Amélioration de la précision d'un système d'alerte de crue par télédétection à l'aide d'une méthode de prétraitement multi-objectifs pour la détection et l'élimination des défauts de signal

Comptes Rendus Géosciences 352 (2020) 73-86.

Hossein Bonakdari ; Amir Hossein Zaji ; Keyvan Soltani ; Bahram Gharabaghi

(traduction post-éditée par N. Bacaër, suggestions d'amélioration : nicolas.bacaer@ird.fr)



Résumé

L'un des principaux objectifs de la gestion des bassins versants est de surveiller et de prévoir de manière proactive les niveaux d'eau des crues afin de fournir une alerte rapide pour les plans d'évacuation en temps opportun et de sauver des vies. L'un des moyens les plus économiques d'atteindre cet objectif consiste à utiliser des signaux satellites de télédétection. Des études antérieures ont indiqué qu'un capteur AMSR (Advanced Microwave Scanning Radiometer) peut être utilisé pour la surveillance du niveau de l'eau des rivières, combiné à quelques jauges hydrométriques in situ pour la collecte de données de terrain. Cependant, les signaux spatiaux sont influencés par de nombreux facteurs naturels induisant des erreurs, tels que la poussière et la couverture nuageuse. Par conséquent, une méthode hybride est proposée, qui comprend un modèle d'optimisation d'essaim de particules multi-objectifs, un algorithme de classification d'arbre de décision, le T^2 de Hotelling pour la détection des valeurs aberrantes et un modèle de régression pour identifier et remplacer les signaux spatiaux inexacts. Cette méthode hybride complexe sera désignée, dans cette étude, par l'acronyme (OCOR). Dans la première phase de cette méthode hybride, les signaux aberrants sont détectés et éliminés de l'ensemble de données, et dans la deuxième phase, les signaux éliminés ainsi que les signaux perdus en raison de problèmes techniques de satellite sont estimés par l'étalonnage des données de vérité terrain à l'aide de l'hydrométrie in situ. stations. Les deux études de cas des rivières White et Willamette démontrent la performance de l'OCOR dans des situations pratiques.

1. Introduction

Malgré le fait que le changement climatique a augmenté la gravité et la fréquence des inondations et que les inondations ont tué plus de personnes que toutes les autres catastrophes naturelles combinées [De Groeve et Riva 2009a], d'autres études sont nécessaires pour fournir des systèmes d'alerte rapide rentables pour une évacuation rapide. plans dans un contexte opérationnel [Zaji et al. 2018]. Des études récentes ont utilisé avec succès des modèles d'intelligence artificielle pour résoudre des problèmes de ressources en eau non linéaires [Akhbari et al. 2017, Bonakdari et al. 2020, Ebtehaj et al. 2018, 2019, Espinoza-Villar et al. 2018, Gholami et al. 2019, Zaji et Bonakdari 2019].

He et coll. [2014] ont comparé les performances de trois des techniques basées sur les données les plus populaires, à savoir les réseaux neuronaux artificiels (ANN), les systèmes d'inférence neuro-floue adaptatifs (ANFIS) et les machines vectorielles de soutien (SVM) pour prédire le débit des rivières. Les auteurs ont conclu que la méthode SVM était plus efficace pour simuler des rivières dans une région montagneuse semiaride en Chine. Chen et coll. [2015] ont appliqué un réseau de neurones hybrides (HNN) basé sur la reconnaissance de formes floues pour simuler la rivière Altamaha. Les auteurs ont comparé trois algorithmes d'optimisation différents pour déterminer les paramètres HNN les plus appropriés et ont conclu que l'évolution différentielle (DE) et l'optimisation des colonies de fourmis (ACO) surpassaient l'algorithme Artificial Bee Colony (ABC).

Darras et coll. [2017] ont utilisé des réseaux de neurones à effet direct et récurrents pour prévoir les crues éclair. Les auteurs ont déterminé que le réseau neuronal récurrent fonctionnait mieux avec un délai plus long dans la prédiction des crues éclair et que le réseau neuronal par anticipation fonctionnait mieux dans la simulation des crues éclair dans la prévision à court terme. Yaseen et coll. [2017] a combiné l'algorithme de luciole avec une technique de calcul neuro-flou adaptatif pour proposer un nouveau modèle d'estimation du flux de flux. Young et coll. [2017] a présenté un modèle hybride formé d'un système de modélisation hydrologique basé sur la physique et de SVM et a déclaré que ce modèle a montré de bonnes performances dans la simulation du ruissellement des pluies pendant les événements de typhon.

Alizadeh et coll. [2017] a combiné la transformée en ondelettes avec ANN afin de prédire le ruissellement des précipitations dans la rivière Tolt. Les résultats ont montré que ce modèle faisait des prévisions appropriées pour un et deux mois à venir. Zeynoddin et coll. [2018] a proposé un nouveau modèle hybride pour prévoir les précipitations d'un bassin à climat tropical en Malaisie. L'un des principaux objectifs de l'utilisation des informations satellitaires est de surveiller et de prévoir les crues dans les bassins non jaugés.

Le problème avec la prévision des crues fluviales est que tous les modèles basés sur des données utilisent uniquement des informations in situ. Par conséquent, en raison de la relation directe entre les données sur les crues et les jauges in situ, les régions dépourvues de stations de mesure ne peuvent fournir aucune information sur la rivière. Néanmoins, la plupart des rivières souffrent d'un manque de stations in situ.

De plus, le nombre de stations in situ existantes dans certains pays est limité [Calmant et Seyler 2006, Khan et al. 2012, Shiklomanov et al. 2002, Sivapalan et al. 2003, Stokstad et al. 1999]. Par conséquent, l'étude et la prévision des inondations ne sont pas possibles sur des régions considérables de la terre. Par conséquent, la prévision dans les bassins non jaugés (PUB) est récemment apparue comme un sujet de très grand intérêt dans la prévision des crues [Salvia et al. 2011]. Les chercheurs dans ce domaine utilisent principalement les informations satellitaires telles que les mesures spatiales des précipitations pour simuler les circonstances hydrologiques [Brakenridge et al. 2007, Jiang et al. 2014, Khan et al.2011, Su et al. 2008, Temimi et al. 2011, 2007].

Un moyen précis de simuler le débit fluvial en appliquant des informations satellitaires consiste à utiliser des capteurs micro-ondes passifs de l'AMSR pour le système d'observation de la Terre (AMSR-E) [Brakenridge et al. 2007]. Les études dans ce domaine ont débuté au Dartmouth Flood Observatory (DFO), puis au Joint Research Centre (JRC). Les capteurs AMSR-E observent toute la surface de la Terre et détectent les zones fluviales humides en calculant les différences de températures de luminosité du sol et de l'eau. Les données brutes sont justifiées par le Global Flood Detection System (GFDS), qui est accessible à http://www.gdacs.org/flooddetection .

Pour prédire les séries chronologiques des niveaux d'eau des rivières à l'aide de données satellitaires, Zaji et al. [2019a] a introduit une nouvelle approche pour minimiser les erreurs horizontales et verticales. Après cela, Zaji et al. [2019b] a développé un nouveau prédicteur de séries chronologiques de régression polynomiale évolutive (EPR-T) et l'a utilisé pour prédire le débit des rivières à l'aide d'informations satellitaires. Bonakdari et coll. [2019] a développé deux extensions de la méthode Markov Chain (MC), à savoir Online-Markov Chain (O-MC) et Extreme Online-Markov Chain (EO-MC) pour améliorer les performances de prévision du débit des rivières à l'aide d'informations satellitaires.

Les signaux reçus par les satellites sont affectés par diverses conditions météorologiques [Kugler et De Groeve 2007]. Par exemple, les intempéries modifient les signaux d'émission à la fois de la terre et de l'eau. L'impact des conditions météorologiques extrêmes est plus critique lorsque la zone entourée d'eau est plus petite, c'est pourquoi la précision du signal spatial est plus faible lorsque le débit de la rivière est faible [Khan et al. 2012]. Ainsi, afin d'améliorer les performances et la fiabilité des données satellitaires, Zaji et al. [2019a] a proposé une approche basée sur des méthodes d'essais et d'erreurs, de classification et de détection des valeurs aberrantes. Les auteurs ont conclu que la méthode développée pourrait augmenter considérablement la précision des données satellitaires.

L'objectif principal de la présente étude est d'introduire une méthode hybride qui comprend un modèle d'optimisation d'essaim de particules multi-objectifs, un algorithme de classification d'arbre de décision, la

détection des valeurs aberrantes avec le T^2 de Hotelling et un modèle de régression (OCOR) pour évaluer les signaux satellites bruts pour obtenir signaux continus précis. Avec cette méthode, les signaux aberrants reçus par des conditions météorologiques extrêmes ou d'autres situations environnementales inconnues sont initialement détectés et éliminés. Ensuite, les signaux supprimés et ceux manqués en raison de problèmes techniques de satellite sont prédits à l'aide d'une méthode de régression.

Des études antérieures se sont principalement concentrées sur la simulation du débit fluvial à l'aide de signaux spatiaux [Frasson et al. 2017]. Cependant, lorsque l'objectif est d'utiliser les informations historiques des capteurs satellitaires pour prévoir le débit futur de la rivière, un ensemble de données spatiales continues sans lacune est nécessaire [Weigend 2018]. Ainsi, la deuxième étape consiste à évaluer un ensemble de données continu de signaux spatiaux sans échantillons manqués. Dans la phase d'apprentissage de la méthode OCOR, les mesures in-situ sont nécessaires afin de déterminer l'une des fonctions objectives de l'algorithme d'optimisation multi-objectifs et de développer la méthode de régression pour remplacer les échantillons bruyants détectés.

2. Signaux spatiaux

Des capteurs hyperfréquences passifs AMSR-E et AMSR2 sont utilisés dans la présente étude pour recueillir des signaux spatiaux [Brakenridge et al. 2007]. Après cela, les signaux obtenus sont utilisés pour mesurer indirectement le débit de la rivière. Afin d'obtenir des observations plus stables, la méthode de l'orbite descendante a été choisie dans cette étude [Kugler et De Groeve 2007, Tekeli et Fouli 2017], et la surface de la Terre a été mesurée au moins une fois par jour.

Les mesures par micro-ondes, qui sont très sensibles à l'eau, sont largement utilisées dans le domaine scientifique de l'estimation de l'humidité du sol [Njoku et al. 2003, Schmugge et al. 1980, Srivastava et al. 2017, Theis et al. 1982, Ulaby et al. 1978, Wang et al. 1982]. En utilisant les signaux micro-ondes, le débit fluvial est estimé en fonction des différences entre l'émission thermique des surfaces sèches et humides. La température de luminosité de la zone humide considérée est appelée Mesure (M), qui est une approximation de l'eau de la rivière et les informations reçues sont calibrées en appliquant la température de luminosité de la zone sèche environnante (C) pour obtenir le signal spatial AMSR final (S) comme suit:

$$S = \frac{C}{M} \,. \tag{1}$$

La température de luminosité du sol est plus élevée que la température de la zone humide [Kugler et De Groeve 2007], de sorte que S est faible lorsque le débit de la rivière est faible et vice versa [Birkinshaw et al. 2010]. Selon Bjerklie et al. [2004] et Brakenridge et al. [2012], dans la plupart des rivières, le débit a une meilleure corrélation avec la largeur d'écoulement qu'avec la vitesse d'écoulement. Ainsi, la largeur de l'écoulement, qui est calculée à l'aide des informations satellitaires, pourrait être considérée comme une approximation robuste du débit fluvial.

3. Zone d'étude

Dans ce travail, deux rivières sont examinées à titre d'études de cas, à savoir la rivière White et la rivière Willamette. Les propriétés de ces deux rivières sont présentées dans le tableau 1 . Les deux sont situés aux États-Unis. Les sources des rivières White et Willamette sont les montagnes Boston et la jonction de la rivière Middle Fork Willamette avec la rivière Coast Fork Willamette, respectivement. La figure 1 présente un aperçu schématique des stations in situ et des emplacements de mesure spatiaux pour les rivières White et Willamette. Sur cette figure, la zone M , la zone C et la station de mesure in situ sont désignées respectivement en bleu, gris et noir.



Figure 1. Vue d'ensemble schématique des emplacements (a) de White River et (b) de Willamette River (images satellites Landsat en composite de fausses couleurs (7,5,2)).

Tableau 1. Propriétés des rivières White et Willamette

Nom de la rivière	Site ID	Station ID	Coordonnées spatiales du site Lat./Long. (jj)	La station in situ coordonne Lat./Long. (jj)	Débit moyen (m ³ / s)
Blanc, Newberry	524	03360500	38.91/-87.07	38.92/-87.011	12137
Willamette	925	14191000	45.18/-123.01	44.94/-123.042	18928

Les informations de la station in situ ont été recueillies auprès de l'USGS (United States Geological Survey) (http://www.usgs.gov), et les informations spatiales ont été obtenues auprès du GFDS (http://www.gdacs.org). Les périodes examinées pour les rivières White et Willamette vont de 2009 à 2011 et de 2002 à 2016, respectivement.

4. Modèles numériques

L'objectif principal de la méthode proposée dans l'étude actuelle est de réévaluer les ensembles de données in situ et spatiales en détectant et en remplaçant leurs erreurs. Comme mentionné dans la littérature, de nombreuses études ont porté sur l'étalonnage des informations spatiales dans le cadre de mesures situ. Cependant, la mauvaise performance des capteurs satellites, dans certains cas, peut conduire à des signaux inexacts. De plus, les jauges satellites ou in situ ne sont pas toujours accessibles. Cependant, la prévision du débit fluvial à l'aide de données spatiales nécessite un ensemble de données sans lacunes, la technique devant remplacer les données éliminées et manquées par des valeurs appropriées.

La méthode OCOR est introduite dans cette étude pour atteindre ces objectifs de manière plus efficace (comparée aux méthodes existantes). La méthode OCOR implique une approche de détection des valeurs aberrantes combinée à un algorithme de classification pour détecter les valeurs aberrantes et les signaux inexacts. L'OCOR utilise par la suite un modèle de régression pour évaluer les informations éliminées et manquées. De plus, le modèle est justifié en utilisant un algorithme d'optimisation multi-objectifs.

La détection des valeurs aberrantes est la science de la détection des échantillons anormaux en considérant les quantités d'autres échantillons de l'ensemble de données [Mitra et al. 2009]. Un modèle de détection des valeurs aberrantes tente d'identifier les valeurs aberrantes globales et contextuelles des signaux spatiaux. Les valeurs aberrantes globales se produisent lorsqu'un échantillon s'écarte considérablement de l'ensemble de données, et les valeurs aberrantes contextuelles se produisent lorsqu'un échantillon s'écarte considérablement de l'ensemble de données, et les valeurs aberrantes contextuelles se produisent lorsqu'un échantillon s'écarte considérablement de l'ensemble de des échantillons voisins au cours de cette saison [Han et al. 2011].

La méthode de détection des valeurs aberrantes avec le T^2 de Hotelling [Hotelling 1992] est utilisée dans la présente étude. Le T^2 de Hotelling a été appliqué comme algorithme de contrôle de la qualité dans de nombreuses études [Mason et Young 2002, Shabbak et al. 2011, Sullivan et Woodall 1996]. Dans le T^2 de Hotelling, l'ensemble de données est transféré vers une hypersphère de rayon α en utilisant la rotation et la normalisation. Suite aux transformations mentionnées, les échantillons qui satisfont à la condition $T^2 \leq \alpha^2$ sont considérés comme normaux et les échantillons restants sont considérés comme anormaux. Par conséquent, α est un critère strict de cette méthode.

Dans la présente étude, sous la direction de détection des valeurs aberrantes est utilisé, qui, conjointement avec le T^2 de Hotelling pour la détection des valeurs aberrantes, nécessite un algorithme de classification [Han et al. 2011]. Dans la détection supervisée des valeurs aberrantes, l'algorithme de classification marque les échantillons comme des experts, après les échantillons dont les échantillons qui agrandissent les classes sont détectés comme des valeurs aberrantes et éliminés de l'ensemble de données.

La méthode OCOR utilise l'algorithme d'arbre de décision, car c'est l'un des algorithmes de classification les plus simples et les plus rapides [Quinlan et al. 1986, 1987, Quinlan et Rivest 1989]. Cet algorithme a une forme arborescente et comprend trois parties. La première partie implique des nœuds internes qui prennent des décisions sur les attributs; la deuxième partie contient des branches qui représentent les décisions prises dans les nœuds internes; et la troisième partie, ou le niveau final des nœuds d'arbre, comprend des nœuds feuilles qui présentent les étiquettes de classe. L'arbre de décision utilisé a été introduit par Breiman et al. [1984] comme algorithme CART pour mesurer l'impureté de l'ensemble de données d'apprentissage comme suit:

$$GI(ext{dataset}) = 1 - \sum_{i=1}^m p_i^2 \quad ext{avec} \quad p_i = rac{|C_{i, ext{dataset}|}}{| ext{dataset}|}$$

où C_i représente l'étiquette de classe p_i est la probabilité que l'échantillon considéré soit dans la classe C_i , et m le nombre de classes considérées. GI conduit à une décision binaire dans chaque nœud interne. La précision de l'algorithme de classification de l'arbre de décision est justifiée à l'aide du paramètre Taille minimale des parents (MPS). MPS détermine le nombre minimum d'échantillons autorisés dans chaque nœud et sert de critère de terminaison. Un MPS plus faible conduit à un arbre de décision plus grand et plus précis et vice versa.

Les paramètres α (critère strict du T^2 de Hotelling), m (nombre de classes) et MPS (critère de terminaison d'arbre de décision) doivent être déterminés pour atteindre un modèle optimal. La position optimale est un modèle avec le plus petit nombre d'échantillons aberrants et le coefficient de détermination le plus élevé (\mathbb{R}^2) entre les signaux satellites et les mesures de décharge in situ. Par conséquent, deux objectifs doivent être atteints et un algorithme d'optimisation multi-objectifs est nécessaire.

Dans la présente étude, l'optimisation des essaims de particules multi-objectifs (MOPSO) [Coello et Lechuga 2002] est utilisée. MOPSO est une combinaison d'optimisation de l'essaim de particules (PSO) [Kennedy et al. 2001 , Shaghaghi et al. 2017 , Zaji et al. 2015] et l'algorithme génétique de tri non dominé II (NSGA-II) [Deb et al. 2000]. L'objectif de MOPSO est de déterminer les valeurs α , *m* et MPS les plus appropriées pour atteindre un modèle qui détecte le plus petit nombre de valeurs aberrantes avec le R^2 le plus élevé .

La dernière étape de la méthode OCOR est la procédure de régression. Ici, les échantillons satellitaires qui ont été éliminés lors des étapes précédentes sont remplacés par des valeurs précises. Cette procédure implique des mesures de débit in situ comme échantillons d'entrée du modèle de régression et des signaux satellitaires comme échantillons de sortie. Après l'élimination, le modèle de régression est entraîné en utilisant les échantillons restants, après quoi il modélise les échantillons éliminés.

La méthode OCOR tente de détecter et d'éliminer le moins d'échantillons inexacts et bruyants que possible en utilisant une combinaison d'algorithmes de détection des valeurs aberrantes, de classification et d'optimisation multi-objectifs, puis de remplacer les échantillons éliminés à l'aide d'une méthode de régression.

L'application de ce modèle avant d'utiliser les informations satellites sur les rivières est nécessaire pour toute application ultérieure de celles-ci. Cependant, cette version de la méthode proposée présente certaines

limites. Dans un premier temps, dans cette méthode, les mesures de débit in-situ sont utilisées pour trouver le R^2 le plus élevé , qui est l'un des objectifs de l'algorithme d'optimisation et également pour remplacer les échantillons éliminés, qui est la dernière phase de la méthode OCOR.

Par conséquent, cette méthode n'est pas applicable aux bassins non jaugés. Deuxièmement, ce modèle n'a été testé que sur les informations des capteurs satellites AMSR. La taille de l'empreinte de l'AMSR est vraiment grande (environ $8 \times 12 \text{ km}^2$). Les informations obtenues par ce capteur sont plus adaptées aux grandes rivières. Troisièmement, afin d'utiliser la méthode de prétraitement proposée sur d'autres capteurs satellites, certaines justifications devraient être apportées à ce modèle. Par exemple, malgré une résolution élevée du spectroradiomètre imageur à résolution modérée (MODIS), les informations collectées sont plus sensibles aux conditions météorologiques nuageuses et ces images doivent être éliminées de l'ensemble de données séparément.

5. Résultats et discussion

Comme mentionné précédemment, les informations satellitaires constituent l'une des alternatives optimales à la mesure du débit fluvial. Pour que les signaux satellites soient utilisables, ils doivent subir une phase de prétraitement. La méthode OCOR comporte deux phases: le prétraitement et l'étalonnage. Les signaux satellites présentent deux inconvénients majeurs. Le premier est l'inexactitude due à de mauvaises conditions environnementales, et le second est des signaux manqués en raison de l'inaccessibilité au satellite.

La figure 2 présente les mesures de débit in situ et les signaux satellites pour les deux rivières considérées. Ce chiffre indique une bonne corrélation entre les signaux satellites et les mesures de décharge in situ. Cependant, il est évident que de nombreuses périodes ne contiennent aucune information et qu'il y a des échantillons manqués. Pour la rivière White, outre les signaux satellites manqués, il y a une période pendant laquelle les mesures in situ ont été perdues.

Selon la section de la zone d'étude, dans la présente étude, trois ans d'informations quotidiennes ont été utilisées pour la rivière White, et 15 ans d'informations quotidiennes ont été utilisées pour la rivière Willamette. Il convient de noter que la sélection de la période la plus appropriée pour chaque étude de cas doit être effectuée en fonction de la complexité du problème considéré. Par exemple, selon la figure 2, pour la rivière Willamette, les signaux satellites ont été manqués pendant une longue période (les zones noires). Dans ce cas, des périodes plus longues de mesures in situ et de signaux satellites sont nécessaires pour évaluer un modèle fiable.

D'autre part, la sélection d'une période de temps plus longue entraîne également plus de temps de calcul.Dans ce cas, en utilisant un MacBook Pro avec un processeur Intel Core i7 à 3,1 GHz et une mémoire de 16 Go, la durée de fonctionnement du modèle Willamette River prend environ 10 h. Cependant, avec le même ordinateur, la durée de fonctionnement du modèle White River prend environ 2 h.

L'un des objectifs les plus critiques des études sur le débit des rivières est la prévision. La prévision du débit fluvial par n'importe quelle méthode nécessite un ensemble de données continu sans échantillons manqués. Le but de la phase de prétraitement dans la méthode OCOR est d'identifier les signaux satellites inexacts et de les éliminer de l'ensemble de données. Le but de la phase d'étalonnage dans la méthode OCOR est de réévaluer les valeurs aberrantes exclues et les signaux manqués afin d'obtenir un ensemble de données continu et précis à utiliser comme série chronologique dans la prévision du débit des rivières à l'aide d'informations satellitaires.



Figure 2. Signaux satellites et mesures de débit in situ pour (a) White River et (b) Willamette River.

Un modèle de détection des valeurs aberrantes optimal atteindra R^2 maximum avec le moins de valeurs aberrantes éliminées. Par conséquent, il y a deux objectifs, et ce problème ne peut être résolu à l'aide d'algorithmes d'optimisation simples. Dans la présente étude, MOPSO a été utilisé pour ajuster trois paramètres, à savoir, MPS, α et m, afin d'obtenir le meilleur modèle. Dans le modèle actuel, la MPS variait de 1 à 500, tandis que α et m variaient de 1 à 10. Chaque étude de cas a subi 100 itérations et le nombre de population considéré était de 300. Avec les méthodes d'optimisation multi-objectifs, le meilleur modèle est sélectionné en utilisant le concept de domination. Selon la définition, j = [Objectif_{1 j},..., Objectif_{nj}] domine i = [Objectif_{1 i},..., Objectif_{ni}] si tous les paramètres de j sont inférieurs ou égaux aux paramètres de i et qu'aucun des paramètres de j n'est inférieur à ceux de i.

Dans la procédure MOPSO, les échantillons non dominés sont enregistrés dans un référentiel. Le référentiel de modèles actuel peut sauvegarder 100 échantillons non dominés. Une fois ce nombre d'échantillons atteint, le meilleur modèle est sélectionné manuellement. La figure 3 illustre les échantillons du dépôt des rivières White et Willamette sous forme de cercles rouges. Sur cette figure, les meilleurs modèles sont désignés par des losanges bleus. Ces modèles représentent évidemment un bon équilibre entre un R^2 élevé et un faible nombre d'échantillons éliminés.





Le modèle le plus approprié de la rivière White avait des MPS de 3, α de 1,49 et *m* de 9, tandis que le meilleur modèle de la rivière Willamette avait des MPS de 4, α de 2,18 et *m* de 10. Tous les échantillons pour ces modèles et les échantillons éliminés sont représentés sur la figure 4, où les classes sont présentées dans des couleurs différentes. De plus, la figure 4 indique que ces ensembles de données n'ont pas été classés avec précision. Cependant, le but de la présente étude n'était pas une classification précise, mais plutôt de déterminer les meilleurs échantillons aberrants qui augmentent le R^2 des signaux satellites et des mesures de décharge in situ.





Jusqu'à présent, une combinaison de méthodes d'optimisation multi-objectifs, de classification et de sélection des valeurs aberrantes a été appliquée pour détecter et éliminer les signaux inexacts de l'ensemble de données. L'ensemble de données présentait également des lacunes en raison de problèmes techniques liés aux satellites. Afin de combler ces lacunes, la deuxième phase OCOR réalisée était l'étalonnage. Dans cette phase, les signaux manqués et aberrants ont été évalués à l'aide de mesures de décharge in situ, qui ont servi d'entrées de modèle de régression non linéaire tandis que les signaux satellites étaient les cibles. Suite à la formation du modèle de régression, les signaux manqués et éliminés ont été modélisés. Les équations de régression d'étalonnage pour les rivières White et Willamette sont données par (3) et (4), respectivement, où S représente les valeurs du signal satellite, et Q désigne des mesures de décharge in situ.

$$S = 4.63 e^{-16} imes Q^3 - 6.16 e^{-11} imes Q^2 + 3.79 e^{-6} imes Q + 1.014$$
 (3)

$$S = 6.52e^{-18} imes Q^3 - 6.86e^{-12} imes Q^2 + 1.15e^{-6} imes Q + 1.017$$
 (4)

Les valeurs manquées en raison de problèmes techniques de satellite ne sont pas spécifiées. Les jauges in situ sont également parfois inaccessibles et l'ensemble de données de mesure de débit in situ a également perdu des échantillons. Dans de tels cas, un modèle de régression peut être utilisé pour modéliser les mesures de décharge in situ à l'aide de signaux satellites.

Les signaux évalués pour les deux études de cas sont présentés à la figure 5 . Ici, les points rouges sont des signaux aberrants éliminés de l'ensemble de données dans la phase de prétraitement OCOR et remplacés lors

de la phase d'étalonnage à l'aide de (3) et (4). Les points bleus représentent les signaux satellites manqués en raison de problèmes techniques. D'après ce chiffre, il est évident que la rivière Willamette a perdu des signaux pendant une longue période.

Ces échantillons perdus ont été réévalués en utilisant les équations de régression mentionnées. Les points roses sur la figure de l'ensemble de données de White River représentent les entrées d'échantillon remplaçant les valeurs de mesure in situ manquées en raison de problèmes techniques de jauge. Pour les rivières White et Willamette, les quantités de R^2 entre les signaux satellites initiaux et les décharges in situ sont respectivement de 0,57 et 0,39 et les quantités de R^2 entre les signaux satellites développés et les décharges in situ sont de 0,86 et 0,7, respectivement. Ainsi, l'utilisation de la méthode OCOR augmente considérablement les performances des signaux satellites. De plus, dans le jeu de données initial, il y a des lacunes dans les signaux satellites et les mesures in situ, qui sont remplacées dans les jeux de données finaux.





Figure 5. L'ensemble de données final des signaux satellitaires pour (a) White River et (b) Willamette River.

Après calibrage, la méthode OCOR a atteint les objectifs. Premièrement, des signaux satellites inexacts ont été détectés et éliminés. Deuxièmement, les signaux manqués et exclus ont été réévalués à l'aide de modèles de régression. Selon la figure 6, suite à ces deux phases, les signaux satellites présentaient une excellente corrélation avec les mesures de décharge in situ; par conséquent, cet ensemble de données peut être utilisé comme série chronologique pour prévoir le débit futur de la rivière.



(b)

Figure 6. Simulations du débit in situ de (a) White River et (b) Willamette River à l'aide de signaux satellites.

Enfin, la méthode OCOR est utilisée pour faire la distinction entre le niveau d'eau normal des rivières et les niveaux d'eau de crue. Dans cette théorie, qui a été utilisée dans le domaine de la détection d'inondations par satellite par DeGroeve et Riva [2009b], la magnitude d'inondation (FM) est définie à l'aide de l'équation suivante:

$$FM = \frac{S - \bar{S}}{\sigma(S)}$$
(5)

où S est les signaux satellites tels que présentés en (1), et $\sigma(S)$ est l'écart type de l'ensemble de données S, qui est défini comme suit:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (S_i - \bar{S})^2}$$
(6)

où N est le nombre d'échantillons de l'ensemble de données S et \overline{S} représente la valeur moyenne de S. Une distribution normale, lorsque la valeur moyenne de l'ensemble de données est égale à zéro et que son écart type est égal à un, la distribution est définie comme suit:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
(7)

où x représente un échantillon de l'ensemble de données S. Les histogrammes des ensembles de données FM pour les rivières Willamette et White sont présentés à la figure 7. Sur cette figure, la distribution normale a été obtenue en utilisant (7) est représentée par une ligne rouge et les distributions réelles des

ensembles de données FM sont indiquées par des cases bleues. DeGroeve et Riva [2009b] et l'organisation GFDS ont proposé de considérer les échantillons FM avec un écart type entre deux et quatre ($2 < \sigma < 4$) comme des crues petites et régulières et des échantillons FM avec un écart type supérieur à 4 ($\sigma \ge 4$) comme des inondations extrêmes.



Figure 7. Histogrammes des ensembles de données FM de Willamette et White Rivers.

6. Conclusion

L'un des objectifs les plus intéressants de la gestion des inondations en cas de catastrophe est de fournir des prévisions en temps réel du niveau des eaux de crue pour les plans d'évacuation d'alerte précoce destinés à sauver des vies. Diverses études ont été menées récemment dans ce domaine scientifique. Les signaux satellitaires sont de plus en plus facilement disponibles et une méthode rentable pour surveiller un vaste réseau de rivières pendant les événements de tempête majeurs. Dans cette étude, une nouvelle méthode est proposée pour améliorer la précision des prévisions de niveau d'eau par satellite, comprenant un modèle d'optimisation d'essaim de particules multi-objectifs, un algorithme de classification d'arbre de décision, la détection des valeurs aberrantes par le T^2 de Hotelling et un modèle de régression (OCOR). La nouvelle méthode OCOR peut identifier efficacement les signaux satellites inexacts et les remplacer par des valeurs raisonnables.

Deux études de cas ont été sélectionnées pour présenter la nouvelle application de la nouvelle méthode OCOR, y compris la White River et la Willamette River, situées aux États-Unis. Au total, 307 échantillons de White River et 1 008 échantillons de Willamette River ont été identifiés comme des valeurs aberrantes ou des données manquantes en raison de problèmes techniques liés aux satellites. Après avoir éliminé ces échantillons, les signaux satellitaires ont été étalonnés avec succès avec les données de jauge hydrométrique in situ, et les signaux supprimés ont été comblés en utilisant le modèle de régression étalonné. Par conséquent, toutes les lacunes de l'ensemble de données ont été comblées avec des valeurs satisfaisantes.

Bibliographie

- 1. [Akhbari et al., 2017] A. Akhbari; H. Bonakdari; I. Ebtehaj Evolutionary prediction of electrocoagulation efficiency and energy consumption probing, Desalin. Water Treat., Volume 64 (2017), pp. 54-63
- [Alizadeh et al., 2017] M. J. Alizadeh; M. R. Kavianpour; O. Kisi; V. Nourani A new approach for simulating and forecasting the rainfall-runoff process within the next two months, J. Hydrol., Volume 548 (2017), pp. 588-597
- 3. [Birkinshaw et al., 2010] S. J. Birkinshaw; G. O'donnell; P. Moore; C. Kilsby; H. Fowler; P. Berry Using satellite altimetry data to augme nt flow estimation techniques on the Mekong River, Hydrol. Process., Volume 24 (2010) no. 26, pp. 3811-3825
- 4. [Bjerklie et al., 2004] D. M. Bjerklie; D. Moller; L. C. Smith; S. L. Dingman Estimating discharge in rivers using remotely sensed hydraulic information, J. Hydrol., Volume 309 (2004) no. 1, pp. 191-209
- 5. [Bonakdari et al., 2019] H. Bonakdari; A. H. Zaji; A. D. Binns; B. Gharabaghi Integrated Markov chains and uncertainty analysis techniques to more accurately forecast floods using satellite signals, J. Hydrol., Volume 572 (2019), pp. 75-95

- [Bonakdari et al., 2020] H. Bonakdari; F. Moradi; I. Ebtehaj; B. Gharabaghi; A. A. Sattar; A. H. Azimi; A. Radecki-Pawlik A non-tuned machine learning technique for abutment scour depth in clear water condition, Water, Volume 12 (2020), 301 pages
- 7. [Brakenridge et al., 2007] G. R. Brakenridge; S. V. Nghiem; E. Anderson; R. Mic Orbital microwave measurement of river discharge and ice status, Water Resour. Res., Volume 43 (2007) no. 4, W04405 pages
- Brakenridge et al., 2012] G. R. Brakenridge; S. Cohen; A. J. Kettner; T. De Groeve; S. V. Nghiem; J. P. Syvitski; B. M. Fekete Calibration of satellite measurements of river discharge using a global hydrology model, J. Hydrol., Volume 475 (2012), pp. 123-136
- 9. [Breiman et al., 1984] L. Breiman; J. Friedman; C. J. Stone; R. A. Olshen Classification and Regression Trees, CRC Press, Boca Raton, Florida, USA, 1984 | Zbl 0541.62042
- [Calmant and Seyler, 2006] S. Calmant; F. Seyler Continental surface waters from satellite altimetry, C. R. Geosci., Volume 338 (2006) no. 14, pp. 1113-1122
- [Chen et al., 2015] X. Y. Chen; K. W. Chau; A. O. Busari A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model, Eng. Appl. Artif. Intell., Volume 46 (2015), pp. 258-268
- [Coello and Lechuga, 2002] C. A. C. Coello; M. S. Lechuga MOPSO: a proposal for multiple objective particle swarm optimization, Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on, Honolulu, USA, IEEE, New York, USA, 2002, pp. 1051-1056
- [Darras et al., 2017] T. Darras; L. Kong-A-Siou; B. Vayssade; A. Johannet; S. Pistre Karst flash flood forecasting using recurrent and non-recurrent artificial neural network models: the case of the Lez Basin (Southern France), EuroKarst 2016, Neuchâtel: Advances in the Hydrogeology of Karst and Carbonate Reservoirs (P. Renard; C. Bertrand, eds.), Springer International Publishing, Cham, 2017, pp. 169-177
- [Deb et al., 2000] K. Deb; S. Agrawal; A. Pratap; T. Meyarivan A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, International Conference on Parallel Problem Solving From Nature, Springer, 2000, pp. 849-858
- 15. [De Groeve and Riva, 2009a] T. De Groeve; P. Riva Early flood detection and mapping for humanitarian response, Proceedings of the 6 International ISCRAM Conference, Gothenburg, Sweden, Information Systems for Crisis Response and Management (ISCRAM), Gothenburg, Sweden, 2009a
- [De Groeve and Riva, 2009b] T. De Groeve; P. Riva Global real-time detection of major floods using passive microwave remote sensing, Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Stresa, Italy, International Center for Remote Sensing of Environment (ICRSE), Stresa, Italy, 2009b, pp. 4-8
- [Ebtehaj et al., 2018] I. Ebtehaj; H. Bonakdari; F. Moradi; B. Gharabaghi; Z. S. Khozani An integrated framework of Extreme Learning Machines for predicting scour at pile groups in clear water condition, Coast. Eng., Volume 135 (2018), pp. 1-15
- [Ebtehaj et al., 2019] I. Ebtehaj; H. Bonakdari; B. Gharabaghi Closure to "An integrated framework of Extreme Learning Machines for predicting scour at pile groups in clear water condition by Ebtehaj, I., Bonakdari, H., Moradi, F., Gharabaghi, B., Khozani, Z. S.", Coast. Eng., Volume 147 (2019), pp. 135-137
- [Espinoza-Villar et al., 2018] R. Espinoza-Villar; J.-M. Martinez; E. Armijos; J.-C. Espinoza; N. Filizola; A. Dos Santos; B. Willems; P. Fraizy; W. Santini; Ph. Vauchel Spatio-temporal monitoring of suspended sediments in the Solimoes River (2000–2014), C. R. Geosci., Volume 350 (2018) no. 1–2, pp. 4-12
- 20. [Frasson et al., 2017] R. P. d. M. Frasson; R. Wei; M. Durand; J. T. Minear; A. Domeneghetti; G. Schumann; B. A. Williams; E. Rodriguez; C. Picamilh; C. Lion; T. Pavelski; P. A. Garamois Automated river reach definition strategies: applications for the surface water and ocean topography mission, Water Resour. Res., Volume 53 (2017) no. 10, pp. 8164-8186
- 21. [Gholami et al., 2019] A. Gholami; H. Bonakdari; P. Samui; M. Mohammadian; B. Gharabaghi Predicting stable alluvial channel profiles using emotional artificial neural networks, Appl. Soft Comput. J., Volume 78 (2019), pp. 420-437
- 22. [Han et al., 2011] J. Han; J. Pei; M. Kamber Data Mining: Concepts and Techniques, Elsevier Inc., Amesterdam, Netherland, 2011 | Zbl 1230.68018
- 23. [He et al., 2014] Z. He; X. Wen; H. Liu; J. Du A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region, J. Hydrol., Volume 509 (2014), pp. 379-386

- 24. [Hotelling, 1992] H. Hotelling The generalization of Student's ratio, Breakthroughs in Statistics, Springer, New York, USA, 1992
- 25. [Jiang et al., 2014] S. Jiang; L. Ren; Y. Hong; X. Yang; M. Ma; Y. Zhang; F. Yuan Improvement of multi-satellite real-time precipitation products for ensemble streamflow simulation in a middle latitude basin in South China, Water Resour. Manage., Volume 28 (2014) no. 8, pp. 2259-2278
- 26. [Kennedy et al., 2001] J. F. Kennedy; J. Kennedy; R. C. Eberhart; Y. Shi Swarm Intelligence, Morgan Kaufmann, San Francisco, California, USA, 2001
- 27. [Khan et al., 2011] S. I. Khan; Y. Hong; J. Wang; K. K. Yilmaz; J. J. Gourley; R. F. Adler; G. R. Brakenridge; F. Policelli; S. Habib; D. Irwin Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: implications for hydrologic prediction in ungauged basins, IEEE Trans. Geosci. Remote Sens., Volume 49 (2011) no. 1, pp. 85-95
- [Khan et al., 2012] S. I. Khan; Y. Hong; H. J. Vergara; J. J. Gourley; G. R. Brakenridge; T. De Groeve;
 Z. L. Flamig; F. Policelli; B. Yong Microwave satellite data for hydrologic modeling in ungauged basins, IEEE Geosci. Remote Sens. Lett., Volume 9 (2012) no. 4, pp. 663-667
- 29. [Kugler and De Groeve, 2007] Z. Kugler; T. De Groeve, 2007 (The global flood detection system. Office for Official Publications of the European Communities, Luxembourg, 45)
- 30. [Mason and Young, 2002] R. L. Mason; J. C. Young Multivariate Statistical Process Control with Industrial Applications, Society for Industrial Mathematics, Philadelphia, USA, 2002 | Zbl 0989.62075
- 31. [Mitra, 2009] A. Mitra Data transformation for normalization, Encyclopedia of Data Warehousing and Mining, IGI Global, Hershey, USA, 2009, pp. 566-571
- 32. [Njoku et al., 2003] E. G. Njoku; T. J. Jackson; V. Lakshmi; T. K. Chan; S. V. Nghiem Soil moisture retrieval from AMSR-E, IEEE Trans. Geosci. Remote Sens., Volume 41 (2003) no. 2, pp. 215-229
- 33. [Quinlan and Rivest, 1989] J. R. Quinlan; R. L. Rivest Inferring decision trees using the minimum description lenght principle, Info. Comput., Volume 80 (1989) no. 3, pp. 227-248 | Zbl 0664.94015
- [Quinlan, 1986] J. R. Quinlan Induction of decision trees, Mach. Learn., Volume 1 (1986) no. 1, pp. 81-106
- 35. [Quinlan, 1987] J. R. Quinlan Simplifying decision trees, Int. J. Man-Mach. Stud., Volume 27 (1987) no. 3, pp. 221-234
- 36. [Salvia et al., 2011] M. Salvia; F. Grings; P. Ferrazzoli; V. Barraza; V. Douna; P. Perna; H. Karszenbaum Estimating flooded area and mean water level using active and passive microwaves: the example of Paran River delta floodplain, Hydrol. Earth Syst. Sci. Discuss., Volume 8 (2011) no. 2, pp. 2895-2928
- 37. [Sattar et al., 2019] M. A. Sattar; H. Bonakdari; B. Gharabaghi; A. Radecki-Pawlik Hydraulic modeling and evaluation equations for the incipient motion of sandbags for levee breach closure operations, Water, Volume 11 (2019) no. 279, pp. 1-22
- [Schmugge, 1980] T. J. Schmugge Microwave approaches in hydrology, Photogramm. Eng. Remote Sens., Volume 46 (1980), pp. 495-507
- [Shabbak et al., 2011] A. Shabbak; H. Midi; M. N. Hassan The performance of robust multivariate statistical control charts based on different cutoff-points with sustained shifts in mean, J. Appl. Sci., Volume 11 (2011) no. 1, pp. 56-65
- 40. [Shaghaghi et al., 2017] S. Shaghaghi; H. Bonakdari; A. Gholami; I. Ebtehaj; M. Zeinolabedini Comparative analysis of GMDH neural network based on genetic algorithm and particle swarm optimization in stable channel design, Appl. Math. Comput., Volume 313 (2017), pp. 271-286
- [Shiklomanov et al., 2002] A. Shiklomanov; R. Lammers; C. J. Vörösmarty Widespread decline in hydrological monitoring threatens pan-Arctic research, EOS Trans. AGU., Volume 83 (2002) no. 2, pp. 13-17
- 42. [Sivapalan, 2003] M. Sivapalan Prediction in ungauged basins: a grand challenge for theoretical hydrology, Hydrol. Process., Volume 17 (2003) no. 15, pp. 3163-3170
- 43. [Srivastava, 2017] P. K. Srivastava Satellite soil moisture: review of theory and applications in water resources, Water Resour. Manage., Volume 31 (2017) no. 10, pp. 3161-3176
- 44. [Stokstad, 1999] E. Stokstad Scarcity of rain, stream gages threatens forecasts, Science, Volume 285 (1999) no. 5431, p. 1199-1200
- 45. [Su et al., 2008] F. Su; Y. Hong; D. P. Lettenmaier Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin, J. Hydrometeorol., Volume 9 (2008) no. 4, pp. 622-640
- 46. [Sullivan and Woodall, 1996] J. H. Sullivan; W. H. Woodall A comparison of multivariate control charts for individual observations, J. Qual. Tech., Volume 28 (1996) no. 4, pp. 398-408

- 47. [Tekeli and Fouli, 2017] A. E. Tekeli; H. Fouli Reducing false flood warnings of TRMM rain rates thresholds over Riyadh City, Saudi Arabia by utilizing AMSR-E Soil moisture information, Water Resour. Manage., Volume 31 (2017) no. 4, pp. 1243-1256
- 48. [Temimi et al., 2007] M. Temimi; R. Leconte; F. Brissette; N. Chaouch Flood and soil wetness monitoring over the Mackenzie River Basin using AMSR-E 37 GHz brightness temperature, J. Hydrol., Volume 333 (2007) no. 2, pp. 317-328
- 49. [Temimi et al., 2011] M. Temimi; T. Lacava; T. Lakhankar; V. Tramutoli; H. Ghedira; R. Ata; R. Khanbilvardi A multi-temporal analysis of AMSR-E data for flood and discharge monitoring during the 2008 flood in Iowa, Hydrol. Process., Volume 25 (2011) no. 16, pp. 2623-2634
- [Theis et al., 1982] S. W. Theis; M. J. McFarland; W. D. Rosenthal; C. L. Jones Microwave Remote Sensing of Soil Moistures, Remote Sensing Center, Texas A&M University, College Station, TX, 1982 (RSC-3458-129)
- 51. [Ulaby et al., 1978] F. T. Ulaby; P. P. Batlivala; M. C. Dobson Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-bare soil, IEEE Trans. Geosci. Elec., Volume GE-16 (1978) no. 4, pp. 286-295
- 52. [Wang et al., 1982] J. Wang; P. ONeill; E. Engman, 1982 (Remote Measurements of Soil Moisture by Microwave Radiometers at BARC Test Site II. 83954, NASA)
- 53. [Weigend, 2018] A. S. Weigend Time Series Prediction: Forecasting the Future and Understanding the Past, Routledge, Taylor and Francis, New York, USA, 2018
- 54. [Yaseen et al., 2017] Z. M. Yaseen; I. Ebtehaj; H. Bonakdari; R. C. Deo; A. Danandeh Mehr; W. H. Melini Wan Mohtar; L. Diop; A. El-shafieh; V. P. Singh Novel approach for streamflow forecasting using a hybrid ANFIS-FFa model, J. Hydrol., Volume 554 (2017), pp. 263-276
- [Young et al., 2017] C. C. Young; W. C. Liu; M. C. Wu A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events, Appl. Soft Comput., Volume 53 (2017), pp. 205-216
- 56. [Zaji and Bonakdari, 2019] A. H. Zaji; H. Bonakdari Robustness lake water level prediction using the search heuristic based artificial intelligence methods, ISH J. Hydraul. Eng., Volume 25 (2019) no. 3, pp. 316-324
- 57. [Zaji et al., 2015] A. H. Zaji; H. Bonakdari; S Shamshirband; S. N. Qasem Potential of particle swarm optimization based radial basis function network to predict the discharge coefficient of a modified triangular side weir, Flow Measur. Instrum., Volume 45 (2015), pp. 404-407
- [Zaji et al., 2018] A. H. Zaji; H. Bonakdari; B. Gharabaghi Remote sensing satellite data preparation for simulating and forecasting river discharge, IEEE Trans. Geosci. Remote Sens., Volume 56 (2018) no. 6, pp. 3432-3441
- [Zaji et al., 2019a] A. H. Zaji; H. Bonakdari; B. Gharabaghi Applying upstream satellite signals and a 2-D Error minimization algorithm to advance early warning and management of flood water levels and river discharge, IEEE Trans. Geosci. Remote Sens., Volume 99 (2019a), pp. 1-9
- 60. [Zaji et al., 2019b] A. H. Zaji; H. Bonakdari; B. Gharabaghi Developing an AI-based method for river discharge forecasting using satellite signals, Theor. Appl. Climatol., Volume 138 (2019b), pp. 347-362
- 61. [Zeynoddin et al., 2018] M. Zeynoddin; H. Bonakdari; A. Azari; I. Ebtehaj; B. Gharabaghi; H. M. Riahi Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate, J. Environ. Manage., Volume 222 (2018), pp. 190-206