



Stage Pratique de l'Ingénierie

TEST DE METHODES D'AIDE A LA CRITIQUE AUTOMATIQUE DES DONNEES
BRUTES REÇUES PAR LES MODELES DE PREVISION DES CRUES

Juillet à Septembre 2018

GUILLOT-GAUTIER CLAIRE

encadrée par **VIATGÉ JULIE ET**

PERRIN CHARLES

Irstea - Centre d'Antony

Unité de Recherche HYCAR, au sein de

l'Équipe Hydrologie des Bassins Versants

Remerciements

Je souhaite chaleureusement remercier mon encadrante Julie Viatgé qui a su être présente à tous les moments clé pour m'aider à atteindre les objectifs que nous nous étions fixés. Je la remercie pour son pragmatisme et la prise de recul que sa vision a apportée. Je suis reconnaissante envers Charles Perrin qui a su nous accompagner tout au long de ce stage, c'est un hydrologue expérimenté et bienveillant qui sait comment aider ceux qui l'entourent. Un grand merci à toute cette équipe, l'équipe HYDRO, qui m'a accueillie et m'a permis de réaliser ce stage dans les meilleures conditions que j'aurais pu imaginer !

Merci à Olivier Delaigue pour la transmission d'une part de ses connaissances en programmation. Merci à François Tilmant et Benoît Génot mes co-bureaux, avec qui les journées étaient instructives les unes après les autres. Merci à Lila Collet pour ses astuces en programmation et nos discussions variées.

Merci aussi à tous les autres membres de l'équipe ! Merci aux doctorants avec lesquels j'ai passé de nombreux bons moments ! J'espère vous revoir prochainement.

Je remercie également Lionel Berthet et Renaud Marty du Service de Préviation des Crues Loire-Cher-Indre pour avoir répondu à l'ensemble de mes questions et pour avoir partagé les données utilisées dans cette étude.

Une pensée toute particulière à mes proches qui ont su m'accompagner et m'aider dans différents moments et grâce auxquels j'ai pu prendre du recul aux bons moments. Merci notamment à Maurine Grivaz et Guillaume Helluy qui m'ont considérablement soutenue lors de la rédaction !

Table des matières

Remerciements	3
Liste des figures	6
Liste des tableaux	6
Introduction	8
I. Contexte de l'étude	9
I.1. Etat des pratiques	9
I.1.1. Présentation du modèle GRP	9
I.1.2. Présentation des données utilisées	9
Tableau 1 : Synthèse des bases de données sources utilisées	10
I.2. Enjeux de l'étude	10
I.2.1. Utilisation dans les SPC	10
I.2.2. Problèmes existants	11
I.3. État des connaissances	11
II. Méthodes et Résultats	14
II.1. Méthodes choisies	14
II.2. Procédure d'évaluation des méthodes	15
II.3. Résultats	16
III. Discussions et analyse critique	20
III.1. Limites de l'étude	20
III.2. Perspectives d'améliorations	21
III.3. Analyse de l'organisation de l'étude	22
Conclusion	23
Bibliographie	24
Annexes	25
Annexe 1 - Décret n°2012-209 du 1 février 2012 relatif à l'Irstea	25
Annexe 2 - Organigramme HYCAR	26
Annexe 3 : Concept de base du filtre de Kalman (schéma personnel de Petteri Aimonen)	27
Annexe 4 - Schéma explicatif du PELT méthode (source article de Xu et al. 2017)	28

Liste des figures

Figure 1 : Schéma de fonctionnement du modèle GRP en temps réel (source : Irstea)	8
Figure 2 : Carte des SPC en France (source : SCHAPI)	9
Figure 3 : Hyétogramme permettant la visualisation de la détection d'erreurs (résultat de l'étude)	16
Figure 4 : Graphique représentant l'évolution des intensités de pluie en mm/ pas de temps (résultat de l'étude)	16
Figure 5 : Hydrogramme issue du traitement automatique développé lors de l'étude (résultat de l'étude)	17
Figure 6 : Zoom sur une zone de débit constant entre le 8 et 9 mai 2003 (résultat de l'étude)	17
Figure 7 : Zoom sur une zone de débit constant le 4 février 2012 au matin (résultat de l'étude)	17

Liste des tableaux

Tableau 1 : Synthèse des bases de données sources utilisées	10
Tableau 2 : Synthèse des articles scientifiques lus et retenus pour l'étude	12
Tableau 3 : Synthèse des typologies d'erreurs à détecter associées à leurs méthodes respectives	14
Tableau 4 : Représentation du schéma type d'une table de contingence ajusté au cas d'étude	15

Introduction

Ce rapport présente les objectifs, la démarche, les expériences, les résultats et les discussions associés de mon Stage Pratique de l'Ingénierie entre le 9 juillet et le 14 septembre 2018. Ce stage s'est déroulé à l'Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (Irstea) sur le site d'Antony au sein de l'équipe Hydrologie des bassins versants (HYDRO¹), sous l'encadrement de Julie Viatgé.

Irstea est un établissement public à caractère scientifique et technologique (EPST) sous la tutelle des ministères en charge de la recherche et de l'agriculture. Créé en 1981 sous le nom de Cemagref, ses activités s'articulent autour de la recherche et de l'expertise. Elles sont précisées dans l'article 4 du Décret² n° 2012-209 du 13 février 2012 (annexe 1). Ces activités sont réparties en trois départements scientifiques : Eaux, Ecotechnologies et Territoires. Irstea regroupe plus de 1500 collaborateurs dont un peu plus de 1100 chercheurs, ingénieurs, doctorants et post-doctorants répartis dans les centres régionaux. Le centre d'Antony est constitué de 94 agents titulaires, 79 agents contractuels, 25 doctorants et environ 40 stagiaires de l'enseignement supérieur chaque année. Il est composé de trois unités de recherches, dont l'unité de recherche Hydrosystèmes Continentaux Anthropisés - Ressources, Risques, Restauration (HYCAR³, voir organigramme en annexe 2). Cette unité de recherche est formée de trois équipes dont l'équipe HYDRO.

Irstea collabore depuis plusieurs années avec le Service central d'hydrométéorologie et d'appui à la prévision des inondations (SCHAPI) afin d'améliorer les systèmes de prévision des crues. Le SCHAPI est un service de la Direction générale de la prévention des risques au Ministère de la transition écologique et solidaire. Il a été créé en 2003, avec les Services de Prévision des Crues (SPC), à la suite d'une série d'inondations catastrophiques et un rapport parlementaire soulignant les lacunes du système d'annonce des crues de l'époque. La convention établie entre le SCHAPI et Irstea a mené au développement du modèle du Génie Rural pour la Prévision des crues GRP (Tangara, 2005 et Berthet, 2010), actuellement utilisé par la majorité des SPC. Le développement et l'amélioration de ce modèle, en accord avec les besoins des SPC, est une des missions de l'équipe HYDRO.

Ainsi, ce stage s'inscrit dans cette démarche d'amélioration, sous l'angle de la critique des données. Concrètement, le modèle GRP utilise des données d'entrée qui proviennent de réseaux de stations de mesures pour fournir une prévision de l'évolution du débit dans les principaux cours d'eau. Ces prévisions présentent un enjeu important en lien avec les coûts des dégâts causés par les crues, et leur qualité est directement liée à la qualité des données d'entrée du modèle. Le travail de ce stage a pour objectif d'identifier une ou plusieurs méthodes qui pourraient servir de base à fournir un module complémentaire au modèle GRP permettant une critique automatique des données d'entrée du modèle.

En pratique, les résultats d'application du modèle GRP par les SPC sont, entre autre, illustrés sur le service d'information Vigicrues⁴ qui propose une carte de vigilance en France. Cette carte est actualisée deux fois par jour et, en période de crise, son actualisation peut être horaire. L'automatisation de la critique des données est intéressante puisque la fréquence d'actualisation des prévisions peut se révéler importante et en conséquence incompatible avec une critique fine des données par les prévisionnistes. La plupart du temps, les prévisionnistes réalisent une analyse rapide des données reçues en temps réel puis une validation plus poussée est réalisée à une échelle mensuelle ou annuelle. L'intérêt de l'outil serait de permettre une analyse en temps réel plus efficace.

¹ <http://www.irstea.fr/la-recherche/unites-de-recherche/hycar/hydrologie-bassins-versants>

² <https://www.legifrance.gouv.fr>

³ <http://www.irstea.fr/hycar>

⁴ <https://www.vigicrues.gouv.fr>

I. Contexte de l'étude

I.1. Etat des pratiques

Cette partie a pour objectif de présenter l'état des pratiques actuelles en termes de critique de données afin de mieux situer dans quel contexte le travail effectué vient s'inscrire. Le modèle de prévision de crues GRP, ainsi que les données utilisées dans le cadre de cette étude sont présentés ici.

I.1.1. Présentation du modèle GRP

Le modèle GRP⁵ est un modèle hydrologique global (à l'échelle du bassin versant), qui fonctionne toute l'année à un pas de temps horaire. Historiquement, ce modèle a été développé pour la prévision des crues. C'est un modèle en constante évolution en lien avec les demandes des Services de Prévision des Crues.

Comme de nombreux modèles de prévision, il demande en entrée des données de pluies, débits et évapotranspiration. L'évapotranspiration n'étant pas un phénomène sensible durant les épisodes de crues (Oudin, 2004), le modèle se contente d'une moyenne interannuelle et ne nécessite pas de données précises. Nous ne nous intéresserons donc qu'aux données observées de précipitations et débits.

Lors de l'utilisation de l'outil en temps réel par les prévisionnistes, les données de forçage utilisées par le modèle sont les données directement mesurées par les pluviomètres et stations hydrométriques, sans post-traitement. En figure 1 se trouve le schéma de fonctionnement détaillé du modèle.

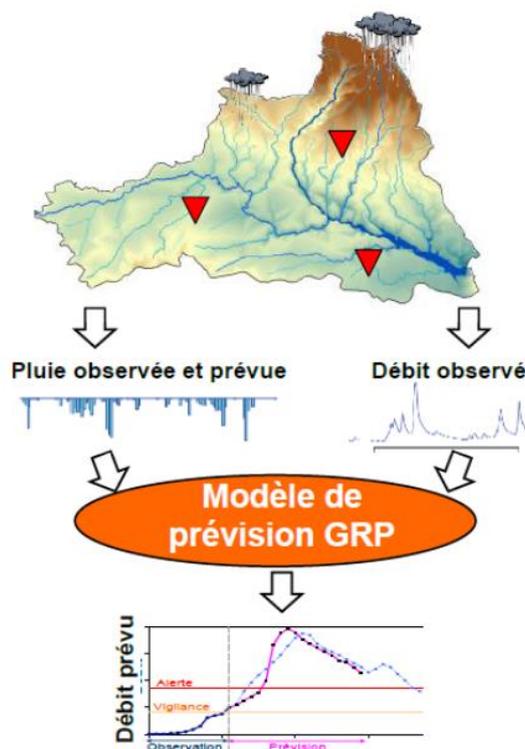


Figure 1 : Schéma de fonctionnement du modèle GRP en temps réel (source : Irstea)

I.1.2. Présentation des données utilisées

Comme cela vient d'être précisé, la qualité des données utilisées en entrée de modèle peut être un facteur d'amélioration des prévisions. En pratique, les données dites "brutes" sont celles relevées par les différentes stations de mesures réparties sur le territoire français et sans post-traitement préalable. C'est dans ces données que peuvent se glisser des erreurs pour diverses raisons qui seront exposées en sous-partie I.2.2. "Problèmes existants". Les données brutes sont ensuite post-traitées plusieurs fois en vue d'améliorer leur qualité et deviennent des données dites "propres". Il a été choisi de travailler sur des jeux de données brutes et propres qui correspondent respectivement aux données enregistrées par les stations de mesures sans post-traitement et aux données post-traitées et validées par les hydromètres.

Les données utilisées pour cette étude (que ce soient des données de pluie ou débit, brutes ou propres) ont été fournies par le SPC Loire-Cher-Indre (figure 2) et s'étendent sur une durée d'au moins quinze ans à partir des années 2000 jusqu'en mars 2018 pour les chroniques les plus longues.

⁵<https://webgr.irstea.fr/modeles/modele-de-prevision-grp/fonctionnement-grp/>

Le tableau 1 ci-après récapitule la provenance de ces données. Le choix des types de données (pluies et débits) est relatif aux paramètres auxquels le modèle est sensible et le choix de leur nature (brutes et propres) permet d'identifier les erreurs des chroniques brutes par comparaison aux chroniques propres, et sera le facteur permettant de valider l'outil automatique de détection de ces erreurs. Les banques de données ont été choisies suite à des échanges avec des agents des Services de Prévision des Crues (SPC). Leur expérience et leurs connaissances sur le sujet ont permis de cibler les données nécessaires à notre étude.

Tableau 1 : Synthèse des bases de données sources utilisées

Jeux de données	Brutes	Propres
Pluviométriques (P)	Cristal	Hydro2 et Météo France
Hydrométriques (Q)	Cristal	Hydro2

La base de données Cristal⁶ (Moulin et Thépot, 1999) s'est construite suite à un souhait d'automatiser le réseau de récupération des données hydrologiques et météorologiques. Dans un premier temps axé sur les données pluviométriques, son champ de collecte s'est par la suite étendu au reste des données de débits, températures etc. Cette base de données est accessible aux gestionnaires du bassin.

Ces données sont une première fois post-traitées mensuellement et deviennent ainsi des données "corrigées" que l'on peut retrouver dans la banque de données SIRCADE au SPC Loire-Cher-Indre par exemple. Les données sont ensuite à nouveau post-traitées, annuellement cette fois, et deviennent des données dites "validées". Elles sont alors stockées dans la base Hydro2⁷ qui est une base de données publique.

Tous les post-traitements sont faits manuellement, ce qui permet d'avoir des données propres a posteriori sur les événements de crue. Néanmoins, en temps réel, il est aussi nécessaire d'avoir des données se rapprochant des données propres, ou tout au moins d'avoir des chroniques non impactées par des problèmes de mesures. C'est là l'intérêt de l'automatisation de la critique des données.

I.2. Enjeux de l'étude

I.2.1. Utilisation dans les SPC

Les données utilisées sont celles d'un seul SPC (le SPC Loire-Cher-Indre encadré en orange sur la figure 1). Cependant les difficultés rencontrées par ce SPC sont considérées représentatives des problèmes rencontrés ailleurs (hors problématique de la neige). Les résultats obtenus devraient ainsi être suffisamment génériques pour qu'ils puissent être étendus à toute la France métropolitaine.

*Figure 2 : Carte des SPC en France
(source : SCHAPI)*



⁶ <http://www.ladocumentationfrancaise.fr>

⁷ <http://hydro.eaufrance.fr>

Le modèle GRP est développé pour et en concertation avec les SPC, qui en sont des utilisateurs expérimentés. Échanger avec eux est, de fait, un moyen d'amélioration judicieux. Une rencontre annuelle intitulée "Groupe Utilisateurs" entre Irstea, le SCHAPI et différents prévisionnistes des SPC, est organisée pour permettre de présenter aux SPC les évolutions du modèle et pour laisser un espace de parole pour les retours d'expérience, souhaits et observations provenant des SPC. C'est ainsi l'occasion de définir les objectifs de développement de l'année suivante en fonction des besoins des utilisateurs.

C'est une convention de partenariat entre le SCHAPI et Irstea qui est à l'origine de cette synergie et de l'évolution du modèle de prévision. Cela permet d'échanger entre les acteurs opérationnels que sont les SPC et les acteurs scientifiques qui développent l'outil (Furusho et al. 2016). Ainsi le travail réalisé est en permanence mis à l'épreuve des besoins réels des usagers et évolue en y répondant.

Les retours d'expériences des prévisionnistes sont à l'origine de cette demande d'automatisation de critique des données. En effet, étant donné l'impact non négligeable de la qualité des données sur la précision des prévisions et les exigences d'actualisations fortes des cartes de vigilances en période de crues, le besoin de produire des prévisions rapidement induit la nécessité d'un post-traitement automatisé.

I.2.2. Problèmes existants

Les données brutes, sensibles à un ensemble de facteurs, peuvent être en partie erronées. En s'appuyant sur la "Charte Qualité de l'Hydrométrie" (Perret et al. 2017)⁸ diffusée par le SCHAPI et sur le "Guide des Pratiques Hydrologiques" (OMM, 1996)⁹ publié par l'Organisation Météorologique Mondiale, il est possible d'établir une liste relativement complète, bien que non exhaustive, des erreurs les plus couramment rencontrées.

Parmi ces erreurs, les plus courantes sont : les pannes ou dérives des capteurs, l'obstruction ou le dérèglement des capteurs (pluviomètre bouché, débitmètre gêné par la présence de végétation ou d'embâcles), un nombre insuffisant de jaugeages pris en compte pour déterminer la courbe de tarage, le gel des capteurs, la fonte tardive de la neige dans les pluviomètres, etc.

Les conséquences de ces erreurs peuvent se ressembler et peuvent se traduire par les symptômes suivants : valeurs aberrantes (statistiquement), valeurs erronées (physiquement), lacunes, pluie constante sur une longue durée (due à la fonte de neige), incohérence dans les valeurs exceptionnelles telles que les dates et durées d'étiage, l'importance des crues etc.

D'autre part, les moyens nécessaires à une identification de la cause des erreurs sont encore trop peu à notre portée et nécessitent un temps de travail trop important pour la durée du stage. Il a donc été choisi de se concentrer sur la détection de certaines erreurs dans les données afin de fournir à l'utilisateur une fiche graphique renseignant sur la qualité des données et sur les anomalies détectées sans pour autant en préciser l'origine. La partie II.1. "Méthodes choisies" détaille les choix des techniques de détection.

I.3. État des connaissances

Au début de cette étude, des recherches bibliographiques ont été menées afin de prendre en main le sujet et de prendre connaissance des méthodes de critiques de données existantes. Ces lectures ont orienté certains choix comme l'idée du test de cohérence qui consiste à comparer les données mesurées et des valeurs statistiques localisées, qui a été adapté (Estevez et al. 2011).

⁸ <http://www.eaufrance.fr/IMG/pdf/Schapi>

⁹ <https://hydrologie.org>

L'utilisation du taux d'accroissement (Robertson et al. 2015) est un autre exemple de ce qui a pu être exploité et adapté à notre étude. Pour rappel, le taux d'accroissement est défini par la relation ci-après :

$$T_{\text{taux d'accroissement}} = \frac{Q_{i+1} - Q_i}{t_{i+1} - t_i}$$

Avec Q les données mesurées de débits,
t les instants auxquels sont mesurés les débits,
i et i+1 les indices auxquels sont évalués Q et t.

L'étude bibliographique s'est concentrée sur la lecture d'articles scientifiques ainsi que sur d'autres types de documents comme des chartes, des guides et des thèses.

Pour résumer les articles les plus intéressants pour l'étude menée, le tableau 2 présente succinctement les méthodes, leurs avantages et inconvénients.

Tableau 2 : Synthèse des articles scientifiques lus et retenus pour l'étude

Références	Résumé	Méthodes	Avantages	Inconvénients
Liu et al. (2018)	Automatisation du contrôle de qualité des données pluviométriques	Comparaison données observées avec une base de référence locale (AWAP ¹⁰)	Contrôle très précis, Amélioration des prévisions fines	Nécessite une importante base de données locale
Hill and Minsker (2010)	Détection en temps réel d'anomalies avec un modèle auto-régressif (compare quatre méthodes)	- Naive predictor - Nearest cluster predictor - Single-layer linear network predictor - Multilayer perceptron predictor (MLP)	Méthode rapide, Applicable à une grande quantité de données	Dans l'article les données sont des données de vitesses du vent
Robertson et al. (2015)	Algorithme de contrôle qualité des précipitations horaires	Méthode doubles masses entre pluviomètres et Algorithme d'identification	Difficile d'identifier les tendances sur du long terme	Rapide contrôle qualité pour de larges échantillons de valeurs
Estevez et al. (2011)	Contrôle qualité des données météorologiques	Test de portée et test de cohérence interne	Tri rapide et calculs simples	Nécessite une forte connaissance des données historiques
Xu et al. (2017)	Algorithme de détection d'erreurs dynamique (combine deux méthodes)	Filtre Kalman et Test du temps linéaire exact élagué (PELT)	Algorithme d'apprentissage automatique	Méthode implémentée sur données industrielles en pétrochimie

¹⁰ [Australian Water Availability Project](#)

L'automatisation du contrôle des données pluviométriques (Liu et al. 2018) repose en partie sur un algorithme de contrôle de qualité préexistant (Robertson et al. 2015). La particularité de cette automatisation réside surtout dans l'application très localisée, en Australie, de la méthode de contrôle de qualité. En effet dans le cadre de cette étude la base de données : AWAP (Australian Water Availability Project) est utilisée. Cette base de données recense des informations météorologiques variées, notamment l'humidité du sol, les pluies, l'évapotranspiration, les écoulements de surface et de drainage. Ces données sont disponibles sur l'ensemble du territoire australien. La réutilisation directe de cette méthode repose donc sur des données non disponibles pour l'étude menée ici. Par ailleurs, les données les plus anciennes collectées par AWAP remontent au début des années 1900, cela permet des traitements statistiques et une analyse très fine.

Le test de cohérence (Estevez et al. 2011) qui consiste à comparer des cumuls horaires entre eux, et par exemple à vérifier que le cumul maximum de pluies tombées au cours de 3h consécutives est bien inférieur à celui des précipitations cumulées maximales sur 6h consécutives. Ce test est un intéressant point de départ dans le choix de faire des calculs d'intensité, pour les pluies, et de taux d'accroissement, pour les débits. En effet, ce test repose sur une connaissance précise de la zone d'étude. Cette connaissance de la zone d'étude, notamment l'observation des crues historiques, est adaptée grâce aux calculs faits sur les données historiques. Ces calculs permettent d'estimer un mode de fonctionnement de la zone étudiée en période de crue. Une partie de ces tests est adaptée dans cette étude, en comparant des pluies en temps réel à des intensités maximales pour les pluies et à des taux d'accroissements en temps réels avec des seuils historiques pour les débits.

L'utilisation de modèle autorégressif (Hill and Minsker, 2010) existe sur des données météorologiques, dans le cas de nos lectures pour les vitesses des vents. Ces modèles autorégressifs permettent la détection en temps réel d'anomalies. Ils reposent sur la prise en compte d'une fenêtre variable de donnée afin de calculer une valeur prévue à l'instant suivant et de coupler cela avec un calcul d'intervalle de prédiction. La diversité de méthodes présentée est intéressante. Néanmoins le transfert de ces méthodes à nos données nécessite d'une certaine façon l'implémentation de modèles simplifiés de prévisions. Concrètement, les quatre méthodes proposées : naive predictor, nearest cluster predictor, single-layer linear network predictor et multilayer perceptron predictor (MLP) se basent sur la prévision de l'instant $t+1$ à partir des données mesurées pendant une durée choisie. Cela reviendrait donc à créer un modèle de prévision annexe et notre l'objectif d'étude est différent.

Une dernière étude ayant mis en place un algorithme de détection dynamique (Xu et al. 2017) apporte des éléments intéressants. Cette méthode dynamique combine deux autres méthodes déjà existantes : celle du filtre Kalman et celle du test du temps linéaire exact élagué (PELT). Ces algorithmes se retrouvent en annexe 3 et 4 où leurs schémas de fonctionnement sont repris. Cette étude a été menée sur des données pétrochimiques donc dans un domaine assez éloigné du nôtre. Toutefois cela pourrait présenter une perspective intéressante pour l'évolution du domaine de l'automatisation de la détection d'erreurs.

II. Méthodes et Résultats

II.1. Méthodes choisies

Après concertation avec les prévisionnistes du SPC Loire-Cher-Indre, nous avons été en mesure de déterminer parmi les données disponibles celles qui auraient le plus d'intérêt pour l'étude menée : en particulier nous nous intéressons aux chroniques "brutes" contenant beaucoup d'erreurs et aux bassins pour lesquels les courbes de tarage n'ont pas beaucoup évolué.

En effet, il est important, au vu des objectifs fixés, de trouver des jeux de données brutes contenant des erreurs d'origines variées. D'autre part, le choix de laisser de côté les décalages induits par les changements de courbes de tarage est fait pour faciliter l'étude. Le choix des stations d'étude s'est donc porté sur les stations hydrométriques de Vic-le-Comte (K2680810) et de Gien aval (K4180010) et sur la station pluviométrique de Vaubarlet (K0459910).

Parmi tous les motifs d'erreurs cités en sous-partie I.2.2. "Problèmes existants", le choix s'est porté sur la détection des typologies d'erreurs reportées dans le tableau 3.

Tableau 3 : Synthèse des typologies d'erreurs à détecter associées à leurs méthodes respectives

Types d'erreurs		Méthodes de détection
P L U I E S	Lacunes	Code erreur : -9.9999 et -99.9 respectivement pour les pluies propres et brutes
	Obstruction pluviomètres	Intensité constante et pluie non nulle pendant un certain temps
	Pluies aberrantes ou erronées	Intensité de pluies : → supérieure à intensité maximale historique → inférieure à intensité minimale historique
	Attention particulière car anomalies statistiques	Calcul du cumul mensuel et comparaison avec le cumul mensuel historique (moyen, maximum et minimum)
D É B I T S	Lacunes	Code erreur (-99.9) ou absence de données
	Capteurs gênés	Taux d'accroissement nul pendant un certain temps
	Débits aberrants ou erronés	Taux d'accroissement entre deux instants : → supérieur au taux maximum historique → inférieur au taux minimum historique → constant pendant un certain nombre de pas de temps (hors étiages)

Pour toutes les méthodes de détection faisant intervenir un "certain nombre de pas de temps" ou un "certain temps" le choix s'est porté sur des durées de 10h. Ce choix est arbitraire car même s'il repose sur des considérations de cohérences, une autre durée aurait aussi permis d'obtenir des détections convenables.

Finalement, de toutes les méthodes présentées dans la partie I.3. "État des connaissances", seuls quelques concepts sont repris et adaptés dans ce travail. Afin d'arriver à un résultat satisfaisant et exploitable pour la suite, l'étude se concentrera sur la production de fiches qualité contenant des informations relatives aux détections des erreurs présentées dans le tableau 3.

II.2. Procédure d'évaluation des méthodes

Afin d'évaluer l'efficacité des méthodes implémentées, plusieurs approches ont été envisagées. Dans notre cas d'étude, comparer les données détectées par la méthode développée au cours du stage avec les données propres permet d'évaluer les correspondances entre les deux séries de données. Une comparaison est faite entre les données post-traitées automatiquement par l'outil développé au cours du stage et les données post-traitées manuellement par les prévisionnistes (données dites propres).

Le coefficient de corrélation de Bravais-Pearson (Kirk et s, 2008) est un indice statistique qui exprime l'intensité et le sens de la relation linéaire qui existe entre deux variables quantitatives. Un coefficient proche de 1 indique une bonne corrélation et un coefficient proche de zéro signifie de grandes variations et différences entre les deux séries. Il s'agit d'un indicateur moins précis que la méthode prévue qui suit.

Une autre façon d'évaluer les méthodes d'automatisation de détection d'erreurs mises en place est d'utiliser une table de contingence. Cela sert notamment d'évaluation du modèle GRP et c'est un outil adéquat et transposable pour comparer les données pour lesquelles on a détecté des erreurs de manière automatique (données "autom") avec les données corrigées manuellement par les prévisionnistes (données propres).

Le tableau 4 présente une table de contingence type.

Tableau 4 : Représentation du schéma type d'une table de contingence ajusté au cas d'étude

		<i>Données, issues des données Propres, considérées</i>	
		.. erronées	.. non erronées
<i>DONNÉES "AUTOM"</i>	Erreurs détectées	erreurs valables (a)	fausses détections (b)
	Erreurs non détectées	erreurs manquées (c)	non-détection valables (d)

L'utilisation des tables de contingence revient donc à distinguer quatre catégories d'erreurs : a, b, c et d. Les valeurs a, b, c et d sont incrémentées à chaque fois qu'une des situations suivantes se présente :

- une erreur est détectée par notre outil et manuellement par les prévisionnistes (a)
- une erreur est détectée par notre outil mais n'a pas été considérée comme telle par les prévisionnistes (b)
- aucune erreur n'est pas détectée par notre outil, alors qu'elle a été détectée par les prévisionnistes (c)
- aucune erreur n'est détectée ni par notre outil, ni par les prévisionnistes (d)

Cela permet ensuite de calculer trois ratios :

- la probabilité de détection d'erreur (Probability Of Detection) : $POD = a / (a+c)$
- le ratio de fausses alertes (False Alert Ratio) : $FAR = b / (a+b)$
- l'indice de succès critique (Critical Success Index) : $CSI = a / (a+b+c)$

Plus les ratios POD et CSI sont élevés et plus le ratio FAR est proche de zéro, plus la proximité entre les données traitées automatiquement et celles traitées manuellement est grande, donc plus le traitement automatique a réussi à reproduire le traitement manuel.

En pratique, pour rigoureusement appliquer ces méthodes, il faut un jeu de données corrigées qui corresponde aux données brutes étant traitées suite aux méthodes de détections mises en place dans cette étude. Pour cela, savoir comment les prévisionnistes font leurs choix de correction est important dans la mesure où cela permet de reproduire les mêmes choix. Cependant sur un échantillon aussi restreint que celui choisi, pour des contraintes temporelles, les corrections sont très particulières et dépendent directement du prévisionniste les ayant corrigées. Cette procédure d'évaluation a donc bien plus de sens sur une étude se déployant sur une plus vaste quantité de données.

Finalement, une évaluation de la méthode est une perspective dans la poursuite de cette étude.

II.3. Résultats

Le logiciel informatique choisi pour implémenter ces méthodes est développé par R Core Team, le manuel d'utilisation ayant servi dans cette étude s'intitule "R : A language and environment for statistical computing". Il s'agit de l'édition de 2018 et la version du logiciel utilisée est la 3.5.1.

Comme précédemment évoqué, l'objectif fixé est de produire des fiches d'une page pour chaque station (pluviométrique et hydrométrique) avant chaque lancement du modèle GRP. Cette fiche renseigne sur la qualité des données brutes reçues par comparaison aux données propres historiquement observées. D'autre part, la temporalité d'utilisation étant courte, la visualisation des résultats se fait à chaque lancement du modèle ou à chaque réception de nouvelles données.

La représentation choisie est pour les données :

→ Pluviométriques :

- un hyétogramme sur les dernières 24h comprenant les informations :
 - de lacunes (en gris),
 - de pluies constantes et non nulles pouvant correspondre à une obstruction (en orange)
 - de pluies aberrantes ou erronées (en rouge)
- un graphique représentant l'évolution des intensités sur les dernières 24h avec :
 - les évolutions au pas de temps horaires,
 - le maximum et le minimum mesurés dans les données propres historiques
- une fenêtre glissante de cumuls de pluie mensuel sur 30 jours par rapport aux cumuls historiques moyen, maximum et minimum, ces trois derniers étant calculés à partir des données historiques propres.

→ Hydrométriques :

- un hydrogramme sur les dernières 24h comprenant les informations :
 - de lacunes (en gris),
 - de débits constants et non nuls pouvant correspondre à une obstruction des capteurs (en orange),
 - de débits aberrants ou erronés (en rouge)

- en ajoutant une information définie de manière statistique à partir des données mensuelles : un diagramme en boîte donnerait les valeurs minimales, maximales et les quantiles 10, 25, 50, 75 et 90 observés historiquement sur le mois en cours.
- un graphique représentant l'évolution des taux d'accroissement sur les dernières 24h avec :
 - les évolutions au pas de temps horaire,
 - le maximum et le minimum calculés à partir des données propres historiques.

Les figures 3 et 4 présentent un exemple de visualisation des résultats. La figure 3 présente un hyétoqramme entre les dates du 27 avril et du 27 mai 2003. Cette période de représentation a été choisie pour illustrer l'évolution des pluies et surtout pour mettre en avant la représentation des zones lacunaires (gris) et des zones de pluies constantes correspondant à une défaillance de mesure (orange). La représentation de pluies aberrantes n'a pas pu se faire sur cette même période, toutefois les outils de détection sont bien opérationnels. En pratique cette représentation se fera sur 24h.

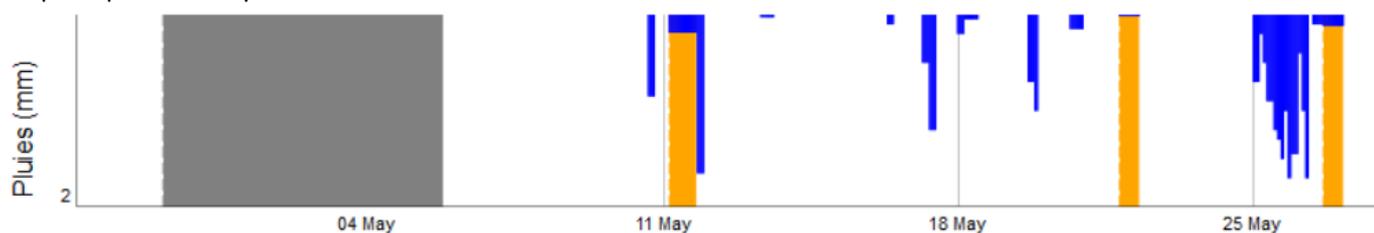


Figure 3 : Hyétoqramme permettant la visualisation de la détection d'erreurs (résultat de l'étude)

La figure 4, elle, représente l'évolution des taux d'accroissement sur la même période. La période durant laquelle les données sont lacunaires, dans la figure 3, correspond à la période où les intensités sont non renseignées, dans la figure 4. La plupart du temps, la pluie étant nulle, elle se visualise par des intensités nulles sur la figure 4, d'où la ligne quasiment continue en abscisse.

Le choix de cette période de visualisation d'un mois permet de visualiser les différents types d'erreurs que notre outil décèle. Le calcul de taux d'accroissement extrêmes (maximum et minimum) pour les débits et l'observation des intensités extrêmes historiques permettent de définir des valeurs de référence relatives aux événements historiques permettant d'alerter les prévisionnistes.

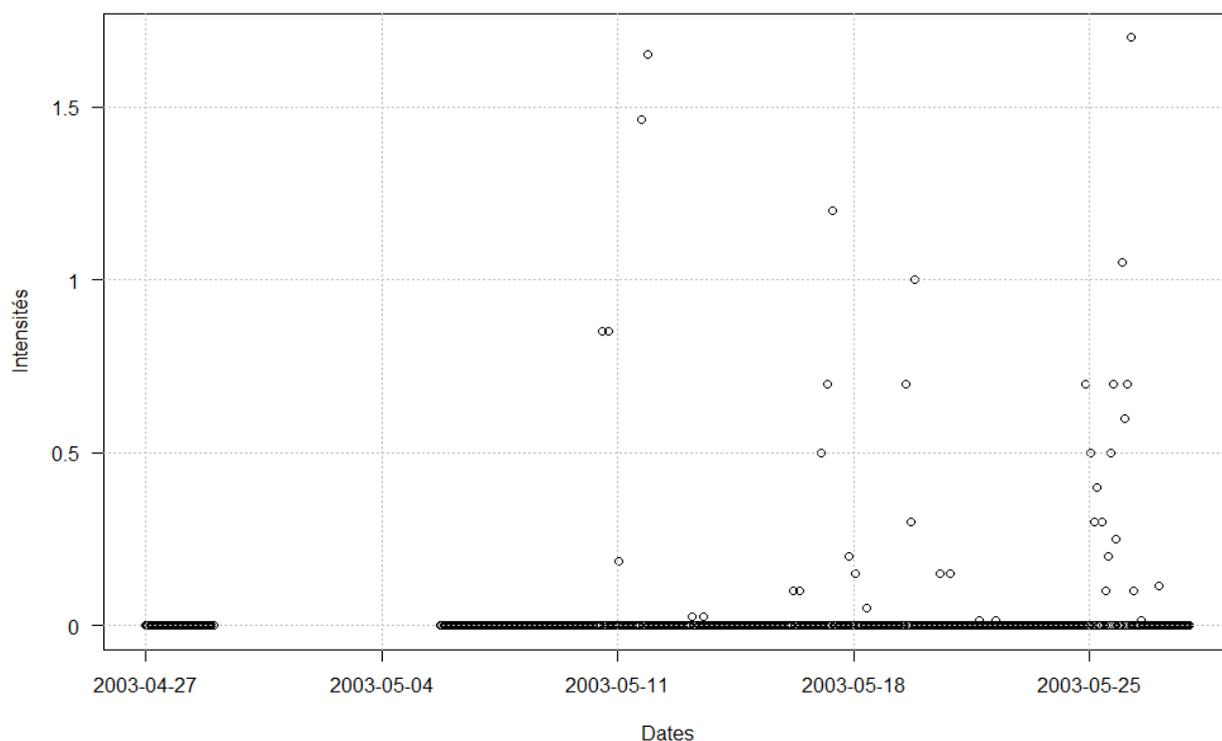


Figure 4 : Graphique représentant l'évolution des intensités de pluie en mm/ pas de temps (résultat de l'étude)

La figure 5 présente l'hydrogramme obtenu après utilisation de notre outil d'automatisation de détection d'erreurs sur une chronique allant de début 2000 à fin 2017 sur la station de Vic-le-Comte, sur la Loire à une trentaine de kilomètres de Clermont Ferrand. Sur cette figure, de nombreuses zones oranges, condensées sur les années 2000 à 2010, indiquent des débits constants pendant des durées excédents 10h consécutives. Un capteur trop peu sensible aux variations a probablement induit ces détections d'erreurs, ce capteur a dû être remplacé par un plus sensible en début 2010 pour permettre d'enregistrer plus finement les débits.

Entouré en orange foncé, un zoom va sur la période du 8 et 9 mai 2003. Au cours de cette période, visualisable en figure 6, un débit constant égal à 21,05 m³/s a été enregistré pendant 15h consécutives.

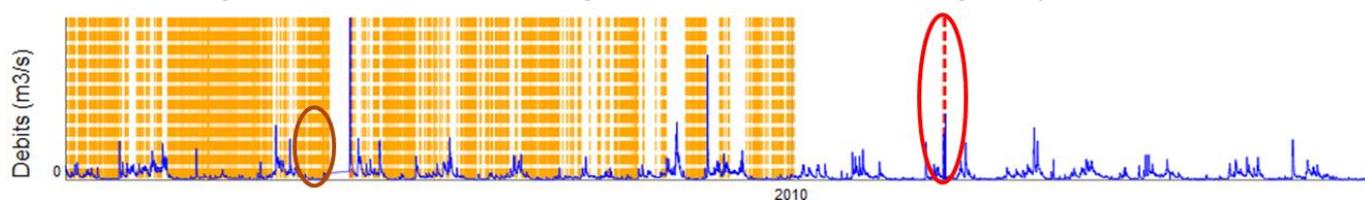


Figure 5 : Hydrogramme issu du traitement automatique développé lors de l'étude (résultat de l'étude)

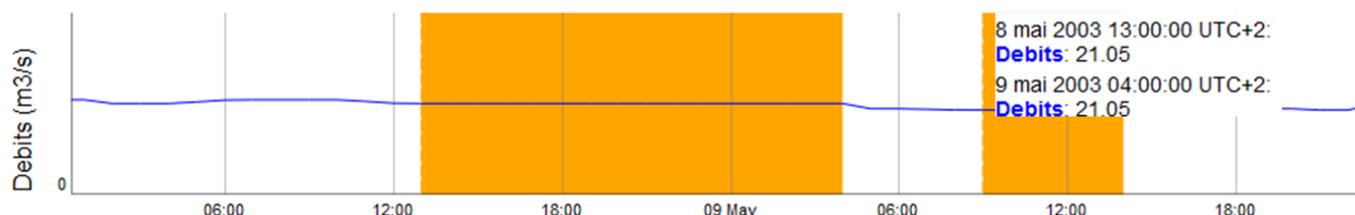


Figure 6 : Zoom sur une zone de débit constant entre le 8 et 9 mai 2003 (résultat de l'étude)

La zone entourée en rouge correspond à une détection de débit aberrant. Sur la figure 7 l'évolution des débits, entre minuit et 2h du matin le 4 février 2012, indique bien une variation anormale du débit.

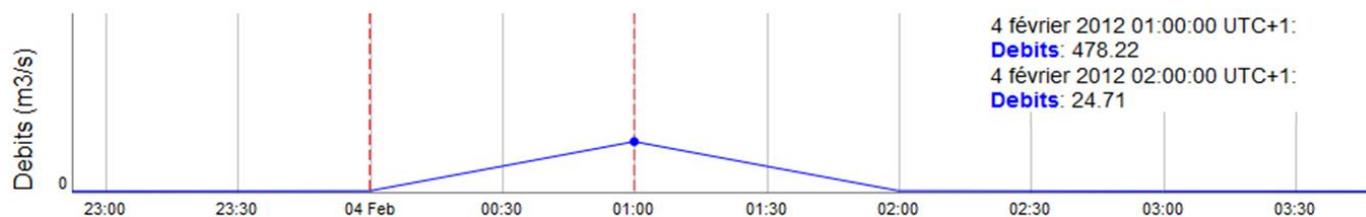


Figure 7 : Zoom sur une zone de débit constant le 4 février 2012 au matin (résultat de l'étude)

En pratique, comme précisé au début de cette partie, tous les résultats se feront sur des durées de 24h, soit un traitement effectué sur les dernières valeurs mesurées de pluies et de débits. C'est pour mieux visualiser l'impact de notre traitement que différents pas de temps ont été choisis ici. Dans le chapitre suivant les limites des méthodes implémentées ainsi que les perspectives d'améliorations sont présentées.

III. Discussions et analyse critique

Dans ce dernier chapitre trois aspects sont abordés : les limites de l'étude, les améliorations potentielles et enfin une analyse rétrospective sur le déroulé du stage avec entre autre une partie sur la gestion du temps.

III.1. Limites de l'étude

Une étude plus poussée permettrait d'améliorer la complexité des méthodes implémentées, l'étendue des données sur lesquelles les méthodes sont testées et ainsi d'accroître la capacité d'analyse des résultats. En effet, l'ensemble de l'étude s'est déroulée sur de petits échantillons qui, même s'ils ont été minutieusement choisis, restent trop restreints pour anticiper le potentiel de généralisation.

Une première limite est donc le nombre de vérifications qui sont nécessaires avant de pouvoir étendre les méthodes mises en place. Il faudrait commencer par déployer ces méthodes sur plusieurs bassins versants. Une fois que l'échantillon est considéré comme suffisamment varié, au regard des critères ci-dessus, il restera à étendre l'étude sur un grand nombre de bassins pour obtenir des résultats applicables à d'autres bassins caractéristiques. Les critères de sélection de ces bassins versants pourraient entre autre être :

- les types de bassins variés : en plaine, en montagne, proche de la Méditerranée avec des crues éclairs, avec et sans aménagements humains de type barrages ;
- la quantité d'informations historiques et validées disponibles : sur les bassins avec de très longues chroniques historiques, les détections d'erreurs relatives à des données statistiques seront potentiellement plus finement détectées, sous réserve que les données soient justes ;
- la taille variée des bassins : voir la ou les différences entre des petits bassins et des grands bassins qui sont finalement plusieurs petits bassins imbriqués, etc.

Une autre limite importante découle du choix fait d'écarter la prise en compte des changements de courbe de tarage. En pratique ce sont les données de hauteurs d'eau qui sont relevées par les appareils de mesure et qui sont interprétées en débit grâce à une courbe de tarage. Une courbe de tarage est un graphique mettant en correspondance les hauteurs d'eau et les débits. Elle est définie empiriquement et s'affine au fur et à mesure que le nombre de jaugeages effectués, sur le lieu de mesure, augmente. Cette relation entre hauteur et débit est directement liée au profil en travers de la rivière à la station de mesure. Or, ce profil varie : en effet, en profondeur et en largeur, une section de rivière évolue dans le temps. Même avec un profil de rivière fixé par un seuil ou tout autre ouvrage en béton armé, des dépôts ou embâcles peuvent gêner. Ainsi les courbes de tarage évoluent dans le temps. Lors des détections relatives à des calculs de taux d'accroissement, il est possible de déceler des erreurs qui sont de fausses alertes. Il s'agit en réalité de variations de la courbe de tarage et non d'erreurs de mesures.

Un autre type de fausse alerte peut être lié au nombre de pas de temps. Plusieurs méthodes de détections sont en effet liées à une notion temporelle : par exemple, la durée à partir de laquelle on considère que le caractère constant des précipitations devient aberrant a été fixée de manière arbitraire. Il a été choisi dans le cadre de l'étude, qu'au-delà de 10h, une pluie constante soit détectée comme une erreur. Or, des périodes de quelques jours de pluie constante ont été retenues et validées par les prévisionnistes dans les données historiques propres.

Parmi les limites de l'étude, il faut aussi éclaircir les raisons expliquant les éventuelles alertes manquées. Il s'agit des erreurs non détectées automatiquement, alors qu'elles devraient l'être puisqu'elles sont retenues comme des erreurs dans les données issues du traitement manuel. Ces alertes manquées sont issues des choix de l'étude de se concentrer sur certains types d'erreurs (cf. tableau 3).

Une autre catégorie d'erreurs non détectées correspond aux types d'erreurs qui n'ont pas été identifiés ou qui ont été écartés lors de l'étude. En effet dans la sous-partie I.2.2. "Problèmes existants", la liste des symptômes d'erreurs possibles n'est pas exhaustive. Il peut s'agir de motifs d'erreurs rares ou involontairement omis à travers les recherches bibliographiques.

III.2. Perspectives d'améliorations

En début de stage, il avait été envisagé d'évaluer l'impact de la critique de données sur les prévisions hydrologiques produites par le modèle GRP. L'idée était d'évaluer les performances du modèle en comparant les observations de débit aux prévisions obtenues, d'une part avec les données brutes, d'autre part avec les données critiquées par l'outil mis en place. Le stage a permis de mettre en avant une première ébauche d'un module de détection automatique des problèmes de données, et si l'évaluation complète des performances de ce module nécessite un travail plus approfondi, cela constitue une perspective d'amélioration future.

Les limites présentées dans la partie précédente III.1. "Limites de l'étude" sont autant de perspectives possibles à cette étude. Notamment :

- l'étude de l'impact de notre traitement automatisé dans GRP, afin d'évaluer de quelle façon l'automatisation du traitement des données proposé ici améliore ou non la qualité des prévisions
- la généralisation de l'étude à un plus grand nombre de stations hydrométriques et pluviomètres
- la prise en compte des détarages, soit en ajoutant un module pour gérer les dates de changement de courbe de tarage, soit en développant un outil de détection de ces changements de courbes
- la gestion des nombres de pas de temps à partir desquels il est choisi que les irrégularités détectées deviennent des erreurs
- des recherches ou tests supplémentaires pour mettre en place ou imaginer des méthodes de détection des erreurs qui ont été consciemment écartées et qui élargirait le champ de détection
- un élargissement des recherches concernant les différentes catégories d'erreurs et l'étude des motifs d'erreurs rares afin d'éviter les omissions involontaires

Pour illustrer simplement un exemple de toutes les améliorations possibles qui viennent d'être listées, il y a le cas des erreurs d'enregistrement dues à la fonte de la neige. Ce type d'erreur se détecte notamment lorsque des précipitations sont enregistrées entre 12h et 14h un jour ensoleillé : la neige accumulée dans le pluviomètre fond car elle est exposée aux rayons solaires. L'appareil de mesure ne transmet quasiment aucune précipitation au moment de l'événement neigeux puis le signal envoyé en retard correspond à une pluie constante le temps de la fonte. Cette information est assez caractéristique et peut assez aisément être considérée comme une erreur afin de compléter le modèle de détection développé au cours de cette étude. Concrètement, une telle information correspond à un taux d'accroissement constant entre deux instants, doublé de précipitations différentes de zéro pendant un certain temps. L'information de température, utilisée par le module neige de GRP, serait une information complémentaire permettant d'identifier une pluie suivant réellement une fonte de neige.

D'autre part, sont citées dans l'étude bibliographique d'autres méthodes de détection, permettant de détecter d'autres types d'erreurs, pourraient constituer des choix améliorations potentielles.

Enfin, au cours de l'étude, d'autres méthodes et moyens de détection de motifs d'erreurs ont émergés :

- la détection de pluies atypiques à l'aide d'une table de correspondance ou de corrélation entre les taux d'accroissement, le temps de retour de l'événement et sa durée pourrait être mises en place à partir d'études statistiques sur les historiques validés manuellement ;
- la détection de corrélations entre pluviomètres d'un même bassin versant : ces corrélations seraient reportées dans une matrice et cela permettrait notamment de détecter les pluies anormalement nulles ou anormalement localisées lorsqu'un pluviomètre n'enregistre pas de pluie alors que ses voisins proches en indiquent ou à l'inverse quand un pluviomètre est seul à donner une information de précipitation ;
- l'affichage des taux d'accroissement de différentes crues passées associées à leurs périodes de retour donnerait une bonne vision de l'événement en cours.

Il est également à noter que le domaine de la détection automatique d'erreurs de mesure pour l'amélioration des prévisions reste encore peu approfondi aujourd'hui.

Un autre modèle de détection d'erreur (Kirk et McCuen, 2008) a attiré notre attention lors de notre étude car il s'intéressait directement aux détections d'erreurs dans les données hydrologiques. Et plus précisément des données hydrologiques en vue d'une utilisation pour de la prévision. Le modèle développé par Kirk et McCuen repose sur une détection de valeurs aberrantes grâce à de nombreuses variables. La complexité de l'implémentation a limité nos possibilités de mise en place de ces détections d'erreurs. Il a semblé important de citer cet article en perspective en vue d'études futures sur des sujets similaires.

III.3. Analyse de l'organisation de l'étude

Cette partie présente une analyse critique relative au déroulement du stage. Dans l'ensemble les conditions de réalisation ont été relativement idéales, quelques aspects auraient pu être améliorés notamment certains aspects de formation antérieure au stage.

Le stage a débuté avec une formation au langage informatique R, une première approche du sujet par la lecture de nombreux articles scientifiques et de quelques guides techniques, le tout parallèlement aux échanges avec les SPC pour choisir et récupérer les données nécessaires à l'étude. Cette phase a donc été une importante période d'adaptation, notamment à des rythmes de travail et de vie nouveaux. Le temps nécessaire à la prise en main des outils et du sujet aurait pu être plus court avec des connaissances et des formations ciblées avant le début du stage. Toutefois, ce fut une occasion d'apprendre, ce qui est aussi un des intérêts de ce type de stage et cela a impliqué une organisation minutieuse du temps restant à la suite de cette période.

Un aspect de gestion du temps qui est assez difficilement anticipable est la prévision du temps nécessaire à l'implémentation des méthodes ainsi que celle du temps de débogage des implémentations. Cette gêne dans la gestion du temps est partagée au sein de l'équipe, il est donc difficile d'estimer comment mieux faire.

Le stage a donc permis de répondre à la problématique fixée, tout en s'adaptant au temps imparti à l'étude, et cela servira de préambule à des analyses plus poussées de méthodes d'aide à la critique de données.

Conclusion

Le stage a finalement permis d'atteindre l'objectif initial fixé qui était de tester des méthodes d'aide à la critique automatique des données brutes reçues par les modèles de prévision des crues. En effet, une automatisation de la critique des données en temps réel a bien été mise en place.

Les résultats du stage pourront être exploités par la suite : les fiches qualités produites pourront servir d'aide à la prévision en temps réel d'une part, et l'ensemble des perspectives d'améliorations permettront d'envisager d'autres méthodes d'automatisation de la critique des données d'autre part.

Pour conclure, le domaine de l'automatisation de la détection d'erreurs présente encore un potentiel pour l'amélioration des prévisions. Cette étude aura donc permis de mettre en place quelques méthodes d'automatisation de détection d'erreurs et d'éclaircir un peu ce domaine qui reste encore à approfondir.

Dans la construction du début de ma vie professionnelle, ce stage m'a apporté une expérience particulièrement enrichissante.

D'une part, j'ai pu m'intégrer dans un organisme de recherche public qui sera probablement une catégorie d'acteur avec laquelle je serai amenée à interagir dans ma vie professionnelle future. J'ai pu y constater des éléments particuliers de gouvernance, de pilotage et de management spécifiques au domaine de la recherche publique. D'autre part, j'ai pu développer un ensemble de compétences liées à la résolution de problématiques de recherche où par essence la solution et le moyen d'y parvenir sont partiellement maîtrisés au démarrage du travail. Cela a donc un impact important sur la gestion du temps d'une activité et une forte valeur ajoutée à long terme.

Enfin, j'ai pu très prosaïquement développer des compétences en algorithmique qui sont un bagage indispensable à tout membre de la fonction publique et de l'industrie du 21ème siècle. Cette expérience m'aidera ainsi à effectuer les choix à venir avec une meilleure conscience et connaissance des enjeux.

Bibliographie

Berthet, L., 2010. "Prévision des crues au pas de temps horaire : pour une meilleure assimilation de l'information de débit dans un modèle hydrologique." Thèse de Doctorat, Cemagref (Antony), AgroParisTech (Paris), 603 pp.

Estevez, J., P. Gavilán et J.V. Giráldez. (2011). "Guidelines on validation procedures for meteorological data from automatic weather stations." *Journal of Hydrology* volume 402 (1-2) : 144-154.

Furusho, C., C. Perrin, J. Viatgé, R. Lamblin et V. Andréassian, 2016. Synergies entre acteurs opérationnels et scientifiques au service de l'amélioration de la prévision des crues. *La Houille Blanche*(4): 5-10.

Hill, D. J. et B. S. Minsker (2010). "Anomaly detection in streaming environmental sensor data: A data-driven modeling approach." *Environmental Modelling & Software* volume 25 : 1014-1022.

Kirk, A. J. et R. H. McCuen (2008). "Outlier detection in multivariate hydrologic data." *Journal of Hydrologic Engineering* volume 13 (7) : 641-646.

Liu, S. L., Li Yuan, Pauwels Valentijn R. N., Walker Jeffrey P. (2018). "Impact of Rain Gauge Quality Control and Interpolation on Streamflow Simulation: An Application to the Warwick Catchment, Australia." *Frontiers in Earth Science* volume 5, DOI=10.3389/feart.2017.00114, 2296-6463.

Organisation Météorologique Mondiale (1996). "Guide des Pratiques Hydrologiques."

Oudin, L., 2004. Recherche d'un modèle d'évapotranspiration potentielle pertinent comme entrée d'un modèle pluie-débit global. Thèse de Doctorat, ENGREF (Paris) / Cemagref (Antony), 495 pp.

Moulin, L. et R. Thepot (1999). "The CRISTAL system of floods and low water management in the Loire basin: the modernization." *Houille Blanche-Revue Internationale De L Eau* (3-4) : 77-81.

Perret C., S. Poligot-Ptisch et R. Puechberty (2017). "Charte Qualité de l'Hydrométrie - Guide de bonnes pratiques", 84 pp.

R Core Team (2018). "R : A language and environment for statistical computing." Version : 5.3.1

Robertson, D. E., Bennett, J. C., and Wanga, Q. J. (2015). "A strategy for quality controlling hourly rainfall observations and its impact on hourly streamflow simulations." in MODSIM2015, 21st International Congress on Modelling and Simulation, eds T.Weber, M. J.McPhee, and R. S. Anderssen (Gold Coast, QLD: Modelling and Simulation Society of Australia and New Zealand), 2110–2116.

Tangara, M., 2005. Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global. Thèse de Doctorat, Cemagref Antony, EPHE, Paris, 374 pp.

Xu, S., Lu, B., Bell, N., et Nixon, M.N. (2017). Outlier Detection in Dynamic Systems with Multiple Operating Points and Application to Improve Industrial Flare Monitoring. *Processes* volume 5 (2) : 28.

Annexes

→ Annexe 1 - Décret n°2012-209 du 1 février 2012 relatif à l'Irstea¹¹

Article 4

L'article R. 832-2 est remplacé par les dispositions suivantes :

« Art. R. 832-2.-L'institut a pour missions :

« 1° D'entreprendre, réaliser, coordonner et soutenir, à moyen et long terme, à son initiative ou à la demande de l'Etat, tous travaux de recherche scientifique et technologique dans les domaines de l'aménagement et de la gestion durables des territoires, en particulier agricoles et naturels, et de leurs ressources.

« Ces travaux de recherche portent notamment sur :

« a) Le traitement des interactions entre les milieux naturels, les zones agricoles et urbaines ;

« b) La gestion de la ressource en eau, l'amélioration de la qualité des eaux et des milieux aquatiques autres que marins ;

« c) La prévention, la prévision et l'atténuation des risques liés à la pollution des écosystèmes, au cycle de l'eau et à la sécurité des ouvrages hydrauliques ;

« d) La gestion et la valorisation des ressources naturelles et de la biodiversité aquatique et forestière ;

« e) La connaissance de l'environnement par l'observation dans ses domaines de compétence ;

« f) La conception et le développement de procédés et technologies liés en particulier au traitement des déchets et rejets, aux activités agricoles, forestières et agroalimentaires et à l'aménagement des territoires ;

« 2° De produire et mobiliser, à partir des résultats de ses travaux, les connaissances scientifiques et technologiques nécessaires aux politiques publiques et de promouvoir la normalisation ;

« 3° De conduire des expertises scientifiques et techniques ;

« 4° De contribuer, dans les conditions déterminées par le [code de la recherche](#), à la valorisation des résultats de ses travaux ;

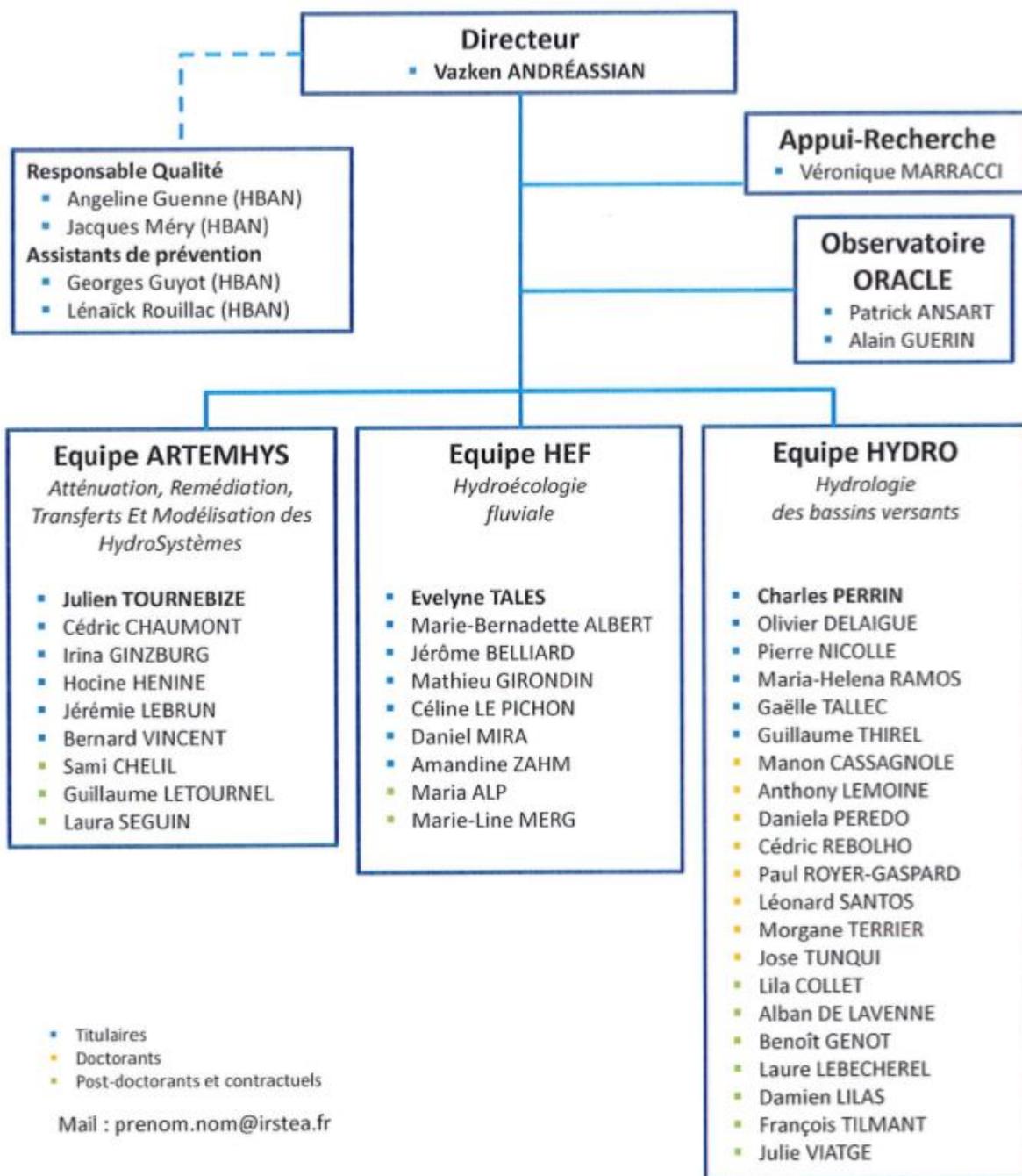
« 5° De favoriser la publication de tous travaux et études liés à ses activités et de contribuer à la diffusion des connaissances scientifiques et techniques ;

« 6° D'apporter son concours à l'enseignement supérieur et à la formation à la recherche et par la recherche dans ses domaines de compétences.»

¹¹ <https://www.legifrance.gouv.fr>

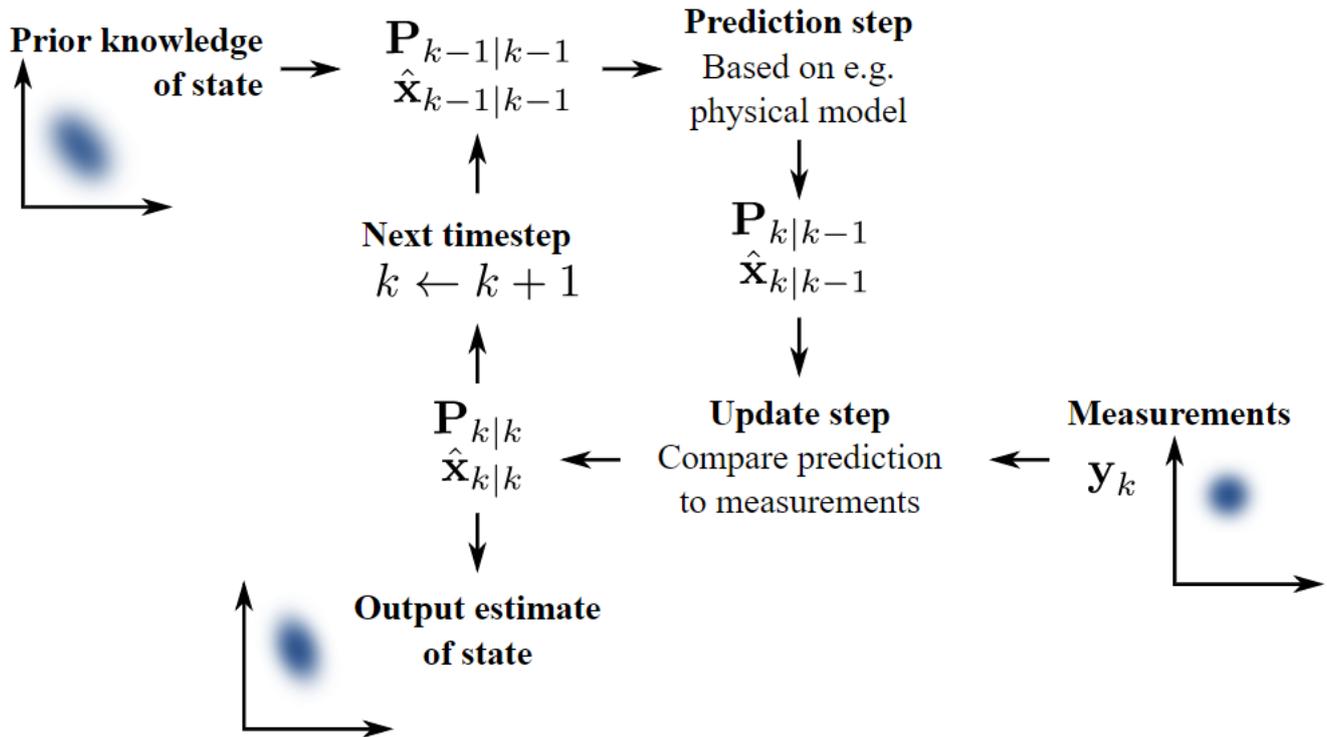


Organigramme Unité de recherche HYCAR
Antony- Hydrosystèmes Continentaux Anthropisés
Ressources, Risques, Restauration



¹² <http://www.irstea.fr/hycar>

→ Annexe 3 : Concept de base du filtre de Kalman (schéma personnel de Petteri Aimonen)



Le filtre Kalman permet d'estimer les différents états d'un système à partir d'une série de mesures incomplètes ou erronées. Ce filtre renvoie une réponse impulsionnelle infinie. Et associé avec le Test du temps linéaire exact élagué (PELT voir annexe suivante) il est utilisé par Xu et al en 2017 afin de développer une algorithmme d'erreurs dynamique.

→ Annexe 4 - Schéma explicatif du PELT méthode (source article de Xu et al. 2017)

