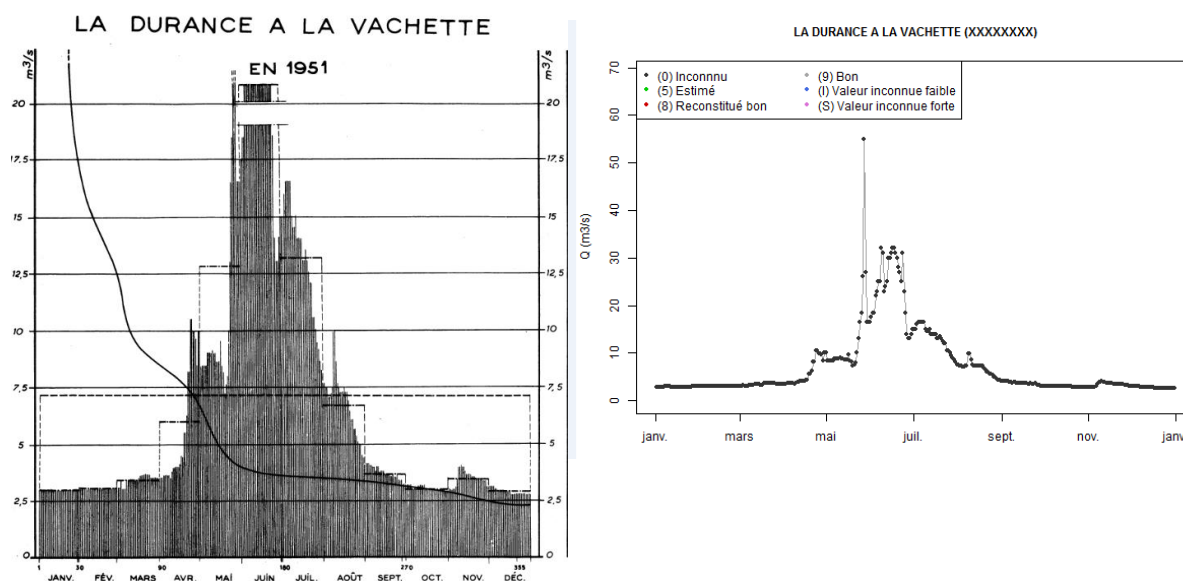


Figure 22 : Exemple d'un cas pour lequel la vérification visuelle des forts débits est impossible. Ici, on ne peut pas vérifier visuellement les débits supérieurs à 20 m³/s.

Le problème des forts débits ne se pose plus à partir de 1964 ; en revanche, les graphes deviennent bien moins précis.

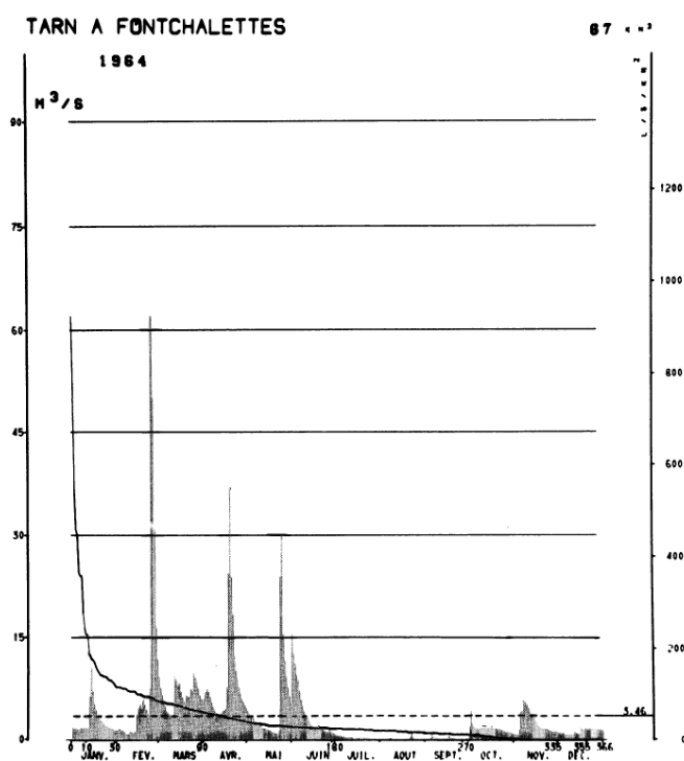


Figure 23 : Exemple de graphe à partir de 1964. On ne distingue plus les données jour par jour.

On a également eu des problèmes de cohérence d'échelles, qui ont rendu les comparaisons impossibles (exemple en Figure 24).

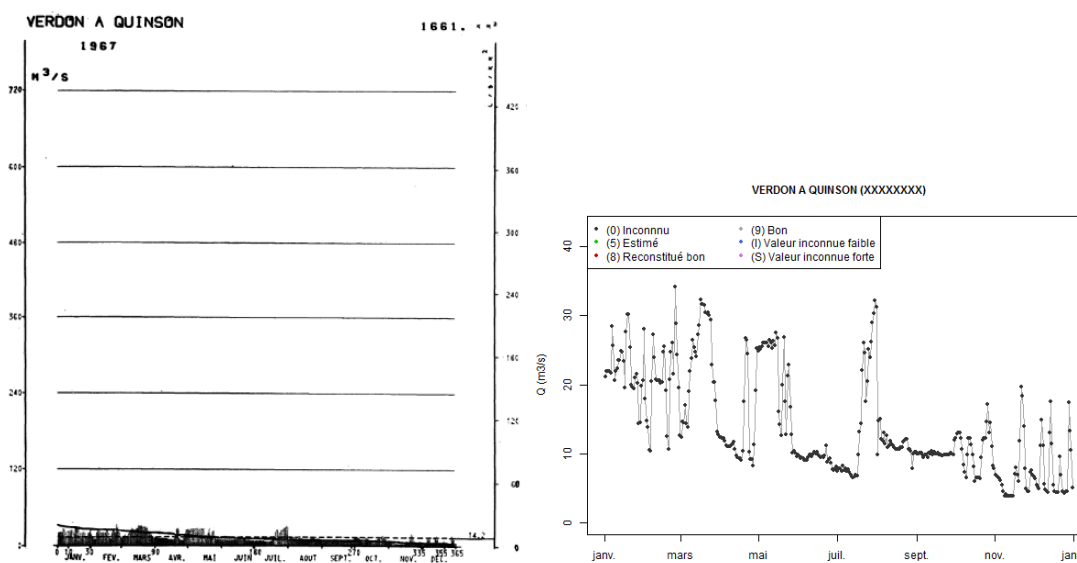


Figure 24 : Le changement d'échelle rend la vérification sur le graphe issu de l'annuaire difficile.

3.3. Continuité des mesures sur les stations des annuaires

Les annuaires hydrologiques présentent des données de débits journaliers sur 120 stations hydrologiques, pour la période 1939-1969. Sur cette période, certaines stations ont été déplacées, remplacées par d'autres ou ont changé de nom car la dénomination n'était plus appropriée...

Malheureusement, les annuaires ne sont pas toujours très explicites sur ces évolutions, et des erreurs peuvent se glisser au milieu d'informations exactes. Il y a ainsi beaucoup d'incohérences temporelles sur les numéros et noms de station, les valeurs d'altitude et de surface du bassin versant, les dates de mise en service, la reconnaissance de la station en tant que usine ou non, les commentaires... Ainsi, de très nombreux annuaires présentent des errata pointant ces incohérences.

Malgré tout, des doutes subsistent. Il est difficile de savoir si des chroniques de débits journaliers mesurés deux années consécutives se rapportent exactement à la même station, ou à deux stations différentes.

Une comparaison des métadonnées pour les stations où la question de continuité se posait a été engagée : elle concerne la surface de bassin versant, l'altitude, la localisation et la date de mise en service des stations. Les chroniques ne sont jamais chevauchantes ; cependant, la comparaison avec la date de mise en service peut être un indicateur pour déterminer si deux stations sont différentes.

Dans les tableaux suivants, nous avons indiqué par un code couleur :

- Rouge : sur la base de ce critère, les stations semblent différentes.
- Vert : sur la base de ce critère, les stations semblent identiques.
- Noir : ce critère ne donne pas d'information dans un sens ni dans l'autre.

Des informations contenues dans le texte des annuaires sont parfois citées pour appuyer les conclusions. Les caractéristiques des stations ont été prises pour la plupart sur l'annuaire de la dernière année d'apparition de la station.

La comparaison avec les stations présentes dans la banque HYDRO peut parfois être éclairante pour conclure sur la continuité d'une station des annuaires. Des tableaux présentant les métadonnées des stations de la banque HYDRO sont donc parfois présents. Lorsque les séries annuelles ont été jugées provenir de la même station ou de stations très proche, nous avons concaténé les débits des différentes années pour fournir une chronique étendue.

3.4. Correspondance avec les stations de la banque HYDRO

Certaines données présentées dans les annuaires hydrologiques sont déjà dans la banque HYDRO. Ces données sont donc déjà exploitables, il n'est pas nécessaire de réaliser le travail d'analyse par OCR. Afin d'identifier ces données qui font doublon, on cherche à identifier quelles stations présentées dans les annuaires correspondent aux stations de la banque HYDRO.

L'évolution du réseau hydrométrique et les incohérences des informations présentées dans les annuaires rendent cette tâche parfois difficile (changement de nom, de surface de bassin versant, d'altitude...). C'est pourquoi il existe un doute sur la correspondance réelle entre certaines stations décrites dans les annuaires et les stations décrites dans la banque HYDRO.

Pour certaines stations, on peut avoir des données de débits journaliers sur une ou plusieurs années communes. On regarde alors l'écart entre les valeurs des débits journaliers de la station de l'annuaire et de la station de la banque HYDRO qui semble correspondre, sur la première année pour laquelle on possède des données communes, puis quelques années plus tard pour des vérifications. En cas de doute, la comparaison est faite sur une période plus longue.

On s'intéresse aux différences absolues entre chaque couple de valeur. Les quantiles de la distribution de ces différences sont inscrits dans un tableau. On considère que lorsque le quantile 90 (Q90) est nul, les données sont similaires, et donc que les deux stations sont les mêmes. Pour certaines stations, le quantile 90 n'est pas nul mais il est faible. On a donc également calculé la moyenne des pourcentages de différence sur tous les couples de valeurs.

- Le premier critère auquel on s'intéresse est le quantile 90 : s'il est égal à zéro, on considère les stations comme similaires. C'est par exemple le cas au Loup Amont en 1963 : le Q90 = 0, mais les forts écarts en fin de quantiles conduisent à un pourcentage de différences de 23%. On considère la station comme correspondant à celle de la banque HYDRO.
- Ensuite, c'est la moyenne du pourcentage de différence. Tant qu'on est inférieur à 5%, on considère que c'est la même station (dans l'onglet « comparaison BH-annuaires » du fichier « données à récupérer.xlsx » = oui ?, pas de justification). Entre 5 et 10%, on essaye de multiplier les années de comparaison pour vérifier (= ?, justification dans un paragraphe du rapport).
- Quand le pourcentage de différence est supérieur à 10%, on a de sérieux doutes, et on compare plusieurs années (=non ? quand proche de 10%, non quand plus éloigné, justification dans un paragraphe du rapport).
- Le rapport entre Q90 et débit moyen de l'annuaire peut aider dans le cas de très faibles débits.

Les résultats de toutes ces comparaisons sont compilés dans un fichier Excel.

Pour certaines stations, la comparaison directe avec les données bancarisées dans la banque HYDRO n'était pas possible (pas de chevauchement entre les chroniques). En plus de la confrontation des noms, surface de bassin versant..., on a appliqué des tests de stationnarité qui ont pu nous éclairer sur l'homogénéité des deux chroniques non chevauchantes. Il s'agit du test de Pettitt, de la statistique U de Buishand, de la recherche de rupture selon Lee et Heighinian, et du test de segmentation de Hubert.

3.5. Analyse par station

Ce paragraphe propose une synthèse des analyses réalisées sur les stations des annuaires. Il sera question de continuité entre stations des annuaires et les recouvrements avec la banque HYDRO. Pour alléger le document, nous ne faisons état que des stations où il y avait ambiguïté.

3.5.1. Station 2/2bis - La Moselle à Epinal

De 1939 à 1941, la station est numérotée n°2. Par la suite, de 1954 à 1969, elle est numérotée n°2bis, puis n°2-1 (à partir de 1957). Ce changement de numérotation indique une modification de cette station.

Tableau 4 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
2		LA MOSELLE A ÉPINAL	1250	320.14				39-41
2bis	2-1	LA MOSELLE A ÉPINAL	1215	324			1952	54-69

1954 : "Nous reprenons cette année la publication des débits à Epinal interrompue depuis 1942. En fait, il s'agit d'une station entièrement nouvelle, les emplacements de l'échelle et de la section de jaugeage ayant été complètement modifiés. Le nombre de jaugeages effectués depuis 1953 a permis d'établir une courbe de tarage et les débits publiés doivent être considérés comme très satisfaisants. En raison de la grosse lacune entre 1941 et 1951 et de l'imprécision des résultats antérieurs, il n'a pas été possible de reconstituer une longue période antérieure à 1951."

1954 : "La publication des débits a été interrompue en 1942 par suite de leur incohérence."

- On peut donc considérer qu'il s'agit de deux stations différentes, portant malgré tout le même nom.
- Les surfaces sont quasiment identiques ; les données sont donc comparables et donc ont été concaténées. Cependant, les données de la première station (1939-1941) sont considérées par le gestionnaire comme incohérentes. La question de leur usage se pose.

3.5.2. Station 3 - La Cure aux Settons/Le Chalaux à Chaumeçon

Deux stations sont numérotée n°3 : « La Cure aux Settons » en 1939 et « Le Chalaux à Chaumeçon » par la suite (1940-1969).

Tableau 5 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
3		LA CURE AUX SETTONS	50					39
3	3	LE CHALAU A CHAUMEÇON	98					40-69

1940: "A partir de 1940, la station n°3 des Settons sur la Cure **est remplacée**, sous le même numéro, par la station du Chaumeçon sur le Chalaux, où la mesure des débits est plus précise".

- Il s'agit de deux stations différentes et les débits de la Cure aux Settons ne sont disponibles que pour 1939.

3.5.3. Station 5ter – La Loire à Montjean

Des années 1954 à 1963, la moyenne des pourcentages de différence entre les couples de valeurs de débits journaliers dépasse les 5%, mais est inférieure à 10%. Le quantile 90 est assez important.

En 1964 et 1966 en revanche, les chroniques sont pratiquement similaires ($Q_{90} = 0$).

Sur l'ensemble de la chronique, les séries se superposent plutôt bien.

Il a pu y avoir une modification des courbes de tarage en 1964, avec correction des années antérieures dans la banque HYDRO.

Tableau 6 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	5-2	LOIRE A MONTJEAN	110000	9	1863
Banque HYDRO	M5300010	La Loire à Montjean-sur-Loire	109930	10	1842

➤ La station M5300010 de la banque HYDRO semble correspondre (du moins pour les années les plus récentes). Les données de l'annuaire n'apporte pas de données complémentaires pour cette station.

3.5.4. Station 10 - La Creuse à Eguzon

Il existe deux dénominations dans les annuaires : « La Creuse à Eguzon » et « La Creuse (Grande Creuse, Petite Creuse et Sédelle) à Eguzon ». Ces deux stations sont probablement les mêmes.

Tableau 7 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
10	10	LA CREUSE A ÉGUZON	2400	148	46°26N	1°34E	1926	39 ; 47-69
10		LA CREUSE (GRANDE CREUSE, PETITE CREUSE ET SEDELLE) A ÉGUZON	2400	146.5			1919	40-46

Tableau 8 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
L4530710	La Creuse à Éguzon-Chantôme	2400	147			01/01/1927	Oui (vérification 1939 et 1941)

➤ Il s'agit vraisemblablement de la même station, les données ont donc été concaténées. Ces données sont identiques aux arrondis près à la station de la banque HYDRO. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.5. Station 12 – La Dordogne à Argentat

Durant le début de la chronique les écarts sont importants (voir très importants), et ils deviennent ensuite très faibles dans les années 1960. Des écarts systématiques se concentrent principalement sur les crues avant 1948.

Tableau 9 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	12	DORDOGNE A ARGENTAT	4418/4420	173	1899
Banque HYDRO	P1350010	La Dordogne à Argentat	4420	173	1929 (Débits validés douteux depuis 1900)

➤ La station P1350010 de la banque HYDRO semble correspondre (du moins pour les années les plus récentes). Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.6. Station 13 – La Dordogne à Cénac (Domme)

Pendant tout le début de la chronique les écarts sont importants (mais restent inférieurs à 10%) et ils deviennent ensuite très faibles dans les années 1960.

On constate des écarts sur les valeurs de crue non systématiques (valeurs des annuaires parfois inférieures et parfois supérieures par rapport à celles de la banque HYDRO).

Tableau 10 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	13	DORDOGNE A CÉNAC (DOMME)	8705/8700	65.14/65	1918
Banque HYDRO	P2380010	La Dordogne à Cénac-et-Saint-Julien [Cénac]	8705	65	1900

➤ La station P2380010 de la banque HYDRO semble correspondre (du moins pour les années les plus récentes). Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.7. Station 15-1 – La Maronne à Basteyroux

L'écart important les premières années semble dû à une correction de certains événements. Deux courbes de tarage semblent avoir été utilisées successivement.

Une valeur étrange apparaît le 14/01/1947.

La convergence semble se faire dans les années 1950.

Tableau 11 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	15-1	MARONNE A BASTEYROUX	821	179	1918
Banque HYDRO	P1592510	La Maronne à Argentat [Pont de Basteyroux]	821	178	1918

➤ **La station P1592510 de la banque HYDRO semble correspondre (du moins pour les années les plus récentes. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.**

3.5.8. Station 17 – La Vézère à Uzerche

Les données correspondent bien, mis à part des écarts de faible/moyenne importance les premières années (45-49). Les deux stations de la banque HYDRO présentent les mêmes données pour les années 1918-1983 (cf. commentaires de la banque HYDRO).

Il y a probablement des erreurs (provenant des annuaires ou de la banque HYDRO) les 17/01/1943 et 24-25/05/1960.

Tableau 12 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	17	VÉZÈRE A UZERCHE	601	286.2	1918
Banque HYDRO	P3131020	La Vézère à Uzerche	601	286	1900
Banque HYDRO	P3131010	La Vézère à Uzerche [aval]	609	286	1900

Commentaires relatifs à la station P3131020 : Du 01/01/1918 au 31/12/1983 les QJ sont issus de la station aval (P3131010) qui est fermée depuis. Débits absents avant 1918, provisoires ensuite.

Commentaires relatifs à la station P3131010 : Débits validés douteux depuis 1918 (absents en 1952 et 1969)

➤ **La station P3131010 de la banque HYDRO semble correspondre. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.**

3.5.9. Station 18 – La Garonne à Valentine

Les écarts sont assez faibles, parfois entre 5 et 10%. Pourtant, la superposition des graphes montre des différences importantes, notamment en ce qui concerne les crues.

Des valeurs étranges sont présentes les 31/10/1939, 02-03/02/1952 et 14/09/1963 (ce ne sont pas des erreurs d'OCR).

Tableau 13 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	18	GARONNE A VALENTINE	2256/2230	354.9/357	1899
Banque HYDRO	O0200030	La Garonne [totale] à Saint-Gaudens [Valentine - ancien]	2230	357	1912 (Débits validés douteux depuis 1912)

➤ **La station O0200030 de la banque HYDRO semble correspondre, mais un doute subsiste.**

3.5.10. Station 18bis – La Garonne à Portet

Malgré une différence de presque 10% en 1956, par la suite les écarts sont plutôt faibles en comparant à la station O1900010. La station d'essai pourrait aussi correspondre mais avec des écarts un peu plus grands. Les métadonnées correspondent à peu près.

Tableau 14 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	18-1	GARONNE A PORTET	9980	139	1918
Banque HYDRO	O1900010	La Garonne à Portet-sur-Garonne	9980	147	1910
Banque HYDRO (station d'essai)	O1900015	La Garonne à Portet-sur-Garonne	9980	147	1910

➤ **Les stations O1900010 et O1900015 de la banque HYDRO peuvent correspondre. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.**

3.5.11. Station 18-2/18-3 - La Garonne à Malause-Lamagistère

L'ajout de la dénomination « Lamagistère » n'apparaît que pendant trois ans (de 1967 à 1969). Les deux premières années, elle est numérotée 18-2, comme la station précédente « La Garonne à Malause », puis 18-3 en 1969. Ces deux stations pourraient être les mêmes, mais la surface de bassin versant est assez différente. L'altitude, la localisation et la date de mise en service sont identiques. Les annuaires ne mentionnent pas l'ajout d'une nouvelle station en 1967 ou 1969. Dans la banque HYDRO, il existe deux stations qui peuvent correspondre : O6000010 « La Garonne à Malause » et O6140010 « La Garonne à Lamagistère »

Tableau 15 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
18ter	18-2	LA GARONNE A MALAUSE	30900	57	44'05'N	0'58'E	1920	56-66
	18-2 18-3	LA GARONNE A MALAUSE- LAMAGISTERE	32350	57	44'05'N	0'58'E	1920	67-69

Tableau 16 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
O6000010	La Garonne à Malause	30920	64			01/01/1915	Correspond à la première station (comparaison en 1956, 1957 et 1962)
O6140010	La Garonne à Lamagistère	32350	48			01/01/1967	Correspond à la seconde

➤ Il s'agit de deux stations différentes mais que l'on peut rapprocher de deux stations de la banque HYDRO. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.12. Station 19 – La Garonne au Mas d'Agenais

Les débits sont assez similaires, il y a probablement eu des modifications de la courbe de tarage qui expliqueraient des écarts supérieurs à 5% dans les années 1940. Les différences sont très faibles ensuite.

Quelques valeurs anormales sont identifiées les 28/02/1952, 17/02/1952, 05-06/02/1952 et 12/02/1944 (ce ne sont pas des erreurs d'OCR).

Tableau 17 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	19	GARONNE AU MAS D'AGENAIS	52000	17.39	1914
Banque HYDRO	O9060010	La Garonne au Mas-d'Agenais	52000	35	1914

➤ La station O9060010 de la banque HYDRO semble correspondre. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.13. Station 20 – Le Salat à Kercabanac

Les écarts sont supérieurs à 5%. Les graphes montrent des débits assez similaires, il semble qu'il y ait eu plusieurs courbes de tarage. La convergence se fait en 1951.

Une valeur étrange est présente le 17/11/1941.

Tableau 18 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	20	SALAT A KERCABANAC	379.3/383	459.8/460	1931
Banque HYDRO	O0362510	Le Salat à Soueix-Rogalle [Kercabanac]	379	464	1931

➤ La station O0362510 de la banque HYDRO semble correspondre. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.14. Station 22 – L'Ariège à Foix

Les écarts sont supérieurs à 10% au début de la chronique. La convergence semble se faire en 1958 (sauf fin 1962 et 1963) avec la station O1252510. Plusieurs courbes de tarages semblent avoir été utilisées. La station O1252515 peut aussi correspondre.

Tableau 19 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	22	ARIÈGE A FOIX	1340	370	1905
Banque HYDRO	O1252515	L'Ariège à Foix	1340	375	1905
Banque HYDRO	O1252510	L'Ariège à Foix	1340	375	

➤ La station O1252510 de la banque HYDRO semble correspondre. Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.15. Stations 25/25bis/25ter - La Neste de Rioumajou

Il existe dans les annuaires quatre stations différentes :

- 1. N°25 La Neste de Rioumajou à Tramezaygues,
- 2. N°25bis La Neste de Rioumajou à Saint Lary,
- 3. N°25bis La Neste de Rioumajou à Saint Lary (Maison blanche),
- 4. N°25ter La Neste de Rioumajou à Maison blanche,
- Il existe aussi la station N° 25-2 « Cap de Long aux Edelweiss ». Cette numérotation laisse supposer qu'il y a seulement deux autres stations sur la Neste. La station 25-2 « Cap de Long aux Edelweiss » est numérotée n°25 bis de 1949 à 1952 (probablement par erreur). Elle

reprend ensuite la bonne numérotation en 1953 (25 ter). Cette station est nommée « La Neste de Cap de Long aux Edelweiss » en 1969, ce qui semble plus adapté.

Les comparaisons des surfaces de bassins versants, des altitudes et des dates de mise en service donnent des interprétations contradictoires. La carte présentée en Figure 25 permet de représenter la configuration des cours d'eau.

Les commentaires dans les annuaires laissent penser qu'il n'y a pas deux mais trois stations en plus de la 25-2 « Cap de Long aux Edelweiss ». Il est probable que la troisième station « La Neste de Rioumajou à Saint Lary (Maison blanche) », qui n'apparaît qu'en 1948, corresponde soit à la seconde soit à la quatrième station.

Tableau 20 : Données des annuaires hydrologiques. Le tableau ne comporte pas le code couleur car les comparaisons doivent se faire entre chacune des lignes.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km ²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
25		LA NESTE DE RIOUMAJOU A TRAMEZAYGUES	67	905.06			1912	39-41
25bis		LA NESTE DE RIOUMAJOU A SAINT-LARY	63.7	1053			1942	43-47
25bis		LA NESTE DE RIOUMAJOU A SAINT-LARY (MAISON BLANCHE)	63.7	1053			1948 (usine)	48
25ter	25-1	LA NESTE DE RIOUMAJOU A MAISON BLANCHE	67	1026	42°47N	0°18 ^E	1948	49-69
25ter	25-2	CAP DE LONG AUX EDELWEISS	5	2182	42°49N	0°08 ^E	1951	56-69

Tableau 21 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
00126210	La Neste de Rioumajou à Tramezaïgues [Maison Blanche]	63.7	1182			01/01/1948	Pourrait correspondre à la station 25bis de 1948 et à la station 25-1 de 1949-1969
00126220	La Neste de Rioumajou à Saint- Lary-Soulan [1]	67	905			01/01/1939	Pas de recouvrement avec les données des annuaires
00105110	La Neste de Cap de Long à Aragnouet [Les Edelweiss]	5	1715			01/01/1948	Semble correspondre à la station Cap de long aux Edelweiss (25ter)

Annuaire 1943 : « Les mesures de débit sur la NESTE de RIOUMAJOU, interrompues en 1942, ont été reprises à l'usine de SAINT-LARY (Bassin versant 66 km²) donc la station n° 25 bis remplacera désormais celle de TRAMEZAYGUES (Bassin versant 67 km²). »

1947 : "Un planimétrage effectué par le Groupe de Production Hydraulique « Pyrénées Centre » conduit à prendre comme surface du bassin versant : 63.7 km² au lieu de 66 km². L'altitude du zéro de l'échelle n'est plus 1.051,85, mais 1.053. Cette cote est celle du déversoir."

Les commentaires donnent la chronologie suivante :

« De 1921 à 1941 : Tramezaygues d'Aure (67 km²) » ; « Pas de relevés en 1942 » ; « De 1943 à 1948 : St Lary (prise de l'usine de St Lary 63.7km²) »

Les données de la dernière station débutent en 1948, un autre changement a donc eu lieu cette année.

On ne sait donc pas à quelle période appartient l'année 48 (43 à 48 ou 48 à 69), mais probablement plutôt à la période 48-69.

Dans la banque HYDRO, il existe une seule station pouvant correspondre : 00126210 « La Neste de Rioumajou à Tramezaïgues [Maison Blanche] », qui pourrait correspondre à la fois à la station 1948 et à la station 1949-1969.



Figure 25 : Carte du secteur de Tramezaïgues. Les surfaces de bassin versant sont indiquées à trois points différents.

Il a existé une première station jusqu'en 1941, qui a été remplacée en 1943.

La correspondance entre stations 25-1 et la station O0126210 est variable, avec des écarts importants en début de chronique qui se réduisent ensuite. Les débits sont, de manière générale, assez faibles, ce qui peut biaiser les conclusions sur les calculs des écarts relatifs.

Les écarts entre les débits de la station 25-2 et ceux de la station O0105110 sont parfois supérieurs à 5%, voire pour l'année 1962 supérieurs à 10%. Toutefois, les débits moyens sont très faibles et on a souvent des valeurs identiques. Il semble qu'il y ait une faute de frappe le 07/12/1967 dans l'annuaire (555 au lieu de 55).

- **Par la suite, un autre changement a eu lieu en 1948. On considère que les données de 1948 correspondent à la station 25-1 Neste de Rioumajou à Maison Blanche.**
- **Une station supplémentaire est située à Cap de Long.**
- **Les numérotations des annuaires pour identifier les stations sont peu fiables. Les surfaces sont peu cohérentes entre sources de données.**
- **Vue leur proximité, les trois stations ne doivent pas être très différentes hydrologiquement.**
- **La station 25bis apporte des données complémentaires sur la période 1943-1947 qui pourraient enrichir les séries des stations O0126210 et O0126220 de la banque HYDRO.**
- **La station 25-2 semble correspondre à la station O0105110 (malgré les écarts constatés).**
- **La station O0105110 semble bien correspondre, mais un doute subsiste.**

3.5.16. Station 27/27bis - Le Tarn au Pont de Montvert/Le Tarn à Fonchalettes

Les annuaires présentent deux stations différentes (« Le Tarn au Pont de Montvert » et « Le Tarn à Fonchalettes »), sans ambiguïté de numérotation (n°27, puis 27bis à partir de 1949). Les surfaces des bassins versants sont très proches.

Dans la banque HYDRO, il existe plusieurs stations pouvant correspondre : O3011010 « Le Tarn au Pont-de-Montvert [Fontchalettes] ». Ce nom suggère qu'il n'existe qu'une seule station dans l'annuaire.

Tableau 22 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
27		LE TARN AU PONT DE MONTVERT	67.7	865.9			1918	39-48
27bis	27-1	LE TARN A FONCHALETES	67	905			1948	49-69

Tableau 23 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
O3011010	Le Tarn au Pont-de-Montvert [Fontchalettes]	67	925			01/01/1948	Oui
O3010101	Le Tarn [corrigé] au Pont-de- Montvert [Fontchalettes]	67	905			01/01/1947	Oui
O3011020	Le Tarn au Pont-de-Montvert [ancien]	67	865			01/01/1918	Oui

1949 : « La Station **est remplacée** par la station de Fontchalette (BV 67 km²), située légèrement à l'amont et équipée d'un limnigraphe ».

- Il s'agit de deux stations différentes, mais très proches donc les données ont été concaténées.
- La comparaison aux données de la banque HYDRO montre que les débits sont quasi identiques à des arrondis près à partir de 1958 (des écarts apparaissent en partie à cause d'un décalage temporel d'un jour autour du 01/01/1958). Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.17. Station 29 - Le Viaur à Thuriès / Thuriès-Roc-Miquel

Dans les annuaires, deux dénominations existent. Pourtant, il semble que ce soient les mêmes stations (surface de bassin versant, altitude, localisation, date de mise en service similaires). Il n'est pas fait mention d'une nouvelle station dans l'annuaire 1964.

Dans la banque HYDRO, il existe deux stations qui pourraient correspondre.

Tableau 24 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km ²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
29	29	LE VIAUR A THURIÈS	1050	263	44'04'N	2'16'E	1921	39-63
	29	LE VIAUR A THURIES-ROC- MIQUEL	1050	263	44'04'N	2'16'E	1921	64-69

Tableau 25 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
O5482930	Le Viaur à Saint-Just-sur-Viaur [Le Cambon]	1010	274			31/12/1990	Une année commune 1962
O5482925	Le Viaur à Saint-Just-sur-Viaur [Le Cambon ancien]	1010	304			01/01/1930	Correspond à la première station
O5482920	Le Viaur à Saint-Just-sur-Viaur [Roc Miquel]	1008	341			01/01/1962	Correspond à la seconde station

- Il s'agit peut-être d'une seule station. La proximité entre les deux stations des annuaires nous a incités à concaténer les séries.
- Les stations O5482925 et O5482930 s'arrêtent en 1940 et permet une comparaison avec la première station sur deux ans (1939-1940). Une comparaison est possible sur l'année 1962 avec la station O5482920. La corrélation entre débits journaliers des stations est élevée mais le coefficient directeur de la droite de régression n'est pas 1 sur 1939-1940 mais s'en approche en 1962.
- Les données de la banque HYDRO sont partiellement recouvrantes avec les données des annuaires. Les annuaires pourraient apporter des débits sur la période 1941-1961.

3.5.18. Station 30 – L'Agout à Clot

Les différences sont supérieures à 10% sur la période 1939-1946. En 1955, les écarts sont très faibles. La relation entre les débits des annuaires de la SHF et les débits de la banque HYDRO n'est pas vraiment linéaire. Des valeurs étranges apparaissent du 10 au 15 février 1954 (ce ne sont pas des erreurs d'OCR). Il semble qu'il y ait eu des corrections nombreuses.

Tableau 26 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	30	AGOUT A CLOT	1028	157.7	1918
Banque HYDRO	O4252510	L'Agout à Castres [Clot]	1028	158	1917

La mise hors service de la station O4252510 le 01/01/1957 correspond à l'arrêt de publication dans les annuaires de la SHF.

- La station O4252510 semble correspondre (du moins pour les années les plus récentes). Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.19. Station 31 – Le Lot à Cajarc

De 1960 à 1966, la station qui correspond est la station O7991510 (écarts nuls).

Cependant sur la période antérieure, la station O7981510 semble bien correspondre, exception faite que les pics de crue sont plus importants dans les annuaires. Cette station O7981510 ne présente pas de données de débits après 1942, contrairement à l'annuaire. Deux erreurs dans la banque HYDRO ont été identifiées le 11/08/1941 et le 02/11/1941 (ce ne sont pas des erreurs d'OCR).

Tableau 27 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	31	LOT A CAJARC	7013/7033	147/137	1913
Banque HYDRO	O7981510	Le Lot à Cajarc [Pont]	7013	145	1912 (Mise hors service en juin 1946)
Banque HYDRO	O7991510	Le Lot à Cajarc	7033	137	1913 (Débits validés douteux depuis 1960, absents avant)

- **La station O7991510 correspond après 1960.**
- **Il est vraisemblable que la station O7981510 corresponde, mais des doutes subsistent.**
- **Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1943-1945 et 1949-1959.**

3.5.20. Station 33-1 - L'Adour à Asté

La station qui correspond le mieux est la station Q0100020, mais la correspondance ne devient vraiment bonne que dans les années 1960.

Les quatre stations de la banque HYDRO qui peuvent correspondre présentent les mêmes métadonnées.

Un évènement ne correspond pas : il s'agit de celui de début février 1952 (aussi bien pour la station Q0100020 que Q0100010)

Tableau 28 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	33-1	ADOUR A ASTE	272	600	
Banque HYDRO	Q0100020	L'Adour [total (1950 à 1984)] à Asté [1]	272	600	
Banque HYDRO	Q0100010	L'Adour [Adour seul] à Asté [Pont d'Asté]	272	600	
Banque HYDRO	Q0105110	Le canal d'Asté à Asté [Pont d'Asté]	0	600	
Banque HYDRO	Q0100025	L'Adour [Adour+canal] à Asté [Pont d'Asté]	272	600	

- La station Q0100020 semble correspondre (du moins pour les années les plus récentes), mais un doute subsiste.
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1942-1949.

3.5.21. Station 34 : Le Gave de Brousset à Miégebat/aux Allias

La station 34 « Le Gave de Brousset à Miégebat » n'est présentée que dans l'annuaire 1939. Dès 1940, la station « Le Gave de Brousset aux Allias » apparaît, avec le même numéro. Il existe une station de la banque HYDRO qui peut correspondre Q6014010 aux métadonnées légèrement différentes.

Tableau 29 : Données des annuaires hydrologiques

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km ²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
34		LE GAVE DE BROUSSET A MIÉGEBAT	54				1912	39
34	34	LE GAVE DE BROUSSET AUX ALLIAS	61	1200	42°54N	0°25E	1912	40-69

1940: "A partir de 1940, l'ancienne dénomination Miegébat de cette station est remplacée par la dénomination plus appropriée des Allias"

Tableau 30 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	34	LE GAVE DE BROUSSET AUX ALLIAS	61	1200	1912
Banque HYDRO	Q6014010	Le Gave de Brousset à Laruns [Allias]	63	1134	1900

- Il s'agit de la même station, qui a changé de nom en 1940 avec correction des métadonnées.
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1940-1959.

3.5.22. Station 35 – Le Gave d'Aspe au Pont d'Escot

En 1948 la correspondance est bonne, en 1961 et 1964 aussi. Il y a peu de données publiées en 1962 : « Station court-circuitée par les installations de l'usine d'ASAP - (Débits assez imprécis pendant la période des essais) »

Les données ne correspondent pas en 1963, mais c'est probablement en raison de la coupure de 1962.

Tableau 31 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	35	GAVE D'ASPE AU PONT D'ESCOT	428.5	311.1	1911
Banque HYDRO	Q6332510	Le Gave d'Aspe à Bedous [Pont d'Escot]	425	311	1948

- La station Q6332510 semble correspondre, mais un doute subsiste.
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1939-1947.

3.5.23. Station 36 – Le Gave d'Oloron à Oloron

Tout le début de la chronique les écarts sont importants (mais inférieurs à 10%), et ils deviennent ensuite très faibles dans les années 60.

Tableau 32 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	36	GAVE D'OLORON A OLORON	1085/1062	196	
Banque HYDRO	Q7002910	Le Gave d'Oloron à Oloron-Sainte-Marie [Oloron-SNCF]	1085	204	

- La station Q7002910 semble correspondre (du moins pour les années les plus récentes), mais un doute subsiste.
- Les annuaires n'apportent de données complémentaires aux données de la banque HYDRO.

3.5.24. Station 37 - Le Gave de Gavarnie à Luz

Il y a une erreur dans l'annuaire 1959 qui peut porter à confusion. La station n°37 « Le Gave de Gavarnie a Luz » est remplacée en 1959 par la station n°37-1 « Le Gave de Gavarnie à Pescadères » (appelée aussi « Le Gave de Pau à Pescadères »). Toutefois, le changement de nom et de numérotation n'a pas été fait dès l'annuaire 1959 : l'erreur est mentionnée dans l'erratum de l'annuaire 1960. En 1959, les données disponibles correspondent donc à la station 37-1 et non à la station 37.

Cette erreur a, a priori, été répercutée dans les données de la banque HYDRO, puisque la station Q4231010 « Le Gave de Pau à Luz-St-Sauveur » qui correspond à la station 37, ne présente aucune différence avec les données de l'annuaire en 1959 (alors qu'il s'agit pour cette année-là de la station 37-1).

- Il s'agit de deux stations différentes selon l'annuaire.
- Une confusion a été faite en 1959, et une erreur est donc présente dans les données de la banque HYDRO.

- Les annuaires apportent des données complémentaires aux données de la banque HYDRO (année 1940).

3.5.25. Station 38 – Gave de Pau à Artiguelouve

La station Q5231010 « Le Gave de Pau à Artiguelouve [Pont de Lescar] » de la banque HYDRO pourrait correspondre, mais les données journalières ne sont pas disponibles (il est mentionné que la station Q5231010 est dédiée à l'annonce de crue). La comparaison n'est donc pas possible.

Tableau 33 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	38	GAVE DE PAU A ARTIGUELOUVE	1905	148	
Banque HYDRO	Q5231010	Le Gave de Pau à Artiguelouve [Pont de Lescar]	0	150	

Néanmoins, pour la représentation sur la carte, on prend les coordonnées de la station Q5231010.

Les annuaires ne présentent que trois années de données pour cette station.

- On ne peut pas conclure sur cette station.
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1939-1941.

3.5.26. Station 38bis- Gave de Pau à Pont de Berenx

Au début de la chronique, les écarts sont importants (entre 5 et 10%), et ils deviennent ensuite très faibles dans les années 60.

Tableau 34 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	38-1	GAVE DE PAU A PONT-DE-BERENX	2576.5	30.06	
Banque HYDRO	Q5501010	Le Gave de Pau à Bérenx [Pont de Bérenx]	2575	38	

- La station Q5501010 semble correspondre (du moins pour les années les plus récentes), mais un doute subsiste.
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1946-1947.

3.5.27. Station 39/39-2/39-ter/39-3 - Le Gave d'Azun à Arras / Lau-Balagnas / Arras-Nouaux

La station 39 « Gave d'Azun à Arras » est supprimée et remplacée en 1948 par la station 39bis « Gave d'Azun à Arras » (il est précisé que la station 39, court-circuitée par l'usine de Nouaux, a été supprimée) drainant 76 km².

En 1960, une station nouvelle reprend ce nom initial « Gave d'Azun à Arras », mais avec la numérotation 39-3. Des précisions sont données en page 10 de cet annuaire, mais il n'est pas fait mention d'une reprise de l'ancienne station. Il est indiqué que la station est en service depuis 1960, ce qui laisse penser que ce n'est pas la même station que pendant les années 1939-1947. Dans les commentaires, les stations antérieures décrites sont Arras (1909-1931) et Lau Balagnas (1932-1959). Ces dates ne correspondent pas aux données publiées dans les annuaires (les années de mise en service fluctuent entre 1920, 1932, 1938 et 1960).

En 1964, le nom est légèrement modifié : « Gave d'Azun à Arras-Nouaux », toujours sous la numérotation 39-3. Tous les éléments de comparaison montrent qu'il s'agit de la même station que celle de 1960-1963. Ce pourrait donc être une précision du nom pour éviter la confusion avec la station des années 1939-1947.

Une seule station de la banque HYDRO de superficie 191 km² correspond. Le test de stationnarité ne met pas en évidence de rupture entre les débits moyens des périodes 1939-48 (débits disponibles dans les annuaires hydrologiques) et 1960-2009 (débits disponibles à la station Q4674010 de la banque HYDRO).

Tableau 35 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km ²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
39		LE GAVE D'AZUN A ARRAS	191	615.07			1920	39-47
39-ter	39-2	LE GAVE D'AZUN A LAU-BALAGNAS	191	635			1932	51-59
	39-3	LE GAVE D'AZUN A ARRAS	191	659	42°59N	0°08E	1960	60-63
	39-3	LE GAVE D'AZUN A ARRAS-NOUAUX	191	659	42°59N	0°08E	1960	64-69

Tableau 36 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
Q4674010	Le Gave d'Azun à Arras-en-Lavedan [Arras-Nouaux]	191	662			01/01/1960	Correspond bien à la station 39-3 (1964-1969)

- Il semble qu'il ait existé une première station de 39 à 47, puis une seconde à partir de 1960 dont le nom a été légèrement modifié en 1964.
- La station Q4674010 correspond bien, sauf en 1961 et 1962 (?). Il s'agirait donc bien de la même station sur toute la période 1960-1969.
- Les tests de stationnarité montrent que les données sur la période 1939-1947 sont plutôt cohérentes avec celles de la station Q4674010 de la banque HYDRO (post 1960).
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1939-1947 et 1951-1959.

3.5.28. Station 40 – Nive à Cambo

Trop peu d'années sont disponibles à la fois dans les annuaires et dans la banque HYDRO pour le test de stationnarité avec les stations Q9312510 et Q9302510.

Tableau 37 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	40	LA NIVE A CAMBO	839.5	14.67	1904
Banque HYDRO	Q9312510	La Nive à Cambo-les-Bains	870	25	1999
Banque HYDRO	Q9302510	La Nive à Itxassou	850	87	1967

La banque HYDRO précise que la station Q9312510 remplace la station Q9302510.

- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1939-1941, même si les surfaces ne correspondent pas strictement

3.5.29. Station 40bis - La Nive à St-Jean-Pied-de-Port/La Nive de Béhérobie à St-Jean-Pied-de-Port

En 1944 puis en 1951, le nom est légèrement modifié : « La Nive de Béhérobie à St Jean Pied de Port ». Tous les éléments de comparaison montrent qu'il s'agit de la même station.

Une seule station de la banque HYDRO correspond. Les écarts sont importants au début de la chronique (> 10%), et deviennent ensuite très faibles en 1964. Il semble que la convergence se fasse en 1953, avec peut-être une erreur le 02/05/1957.

Tableau 38 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
40bis		LA NIVE A SAINT-JEAN-PIED-DE-PORT	115	159.7			1912	42-43;45-50
40bis	40-1	LA NIVE DE BÉHÉROBIE A SAINT-JEAN-PIED-DE-PORT	115	160	43°10N	1°14W	1912	44;51-69

Tableau 39 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
Q9034610	La Nive de Béhérobie à Saint-Jean-Pied-de-Port	115	167			01/01/1918	Les écarts sont assez importants pour la première station et au début de la chronique pour la seconde. Les écarts sont ensuite très faible dans les années 1960

- Il semble que le nom ait simplement été modifié. Les données ont été concaténées.
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur les années 1952 et 1969.

3.5.30. Station 41bis – Têt à Vinça

La comparaison avec la station Y0464040 « La Têt [totale] à Vinça » n'est pas possible car la chronique débute dans la banque HYDRO en 1968, et seule l'année 1941 est présentée dans les annuaires hydrologiques (le test de stationnarité n'est pas applicable avec une seule année).

La comparaison n'est pas non plus possible avec la station Y0464020 (début de la chronique en 1967) ni avec la station Y0464010 (pas de données journalières).

Tableau 40 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	41-bis	TET A VINCA	900	206.4	
Banque HYDRO	Y0464040	La Têt [totale] à Vinça	942	191	
Banque HYDRO	Y0464020	La Têt [partielle] à Vinça	942	191	
Banque HYDRO	Y0464010	La Têt à Rodès [Barrage de Vinça]	0	100	

- Les annuaires apportent vraisemblablement des données complémentaires aux données de la banque HYDRO sur l'année 1941.

3.5.31. Station 41-2 – Le Tech à Pas du Loup

Le test de stationnarité ne montre pas de différence significative, mais le nombre d'années disponibles dans la banque HYDRO est faible (7 ans). Les métadonnées, sans être très différentes, ne sont pas un indice dans le sens de la correspondance.

Tableau 41 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	41-2	LE TECH A PAS-DU-LOUP	240/236	402/404	1935
Banque HYDRO	Y0224011	Le Tech à Montferrer [Pas du Loup]	250	343	2000

➤ Les annuaires apportent vraisemblablement des données complémentaires aux données de la banque HYDRO sur la période 1942-1969.

3.5.32. Station 42 – L'Aude à Belvianes

En comparant avec la station Y1112010, on a un pourcentage de différences moyen autour de 10%. Le graphe montre que les pics de crue ont été systématiquement relevés dans la banque HYDRO.

Un test de stationnarité a été fait avec la station Y1112020, mais il y a un écart trop grand entre les périodes pour que les tests soient concluants. On observe une baisse assez constante des débits entre 1939 et 1948.

Tableau 42 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	42	L'AUDE A BELVIANES	696.5/683.5	303	1914
Banque HYDRO	Y1112010	L'Aude à Belvianes-et-Cavirac	692	303	1921 (Débits validés douteux depuis 1914)
Banque HYDRO	Y1112020	L'Aude à Quillan [OA RD 118]	698	280	1985

➤ La station Y1112010 semble correspondre, mais un doute subsiste.
 ➤ Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.33. Station 42-2 – L'Angoustrine à Angoustrine

Le test de stationnarité ne montre pas de différence significative. Les métadonnées correspondent plutôt bien. Il pourrait donc s'agir de la même station.

Tableau 43 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	42-2	L'ANGOUSTRINE A ANGOUSTRINE	46	1351	1913
Banque HYDRO	Y0025010	L'Angoustrine à Angoustrine-Villeneuve-des-Escalades	45.9	1355	1911 (pas de débits disponibles avant 1970)

➤ La station Y0025010 pourrait correspondre, mais un doute subsiste.

- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur la période 1956-1969.

3.5.34. Station 43 - L'Ain à la Chartreuse de Vaucluse/Ain à Chartr. Voulgans

Il semble qu'en 1968, le nom ait été modifié. Le nom est tronqué : « Ain à Chartr. Voulgans ».

En 1969, la numérotation change. Toutes les autres caractéristiques semblent montrer qu'il s'agit d'une seule et même station. Le barrage du Voulgans a été construit en 1965 et le lac, mis en eau en 1968, a englouti la Chartreuse de Vaucluse (<http://claoli.free.fr/html/chartreuse.html>).

Une station de la banque HYDRO correspond. Les écarts aux débits de la station de la banque HYDRO sont faibles dans les années 1950 et en 1969 mais pas pour l'année 1968. Il y a probablement une faute de frappe dans les annuaires hydrologiques le 07/03/1946 : la banque HYDRO indique une valeur de 22.1 m³/s, l'annuaire de 221 m³/s. De plus le graphe de l'annuaire ne fait pas apparaître cette valeur élevée.

Tableau 44 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
43	43	L'AIN A LA CHARTREUSE DE VAUCLUSE	1115	342	46°23N	5°29E	1933	39-67
	43	AIN A CHARTR. VOUGLANS	1120	342	46°23N	5°29E	1933	68
	43-1	AIN A CHARTR. VOUGLANS	1120	342	46°23N	5°29E	1933	69

Tableau 45 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
V2322010	L'Ain à Cernon [Voulgans]	1120	392			01/01/1933	Les débits entre la station de la BH et la première station de l'annuaire correspondent. Ecart >10% en 1968 mais bonne correspondance en 1969

- Il s'agit probablement d'une seule station, avec une erreur dans la numérotation en 1969. Le nom a changé en raison de la mise en eau du barrage de Voulgans.
- Des écarts sont observés en 1968 (associés vraisemblablement au comblement de la lacune du 1^{er} janvier au 15 avril).
- Les commentaires sur la station de la banque HYDRO indiquent que le barrage a noyé à partir de 1968 l'usine de la Chartreuse de Vaucluse.

➤ Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.35. Station 47 - Le Rhône à Génissiat

Les données sont similaires au début de la chronique.

Le contrôle de cette correspondance réalisé dans les années 1960 montre cependant un écart de 5%. Les données redeviennent très similaires en 1969.

3.5.36. Station 47-1 - Le Rhône à Lyon Poincaré/ Saint Clair

Le nom change en 1965, mais toutes les autres caractéristiques restent inchangées.

Tableau 46 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
	47-1	LE RHONE A LYON-POINCARE	20300	163	45°46N	4°50E	1957	57-64
	47-1	LE RHONE A LYON SAINT CLAIR	20300	163	45°46N	4°50E	1957	65-69

Tableau 47 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
V3000010	Le Rhône à Caluire-et-Cuire [Saint-Clair]	20300	168			01/01/1920	Pour les stations Lyon-Poincaré et Lyon St Clair

➤ Il s'agit d'une seule station dont le nom change en 1965.

3.5.37. Station 47-2 - Le Rhône à la Mulatière / Mulatière-Givors

En 1965 le « -Givors » est rajouté au nom de station. La surface de bassin versant est assez différente, mais les autres caractéristiques sont inchangées.

Tableau 48 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
47bis	47-2	LE RHONE A LA MULATIERE	50200	154	45°42N	4°49E	1920	56-64
	47-2	LE RHONE A LA MULATIERE-GIVORS	51080	154	45°42N	4°49E	1920	65-69

Tableau 49 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
V3130010	Le Rhône à Givors	51100	156			01/01/1920	Oui pour les deux stations

La citation suivante semble indiquer que la station de la Mulatière devait être remplacée aux environs de l'année 1965.

1957 : « ... le temps que durera la station de LA MULATIÈRE, c'est-à-dire au plus une dizaine d'années, jusqu'à la mise en service de l'usine de Pierre-Bénite »

La station de la banque HYDRO correspond aux deux stations des annuaires.

➤ Il semble s'agir d'une seule station qui a changé de nom.

3.5.38. Station 49 - Le Rhône au Teil/ Teil (Pouzin)

De 1958 à 1963, le nom change avec ajout du terme « Pouzin ».

Tableau 50 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
49	49	LE RHONE AU TEIL	70141	62.93	44°33N	4°40E	1910	39-57
	49	LE RHONE AU TEIL (POUZIN)	69500	84/62.93	44°33N	4°40E	1910	58-63
49	49	LE RHONE AU TEIL	69500	84	44°33N	4°40E	1910	64-69

1957: « Par suite de l'aménagement de Montélimar qui a détourné les eaux du Rhône en amont du limnigraphe de Joviac (Le Teil) les débits sont mesurés depuis le 12 juin 1957 au limnigraphe du Pouzin ».

1958: « Nous avons conservé le nom du Teil bien que l'aménagement de Montélimar ait court-circuité cette station en 1957. Celle du Pouzin qui la remplace est elle-même provisoire 1957-1960. Les débits de longues périodes sont donc hétérogènes, mais à un degré négligeable : faible différence de BV (480 km², moins de 1%), brève durée de fonctionnement du Pouzin ».

Il y a des incohérences dans les valeurs des surfaces des bassins versants et d'altitude, mais a priori la localisation de la station aurait changé en 1958 et peut-être par la suite, contrairement à ce qui est indiqué dans l'annuaire 1958.

➤ Il s'agit donc de deux stations différentes, mais les différences sont jugées comme négligeables.

3.5.39. Station 52 – Le Fier à Motz – Val de Fier

Deux stations sur le Fier proposent des débits dans les annuaires avec des superficies équivalentes (1350 et 1370 km²). Il est indiqué dans l'annuaire 1943 que « La station de MOTZ-VAL-DE-FIER, n°52 (Bassin versant 1.370 km²) a été remplacée par celle de VALLIERES, n°52 bis (Bassin versant 1.350 km²) en raison des meilleures conditions de mesure à cette dernière station ». Les données des annuaires de ces deux stations ont été concaténées.

Tableau 51 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	52	LE FIER A MOTZ – VAL DE FIER	1370	256	1911
Annuaire SHF	52-1	LE FIER A VALLIERES	1350	295	1943
Banque HYDRO	V1264010	Le Fier à Vallières	1350	253	1948
Banque HYDRO	V1264021	Le Fier à Motz	1379	259	1997

- Les stations présentent des surfaces très similaires.
- Les débits des stations des annuaires sont égaux à ceux de la station V1264010 pour ce qui est de la période de recouvrement (post 1948).
- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur les périodes 1939-1942 et 1943-1947.

3.5.40. Station 54 – L'Isère à Beaumont - Monteux

Deux stations des annuaires mesurent les débits sur la partie aval de l'Isère : l'Isère à Beaumont-Monteux (n°54) et à Pizançon (n°54 et 54bis) dont les débits sont publiés à partir de 1952. L'annuaire de 1952 précise : « Le G.R.P.H. Rhône a renoncé à publier le tableau des débits journaliers de l'Isère à Beaumont-Monteux. Les débits" figurant à ce tableau, qui fait d'ailleurs pratiquement double emploi avec celui de Pizançon, sont en effet sujet à caution depuis la suppression de l'échelle de Vernaison maintenant noyée dans la retenue de La Vanelle. ». Les séries ont été concaténées.

Une station de la banque HYDRO pourrait correspondre aux stations des annuaires. Les métadonnées semblent cohérentes (mis à part l'année de mise en service), il s'agit probablement de la même station.

Tableau 52 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	54	L'ISÈRE A BEAUMONT-MONTEUX	11500	120/129.94	1921
Annuaire SHF	54-1	L'ISÈRE A BEAUMONT-MONTEUX	11320	150	1932
Banque HYDRO	W3540010	L'Isère à Beaumont-Monteux	11800	120	1955

- La station W3540010 pourrait correspondre.

- Les annuaires apportent des données complémentaires aux données de la banque HYDRO sur les périodes 1939-1951 et 1952-1955.

3.5.41. Station 53bis - L'Isère au Pont du Laisinant/ à Val d'Isère - Laisinant

Le nom change en 1962, les autres caractéristiques indiquent qu'il s'agit de la même station.

Une station de la banque HYDRO correspond.

En 1963, la station reprend l'ancien nom « Isère au Pont du Laisinant ».

Tableau 53 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
53bis	53-1	L'ISERE AU PONT DU LAISINANT	45	1860			1951	52-61
	53-1	L'ISERE A VAL-D'ISERE – LAISINANT	46	1830	45°27N	7°0E	1951	62-69

Tableau 54 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
W0000010	L'Isère à Val-d'Isère	46	1852			01/01/1948	Oui pour les deux stations

- Il s'agirait de la même station.
- Les annuaires n'apportent pas de données complémentaires aux données de la banque HYDRO.

3.5.42. Station 56bis – L'Arc à Aiguebelle

Une station de la banque HYDRO correspond.

Seulement quatre années de données sont publiées dans les annuaires (1941-1944), ce n'est pas suffisant pour faire un test de stationnarité. Mais les informations semblent cohérentes, il pourrait s'agir de la même station.

Tableau 55 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km²)	Altitude (m)	Année de mise en service
Annuaire SHF	56bis	L'ARC A AIGUEBELLE	1928	316.3	1930
Banque HYDRO	W1074030	L'Arc à Aiguebelle	0	310	2011

- Les stations présentent des caractéristiques similaires.
- On considère qu'il s'agit bien de la station W1074030 et les annuaires pourraient apporter des données complémentaires aux données de la banque HYDRO sur les périodes 1941-1944.

3.5.43. Station 56ter – L'Averole à Avrole

Une station de la banque HYDRO correspond.

Une seule année de comparaison est possible (1969) : les écarts sont très faibles et les métadonnées semblent cohérentes.

Tableau 56 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	56-2	AVEROLE A AVEROLE	48	1950	
Banque HYDRO	W1006010	L'Avérole à Bessans [Avérole]	45.4	1950	

- On considère qu'il s'agit bien de la station W1006010.
- Les annuaires pourraient apporter des données complémentaires aux données de la banque HYDRO sur les périodes 1952-1968.

3.5.44. Station 57 – Le Drac au Sautet

Une station de la banque HYDRO correspond.

Une seule année de comparaison est possible (1969) : les écarts sont assez faibles. Les métadonnées ne correspondent pas tout à fait, mais les caractéristiques des bassins sont similaires.

Tableau 57 : Caractéristiques des stations potentiellement identiques.

	N° de station	Nom de station	Surface du bassin versant (km ²)	Altitude (m)	Année de mise en service
Annuaire SHF	57	DRAC AU SAUTET	1020	750	
Banque HYDRO	W2222010	Le Drac à Corps [Le Sautet]	990	671	

- La station W2222010 pourrait correspondre.
- Les stations présentent des caractéristiques similaires et les annuaires pourraient apporter des données complémentaires aux données de la banque HYDRO sur les périodes 1939-1968.

3.5.45. Station 64/64-1 – La Durance à Mirabeau / La Durance à Jouques-Cadarache

En 1959, la station 64-1 « Durance à Jouques-Cadarache » remplace la station 64 « Durance à Mirabeau ». Il est précisé que le remplacement se fait « à bassin versant équivalent ». Les métadonnées sont les mêmes, et les deux stations correspondent à une même station de la banque HYDRO aux caractéristiques quelque peu différentes. La comparaison des débits de la Durance à Mirabeau aux débits de la station X3000010 « La Durance à Saint-Paul-lès-Durance [Jouques-Cadarache] » montre un écart de 4% sur la période complète, avec une différence plus notable sur les pics de crue. Une seule année de comparaison est possible pour la station 64-1 et les écarts sont très faibles.

Tableau 58 : Données des annuaires hydrologiques.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface BV (km²)	Altitude (m)	Latitude	Longitude	En service depuis	Chronique disponible
64	64	LA DURANCE A MIRABEAU	11917/1 1920	228			1904	39-58
	64-1	DURANCE A JOUQUES- CADARACHE	11920	228			1904	59-69

Tableau 59 : Données de la banque HYDRO.

Code BH	Nom (BH)	Surface BV (BH)	Altitude (MNT)	Latitude	Longitude	Début de la chronique disponible (BH)	Correspondance entre les stations
X3000010	La Durance à Saint-Paul- lès-Durance [Jouques- Cadarache]	11700	247			01/01/1918	Oui pour les deux stations

- Il a été décidé de concaténer les séries des annuaires.
- La station X3000010 semble correspondre.
- Les annuaires pourraient apporter des données complémentaires aux données de la banque HYDRO sur les périodes 1959-1968.

4. Synthèse

Nous avons identifié 2293 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 1638, 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 mise 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 (Figure 26 ; Tableau 60), et 41 séries portent des données complémentaires à des stations de la banque HYDRO (Figure 26 ; Tableau 61). Un total de près de 600 années stations inédites ont été exhumées (209 sont associées aux stations inédites).

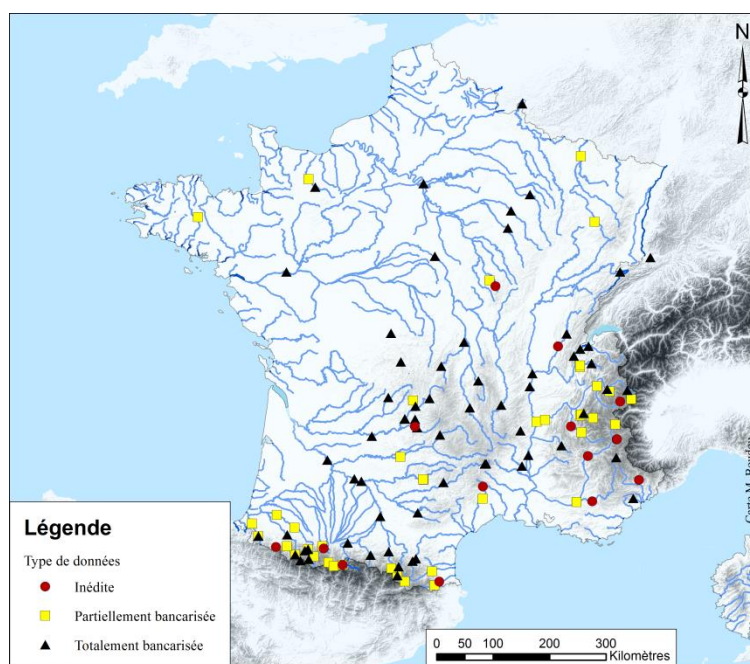


Figure 26 : Carte des stations présentées dans les annuaires de la SHF selon leur statut vis-à-vis des stations de la banque HYDRO.

Tableau 60 : Descriptif des stations inédites.

N° (jusqu'en 1956)	N° (à partir de 1957)	Nom (annuaire)	Surface (km²)	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

Tableau 61 : Descriptifs des stations bancarisées partiellement.

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
2		LA MOSELLE A ÉPINAL	1250	1891	1939-1941	A4250640	LA MOSELLE A ÉPINAL	1217	1960-2015
2bis	2-1		1215	1952	1954-1969				
	2-2	LA MOSELLE A HAUCONCOURT	9400	1956	1957-1969	A7930610	LA MOSELLE A HAUCONCOURT	9387	1960-2013
3	3	LE CHALAUX A CHAUMEÇON	98/100	1935	1940-1969	H2133010	LE CHALAUX A SAINT-MARTIN-DU-PUY	98	1946-1981
4	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'ARIEGE 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 TRAMEZAÏGUES [MAISON BLANCHE]	63.7	1948-2011
						O0126220	LA NESTE DE RIOUMAJOU A SAINT-LARY-SOULAN [1]	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 THURIÈS / 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
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-SAINT-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 LESCAR]		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	191	1920	1939-1947	Q4674010	LE GAVE D AZUN A ARRAS-EN-LAVEDAN [ARRAS-NOUAUX]	191	1960-1986 ; 1989-2011
	39-2	LE GAVE D'AZUN A LAU-BALAGNAS			1951-1959				

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
40		LA NIVE A CAMBO	839.5	1904	1939-1941	Q9302510	LA NIVE A ITXASSOU	850	1967-1973 ; 1975-1999
						Q9312510	LA NIVE A CAMBO-LES-BAINS	870	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
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 PERRIÈRE	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				

5. Conclusion

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^{3 4 5}. 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 risque⁶.

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 numérique⁷. Nous avons souhaité engager une démarche similaire en nous concentrant 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 de près de 600 années stations), soit près de 50% des stations mentionnées dans les annuaires. Les données redondantes n'ont pas forcément été numérisées ; certaines séries exhumées ne contiennent pas les années déjà présentes dans la banque HYDRO.

³ 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.

⁴ 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.

⁵ Minvielle, M., Pagé, 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*, 1, 189–192, doi:10.1007/978-3-319-09300-0_36, 2015.

⁶ Boudou, M. (2015). Caractérisation de neuf événements remarquables d'inondations en France au XXe siècle : contribution à la valorisation de l'information historique pour la gestion prospective des risques. Thèse, Montpellier 3, en cours.

⁷ 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, 268-271 (<http://digital.onwindows.com/climate-exchange/>).

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), sur l'expertise des gestionnaires actuels des séries potentiellement enrichies et au regard des techniques de mesure employées en s'inspirant des travaux de Kuentz⁸. 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.

⁸ Kuentz, A. (2013). 100 ans de variabilité climatique observée et reconstituée sur le bassin de la Durance. Thèse, AgroParisTech, Paris, 373 pages.

ANNEXES

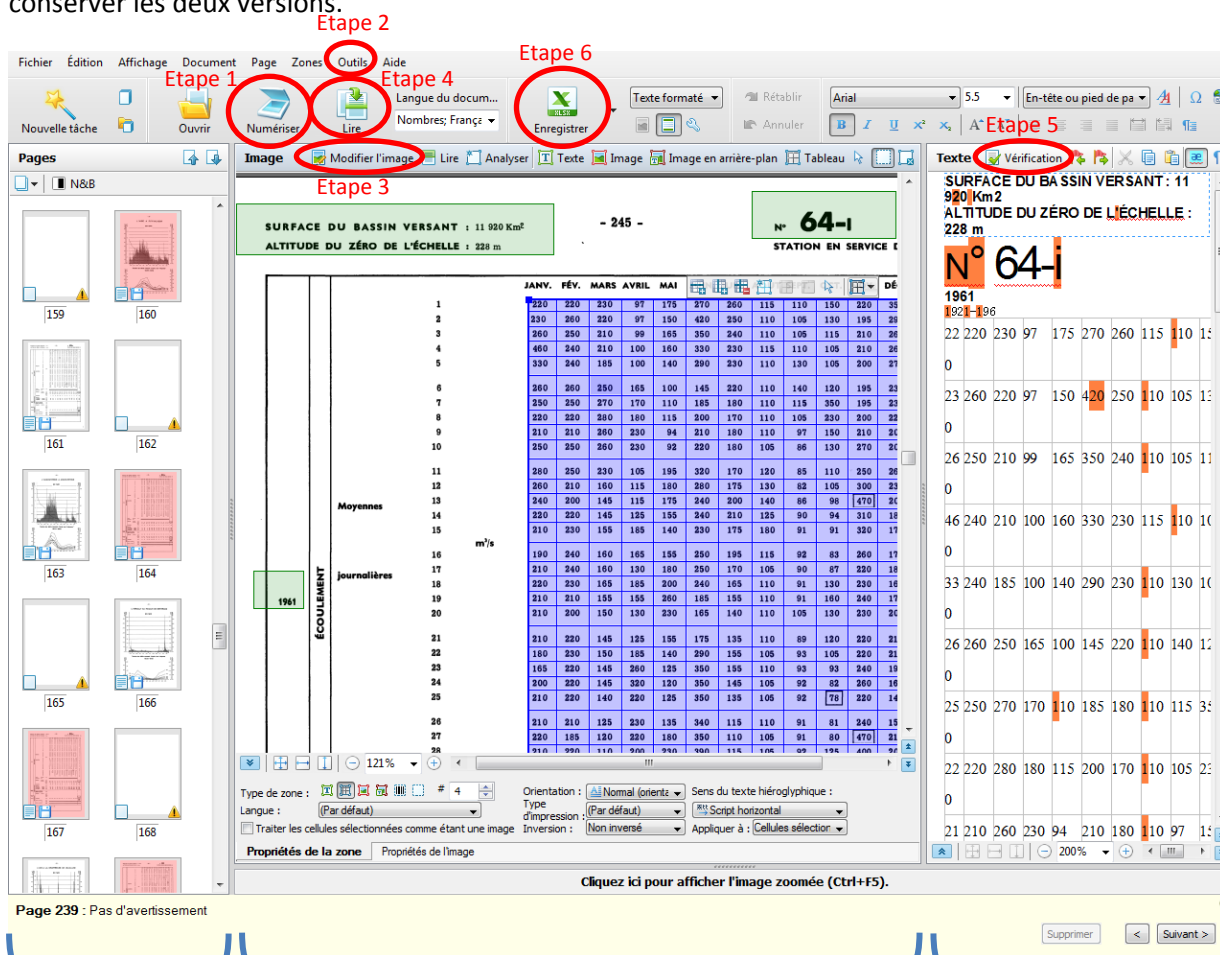
1. Annexe 1 : Fonctionnement du logiciel ABBYY Fine Reader

Un logiciel d'OCR fonctionne selon le principe général suivant : on donne au logiciel un fichier d'entrée en format image et on obtient un fichier de sortie en format texte. Le processus comporte plusieurs étapes :

- Numérisation du document original
- Définition des paramètres du logiciel
- Retouches de l'image
- Lecture de l'image
- Enregistrement du résultat

Chacune de ces étapes est détaillée ci-après, en se basant sur l'exemple de l'utilisation du logiciel ABBYY Fine Reader 11 (un aperçu de l'interface du logiciel est présenté en Figure 27).

NB : ABBYY est configuré pour sauvegarder automatiquement le projet en cours. Attention donc à créer une copie **avant** toute modification dans le cas où l'on souhaite créer un second projet à partir du projet initial et conserver les deux versions.



Aperçu des pages du document

Aperçu de la page sélectionnée. Les zones de texte apparaissent en vert, les zones de tableau en bleu, les zones d'image en rouge

Aperçu de la reconnaissance des caractères : les caractères incertains sont indiqués en orange

Figure 27 : Aperçu de l'interface graphique du logiciel ABBYY FineReader, et localisation des commandes servant aux différentes étapes de la numérisation par OCR.

1.1. Etape 1 : Numérisation du document original

Cette étape de numérisation s'effectue classiquement avec un scanner. Elle doit être effectuée avec soin, en veillant à bien paramétrer le logiciel qui pilote le scanner (luminosité, contraste, niveaux, résolution) afin d'obtenir un fichier d'entrée (en format image) de la meilleure qualité possible.

La résolution minimum doit être de 300 dpi. Dans notre cas, la numérisation a été réalisée avec une résolution de 600 dpi. Les images sont ainsi plus détaillées, mais les temps de traitement sont allongés. Il est recommandé d'utiliser plutôt les niveaux de gris pour une meilleure reconnaissance des caractères.

Le logiciel ABBYY peut réaliser l'étape de numérisation si le scanner est connecté au même ordinateur. Dans ce cas, cliquer sur le bouton Numériser (Figure 27).

1.2. Etape 2 : Définition des paramètres du logiciel

En ouvrant le logiciel, il est utile d'aller dans l'onglet « Outil/Options » afin de définir les paramètres généraux (Figure 28). Passer un peu de temps à voir quelles sont les options possibles pour chacune des étapes peut permettre de gagner du temps ensuite et d'obtenir un résultat de meilleure qualité. On peut notamment régler la langue à reconnaître (ou indiquer qu'il s'agit de données numériques, comme dans notre cas), les polices utilisées, demander de séparer des pages en vis-à-vis, afficher différemment les caractères incertains, modifier le dictionnaire...

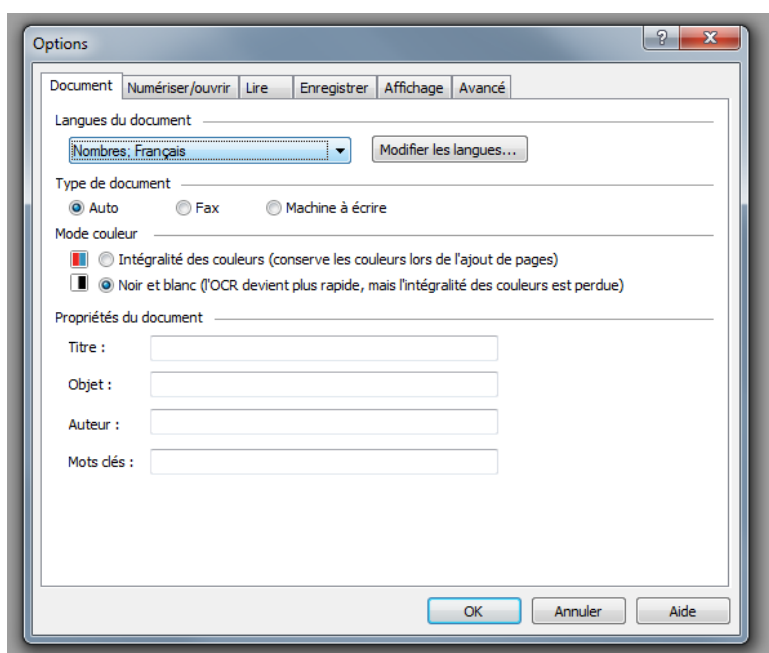


Figure 28 : Aperçu des options du logiciel.

1.3. Etape 3 : Retouches de l'image

On importe ensuite le fichier en format image dans ABBYY (bouton « Ouvrir », Figure 27). Cette opération peut prendre du temps selon la taille du fichier à ouvrir (entre 3 et 4 minutes par document de 20 000 Ko).

Si le fichier premier dont l'on dispose est déjà un document scanné, le logiciel d'OCR permet d'éditer un certain nombre de paramètres (luminosité, contraste) qui n'auraient pas été optimisés lors du scan. On peut

aussi corriger des défauts de l'image : présence de taches, de points parasites, mauvaise rotation de l'image, déformations de l'écriture sur les bords à cause de l'épaisseur du livre, éclairage non homogène, écriture aux contours flous, écriture fragmentée... Il peut aussi être utile de faire disparaître (« gommer ») certains éléments de l'image (annotations manuscrites par exemple) pour éviter de perturber la lecture par le logiciel.

Cette étape peut permettre d'améliorer la qualité optique de l'image, même si le document original n'est plus disponible pour une nouvelle numérisation à l'aide d'un scanner. Le logiciel ne pourra par contre rien changer à la résolution, qui reste un paramètre important pour obtenir une bonne reconnaissance des caractères.

ABBY ne peut gérer que certains de ces problèmes (Figure 29).

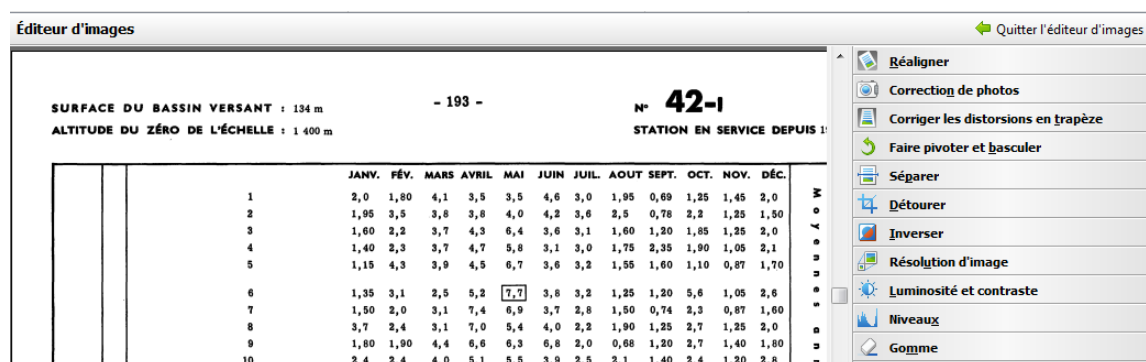


Figure 29 : Aperçu des modifications pouvant être apportées aux images.

1.4. Etape 4 : Lecture de l'image

Cette étape est effectuée lorsqu'on active le bouton « Lire » (Figure 27). Elle correspond en réalité à deux actions. Le logiciel analyse d'abord la composition de l'image, afin d'identifier le type d'éléments présents (texte, tableaux, images, codes-barres...) ainsi que leur position sur la page. Il lit ensuite dans chacun de ces blocs les caractères présents.

Il est important de vérifier que l'analyse de la structure de la page a été faite correctement. Des rectifications sont possibles manuellement. Il faut pour cela modifier la forme ou l'emplacement des zones, ou bien utiliser les différents outils qui s'affichent sur la barre des tâches, présente en haut à droite de chaque zone lorsque l'on clique dessus. Cette étape peut s'avérer assez fastidieuse, comme ça a été le cas dans le cadre de la récupération des données hydrologiques (nombreuses lignes à rajouter/supprimer, cellules fusionnées à séparer..., Figure 30)

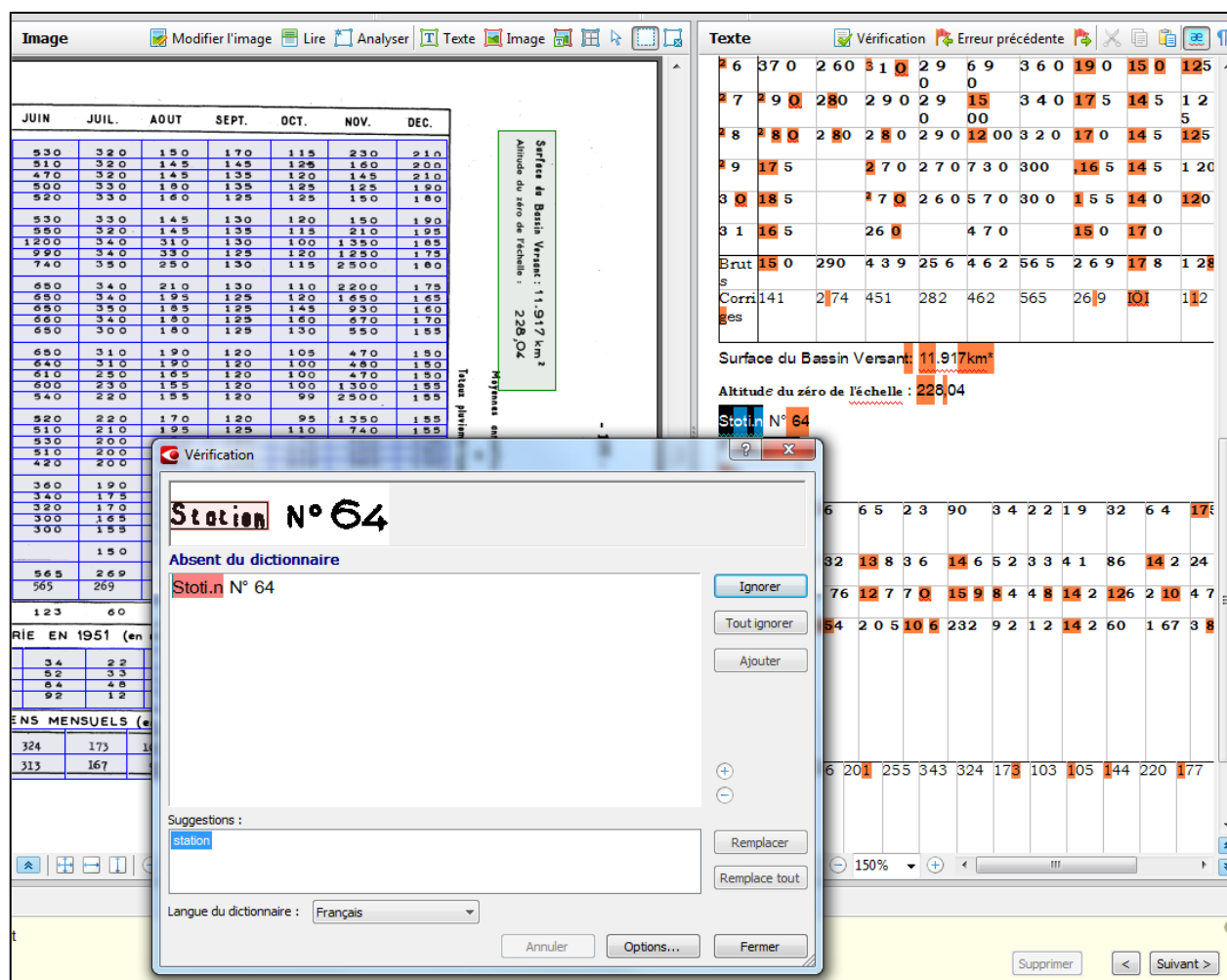


Figure 31 : Aperçu de l'outil « Vérification ».

Cet outil n'a pas été utilisé lors de la récupération des données des annuaires hydrologiques. En effet, c'est un processus assez long et qui ne permet pas un apprentissage des erreurs commises couramment. De plus, certains caractères sont mal identifiés par le logiciel mais ne sont pas identifiés comme tels (pas de surlignage orange, donc on ne reviendra pas sur ces caractères avec l'outil « Vérification »). On a préféré utiliser des outils de correction automatique (Paragraphe 2.2.7.2) couplés à l'utilisation de gabarits utilisateurs (Paragraphe 1.8.4 de l'Annexe 1).

1.6. Etape 6 : Enregistrement du résultat

Une fois la lecture terminée, il faut générer un fichier de sortie. Pour cela, choisir le format du fichier de sortie que l'on souhaite dans la liste déroulante associée au bouton « Enregistrer » (Figure 27). Le logiciel ABBYY permet de créer un fichier unique pour l'ensemble du document, des fichiers séparés pour chaque page du document, ou des fichiers séparés à chaque page vide du document.

1.7. Outils de correction et vérification

La cinquième étape permet d'obtenir les fichiers de sortie souhaités. On peut se contenter de cela, mais il faut bien garder à l'esprit que le logiciel d'OCR commet toujours des erreurs de reconnaissance des caractères, sauf peut-être s'il s'agit d'un document très récent avec une police très claire. Le nombre d'erreurs augmente aussi sensiblement lorsque le document mélange des données chiffrées et du texte. Sans

vérification, la qualité des données du fichier de sortie est à relativiser. Une vérification « manuelle » est donc recommandée pour contrôler la qualité des données numérisées.

On a montré dans ce document les vérifications possibles dans le cas de la récupération des données de débits journaliers des annuaires hydrologiques (Paragraphe 2.2.7). De telles vérifications ne sont pas toujours possibles, selon le type de données disponibles.

1.8. Difficultés et solutions

Sans prétention d'être exhaustif, on présente ici quelques exemples de difficulté que l'on peut rencontrer lors de la reconnaissance optique des documents et des solutions qu'on a pu y apporter.

1.8.1. Vérifier la qualité des documents disponibles

La qualité des données obtenues en fichier de sortie dépend en tout premier lieu de la qualité originale des documents à numériser. Un document froissé, taché, annoté... posera forcément des problèmes pour la reconnaissance optique des caractères. De même, certains documents ne pourront pas être exploités, par exemple s'il s'agit de papier millimétré : la présence de trop nombreuses lignes parasites va complètement empêcher la reconnaissance des caractères.

Il faut donc tout d'abord s'assurer que la qualité des documents originaux est correcte avant de se lancer dans une analyse par OCR de ces documents.

1.8.2. Structure de page se répétant : utiliser un modèle de zone

Lors de l'étape 4, le logiciel reconnaît automatiquement les différents types de zones présentes sur chaque page. La vérification de la structuration des pages en différentes zones peut être longue et fastidieuse. L'utilisation de modèles de zone peut alors s'avérer très utile, dans le cas où la même structure de page se répète à l'identique au fil du document. Le principe est de mémoriser la structure d'une page, et de l'appliquer aux autres pages. Pour cela, lorsqu'on est satisfait de la reconnaissance de la structure de la page servant de modèle, on enregistre ce modèle de zone grâce à l'onglet « Zones ». On le charge ensuite sur toutes les pages préalablement sélectionnées. Cette méthode peut faire gagner beaucoup de temps. Il faut néanmoins vérifier que les différentes zones sont bien placées sur toutes les pages, car même si la structure des pages est identique il peut y avoir un léger décalage, lié notamment à l'étape de scan du document (Figure 32).

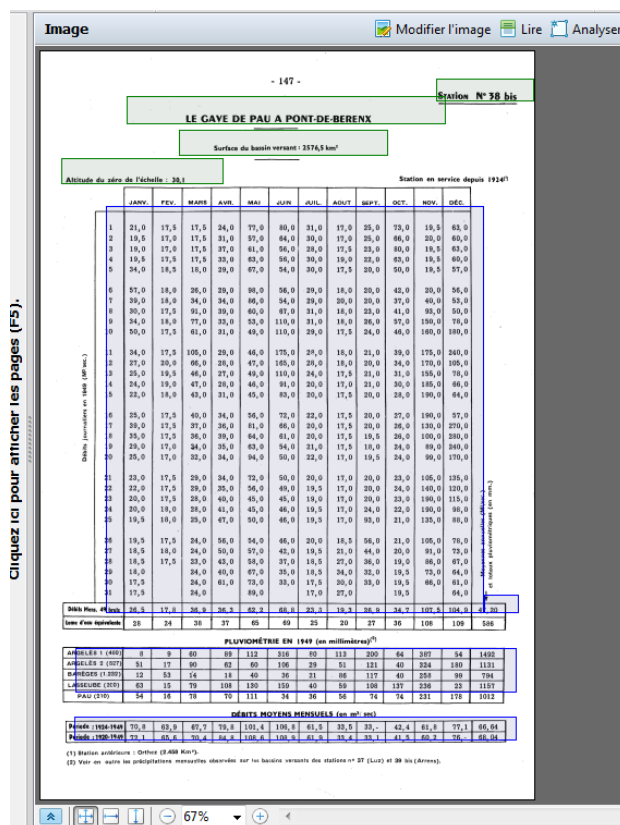


Figure 32 : Le modèle de zone permet de définir la même position de zones de texte (en vert) et de zones de tableaux (en bleu) sur toutes les pages du document. On constate ici qu'un ajustement manuel est nécessaire pour caller les zones correctement sur la page.

Il faut ensuite relancer la lecture du document, afin de prendre en compte ces modifications.

L'utilisation de modèle de zone a également été utile dans notre cas, car toutes les données présentes sur la page n'étaient pas intéressantes à récupérer (numéro en haut de page, données mensuelles de lames d'eau...). Une reconnaissance automatique aurait inclus l'intégralité des caractères présents sur la page.

1.8.3. Les zones présentes sur une même page ont des caractéristiques différentes

Dans le cas des annuaires hydrologiques, certaines pages présentaient le texte dans le sens de la lecture et les tableaux de chiffres décalés d'un quart de tour (Figure 33). Il faut alors appliquer une rotation à la page entière, afin que la majorité des caractères à reconnaître soient dans le bon sens. Il est ensuite possible de définir pour chacune des zones l'orientation du texte (cliquer sur la zone et afficher les propriétés de la zone en cliquant sur la double flèche dans la barre des tâches en bas de l'écran).

Cela peut également permettre de préciser par exemple une langue différente pour certaines zones.

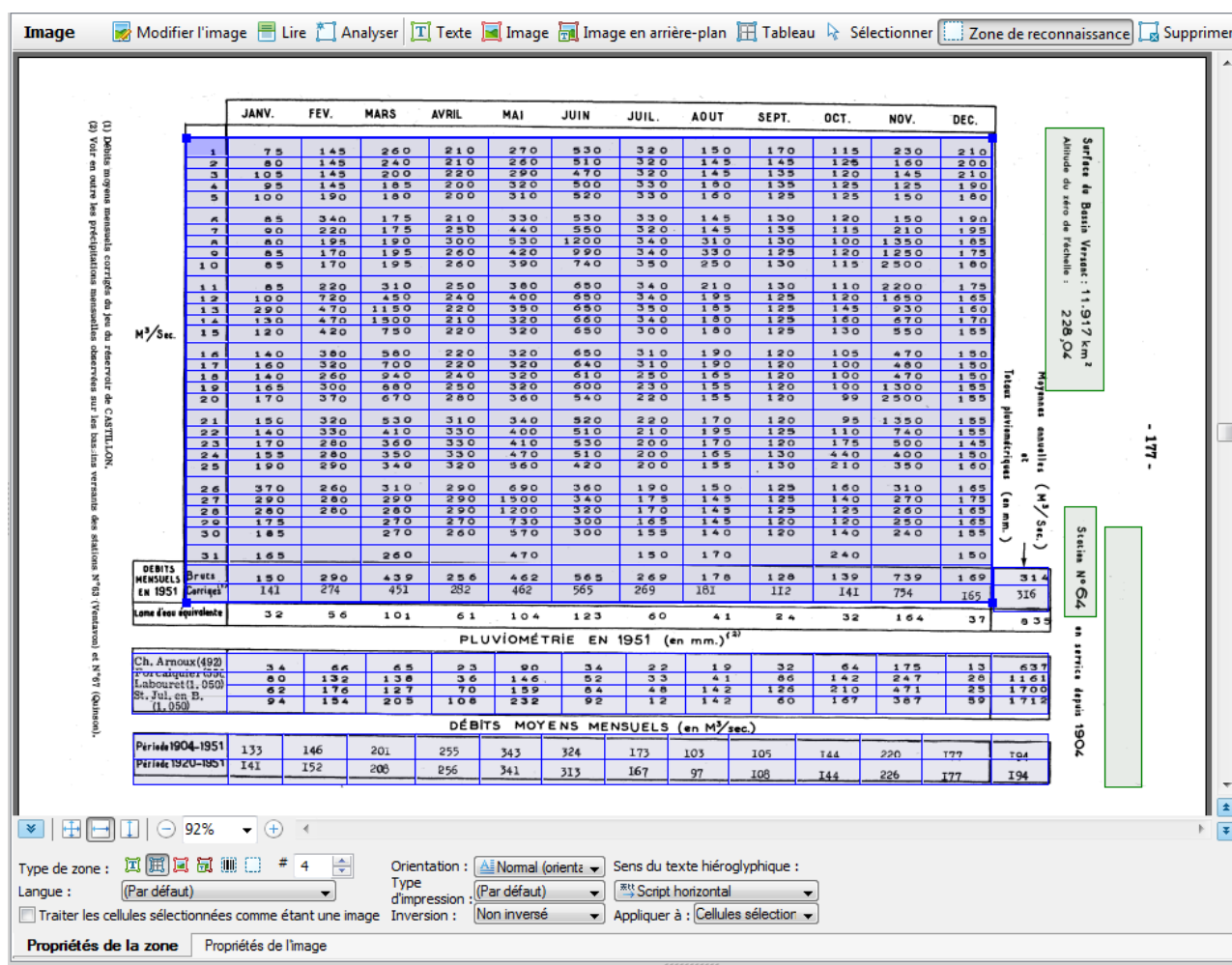


Figure 33 : Exemple d'un cas où il faut spécifier une orientation du texte différente selon les zones.

1.8.4. Police de caractère inconnue : lecture avec apprentissage

La reconnaissance des caractères se fait sur la base d'une liste de polices de caractères intégrée dans le logiciel. Si l'on connaît la police de caractères utilisée dans le document à reconnaître, il peut être utile d'aller vérifier dans cette liste si la police en question est active (la liste est accessible dans les options, Etape 2).

Le logiciel peut également traiter des polices inconnues grâce à la lecture avec apprentissage. L'utilisation du mode apprentissage n'est intéressante que si le document est volumineux et de mauvaise qualité, ou contient régulièrement des symboles spéciaux ou non reconnus avec le mode par défaut, car l'apprentissage est long. Cela consiste en une lecture pendant laquelle le logiciel s'arrête sur chaque caractère qu'il ne reconnaît pas, et demande à l'utilisateur de lui indiquer de quel caractère il s'agit à travers une boîte de dialogue (Figure 34). Le logiciel garde ensuite en mémoire chaque couple image/caractère dans un « gabarit utilisateur ». Il utilise ensuite ce gabarit pour améliorer le reste de la lecture. La lecture avec apprentissage peut se faire uniquement sur une petite partie du document, mais il faut s'assurer que tous les symboles pouvant poser problème ont été intégrés dans le gabarit. L'utilisation du mode apprentissage n'est pas très intuitive. C'est pourquoi le détail du mode opératoire est détaillé ici.

Une fois que la structuration en zones de la page est satisfaisante, aller dans outil/options/lire, cocher « Utiliser les gabarits intégrés et les gabarits utilisateur » (ou « N'utiliser que le gabarit utilisateurs »), cocher « Lire avec apprentissage », créer un nouveau gabarit utilisateur (il doit apparaître comme actif dans la liste),

lancer la lecture avec apprentissage (« Lire »), donner la correspondance pour les symboles inconnus (on peut s'arrêter avant la fin du document pour apprendre une partie des symboles inconnus uniquement), enregistrer les modifications, aller dans « Editeur de gabarit », modifier le gabarit que l'on vient de créer, vérifier que les correspondances sont bonnes ou supprimer certains symboles que l'on a appris par erreur, relancer la lecture (sans apprentissage).

La gestion des gabarits utilisateur par le logiciel est parfois approximative, il faut donc veiller à enregistrer le gabarit que l'on vient de créer (« Enregistrer dans un fichier »), et si par la suite ce nouveau gabarit n'apparaît pas dans la liste des gabarits utilisateurs, il faut charger le fichier en question. Si on garde l'option « Utiliser les gabarits intégrés et les gabarits utilisateur » pour la lecture, il faut bien vérifier avant chaque lecture quel gabarit utilisateur est actif (gabarit noté actif dans la liste qui apparaît dans « Editeur de gabarit »).

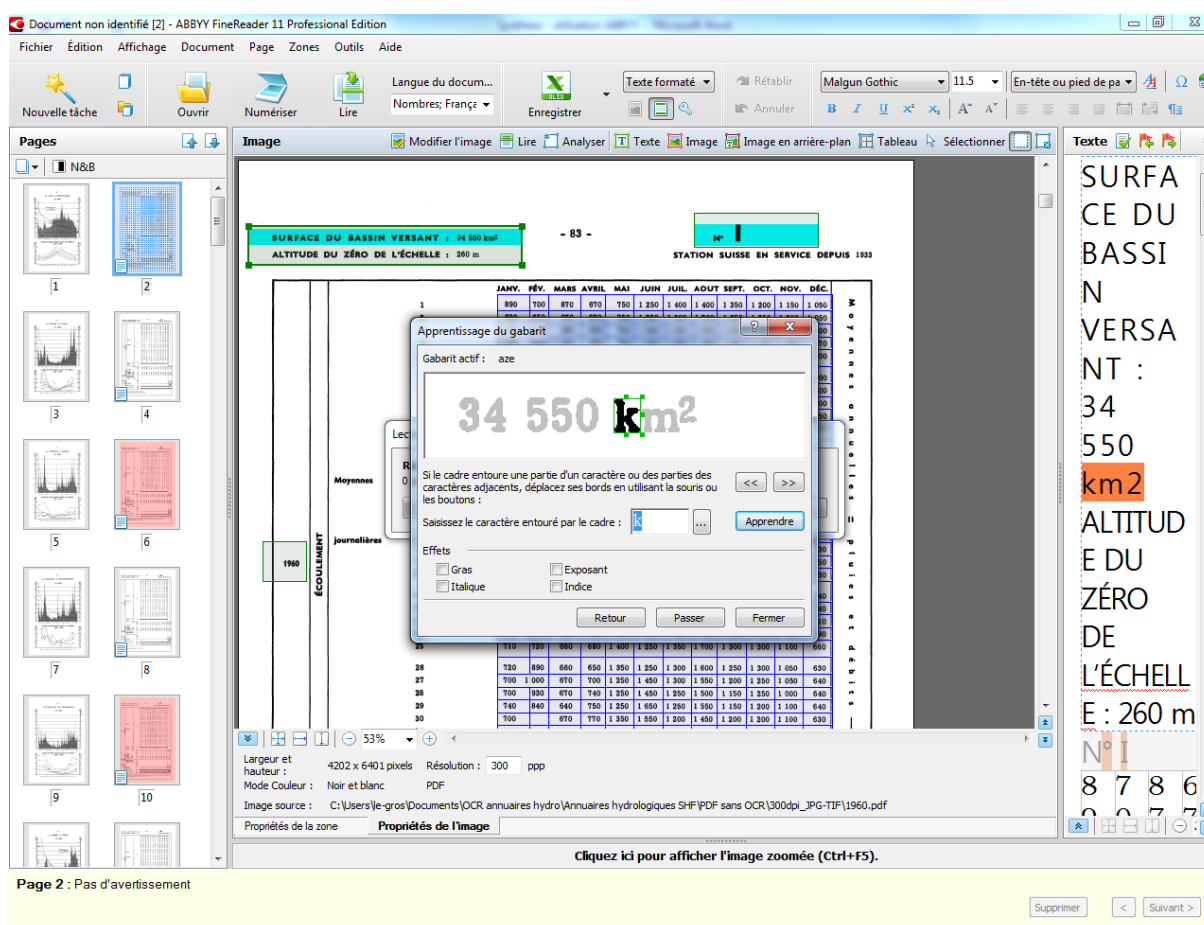


Figure 34 : Aperçu de la boîte de dialogue pour la création d'un gabarit utilisateur.

1.8.5. Modification de l'ordre des zones dans le fichier de sortie

Par défaut, le logiciel lit les zones dans l'ordre d'apparition, et produit un fichier de sortie dans lequel les caractères reconnus apparaissent dans ce même ordre. On peut modifier l'ordre de lecture des zones (onglet « Zones/Organiser les zones »), afin que les caractères de certaines zones apparaissent avant ou après les autres (Figure 35).

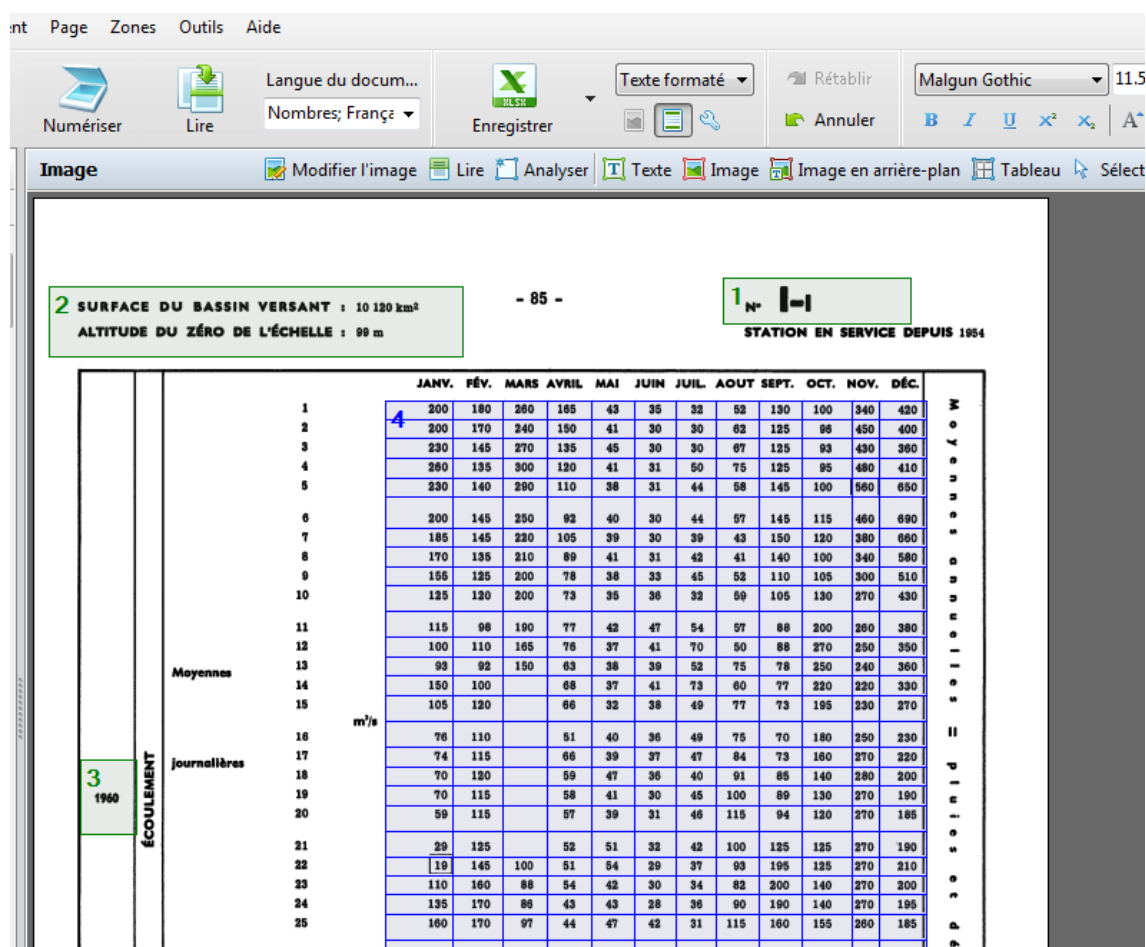


Figure 35 : L'ordre des zones est indiqué par un chiffre dans chacune des zones. Pour le modifier, cliquer sur chaque zone dans le nouvel ordre voulu.

1.8.6. Autres problèmes

Les logiciels d'OCR sont adaptés pour la reconnaissance des documents assez récents, présentant des polices de caractères et des façons d'écrire modernes (fin du XXe siècle). Ils ne reconnaissent pas :

- l'écriture manuscrite (police pas assez uniforme),
- les formules de mathématiques ou de chimie (caractères et organisation sur plusieurs lignes non gérés par le logiciel),
- les caractères imprimés des siècles passés,
- les abréviations anciennes,
- le papier millimétré (trop de lignes parasites).

Au cours de la numérisation des annuaires de la SHF, on a aussi identifié d'autres sources de problème de reconnaissance optique de caractères :

- Les fonds colorés formés par des points (Figure 36) : La plupart du temps, le logiciel était incapable de distinguer les chiffres. Mais ce n'était pas toujours le cas : la Figure 37 montre une image de quatre lignes de l'annuaire et la transcription qui en a été faite par ABBYY. Certaines valeurs sur fond gris ont été reconnues, d'autres pas. Dans certaines cases le logiciel a ajouté des « i » ou des « | » alors qu'il n'y a pas de colonnes.

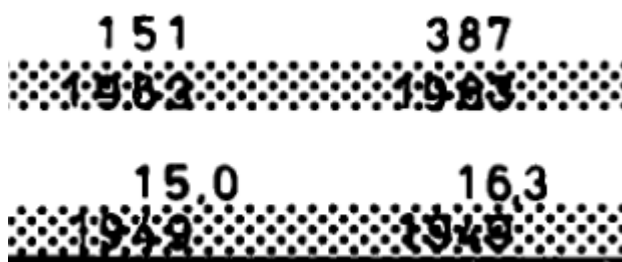


Figure 36 : Fond grisé vu de près.

1886.0	4319.0	1458.0	1500.0	1822.0	1300.0	1158.0	675.0	1910.0	1004.0	1300.0	1850.0	MODULES EXTR.
23-1955	3-1952	14-1930	18-1959	5-1948	2-1962	14-1932	31-1938	14-1963	25-1930	29-1931	12-1965	273
36.00	35.00	40.00	74.00	96.00	77.00	18.00	23.00	34.00	35.00	36.00	42.00	1930
27-1933	1-1949	1-1949	1-1949	3-1938	36-1943	11-1935	16-1947	18-1918	10-1947	23-1933	30-1942	103
1886.0	4319.0	1458.0	1500.0	1822.0	1300.0	1158.0	675.0	1910.0	1004.0	1300.0	1850.0	MODULES EXTR.
25-1955	3-1952	14-1930	18-1959	5-1948	2-1962	14-1932	31-1938	14-1963	25-1930	29-1931	12-1965	273
36.00	35.00	40.00	74.00	96.00	77.00	18.00	23.00	34.00	35.00	36.00	42.00	1930
27-1933	1-1949	1-1949	1-1949	3-1938	36-1943	11-1935	16-1947	18-1918	10-1947	23-1933	30-1942	103

Figure 37 : Transcription d'une portion de tableau grisé par ABBYY. Certaines cases sont bien reconnues, d'autres non.

- La présence de cadres autour des chiffres (les valeurs maximales ou minimales étaient parfois signalisées de cette manière, Figure 38) empêche la reconnaissance optique des caractères.

2.2	1.5	4.5
2.6	1.3	4.6
2.5	1.1	4.6
2.4	1.6	5.8
3.1	1.3	5.8
3.3	1.0	4.5
2.6	2.0	3.1
2.8	1.6	3.5
2.7	2.5	4.2
2.1	2.5	2.5

Figure 38 : Cadre non reconnu par l'OCR.

- Les « + » ont pratiquement toujours été reconnus comme des losanges, malgré la lecture avec apprentissage.
- De même, les points évidés n'ont jamais pu être considérés comme des points malgré la lecture avec apprentissage (Figure 39).

JAN	FÉV	MARS	AVRIL
670.	700.	625.	545.
620.	540.	497.	630.
625.	474.	457.	570.
600.	600.	610.	520.
550.	600.	610.	555.
450.	595.	575.	492.
575.	585.	570.	433.
525.	530.	535.	458.
555.	429.	406.	498.
535.	361.	374.	545.

Figure 39 : Les points sont représentés par les cercles, qui ont toujours été reconnus comme des « 0 » malgré la lecture avec apprentissage.

- Le non-respect de l'horizontalité des lignes n'était pas gérable par ABBYY (Figure 38).

319	276	160	186	97.5	48.1	53.5	83.0	58.0	50.0	79.5	441	154
319	276	160	186	97.5	48.1	53.5	83.0	58.0	50.0	79.5	441	154
33.9	29.4	17.0	19.8	10.4	5.1	5.7	8.8	6.1	5.3	8.4	46.9	16.4
855	670	424	482	261	125	143	222	150	134	206	1180	4860
91	71	45	51	28	13	15	24	16	14	22	126	515

Figure 40 : Exemple d'une portion de tableau qui n'est pas parfaitement droite. Le modèle de zone respecte quant à lui l'horizontalité, ce qui empêche la reconnaissance (chiffres situés dans 2 cases différentes).

- En 1966, en raison d'une police d'écriture particulière, la surface du bassin versant n'est pas souvent reconnue (ligne vide, qui n'est pas transcrite dans Excel).
- Souvent, le décalage entre l'écriture « latitude » ou « longitude » et les coordonnées conduit le logiciel à interpréter cela comme deux lignes différentes, ce qui a ensuite posé des problèmes pour le traitement Excel (Figure 41).

LONGITUDE : 2°29'E

Figure 41 : Décalage suffisant pour que le logiciel considère qu'il s'agit de deux lignes différentes.

2. Annexe 2 : Liste détaillée des étapes du protocole (valable pour les années 1964-1969)

- à l'aide du fichier "données à récupérer.xls", signaler les fichiers à ne pas traiter (stations déjà présentes dans la banque HYDRO) sur la feuille "avancement et modifications.xls" (en gris foncé) et les fichiers qu'il faudra comparer aux données de la banque HYDRO en fin de protocole (en violet)
- identifier les stations pour lesquelles les données suivantes sont disponibles: valeurs de débit instantanées, débits max/min, moyennes mesurées et naturelles. Renseigner les colonnes correspondantes dans le fichier d'avancement.
- reporter les commentaires de bas de page dans les remarques de la feuille "avancement et modifications.xls" et "liste_stations-5.xls" s'il s'agit d'un commentaire concernant la station en général. Si le commentaire nécessite de changer le code de validité des données ("données approximatives", "données estimées", ...) renseigner la colonne.
- ouvrir le fichier de l'année XXXX dans ABBYY Fine Reader (Données de base: PDF sans OCR/600dpi binaire/année/)
- supprimer toutes les pages qui ne contiennent pas les graphes et les tableaux de débits journaliers
- ouvrir le fichier "pages_blanches.pdf" et intercaler une page blanche toute les trois pages (page de graphe, page de tableau, page blanche, page de graphe, page de tableau, page blanche, ...)
- sélectionner toutes les pages de graphe et ouvrir le modèle de zone "nom_st_XXXX" (éventuellement créer un nouveau modèle de zone)
- sélectionner toutes les pages de tableau, et ouvrir le modèle de zone "debits_XXXX" (éventuellement créer un nouveau modèle de zone)
- page de graphe: 1. nom de la station
 - 2. surface du BV
 - 3. numéro de station
- page de tableau: 1. altitude
 - 2. surface du bassin versant
 - 3. année
 - 4. débits journaliers
 - 5. valeurs instantanées + moyennes mensuelles mesurées et naturelles
 - 6. années période longue n°1
 - 7. années période longue n°2
 - 8. débits mensuels pour la période longue n°1 + n°2 + max/min journaliers
 - 9. pluviométries mensuelles
- vérifier quel gabarit utilisateur est actif
- lire
- vérifier l'ajustement des zones et la bonne reconnaissance de la structure des tableaux: ajuster les délimitations, ajouter/supprimer des lignes (options dans la barre d'outil qui s'affiche en sélectionnant un tableau), vérifier que des colonnes/lignes vides ne se sont pas ajoutées en bordure des tableaux
- relire
- enregistrer le résultat de l'analyse en format Excel, avec un fichier séparé par page (cocher l'option créer un fichier séparé à chaque page vide) dans "chroniques debit/19XX/tableaux debit OCR"
- enregistrer le gabarit s'il a été modifié
- enregistrer le doc fine reader "XXXX_tableaux_debit" dans le dossier "ABBY"
- déplacer les fichiers des stations correspondant à celles de la banque HYDRO vers un sous dossier "données dispo dans la BH"
- au fur et à mesure de la vérification Excel, noter les fichiers qui comportent des lacunes ("lacunes.xls")
- depuis le fichier macro.xls, lancer la macro "OCR_1_Tous_fichiers" en renseignant le répertoire contenant les tableaux issus de l'OCR. La macro remplace les "." par ".", supprime des caractères spéciaux (: ; | ...), ajoute les lignes de calcul et enregistre le fichier sous le nom "19XX - 00XX_cor1.xlsx". Tous les fichiers s'ouvrent automatiquement.
- Pour chaque fichier:
 - vérifier que la latitude et la longitude sont bien indiquées et que les 4 stations pluviométriques sont bien traitées
 - si ce n'est pas le cas, il faut le corriger sur le fichier de base et relancer la macro OCR_1 pour ce fichier (il y a un décalage). Vérifiez que:
 - le nom de la station est en ligne 1
 - la surface du bassin versant est en ligne 3
 - le n° de station est en ligne 5
 - l'altitude est en ligne 7

l'année est en ligne 9
 le tableau de débits journaliers est en ligne 11 à 41
 les valeurs instantanées sont en lignes 43-44
 les débits mensuels sont en lignes 45-46
 les années longues périodes sont en lignes 48 et 50
 les débits longue périodes sont en lignes 52-53
 les débits max/min sont en lignes 54-57
 les pluviométries sont en lignes 59-62
 les latitudes/longitude sont en lignes 64 et 66

- vérifier le nom de station, le numéro, la surface du BV, l'altitude, la latitude et la longitude.
- éventuellement convertir en nombre les cellules du tableau
- vérifier l'année et les années longues périodes (moy_mens_19XX-19XX, ...), éventuellement modifier le texte dans la colonne "année"
- vérifier les moyennes annuelles et longues périodes
- si un caractère spécial a été supprimé, la case est affichée en violet. Vérifier le contenu de cette case en le comparant au tableau de l'annuaire.
- s'il reste beaucoup d'erreurs similaires, faire un remplacer (Ctrl+F). Eventuellement fermer excel et rouvrir le fichier si ça ne marche pas
- pour voir si une erreur de frappe est présente, vérifier les lignes "moyenne mensuelle", "différence_absolue", et "différence_relative", ainsi que la colonne "contrôle OCR" (pour les débits mensuels et la pluviométrie). S'il y a une faute de frappe, le calcul ne se fait pas et des ##### sont indiqués : corriger la case manuellement et remplir la case en couleur verte
- lorsque des données de débits max/min sont disponibles vérifier les valeurs et les années correspondantes (beaucoup d'erreur sont commises sur ces lignes)
- vérifier que la différence relative entre les débits ne dépasse pas 1% (affichage en rouge de la case différence_relative) et/ou que l'arrondi est correct (par rapport à la moyenne recalculée)
- si c'est le cas, vérifier visuellement les valeurs dans la colonne concernée, les modifier éventuellement (mettre un fond vert). S'il n'y a pas d'erreur mais qu'il y a quand même une différence, rajouter une ligne "vérif_visuelle" et mettre "ok"
- vérifier que les moyennes annuelles et longue périodes sont similaires à celles indiquées en colonne N (à quelques unités près...).
- vérifier que les sommes de pluviométrie sont similaires à celles indiquées en colonne N (à quelques unités près...).
- vérifier que le nom de la station pluviométrique en colonne A est correct, sinon le corriger
- exécuter la macro "OCR_2", qui enregistre les éventuelles modifications, supprime les lignes de calcul, remplace les cases vides par "NA" et enregistre le fichier sous le nom "19XX - 00XX_cor2.txt"
- vérifier quels fichiers sont concernés par des erratum et les corriger si besoin. Dans ce cas, dans le fichier d'avancement, écrire ":ok" après l'année de l'erratum
- mettre dans un dossier à part les fichiers qui contiennent des lacunes et les traiter séparément
- dans R, ouvrir le script "protocole_OCR.R" (modifier le chemin d'accès et l'année)
- exécuter le script
- s'il y a un message d'erreur à la fin, vérifier qu'il n'y a pas de valeurs manquantes dans les débits. Si c'est le cas, mettre ces fichiers dans un dossier à part pour les traiter séparément. Utiliser le script "protocole_OCR_1fichier.R" et modifier le passage de data372 à data 365
- traiter les fichiers du dossier "lacunes". Vérifier pour ceux-ci s'il faut modifier le code validité en même temps
- modifier le code validité si les commentaires le demandent
- copier les fichiers "chronique_19XX-00XX.txt" dans le dossier "graphes"
- lancer le script "graphes.R" pour générer tous les graphs
- copier les graphes dans le dossier "19XX/graphes chroniques"
- vérifier que chaque graphe corresponde à celui de l'annuaire. Si ce n'est pas le cas, vérifier qu'il s'agit d'une incohérence entre le tableau et le graph de l'annuaire et non pas d'une erreur d'OCR, ou d'une faute de frappe
- si c'est une erreur d'OCR, ou une faute de frappe, corriger dans le fichier XXXX_00XX_cor1.xls, dans le fichier XXXX_00XX_cor2.txt et relancer R (ou s'il y a des lacunes ou des modifications du code validité, ne pas relancer R mais modifier le fichier "chronique_19XX-00XX")
- comparer les données indiquées en violet dans le fichier d'avancement avec la banque HYDRO:
- créer un dossier au nom de la station
- y copier le fichier complet de la BH
- faire un second fichier qui ne contient que l'année à comparer (ou une combinaison de toutes les années disponibles)
- y copier la chronique de l'annuaire à comparer

- lancer le script Comparaison_BH.R (modifier les 2 noms de fichier)
- remplir le tableau présent dans la deuxième feuille du fichier "données à récupérer.xls"
- reprendre la conclusion dans la feuille 1
- si la chronique correspond, passer la couleur pour les années à venir en bleu, si la chronique ne correspond pas, laisser en vert
- vérifier que la chronique disponible dans la BH couvre toutes les années et remplir la colonne
- voir si pour les stations présentant un doute, une autre station de la BH ne pourrait pas correspondre mieux

Code couleur des fichiers de débit en version Excel :

- vert: modification manuelle
- violet: modification automatique (suppression de caractères spéciaux)
- orange: modification manuelle pour prendre en compte l'erratum d'un annuaire ultérieur

3. Annexe 3 : Offre commerciale de la société IMSO



Micrographie
Documentaire
Archivage
électronique
Traitement
d'images

IRSTEA
VILLEURBANNE
anne-laure.achard@irstea.fr

N/REF : D13/3858b/SC/lm

Neuilly-sur-Mame, le 28 octobre 2013

A l'attention de Madame ACHARD

Madame,

Pour faire suite nos différents échanges, nous avons le plaisir de vous remettre la réactualisation de notre offre commerciale concernant la numérisation de vos dossiers « annuaire hydrologique »

Nous sommes à votre disposition pour toutes informations complémentaires.

Nous vous prions de croire, Madame, en l'assurance de notre considération distinguée.

Linda MARTINEZ
Assistante commerciale

Tél. : 01 41 54 10 10
Fax : 01 41 54 13 00
18, Bd Aristide Briand
93330 Neuilly sur Mame
Web : www.imso.fr
E.mail : imso@imso.fr

SARL AU CAPITAL DE : 150.000 €
RCS Bobigny B 353 001 183
Siret 353 001 183 00022
APE 723 Z - Fr 473530011830022

PROPOSITION COMMERCIALE

I - OBJET

Numérisation de documents papier

II - DESCRIPTION DES SUPPORTS ET VOLUME

Environ 30 fascicules « annuaire hydrologique » d'environ 300 pages chacun en couleur recto-verso au format A4 environ, soit un total d'environ 9 000 pages.

III - PROCEDURE

3.1 – Transport et logistique

Le transport aller et retour peut être assuré par nos soins ou par notre transporteur UPS.

- Dans les malles métalliques fermées par cadenas fournies par nos soins,
- Dans notre propre véhicule banalisé et doté de tous les équipements nécessaires à la sécurité (extincteur,....) et sans arrêt entre les deux sites.

Le déchargement est réalisé dans l'enceinte de nos locaux sous caméra de surveillance.



Stockage des documents :

Les dossiers sont stockés à l'abri de la lumière dans une pièce de stockage protégée dans vos conditionnements en attente de traitement.

Les pièces de stockage sont maintenues aux normes d'hygrométrie (45% + ou – 5%) et de température (18°C + ou – 2°).

3.2 - Préparation et reconditionnement

Les « annuaires » sont massicotés par nos soins et vous sont rendus entourés d'un élastique.

3.3 - Numérisation

La numérisation est réalisée en 600 dpi en détection automatique de couleur au format TIF GIV pour les pages en noir & blanc et en JPEG pour les pages en couleur.

Une dégradation en 300 DPI est appliquée pour la consultation au format PDF.

Les pages en 600 DPI seront binarisées (noir et blanc) pour l'application de l'OCR par vos soins.

Des traitements d'images en automatique sont appliqués sur le noir & blanc tels que:

- détourage des bordures
- suppression des points isolés
- lissage de caractères

Les pages comportant des tableaux doivent être orientées en sens lecture des tableaux. La rotation est effectuée par nos soins.

3.4 - Indexation

Un répertoire par dossier indexé sur le titre de l'annuaire et l'année, puis les fichiers sont nommés de 01 à n.

3.5 Reconnaissance de caractères

Celle-ci est effectuée au format PDF texte caché sur les fichiers en 300 DPI.

Aucune reconnaissance de caractères n'est appliquée sur le 600 DPI.

Le résultat de l'OCR est totalement lié à la qualité d'origine du document.

3.6 - Format des fichiers

Les fichiers sont livrés au format PDF et TIF GIV par « annuaire ».

3.7 - Support

Les fichiers sont livrés sur support DVD non réinscriptible ou disque dur externe USB.

IV – DELAI

Début de prestation : à définir entre les parties et en fonction de notre charge de travail au moment de la commande

Délai de prestation : environ 3 semaines en fonction de notre charge de travail lors de votre commande et pouvant être réduit selon les impératifs de votre client.

V – TARIFICATION (facturation au réel)

Mise en place d'application	Forfait	150,00 € HT
Le massicotage	L'ouvrage	2,50 € HT
La prestation comprend : <ul style="list-style-type: none"> • La numérisation en détection automatique de couleur en 600 DPI • L'indexation • La rotation des pages en sens lecture des tableaux • L'OCR • Le contrôle • Les traitements d'images en automatique 	La page	0,12 € HT
La dégradation des fichiers en 300 DPI et la binarisation des fichiers 600 DPI	Forfait	100,00 € HT
Transfert des fichiers sur support DVD	Le DVD	30,00 € HT
Transfert des fichiers sur DD externe fourni par nos soins – 500 GO	Le disque	150,00 € HT
Transfert des fichiers sur support DD externe fourni par vos soins	Le transfert	30,00 HT
Transport paris petite couronne – Ile de France	La course aller	45,00 € HT
Transport	L'envoi par UPS	Tarif en vigueur

ENVELOPPE BUDGETAIRE ESTIMATIVE**Facturation au réel**

Mise en place d'application	1 x 150,00	150,00 € HT
Le massicotage	30 x 2,50	75,00 € HT
La prestation comprend : <ul style="list-style-type: none"> • La numérisation en détection automatique de couleur en 600 DPI • L'indexation • La rotation des pages en sens lecture des tableaux • L'OCR • Le contrôle • Les traitements d'images en automatique 	9 000 x 0,12	1 080,00 € HT
La dégradation des fichiers en 300 DPI et la binarisation des fichiers 600 DPI	1 x 100,00	100,00 € HT
Transfert des fichiers sur DD externe fourni par nos soins – 500 GO	1 x 150,00	150,00 € HT
Transport paris petite couronne – Ile de France	2 x 45,00	90,00 € HT
	TOTAL HT	1 645,00 € HT

VI- SECRET PROFESSIONNEL

La société I.M.S.O. s'engage pour elle et l'ensemble de son personnel, à observer le secret le plus absolu sur la nature des travaux qui lui sont confiés et à ne divulguer aucune information figurant sur les supports en sa possession.

VII – CONDITIONS GENERALES

- Conservation des données : IMSO s'engage à conserver une copie de sauvegarde des fichiers traités pas ses soins durant deux mois après la date de livraison ; au delà, les fichiers sont détruits, sauf mention particulière définie entre les deux parties.
- Validité de l'offre : 6 mois
- Règlement : 30 jours date de facture
- TVA : 19,6 %

Irstea – centre de Lyon-Villeurbanne
UR HHLY
5 rue de la Doua
CS 70077
69626 Villeurbanne Cedex
tél. +33 (0)4 72 20 87 87
www.irstea.fr

