

Université du Québec  
Institut National de la Recherche Scientifique Centre Eau Terre Environnement

# **ESTIMATION DU DÉBIT EN RIVIÈRE À L'AIDE DE L'IMAGERIE RADARSAT EN UTILISANT L'APPRENTISSAGE PROFOND**

Par  
Samar Ziadi

Mémoire présenté pour l'obtention du grade de  
Maître ès Sciences (M.Sc.)  
en sciences de l'eau

## **Jury d'évaluation**

Président du jury et  
examineur interne

Sebastien Raymond  
INRS Centre Eau Terre Environnement

Examineur externe

Yacine Bouroubi  
Professeur, Université de Sherbrooke

Directeur de recherche

Karem Chokmani  
INRS Centre Eau Terre Environnement

## REMERCIEMENTS

Je remercie grandement mon directeur de mémoire, Karem Chokmani, pour m'avoir accueilli au sein de son équipe de recherche, pour le temps qu'il m'a accordé et ses nombreux conseils. Je salue son amabilité, sa disponibilité et son assistance ont été d'une grande utilité.

Je remercie également Chayma Chaabani et Anas El Alem pour leurs supports, leurs rigueurs professionnelles et leurs critiques constructives qui ont contribué à me guider dans mes recherches.

Mes remerciements vont également à toute l'équipe de TENOR qui a rendu cette expérience particulièrement agréable et enrichissante.

Et enfin, je remercie ma famille pour leur confiance et leur soutien tout au long de mon projet et mes études.



## RÉSUMÉ

L'estimation du débit en rivières est un paramètre clé pour la gestion des ressources hydriques, la prévention des risques liés aux inondations et la planification des équipements hydroélectriques. Lorsque le débit d'eau est très élevé lors d'évènements extrêmes, les méthodes de jaugeage traditionnelles ne peuvent pas être utilisées. De plus, les stations du réseau hydrométrique sont généralement éparses et leur répartition spatiale n'est pas optimale. Par conséquent, de nombreuses sections de rivières ne peuvent être suivies par des mesures et observations du débit. Pour ces raisons, pendant la dernière décennie, les capteurs satellitaires ont été considérés comme une source d'observation complémentaire aux observations traditionnelles du niveau d'eau et du débit en rivières. L'utilisation d'une telle approche a fourni un moyen de maintenir et d'étendre le réseau d'observation hydrométrique.

L'approche avec télédétection permet d'estimer le débit à partir des courbes de tarage qui met en relation le débit instantané ( $Q$ ) et la géométrie d'une section transversale du chenal (la largeur ou la profondeur effective de la surface d'eau). En revanche, cette méthode est associée à des limitations, notamment, sa dépendance aux courbes de tarage. En effet, en raison de leurs natures empiriques, les courbes de tarage sont limitées à des sections spécifiques et ne peuvent être appliquées dans d'autres rivières. Récemment, des techniques d'apprentissage profond ont été appliquées avec succès dans de nombreux domaines, y compris en hydrologie.

Dans le présent travail, l'approche d'apprentissage profond a été choisie, en particulier les réseaux de neurones convolutifs (CNN), pour estimer le débit en rivière. L'objectif principal de ce travail est de développer une approche d'estimation du débit en rivières à partir de l'imagerie RADARSAT 1&2 à l'aide de l'apprentissage profond. La zone d'étude se trouve dans l'écozone du bouclier boréal à l'Est du Canada. Au total, 39 sites hydrographiques ont fait l'objet de cette étude. Dans le présent travail, une nouvelle architecture de CNN a été proposée, elle s'adapte aux données utilisées et permet d'estimer le débit en rivière instantané. Ce modèle donne un résultat du coefficient de détermination ( $R^2$ ) et de Nash-Sutcliffe égale à 0.91, le résultat d'erreur quadratique moyenne égale à  $33 \text{ m}^3/\text{s}$ . Cela démontre que le modèle CNN donne une solution appropriée aux problèmes d'estimation du débit avec des capteurs satellites sans intervention humaine.

Mots-clés : Apprentissage profond, CNN, courbe de tarage, débit, hauteur d'eau, images radar.



## ABSTRACT

Estimating river flow is a key parameter for effective water resources management, flood risk prevention and hydroelectric facilities planning. In cases of very high flow of water or extreme events, traditional gauging methods cannot be reliable. In addition, hydrometric network stations are often sparse and their spatial distribution is not optimal. Therefore, many river sections cannot be monitored using traditional flow measurements and observations. For these reasons, satellite sensors are considered as a complementary observation source to traditional water level and flow observations in the last decades. The use of this kind of approach has provided a way to maintain and expand the hydrometric observation network.

Remote sensing data can be used to estimate flow from rating curves that relate the instantaneous flow ( $Q$ ) to the geometry of a channel cross-section (the effective width or depth of the water surface). On the other hand, remote sensing is also associated with limitations, notably its dependence on the rating curves. Indeed, due to their empirical nature, rating curves are limited to specific sections and cannot be applied in other rivers. Recently, deep learning techniques have been successfully applied in many fields, including hydrology.

In the present work, the deep learning approach has been chosen, in particular convolutional neural networks (CNN), to estimate river flow. The main objective of this work is to develop an approach to estimate river flow from RADARSAT 1&2 imagery using deep learning. In this study, 39 hydrographic sites of the Boreal Shield ecozone in Eastern Canada were considered. A new CNN architecture was developed to provide a straightforward estimation of the instantaneous river flow rate.

The achieved results demonstrated a coefficient of determination ( $R^2$ ) and Nash-Sutcliffe values of 0.91, and a root mean square error of  $33\text{m}^3/\text{s}$ . This indicates the effectiveness of CNN in automatic flow estimation with satellite sensors.

**Key words:** Deep learning, CNN, rating curve, flow, water level, radar images.



# TABLE DES MATIÈRES

<b>REMERCIEMENTS</b> .....	<b>III</b>
<b>RÉSUMÉ</b> .....	<b>V</b>
<b>ABSTRACT</b> .....	<b>VII</b>
<b>TABLE DES MATIÈRES</b> .....	<b>IX</b>
<b>LISTE DES FIGURES</b> .....	<b>XI</b>
<b>LISTE DES TABLEAUX</b> .....	<b>XIII</b>
<b>LISTE DES ABRÉVIATIONS</b> .....	<b>XVII</b>
<b>PARTIE 1 : INTRODUCTION GÉNÉRALE</b> .....	<b>1</b>
<b>1 INTRODUCTION</b> .....	<b>1</b>
1.1 CONTEXTE .....	1
1.2 PROBLEMATIQUE .....	1
1.3 OBJECTIFS ET HYPOTHESES DE LA RECHERCHE .....	3
1.4 STRUCTURE DU MEMOIRE .....	4
<b>2 REVUE DE LITTÉRATURE</b> .....	<b>5</b>
2.1 DEFINITION DU DEBIT .....	5
2.2 ESTIMATION DU DEBIT DES RIVIERES PAR LES METHODES CONVENTIONNELLES .....	5
2.3 ESTIMATION DU DEBIT DES RIVIERES A PARTIR DE LA TELEDETECTION SATELLITAIRE .....	8
2.4 APPRENTISSAGE PROFOND .....	12
2.4.1 Réseaux de neurones .....	12
2.4.2 Fonctionnement des réseaux de neurones artificiels .....	12
2.4.3 Réseaux de neurones convolutifs.....	15
2.4.4 Méthode d'apprentissage profond en hydrologie.....	16
<b>3 MATÉRIEL ET MÉTHODES</b> .....	<b>19</b>
3.1 SITES A L'ETUDE .....	19
3.2 DONNEES SATELLITAIRES.....	19
3.3 DONNEES HYDROMETRIQUES .....	19
3.4 DONNEES TOPOGRAPHIQUES .....	19
3.5 METHODOLOGIE PROPOSEE .....	20
3.5.1 Prétraitement des données d'entrée.....	20
3.5.2 Entraînement et évaluation du modèle d'apprentissage profond .....	21
3.5.3 Évaluation de la performance du modèle .....	21



<b>PARTIE 2 : ARTICLE .....</b>	<b>23</b>
<b>4 RIVER FLOW ESTIMATION WITH RADARSAT IMAGERY USING DEEP LEARNING.....</b>	<b>25</b>
1. Introduction .....	27
2. Study area .....	30
3. Material and Methods.....	31
3.1 In-situ and satellite dataset .....	31
3.1.1. Satellite data .....	31
3.1.2. In-situ gauging dataset.....	32
3.1.3 Topographic dataset .....	33
3.2 CNN approach for river flow estimation.....	34
3.2.1 Input data pre-processing .....	35
3.2.2 Training and evaluation of the deep learning model.....	37
3.2.3 The proposed CNN architecture.....	38
3.2.3 Model libraries.....	40
3.2.4 Evaluation of model performance.....	40
4. RESULTS AND DISCUSSION .....	41
4.1 Analysis based on input data .....	41
4.2 Analysis according to the morphometric characteristics of the estimation site .....	44
4.3 Analysis by individual station of the flow estimate.....	47
5. Conclusion .....	51
Credit author statement.....	52
<b>5 DISCUSSION GÉNÉRALE ET CONCLUSION.....</b>	<b>53</b>
<b>6 BIBLIOGRAPHIE.....</b>	<b>59</b>
<b>7 ANNEXE.....</b>	<b>79</b>

## LISTE DES FIGURES

FIGURE 2.1	PRINCIPAUX METHODES ET OUTILS DE MESURE DE DEBIT. ....	6
FIGURE 2.2	REPRESENTATION D'UN PERCEPTRON SIMPLE .....	13
FIGURE 2.3	DIFFERENTS TYPE DE FONCTIONS D'ACTIVATION .....	14
FIGURE 2.4	ARCHITECTURE D'UN RESEAU DE NEURONES .....	15
FIGURE 2.5	REPRESENTATION GENERIQUE D'UN RESEAU DE NEURONES CONVOLUTIFS (SENGUPTA ET AL., 2020) .....	16
FIGURE 4.1	OVERVIEW OF THE STUDY AREA AND LOCATION OF THE HYDROMETRIC STATIONS.....	30
FIGURE 4.2	RADAR IMAGE COVERAGE IN THE STUDY ZONE .....	32
FIGURE 4.3	HAND COMPUTATION WORKFLOW, SOURCE : (NOBRE ET AL., 2011) .....	34
FIGURE 4.4	GENERAL FLOWCHART OF THE RIVER FLOW ESTIMATION APPROACH .....	34
FIGURE 4.5	CLIPPING DATA.....	36
FIGURE 4.6	DATA AUGMENTATION.....	37
FIGURE 4.7	AN OVERVIEW OF A CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURE.....	40
FIGURE 4.8	AN ILLUSTRATION OF THE EFFECT OF ADDING HAND ON THE ACCURACY OF RIVER FLOW ESTIMATION.....	43
FIGURE 4.9	ILLUSTRATION OF SOME EXAMPLES OF OVERESTIMATED OR UNDERESTIMATED FLOW PREDICTIONS.....	44
FIGURE 4.10	FLOW PREDICTION USING DATA THAT IS WITHIN A TWO KM BUFFER FROM THE HYDROMETRIC STATION.....	46
FIGURE 4.11	FLOW ESTIMATION USING DATA PASSING THROUGH THE HYDROMETRIC STATION .....	46
FIGURE 4.12	OCCURRENCE OF NSE IN THE 15 HYDROMETRIC STATIONS .....	48
FIGURE 4.13	ANALYSIS OF FLOW ESTIMATION STATION BY STATION AT DIFFERENT DATES AND DIFFERENT FLOW CONDITIONS. RIVER FLOW VARIES FROM 33 TO 650 M <sup>3</sup> /S.....	50
FIGURE 4.14	ANALYSIS OF FLOW ESTIMATION STATION BY STATION AT DIFFERENT DATES AND DIFFERENT FLOW CONDITIONS. RIVER FLOW VARIES FROM 12 TO 22 M <sup>3</sup> /S.....	50



## LISTE DES TABLEAUX

TABLE 4-1 STATISTICAL INFORMATION ON HYDROMETRIC STATIONS USED FOR RIVER FLOW ESTIMATION. .... 32







## LISTE DES ABRÉVIATIONS

ADCP	Acoustic Doppler Current Profiler
ANN	Réseau de neurone artificiel
CNN	Réseau neuronal convolutif
CEHQ	Centre d'expertise hydrique du Québec
CPU	Processeur central d'ordinateur
DBN	Deep belief net
DL	Deep Learning
GAP	Global Average Pooling Layers
GPU	Processeurs graphiques
HAND	Height Above Nearest Drainage
LSTM	Long short-term memory
LSPIV	Large Scale Particle Image Velocimetry
MAPE	Erreur moyenne en pourcentage absolu
MLP	Perceptron multicouche
MNT	Modèles numériques de terrain
MSE	Erreur quadratique moyenne
NSE	L'efficacité de Nash-Sutcliffe
$R^2$	Coefficient de détermination
ReLU	Rectified linear unit
RMSE	Racine de l'erreur quadratique moyenne
RSO	Radar à synthèse d'ouverture
RRMSE	Erreur quadratique moyenne relative



---

**PARTIE 1 : INTRODUCTION GÉNÉRAL**

# 1 INTRODUCTION

---

## 1.1 Contexte

L'estimation de débit d'eau est une opération qui se fait depuis l'Antiquité. Cette mesure est une variable clé pour la gestion optimale des ressources hydriques, pour l'atténuation des risques d'inondation et la planification des équipements hydroélectriques. Prédire et gérer les débits des cours d'eau est d'une importance capitale pour la maîtrise des crues, les études d'impact environnemental et la modélisation hydrologique de bassin versant ( Smith et Pavelsky, 2008a ; Dottori *et al.*, 2009; Barbetta *et al.*, 2012a; Guerrero *et al.*, 2012a; F; Gleason et Smith, 2014). Les méthodes de jaugeage des débits des rivières sont nombreuses et variées. Cependant, plusieurs études ont montré que de nombreuses sections de rivière se trouvent non jaugées que même le nombre de stations de jaugeage décline ( Bjerklie *et al.*, 2003a; Trambly *et al.*, 2010a; Baldassarre *et al.*, 2012; Hossain *et al.*, 2013, s.d.; Tourian *et al.*, s.d., 2017; D Feng *et al.*, 2019; C. Gleason *et al.*, 2018). Ceci constitue une difficulté d'accès à l'information de débit en rivières depuis les années 1980.

Les méthodes traditionnelles d'estimation des débits en rivières reposent généralement sur l'utilisation de courbe de tarage. Ces méthodes sont basées sur des instrumentations assez simples, mais permettent de fournir des résultats satisfaisants et de donner des mesures assez précises de niveau d'eau. Ceci explique leurs popularités et leur utilisations qui remontent à plusieurs décennies (F Dottori *et al.*, 2009). En revanche, les techniques traditionnelles de mesure de débit ont montré leurs limites (Dai et Trenberth, 2002).

Il est difficile d'obtenir des mesures représentatives des débits et de plus de réaliser des mesures sur le terrain, notamment en présence d'un évènement extrême tel qu'inondations, érosion, présence de nappe de glace, période de crue et de décrue ( Dai et Trenberth, 2002; F. Dottori *et al.*, 2009; Barbetta *et al.*, 2012). En outre, le coût d'installation d'une station hydrométrique est très élevé et elle mobilise beaucoup de temps.

## 1.2 Problématique

La difficulté d'estimation du débit dans plusieurs sections de rivières à cause de la pénurie et même parfois l'absence des stations de mesures de débit crée un grand impact sur la gestion des ressources en eau. Pendant la dernière décennie, les capteurs satellitaires ont fourni une

source supplémentaire permettant de mesures de débits pour suivre efficacement les ressources en eau (Koblinsky et al., 1993 ; Smith, 1997 ; Alsdorf et Lettenmaier, 2003 ; Syed et al., 2010 ; Huang et al., 2018 ; Ahmad et Kim, 2019). L'estimation de débit par des méthodes de télédétection dépend toujours de la précision des paramètres estimés qui peuvent être obtenus à partir de capteurs satellitaires d'une manière directe (largeur, pente, hauteur) ou indirecte (vitesse et profondeur bathymétrique) (Pan et Nichols, 2013 ; Pan et al., 2016 ; Sichangi et al., 2016, 2018). La plupart des techniques de la télédétection par satellite pour les analyses du débit des cours d'eau sont basées sur la relation traditionnelle entre le débit instantané  $Q$  et la géométrie d'une section transversale du chenal, précisément, la largeur  $w$  (m), la profondeur  $d$  (m), et la vitesse d'écoulement  $v$  (m/s). Plusieurs études ont mesuré la variation du niveau d'eau à l'aide des données altimétriques satellitaires. Ces mesures sont ensuite converties en débit fluvial en se basant sur la courbe de tarage entre le niveau d'eau dérivé du satellite et le débit mesuré in situ (Birkett et al., 2002; Kouraev et al., 2004; Nathanson et al., 2012; Pavelsky, 2014; Paris et al., 2016; Schneider et al., 2017). D'autres études se sont reposées sur l'hypothèse de changement de la largeur de la rivière. Elles se sont basées sur le principe de la corrélation de la surface de l'eau dérivée du satellite avec le débit mesuré in situ, puis elles ont déduit le débit de la rivière à partir des données satellitaires en se basant sur la courbe de tarage de la surface de l'eau-débit. D'autres études ont utilisé les données radar à synthèse d'ouverture (SAR) et les images optiques (Smith, 1997 ; Gleason et Smith, 2014; Pavelsky, 2014; Sichangi et al., 2016 ; Feng et al., 2019; Fulton et al., 2020). Néanmoins, il était confirmé que l'utilisation de données satellitaire présente plusieurs limitations. En effet, l'inconvénient de l'utilisation des données optiques pour estimer la largeur effective réside dans le fait que l'acquisition des données optiques est souvent très réduite par la présence de couvert nuageux.

Les données radar, un des problèmes récurrents à l'utilisation des données RSO pour estimer la largeur effective est le haut niveau de bruit qui dégrade la qualité des images et rend difficile leur interprétation. Les données de satellite Altimètre radar ont une faible résolution spatiale (Birkett et al., 2002) et une grande résolution temporelle (10 à 35 jours). Il s'en suit une limitation importante dans le suivi de niveau d'eau et l'estimation de débit.

En effet, que ce soit au sol ou dans l'espace, une limitation commune de la plupart des méthodes satellitaires dans l'estimation du débit, qu'elles dépendent toutes à la courbe de tarage. Cependant, en raison de leurs natures empiriques, les courbes de tarage sont limitées à des sections spécifiques et ne peuvent être appliquées dans d'autres rivières (Bjerklie et al., 2003b). De plus, la qualité des résultats donnée par la courbe de tarage est très dépendante de la qualité

de la calibration, laquelle dépend de la disponibilité et de la représentativité des données d'observation dérivées par le satellite ayant servi à calibrer le modèle. Également, les mesures dérivées par le satellite ne sont pas toujours disponibles et de bonne qualité. Donc, l'incertitude liée aux variables mesurées par le satellite, aux paramètres physiques, et aux équations de la courbe de tarage donne lieu à de nombreuses erreurs dans l'estimation de débit de la rivière.

Récemment, des techniques d'apprentissage profond ont été appliquées avec succès en hydrologie et l'analyse des ressources en eau (Zhang et al., 2016 ; Shen, 2018). Pour surmonter la complexité et la difficulté de l'utilisation des approches traditionnelles ainsi que pour diminuer les inconvénients d'utilisation des données de télédétection, une approche d'apprentissage profond, en particulier un réseau de neurone convolutif (CNN) a été déployé pour développer un modèle de prédiction de débit en rivière. L'objectif principal de notre travail est de développer une approche d'estimation du débit en rivières sur l'écozone boréal du Canada basée sur l'utilisation d'un modèle CNN et de données RSO.

En effet, le CNN semble une solution attrayante et appropriée aux problèmes d'estimation du débit avec des capteurs satellites. La dynamique de l'eau ainsi que le mécanisme qui gère le débit sont complexes et hautement non linéaires. L'estimation du débit dépend de plusieurs paramètres tels que la pente longitudinale et latérale, le niveau d'eau, la forme de terrain, etc. (Bjerklie et al. 2003). L'approche avec le CNN, grâce à ses nombreuses couches empilées, est capable de représenter des caractéristiques des données spatiales complexes et de haute dimension (Krizhevsky et al., 2012), ainsi que de déduire des informations robustes et évolutives tout en minimisant le taux d'intervention manuel (Goodfellow et al., 2016; Saba et al., 2019; Sengupta et al., 2020). De plus, les CNN peuvent représenter des caractéristiques non linéaires et traiter des données brutes et bruitées (Sahiner et al., 2019). Cette étude est une contribution précieuse à la prédiction du débit des rivières par la méthode d'apprentissage profond.

### **1.3 Objectifs et hypothèses de la recherche**

Le présent travail de recherche a pour objectif principal de développer une approche d'estimation du débit en rivières sur l'écozone boréal à partir d'un modèle CNN et de données RSO.

Les objectifs spécifiques sont les suivants :

- Calibrer et valider un modèle du réseau de neurones convolutifs (CNN) en utilisant uniquement les images radars;

- Étudier l'effet de la forme du chenal, de la distance de la station et de la taille de la rivière sur la qualité de l'estimation;
- Étudier l'effet de l'ajout des variables auxiliaires (les données de la hauteur au-dessus du drainage le plus proche (HAND)) au modèle et évaluer la plus-value de son utilisation dans le modèle CNN.

La présente étude cherche à valider les hypothèses suivantes :

- Étant donné qu'un changement dans le débit se traduit par un changement de la largeur de la rivière et étant donné que le RSO permet la détection de l'étendue de l'eau, les images radar permettent donc de détecter le changement dans le débit de rivière.
- L'ajout des variables auxiliaires au modèle CNN (comme le HAND) améliore les performances de prédiction de débit en rivières.

#### **1.4 Structure du mémoire**

Le mémoire est composé en deux parties. La première partie, appelée « Introduction générale » et la deuxième partie appelée « article ». Au total, nous avons créé cinq chapitres. Le premier chapitre présente l'introduction et la problématique, le deuxième chapitre expose une revue de littérature qui détaille les différentes méthodes d'estimation de débit des rivières par les méthodes conventionnelles et à partir de la télédétection satellitaire. On va évoquer aussi dans ce chapitre le développement de l'intelligence artificielle et de l'apprentissage profond. Par la suite, une présentation de la structure du mémoire.

Dans le troisième chapitre, on va présenter les matériels et méthodes. Dans cette partie, on va exposer succinctement la zone d'étude, les données hydrométriques utilisées, les données satellitaires acquises et le modèle d'apprentissage profond développé pour estimer le débit des rivières. Les détails sont à consulter dans le chapitre 2 de l'article.

La deuxième partie du mémoire, qui est le quatrième chapitre, présente l'article scientifique qui est sous révision et sera soumis pour une publication dans la revue Remote Sensing.

Dans le chapitre 5, nous allons mettre l'accent sur les discussions générales et conclusions.

Les références de la partie introduction générale et de l'article seront ajoutées à la suite de ce cinquième chapitre.

## 2 REVUE DE LITTÉRATURE

---

### 2.1 Définition du débit

Le débit désigne le volume d'eau traversant une section de cours d'eau pendant une unité de temps, il s'exprime ainsi en mètres cubes par seconde (m<sup>3</sup>/s).

$$Q = \frac{V}{t} \quad (1)$$

Avec :

$Q$  : Débit en m<sup>3</sup>/s

$V$ : Volume d'eau

$t$  : temps en s

Le débit peut être obtenu aussi par l'équation générale suivante

$$Q = v \times A \quad (2)$$

Avec

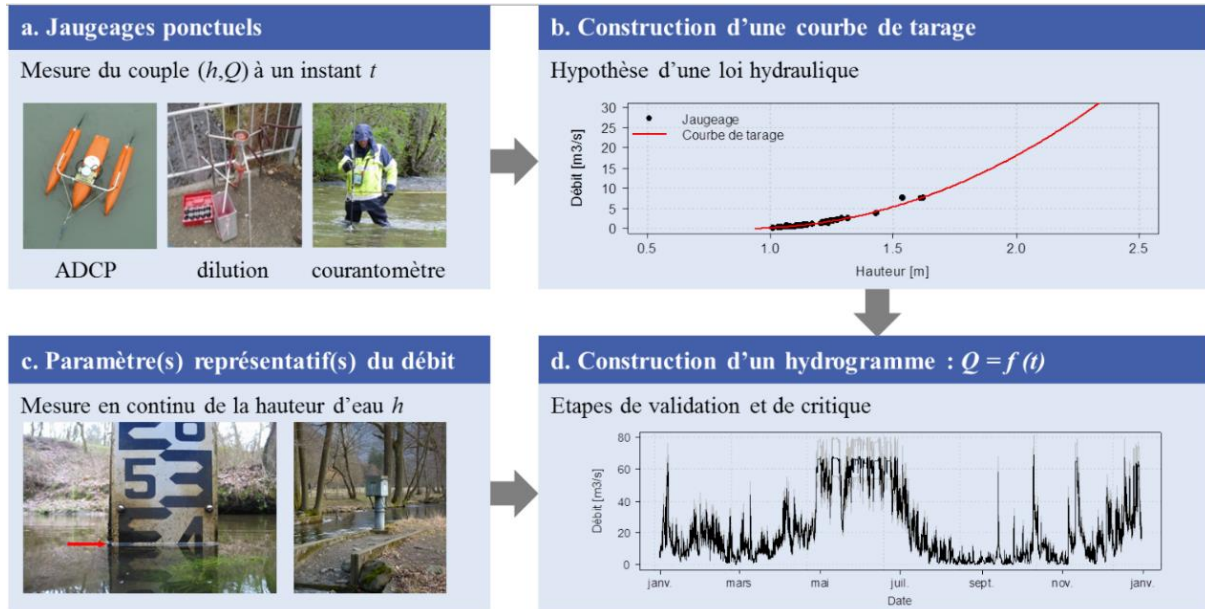
$A$  : aire de la section mouillée perpendiculaire à l'écoulement, exprimé en m<sup>2</sup>. Elle est égale au produit de la largeur transversale de l'écoulement d'un cours d'eau et de la hauteur de ce cours d'eau.

$v$  : vitesse d'écoulement en m/s

### 2.2 Estimation du débit des rivières par les méthodes conventionnelles

Il existe plusieurs approches de mesure du débit en rivière et qui peuvent être divisées en deux catégories : les méthodes directes et les méthodes indirectes ( Dottori et al., 2009). Les méthodes directes appelées aussi méthodes de jaugeage (Figure 2.1) consistent à effectuer des estimations ponctuelles du débit mesuré à un instant  $t$ . Les méthodes indirectes consistent à estimer la valeur de débit à travers la mesure d'une grandeur intermédiaire qui est généralement la hauteur d'eau. Les méthodes indirectes ont tendance à être utilisées de manière plus opérationnelle dans les différentes sections des rivières. Cette approche repose sur l'utilisation d'une courbe de tarage (Herschy, 1993) qui permet d'établir la relation entre la hauteur d'eau et le débit. Pour ce faire, il faut effectuer une campagne sur le terrain pour mesurer ponctuellement le couple couple hauteur-

débit. Une mesure de la hauteur est ensuite directement convertie en débit au moyen de cette courbe de tarage développée, Les principales techniques de jaugeage sont décrites succinctement ci-après (Fourquet, 2005; Le Coz et al., 2007).



Source : (Despax, 2017)

**Figure 2.1** principaux méthodes et outils de mesure de débit.

- **Les jaugeages par dilution**

Les jaugeages par dilution sont des méthodes volumétriques qui se base sur l'injection d'un traceur en solution dans le cours d'eau et à en suivre sa concentration au cours du temps. Cette méthode est basée sur la loi de conservation de la masse, où le facteur de dilution est directement proportionnel au débit de la rivière.

- **Le jaugeage par exploration du champ de vitesses**

Ces méthodes permettent de mesurer les vitesses de l'écoulement. Le jaugeage peut être effectué avec plusieurs outils tels que la méthode ultrasonique et la méthode du flotteur (bâtons lestés suivant le courant). D'autres outils sont largement utilisés tels que le courantomètre ou vélocimètre. Ce dernier se base sur l'utilisation des moulinets à hélice dont la vitesse de rotation est proportionnelle à la vitesse de l'écoulement.

Depuis les années 1990, des méthodes qui se basent sur les profileurs acoustiques à effet Doppler (Acoustic Doppler Current Profiler: ADCP) ont créé une véritable avancée technologique dans le domaine de l'hydrométrie (Le Coz et al., 2007). Cet outil permet d'estimer la vitesse

d'écoulement ainsi que la profondeur de lit de chenal d'écoulement. Ils ont permis de réduire le temps de mesure sur les grandes rivières et ils sont aussi adaptés à des petites rivières.

D'autres nouvelles méthodes de jaugeage hydrométriques sont utilisées aussi. Elles se basent sur la technologie d'imagerie (LSPIV : large Scale Particle Image Velocimetry) et permettent en retour d'augmenter le nombre de données collectées (Hauet, 2009 ;Dramais et al., 2011).

Bien que ces méthodes soient largement utilisées et ont montré un grand potentiel pour mesurer les débits, elles présentent plusieurs limitations. Pour la méthode ADCP, il peut être difficile de mesurer avec précision la vitesse près d'une surface d'eau lorsque la hauteur de mesure se situe entre 30 et 50 cm par rapport à la profondeur d'immersion des transducteurs. (Le Coz *et al.*, 2008). En outre, le signal acoustique émis par l'appareil peut être dissipé dans l'eau, l'écho devenant trop faible à partir d'une certaine profondeur. Il est par conséquent difficile de l'analyser. Les méthodes de jaugeage ont d'autres limitations majeures comme les coûts en temps, les coûts de mise en place et les coûts d'entretien qui sont très élevés (Smith et Pavelsky, 2008 ; Chokmani et al., 2014).

En ce qui concerne la courbe de tarage, cette dernière présente aussi plusieurs limitations. En effet, la relation hauteur-débit peut évoluer et changer avec l'évolution brusque des paramètres hydrauliques qui se manifestent par plusieurs événements : crues, décrues, présence de glace (Wolfs et Willems, 2014). Ces phénomènes naturels ont conduit au changement de la hauteur d'eau qui ne sera plus représentative. Ces modifications provoquent de fortes incertitudes d'estimation de débit avec la courbe de tarage en cas d'évènement extrême (Pelletier, 1988). Par conséquent, cette opération peut conduire à des résultats erronés, car le calcul du débit avec la courbe de tarage nécessite la mesure de plusieurs grandeurs. Comme toutes les grandeurs physiques, elle est sujette à des erreurs, qui doivent être quantifiées par une analyse d'incertitude. Ces erreurs peuvent provenir de plusieurs sources, telles que la sensibilité de la station, la manière dont la courbe de tarage est tracée et la précision du capteur utilisé pour estimer la hauteur d'eau ( Schmidt, 2003; Morlot, 2014; Le Coz *et al.*, 2016).Autre limitation à mentionner aussi, de nombreuses sections de rivières ne peuvent pas être suivies par des mesures et d'observation de débit à cause de la répartition éparse des stations du réseau hydrométrique (Chokmani *et al.*, 2014a). Elles sont généralement localisées dans des zones habitées ; alors que les endroits éloignés et/ou à faible population, ils se retrouvent peu ou pas observées. Donc, la répartition des stations hydrométriques n'est pas optimale (Chokmani *et al.*, 2014a).



### 2.3 Estimation du débit des rivières à partir de la télédétection satellitaire

L'estimation du débit d'une rivière avec les données de télédétection satellitaire a vu un intérêt croissant durant la dernière décennie (Alsdorf, 2003; Bjerklie *et al.*, 2003; Alsdorf *et al.*, 2007). L'estimation du débit de rivières à l'aide d'une telle approche a permis de fournir un moyen pour maintenir et étendre le réseau d'observation hydrométrique (Bjerklie *et al.*, 2003). Sur la base des études précédentes (Bjerklie *et al.*, 2003; Alsdorf *et al.*, 2007), ils existent plusieurs approches pour estimer le débit avec des données satellitaires. Ces approches sont basées notamment sur la relation traditionnelle entre le débit instantané  $Q$  et la géométrie d'une section transversale du chenal, précisément, la largeur  $w$  (m), la profondeur  $d$  (m), et la vitesse d'écoulement  $v$  (m/s) :

$$Q = w \times d \times v$$

Les valeurs  $w$ ,  $d$  et  $v$  varient avec le débit  $Q$ , ceci permet d'estimer le débit à partir des courbes de tarage distinctes, qui sont construites à partir d'une seule caractéristique du chenal, que ce soit en utilisant la relation  $w$ - $Q$ , la relation  $d$ - $Q$  ou avec  $v$ - $Q$  (Martin et Ham, 2005)

Par ailleurs, ces approches pour estimer le débit des rivières peuvent être résumées comme suit :

- Mesurer la variation du niveau d'eau en utilisant les données altimétriques satellitaires. Ces mesures sont ensuite converties en débit d'eau en se basant sur la courbe de tarage entre le "niveau d'eau" dérivé du satellite et le débit mesuré in situ ( Koblinsky *et al.*, 1993; Calmant *et al.*, 2008; Birkinshaw *et al.*, 2010; Tarpanelli *et al.*, 2013; Young *et al.*, 2015; Paris *et al.*, 2016; Schneider *et al.*, 2017).
- Corréler la surface de l'eau dérivée du satellite avec le débit mesuré in situ. Cette méthode repose sur l'hypothèse de changement de la largeur de la rivière. Elle est basée sur le principe de la corrélation de la surface de l'eau dérivée du satellite avec le débit mesuré in situ, puis elle a déduit le débit de la rivière à partir des données satellitaires en se basant sur la courbe de tarage de la surface de l'eau-débit ( Smith et Pavelsky, 2008b ; Pavelsky, 2014).
- Utiliser des équations hydrauliques pour estimer le débit des rivières en se basant sur la mesure de variables hydrauliques provenant de données satellitaires (Bjerklie *et al.*, 2003b, 2005; LeFavour *et al.*, 2005; Negrel *et al.*, 2011).

Cette méthode a constitué une alternative intéressante et puissante pour suivre les variations de niveau de l'eau et par ailleurs estimer le débit d'eau. Birkett *et al.* (2002) ont réalisé une étude sur l'estimation du débit dans les affluents du bassin versant de l'Amazonie en utilisant l'altimètre TOPEX/Poseidon. Ils ont sélectionné des cours d'eau dont la largeur était supérieure à 500 m.

les résultats de ce travail a montré qu'entre 30 - 50% des zones choisies donnent une bonne qualité d'observation. En revanche, les auteurs ont confirmé aussi que le satellite a donné de mauvais résultats pour les rivières qui ont une largeur inférieure à 1 km. Ce résultat a été confirmé dans le travail de Coe et Birkett (2004) où ils ont utilisé le même satellite pour estimer le débit avec la courbe de tarage dans les tributaires au niveau du bassin du lac Tchad. Les auteurs ont montré que cette méthode est opérationnelle que sur des cours d'eau qui ont une largeur supérieure à 1 km.

Michailovsky *et al.* (2012) ont utilisé des données altimétriques Envisat pour suivre le niveau d'eau et estimer le débit des rivières dans le bassin du fleuve Zambèze. L'objectif principal de cette étude était d'extraire des mesures précises de niveau d'eau dans les rivières qui ont des largeurs de 80 m et de 40 m. Les résultats de ce travail ont donné un RMSE relatif de niveaux d'eau variant de 0,32 à 0,72 m à différents endroits. Ces niveaux d'eau avec les données altimétriques ont ensuite été convertis en débit à l'aide de la courbe de tarage. Les résultats ont montré que les valeurs RMSE varient entre de 4,1 à 6,5% de l'amplitude moyenne annuelle jaugée in situ. Cette étude a montré le fort potentiel de mesure du niveau et du débit avec les données Envisat. Toutefois, les auteurs ont précisé aussi que la résolution temporelle d'Envisat est de 35 jours, ce qui implique une difficulté de suivi du débit d'eau d'une manière continue. Autre limitation de l'utilisation de l'altimétrie radar est que les signaux altimétriques sont souvent applicables uniquement sur les grands fleuves pour estimer le débit d'eau. La plupart des altimètres ont été conçus pour des applications océaniques et le traitement de ces signaux sur de petites rivières est souvent difficile.

Sichangi *et al.* (2016) dans leurs études ont estimé le débit quotidien des rivières à partir des données altimétriques satellitaires et des données de radiomètre spectral pour imagerie de résolution moyenne (MODIS). Ces chercheurs ont utilisé ces données pour fournir une série chronologique de l'évolution de la rivière et de sa largeur effective. L'indice d'efficacité Nash-stucliffe a été calculé et a donné un résultat entre 0.6 et 0.97, ce qui indique la puissance de la méthode dans l'estimation de débit fluviale. Les auteurs ont constaté aussi que les estimations de débit en utilisant à la fois la largeur effective de la rivière et les informations sur le niveau d'eau donnent de meilleurs résultats que celles qui n'utilisent que des données sur le niveau.

En ce qui concerne les études qui ont utilisé la largeur effective pour estimer les débits d'une rivière, on note par exemple l'étude de Brakenridge *et al.* (2007) qui illustre parfaitement ce concept. Les auteurs ont montré que la largeur effective estimée par des données satellitaires en micro-ondes passives est fortement corrélée avec le débit de la rivière. Tarpanelli *et al.* (2017)

ont estimé le débit de rivière à partir de la fusion entre l'imagerie optique et l'altimétrie. La recherche dans ce domaine va également au-delà de l'utilisation des images satellitaire, Huang *et al.* (2018) ont utilisé des drones avec des données satellitaires pour estimer la largeur effective des rivières sur le plateau tibétain en Asie centrale. Ces applications à de petites superficies fournissent des solutions innovantes aux problèmes de terrain. Certaines études ont opté pour l'utilisation des images à haute résolution spatiale pour estimer le débit sur des rivières aussi étroites que 12 m (D Feng *et al.*, 2019).

Junqueira *et al.* (2021) ont utilisé une nouvelle approche d'estimation de débit dans la rivière Araguaia (Brésil) à l'aide de la constellation Planet CubeSats. Les auteurs ont calculé la zone de plans d'eau à partir des images Planet CubeSat. Ils ont aussi utilisé les mesures de niveau d'eau calculé par des données altimétriques JASON-2 et des mesures du débit de la rivière à l'aide de données in situ de la station conventionnelle. Ces données ont été utilisées pour analyser leurs corrélations avec les zones d'eau de CubeSat. Les résultats de cette recherche ont montré une très bonne corrélation avec un R2 égale 88.83% entre les données extraites de CubeSat et les données extraites de JASON-2. Une corrélation significative a été trouvée aussi entre les données de CubeSat et les données in situ avec un R2 égal à 96.49. Ceci a montré que les images CubeSat peuvent être utilisées comme une station virtuelle CubeSat (CVS) pour estimer le débit de la rivière. L'inconvénient des données CubeSat est qu'ils ont souvent de moins bonnes qualités de résolution radiométrique par rapport aux satellites conventionnels.

Des études ont utilisé aussi des données radars pour estimer le débit à partir de la largeur effective, Smith *et al.* (1995) qui ont été parmi les premiers qui à utiliser des données RSO de ERS-1 pour estimer le débit à partir de la largeur effective. L'étude a été effectuée en Colombie-Britannique sur la rivière tressée 'Iskut' sur un tronçon de 10 km. Les résultats de cette étude ont montré une forte corrélation entre la largeur effective et le débit avec un  $R^2 = 0.89$ . Chokmani *et al.* (2014) ont utilisé des données RSO de RADARSAT 1&2 pour estimer le débit des rivières à partir de la largeur effective à l'est du Canada. La comparaison des résultats avec des données in situ a donné une bonne corrélation  $R^2 = 0.72$ .

D'autres méthodes ont été développées en utilisant des modèles hydrauliques pour estimer le débit des rivières. Les paramètres des modèles sont calculés à partir de données de télédétections (Fleischmann *et al.*, 2016; Kebede *et al.*, 2020) . Kebede *et al.* (2020) ont estimé le débit quotidien des rivières dans les régions de haute montagne du plateau tibétain avec trois modèles hydrologiques différents tout en combinant des variables hydrauliques de surface des rivières directement dérivées à partir des données satellitaires (pente et largeur combinées à un

coefficient de rugosité estimé rationnellement) et d'autres variables indirectement dérivées d'équation empiriques. L'équation de Manning fait partie de ces modèles. La performance globale des résultats de ce modèle par rapport au débit in situ montre une très bonne corrélation avec la valeur NSE de 0,996 et RMSE de  $20,2 \text{ m}^3\text{s}^{-1}$  pour le modèle. Cette méthode a contribué grandement à l'amélioration des informations de mesure du débit fluvial.

Bien que les différentes méthodes d'estimation de débit d'eau avec la largeur effective ont montré une forte corrélation entre les données observées et les données in situ, ces méthodes possèdent leurs propres limitations. En effet, l'inconvénient de l'utilisation des données optiques pour estimer la largeur effective réside dans le fait que l'acquisition des données optiques est souvent très réduite par la présence de couvert nuageux (Horritt *et al.*, 2003). En outre les images optiques ne permettent pas de détecter les plans d'eau sous un couvert végétal dense (Sanyal et Lu, 2004). Les limites de l'utilisation des données de micro-ondes passives pour estimer la largeur effective sont expliquées par leur faible résolution spatiale (de l'ordre de 20 à 60 km, selon les capteurs). Pour les données radar, un des problèmes récurrents à l'utilisation des données RSO pour estimer la largeur effective c'est le haut niveau de bruit qui dégrade la qualité des images et rend difficile leur interprétation (Hahmann *et al.*, 2008). Une autre limitation des méthodes d'estimation de débit en utilisant les paramètres mesurés à distance ( $w$ ,  $d$ ,  $v$ ) et le débit mesuré in situ est que ces approches ne sont pas applicables à tous les environnements fluviaux (LeFavour et Alsdorf, 2005). Pour certains types de rivières, par exemple une section encaissée, les changements de hauteur de l'eau entraîne des changements négligeables de la largeur, mais des changements importants dans le débit. Ceci exclut l'utilisation de la méthode de mesure du débit à partir de la largeur effective. Une autre limitation majeure de ces méthodes est que l'estimation de débit de la rivière est basée sur la construction d'une courbe de tarage qui nécessite la mesure de plusieurs grandeurs qui sont source d'erreur et d'incertitude.

C'est pour cette raison que dans le présent travail, le débit de la rivière est estimé en se basant sur une approche d'apprentissage profond et sans avoir recours à la construction de la courbe de tarage. Dans la section ci-dessous, le potentiel de l'apprentissage profond en hydrologie sera présenté.

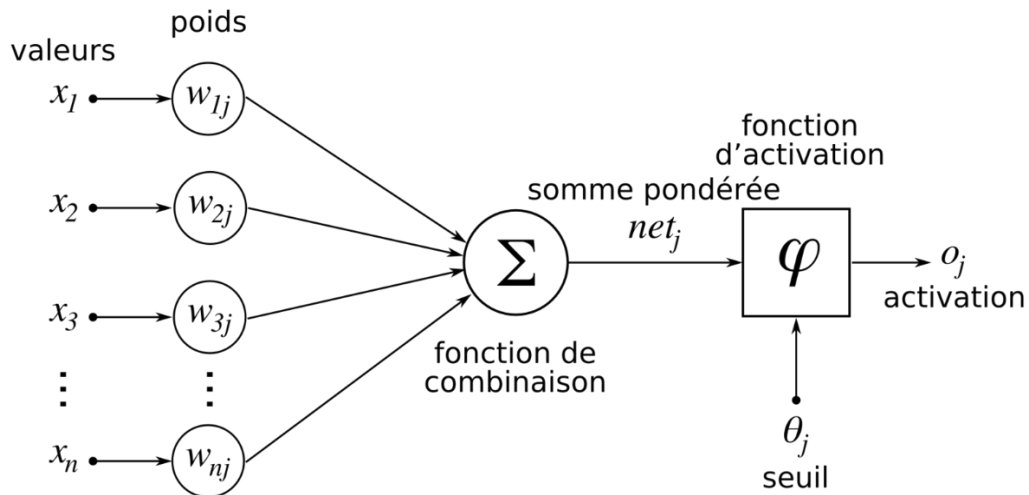
## **2.4 Apprentissage profond**

### **2.4.1 Réseaux de neurones**

L'algorithme de l'apprentissage profond (DL) est un sous-domaine du *Machine Learning* et du réseau de neurones artificiel (ANN). Les ANN ne sont pas un nouveau concept. Ils sont apparus dans les années 1940 avec le neurophysiologiste Warren McCulloch et le mathématicien Walter Pitts (McCulloch et Pitts, 1943). Mais c'est seulement depuis quelques années que le domaine de l'intelligence artificielle a attiré l'attention des chercheurs de diverses disciplines scientifiques pour plusieurs raisons notamment, grâce au développement récent dans les processeurs graphiques (GPU) qui a rendu le traitement des algorithmes d'apprentissage profond extrêmement rapide par rapport à un processeur central d'ordinateur (CPU) ( Goodfellow *et al.*, 2016; Chollet, 2018), sachant qu'avant ce développement, il était impossible d'utiliser une architecture de réseau de neurones en raison de la lenteur temporelle des CPU. De plus, lors de la compétition ImageNet en 2012 pour la reconnaissance d'images, Krizhevsky *et al.*, (2012) ont proposé un algorithme de DL efficace appelé AlexNet qui a permis de classer les images à haute résolution de l'ensemble d'entraînement ImageNet LSVRC-2010. Depuis ce moment-là, les algorithmes d'apprentissage profond n'ont pas cessé de se développer et de s'améliorer favorisant l'application de la DL à la résolution de problèmes concrets.

### **2.4.2 Fonctionnement des réseaux de neurones artificiels**

La forme traditionnelle d'un ANN est le perceptron multicouche (MLP). Un MLP est composé d'une couche d'entrée, d'une ou plusieurs couches cachées intermédiaires et une couche de sortie qui sont connectées ensemble. Ces couches reçoivent des informations provenant des entrées et les transmettent jusqu'à la dernière couche (Figure 2.2). Chaque couche est composée d'un neurone artificiel auquel un poids est associé. Ce dernier décide quel signal est important et à quelle force il doit avoir un impact sur la sortie. Une fonction d'activation est utilisée dans le MLP qui multiplie l'entrée par le poids et la transmet au neurone suivant en ligne.



Src : (Kapoor *et al.*, 2022)

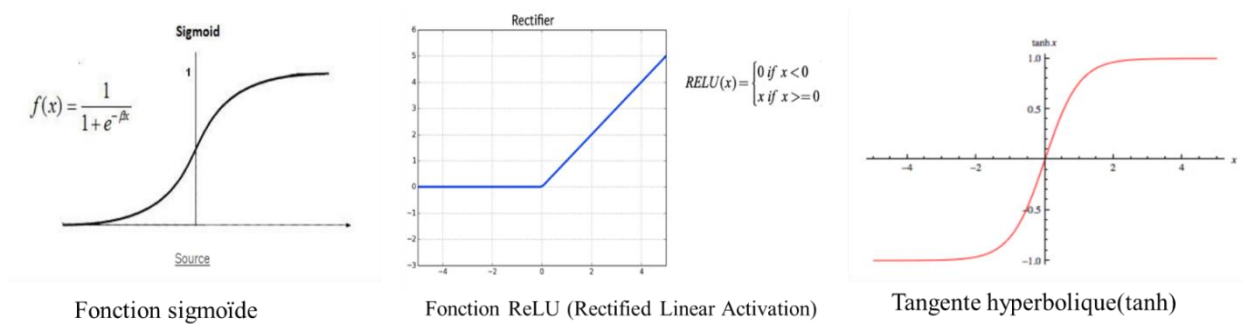
**Figure 2.2** Représentation d'un perceptron simple

Le but de la fonction d'activation est d'introduire une non-linéarité dans un modèle qui pourrait seulement apprendre des transformations linéaires. Les fonctions d'activations les plus utilisées (Figure 2.3) sont :

**La fonction sigmoïde** : elle retourne une valeur comprise entre 0 et 1. Cette fonction est très utile dans la couche de sortie, elle permet d'obtenir un score qui pourrait se traduire comme une probabilité.

**La fonction ReLU** : les scientifiques ont découvert que les couches ReLU (rectified linear unit) fonctionnent beaucoup mieux que les autres fonctions (Hinton *et al.*, 2006) parce que le réseau est capable de s'entraîner beaucoup plus rapidement. En effet, la couche ReLU applique la fonction  $f(x) = \max(0, x)$  à toutes les valeurs d'entrée. Elle permet d'éviter des valeurs négatives à la sortie du neurone puisqu'elle met à zéro toutes les valeurs négatives.

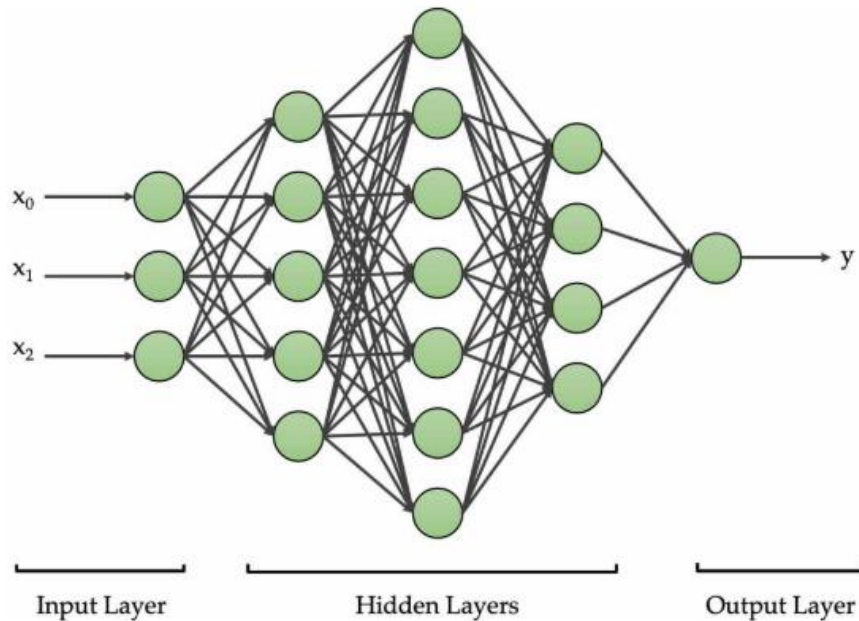
**La fonction tangente hyperbolique** : très similaire à la fonction sigmoïde. Cependant ici les valeurs vont de 1 à 0 et de 0 à -1.



**Figure 2.3** Différents type de fonctions d'activation

En effet, pendant la phase du développement du modèle ANN et à travers ses différentes couches (Figure 2.4), les données se transforment et vont avoir une nouvelle présentation. La dernière couche du ANN permettra d'obtenir une valeur prédite. Grâce à une fonction de perte ou fonction de coût qui mesure la performance de l'algorithme, cette valeur prédite sera ensuite comparée à la vraie valeur des données. Cet exercice permettra d'obtenir un score de perte qui indique l'écart d'erreur entre la valeur prédite et la vraie valeur. Si l'écart d'erreur est grand, une fonction d'optimisation sera utilisée pour ajuster les poids et minimiser l'écart d'erreur. La technique la plus courante consiste à utiliser l'algorithme de rétro-propagation (Rumelhart *et al.*, 1985) qui permet de calculer le gradient de la fonction d'erreur par rapport à chaque poids. L'algorithme de descente du gradient stochastique et ses variations telles que AdaGrad, RMSprop ou Adam sont couramment utilisées pour mettre à jour les poids et minimiser l'écart de l'erreur.

Les données d'entrée seront donc de nouveau exposées au réseau d'ANN avec des poids ajustés. Les étapes ci-dessus seront répétées jusqu'à on obtient un écart d'erreur minimal entre la valeur prédite et la vraie valeur (Figure 2.4). Il est probable aussi que la précision obtenue après entraînement n'est pas jugée suffisante. Dans ce cas, des modifications de l'architecture développée sont nécessaires. Cela signifie de modifier, notamment les paramètres, ou ajouter/enlever les couches afin de trouver une architecture qui modélise correctement le jeu de données en question.



Source : (Sit *et al.*, 2020)

Figure 2.4 Architecture d'un réseau de neurones

### 2.4.3 Réseaux de neurones convolutifs

Un réseau de neurones convolutifs (CNN) est type d'algorithme très utile pour traiter les images satellites. La structure typique d'un CNN (Figure 2.5) se compose principalement d'une couche d'entrée, d'une ou plusieurs couches cachées, d'une couche de sortie (Anwar *et al.*, 2017). Les couches cachées d'un CNN sont composées de couches convolutionnelles, de couche de non-linéarité ou appelées aussi fonction d'activation (Krizhevsky *et al.*, 2012), de couches de regroupement (Nagi *et al.*, 2011) et des couches entièrement connectées (Wu, 2016).

**La couche d'entrée (Input layer)** permet de manipuler facilement les données d'entrée de l'ensemble du CNN sans détruire sa structure interne. Elle représente généralement la matrice de pixels de l'image.

**Une couche de convolution (Convolutional layer)** est responsable de l'extraction des caractéristiques (features) dans les images d'entrée de la couche précédente et transmet le résultat à la couche suivante. Cela est effectué grâce à l'utilisation d'un filtre qui balaye toute l'image d'entrée et calcule le produit de convolution entre les poids du neurone convolutif (filtre) et chaque portion de l'image balayée. L'étape de convolution est généralement suivie par une opération de transfert non linéaire qui est la fonction d'activation.



**La couche de regroupement (pooling layer)** permet d'alléger et simplifier le calcul, en réduisant la taille des images d'entrée en ne conservant que les caractéristiques (feature) les plus importantes.

La sortie des couches de convolution et des couches de regroupement est l'entrée des **couches entièrement connectées (Fully connected layer)**. Cette dernière se connecte à tous les éléments de la couche précédente. L'image dans cette étape est mise à plat dans un vecteur vertical et envoyé dans un réseau de neurones artificiel qui va alors optimiser la combinaison de features.

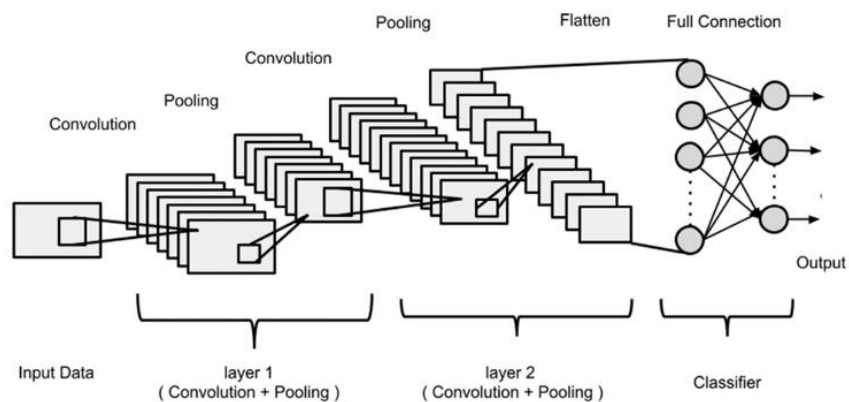


Figure 2.5 Représentation générique d'un réseau de neurones convolutifs (Sengupta et al., 2020)

#### 2.4.4 Méthode d'apprentissage profond en hydrologie

Depuis plusieurs années, le concept d'apprentissage profond (DL) a connu une très forte croissance et popularité dans tous les domaines scientifiques. En hydrologie, les techniques de réseaux de neurones et DL ont été appliquées avec succès notamment pour l'extraction d'information hydrométéorologique et hydrologique à partir d'image (Xiaodong *et al.*, 2016; Ling *et al.*, 2019) et dans la modélisation des variables hydrologiques (Bai *et al.*, 2016; Fang et Shen, 2017; Zhang *et al.*, 2017). Une conclusion commune de ces études est que les modèles DL ont dépassé les méthodes statistiques traditionnelles. On note par exemple, Ling *et al.* (2019) ont proposé une nouvelle approche pour améliorer l'estimation de la largeur d'une section d'une rivière mouillée (RWW) à partir d'image de télédétection à résolution grossière. D'abord, les auteurs ont appliqué la méthode de démixage spectrale à l'imagerie pour obtenir une carte de fraction d'eau indiquant les proportions de pixels de l'image correspondant à la classe d'eau. Ensuite, une carte de rivière à résolution fine est produite à partir de l'image de fraction d'eau à résolution grossière en utilisant une méthode de cartographie à super résolution (SRM) basée

sur un réseau de neurone convolutif (CNN). Finalement, les auteurs ont pu estimer la RWW à partir de la carte fluviale à résolution fine résultante. Les résultats de cette méthode ont montré que le modèle SRM basé sur un réseau neuronal convolutif peut estimer efficacement les détails à l'échelle des sous-pixels des rivières et que la précision de l'estimation de RWW est considérablement plus élevée (avec un  $R^2= 0.8842$ ) que celle obtenue à partir de l'utilisation d'une classification des images (avec un  $R^2= 0,6917$ ). Dans une autre étude, le DL a été utilisé pour améliorer la précision de l'estimation de la précipitation à partir de l'imagerie satellitaire (Tao *et al.*, 2017). Dans cette étude, les auteurs ont utilisé une méthode de DL qui est l'algorithme SDAE (Stacked Denoising Auto-Encoders) avec quatre couches et 1000 nœuds cachés pour extraire les précipitations à partir d'images du satellite géostationnaire opérationnel pour l'environnement (GOES) avec nuages de  $15 \times 15$  pixels. Les résultats de cette étude ont montré des améliorations significatives à la fois dans la détection de la pluie/non-pluie et dans la quantification du taux de précipitation. Les auteurs ont trouvé que la méthode SDAE a apporté des corrections de 33 % à 43 % sur les pixels de fausse alarme par rapport à une autre méthode qui s'appelle PERSIANN-CCS (PERSIANN Cloud Classification System).

Bai *et al.* (2016), ont montré dans leur étude le potentiel de la méthode de DBN (Deep belief net) qui est une méthode de DL dans la prévision des flux de réservoirs. L'étude a été basée sur une série de flux quotidiens historiques (du 1/1/2000 au 31/12/2012) du barrage des Trois Gorges au centre de la Chine. À titre de comparaison, les auteurs ont utilisé quatre autres modèles conventionnels pour effectuer le même travail. Les résultats montrent que le modèle avec DL dépasse tous les autres modèles homologues en termes d'erreur moyenne en pourcentage absolu (MAPE = 11,2896%) et selon le coefficient de détermination ( $R^2 = 0,8905$ ). Il convient de noter que ce travail a montré la capacité de DL dans la conception des caractéristiques multiéchelles et à effectuer des prédictions raffinées concernant les flux de réservoirs. Zhang *et al.* (2017) ont construit différents modèles de réseaux neuronaux, y compris le LSTM (long short-term memory) et un réseau récurrent fermé pour simuler et prédire le niveau d'eaux usées provenant de la structure de surveillance des débordements d'égouts (CSO). Les auteurs ont montré les avantages du LSTM par rapport aux méthodes plus simples. Plusieurs autres études scientifiques ont utilisé la méthode de l'apprentissage profond dans d'autres volets de l'hydrologie, telles que la prédiction de ruissellement et la prévision des crues (Bhola *et al.*, 2019; Damavandi *et al.*, 2019; Kratzert *et al.*, 2019; Kumar *et al.*, 2019; Qin *et al.*, 2019), l'étude des eaux souterraines et l'estimation de niveau d'eau de la nappe phréatique (Jeong *et al.*, 2020; Jeong et Park, 2019; Mo *et al.*, 2019; Zhou *et al.*, 2020) et dans la surveillance et la prévision de

la qualité de l'eau (Banerjee *et al.*, 2019; Yu *et al.*, 2020). Toutes ces études ont montré le potentiel de DL par rapport à l'utilisation des méthodes traditionnelles.

## 3 MATÉRIEL ET MÉTHODES

---

### 3.1 Sites à l'étude

Dans ce travail, la zone d'étude est située dans l'écozone du bouclier boréal à l'Est du Canada (Figure 4.1, article). Le bouclier boréal est la plus grande écozone terrestre du Canada. Il couvre 1,8 million de kilomètres carrés et représente 18,2 % de la surface terrestre du pays (Eamer, 2015). La zone d'étude s'étend sur deux provinces : le Québec et le Labrador (Figure 4.1, article).

### 3.2 Données satellitaires

Les données RADARSAT-1 et RADARSAT-2 en orbite ascendant et descendant ont été utilisées dans cette étude. Les coefficients de rétrodiffusion radar, notés  $\sigma^0$  d'images radar à synthèse d'ouverture (RSO) dans cette étude sont calculées à partir des produits Ground-Range-Detected.

Les images choisies sont en mode d'acquisition fin, avec une résolution spatiale de 8 m est une polarisation HH. Un total de 600 images (entre 1997 et 2013) a été téléchargé et prétraité pour fin de calibration et validation du modèle CNN pour l'estimation du débit des rivières. Ces images couvrent toutes les stations hydrométriques de la zone d'étude (Figure 4.2, article).

### 3.3 Données hydrométriques

Un total de 39 stations hydrométriques (Table1, annexe1) qui s'étalent sur la zone d'étude ont été utilisées. Des mesures des débits instantanés pour la période 1997 à 2013 ont été utilisées pour cette étude. Au niveau de ces stations de mesure, les largeurs de rivières varient entre 16 et 431 mètres et les débits d'eau varient entre 0.1 m<sup>3</sup>/s et 750 m<sup>3</sup>/s.

Au niveau de la province du Québec, les données ont été acquises auprès du direction de l'expertise hydrique (DEH). Quant aux stations situées dans le Labrador, elles ont été obtenues via Environnement Canada website.

### 3.4 Données topographiques

Dans cette étude, les données topographiques sont extraites de modèles numériques de terrain (MNT) SRTM à résolution spatiale de 30 x 30 m. Les données couvrant la zone d'étude ont été téléchargées du serveur web de la NASA (<https://earthexplorer.usgs.gov>) et traités pour calculer

la Hauteur au-dessus du drainage le plus proche (Height Above Nearest Drainage (HAND)) (Nobre *et al.*, 2011; Huang *et al.*, 2017).

Le modèle HAND permet de calculer l'élévation de chaque point du bassin versant au-dessus du cours d'eau le plus proche de sa direction d'écoulement. Il normalise la topographie par rapport aux hauteurs relatives locales qui se trouvent sur le réseau de drainage (Huang *et al.*, 2017; Nobre *et al.*, 2011).

Pour ce travail, le produit de HAND a été ré-échantillonné de 30 m à la même résolution spatiale que les images radar (8 m). Pour effectuer ces étapes, l'outil pc raster (une librairie Qgis) et jupyter notebook ont été utilisés pour automatiser le traitement.

### **3.5 Méthodologie proposée**

#### **3.5.1 Prétraitement des données d'entrée**

Afin d'entraîner un algorithme CNN, il est nécessaire de séparer les données en trois jeux de données : données d'entraînement qui servent à calibrer le modèle et ajuster les poids, données de validation qui servent à évaluer ce modèle à chaque itération de calibration et données test qui sont entièrement indépendante, sont utilisées afin de mesurer la précision des modèles et évaluer sa performance. Le ratio optimal qui a été utilisé pour séparer les données est comme suit : 80 % pour l'entraînement, 10 % pour la validation et 10 % pour le test.

Le choix de ce ratio a été sélectionné en se basant sur les performances obtenues lors des essais.

Les données d'entrée considérées pour cette approche sont les données RSO et le HAND. Une série de prétraitement a été effectuée sur ces données. D'abord, une étape de découpage a été appliquée sur ces données (Figure 4.5, article). Sur une zone tampon de 5 km autour des stations hydrométriques, les images radar ont été découpées en des petites imagerie tout au long des rivières de largeur de 1 km. La même valeur du débit a été attribuée à toutes les imagerie. Ceci a permis d'augmenter le nombre des imagerie pour l'étape de l'entraînement du modèle. La taille de la base de données est alors passée de 600 images à 10 000 images.

Ensuite, une fois le découpage des images effectué, il a été nécessaire de procéder à la normalisation des imagerie. Cette étape est importante, car elle permet d'uniformiser les données d'entrée au modèle ainsi que la variance unitaire. Enfin et pour faire face aux problèmes du surapprentissage (overfitting), une augmentation de données a été appliquée. Le nombre total

de la base de données est passé de 10 000 à entraîner à 82 000 imageries. L'augmentation de données a été appliquée sur les données d'apprentissage et les données de validation.

### **3.5.2 Entraînement et évaluation du modèle d'apprentissage profond**

La Figure 4.9 (article), le tableau 2 et 3 (annexe) montrent la meilleure conception selon nos tests en termes de nombre de couches utilisées et des valeurs des hyperparamètres, qui sont appropriées aux données d'entrées.

Le modèle CNN dans ce travail est composé de 7 couches de convolutions, 3 couches de regroupement. La sortie des couches de convolution et des couches de regroupement est l'entrée des couches entièrement connectées (Fully connected layer). Dans ce travail, les couches de mise en commun de la moyenne globale (Global Average Pooling Layers (GAP)) ont été utilisées à la place des couches entièrement connectées. Les couches GAP sont une opération de mise en commun, conçue pour remplacer les couches entièrement connectées (Al-Sabaawi *et al.*, 2021). Pour améliorer la généralisation de l'entraînement et éviter le surapprentissage, la technique de Couches d'abandon (Dropout Layers) (Srivastava *et al.*, 2014) a été utilisée.

Cette architecture est le résultat de nombreux essais et tests. Au total, 3,536,617 paramètres ont été entraînés par ce modèle.

### **3.5.3 Évaluation de la performance du modèle**

Après l'entraînement du modèle, les paramètres du réseau ont été utilisés pour tester la performance et la généralisation du modèle avec un jeu de données de test, qui n'ont jamais été utilisées avant. Les indices de performances statistiques choisis sont : coefficient de détermination ( $R^2$ ), la racine de l'erreur quadratique moyenne (RMSE) et l'efficacité de Nash-Sutcliffe (NSE).



**PARTIE 2 : Article**





## 4 RIVER FLOW ESTIMATION WITH RADARSAT IMAGERY USING DEEP LEARNING

---

### Auteurs :

Samar Ziadi<sup>1</sup>, Karem Chokmani<sup>1</sup>, Chayma Chaabani<sup>1</sup>, Anas El\_Alem<sup>1</sup>

<sup>1</sup> *Centre Eau Terre Environnement, Institut national de la recherche scientifique (INRS), 490 rue de la Couronne, Québec, QC G1K 9A9, Canada ; [samar.ziadi@inrs.ca](mailto:samar.ziadi@inrs.ca) (SZ),*

*[karem.chokmani@inrs.ca](mailto:karem.chokmani@inrs.ca) (KC), [anas.el\\_alem@inrs.ca](mailto:anas.el_alem@inrs.ca) (AE), [chayma.chaabani@inrs.ca](mailto:chayma.chaabani@inrs.ca) (CC)*

### Corresponding author:

Samar ZIADI : [samar.ziadi@inrs.ca](mailto:samar.ziadi@inrs.ca)

Tel: +1873-200-0235

INRS-ETE, 490 de la couronne, Québec, QC, G1K9A9, Canada

(Prêt à soumettre)

### Contribution des auteurs :

Conceptualisation, SZ et KC ; méthodologie, SZ et KC ; Analyse formelle et interprétation, SZ et KC ; Développement du code, SZ ; Rédaction de l'article, SZ ; Supervision, KC ; Soutient technique au développement d'une partie du code, CC ; Révision et édition du document, AE, KC, CC

Tous les auteurs ont approuvé la version publiée du manuscrit

## **Abstract**

Estimating river flow is a key parameter for effective water resources management, flood risk prevention and hydroelectric facilities planning. In cases of very high flow of water or extreme events, traditional gauging methods cannot be reliable. In addition, hydrometric network stations are often sparse, and their spatial distribution is not optimal. Therefore, many river sections cannot be monitored using traditional flow measurements and observations. For these reasons, satellite sensors are considered as a complementary observation source to traditional water level and flow observations in the last decades. The use of this kind of approach has provided a way to maintain and expand the hydrometric observation network.

Remote sensing data can be used to estimate flow from rating curves that relate the instantaneous flow ( $Q$ ) to the geometry of a channel cross-section (the effective width or depth of the water surface). On the other hand, remote sensing is also associated with limitations, notably its dependence on the rating curves. Indeed, due to their empirical nature, rating curves are limited to specific sections and cannot be applied in other rivers. Recently, deep learning techniques have been successfully applied in many fields, including hydrology.

In the present work, the deep learning approach has been chosen, in particular convolutional neural networks (CNN), to estimate river flow. The main objective of this work is to develop an approach to estimate river flow from RADARSAT 1&2 imagery using deep learning. In this study, 39 hydrographic sites of the Boreal Shield ecozone in Eastern Canada were considered. A new CNN architecture was developed to provide a straightforward estimation of the instantaneous river flow rate.

The achieved results demonstrated a coefficient of determination ( $R^2$ ) and Nash-Sutcliffe values of 0.91, and a root mean square error of  $33\text{m}^3/\text{s}$ . This indicates the effectiveness of CNN in automatic flow estimation with satellite sensors.

**Key words:** Deep learning, CNN, rating curve, flow, water level, radar images.

## 1. Introduction

The accurate Prediction and management of stream flows is critical for flood monitoring, environmental impact assessments and watershed hydrologic modeling (Barbetta *et al.*, 2012b; F. Dottori *et al.*, 2009b; Guerrero *et al.*, 2012b; Smith et Pavelsky, 2008b).

The traditional approach of estimating river flows is based on the continuous measurements of water levels that are fitted to rating curves. However, this approach has its limitations. It is difficult to obtain representative measurements of flows in extreme situations because there is no field staff that can go out and take measurements when an extreme event occurs (Dai et Trenberth, 2002). In addition, the stage-flow relationship can be significantly altered when river morphology changes due to erosion or solids input. This requires new gauging to calibrate the rating curve (Barbetta *et al.*, 2012b; F. Dottori *et al.*, 2009b).

It is also necessary to mention that the rating curve-based method is time-consuming and costly as it requires the maintenance of flow measurement stations. Moreover, the spatial distribution of the hydrometric network stations is generally sparse and their distribution is not optimal (Chokmani *et al.*, 2014b). They are generally located in populated areas while distant and/or sparsely populated locations are found with limited or no observations. In addition, areas with high latitudes as well as log-jams and the dominance of braided gravel bed rivers can limit river gauging (Smith et Pavelsky, 2008b). As a result, many river sections cannot be monitored by flow measurements and observations. This is reflected in the decline and decrease in the number of gauging stations and the difficulty in accessing river flow information since the 1980s (Bjerklie *et al.*, 2003b; Di Baldassarre et Uhlenbrook, 2012; Dongmei Feng *et al.*, 2019; C. J. Gleason *et al.*, 2018; Hossain *et al.*, 2013; Tourian *et al.*, 2017; Trambly *et al.*, 2010b). The constraints and limitations of instantaneous flow measurement stations affect Canada, which has 9% of the world's renewable freshwater, thus alternative flow monitoring approaches with a larger spatial coverage and scale are required (Chokmani *et al.*, 2014b).

For these reasons, during the last decades, satellite sensors are considered as a complementary observation source to traditional water level and flow observations (Ahmad et Kim, 2019; Alsdorf *et al.*, 2007b; Alsdorf et Lettenmaier, 2003; Huang *et al.*, 2018; Koblinsky *et al.*, 1993; Syed *et al.*, 2010). These data provide a potential solution to water monitoring problems (Smith, 1997). Indeed, with the increasing availability of remotely sensed data and repeated, consistent, and global measurements from space-based sensors, the integration of satellite data provides a way to maintain and expand the hydrometric observation network (Bjerklie *et al.*, 2003b). The precision

of flow estimation using remote sensing methods still depends on the accuracy of the estimated parameters obtained from satellite sensors directly (width, slope, and height) or indirectly (velocity and bathymetric depth) (Pan *et al.*, 2016; Pan et Nichols, 2013; Sichangi *et al.*, 2016, 2018). Early applications of satellite data in hydrology involved radar altimeters, namely Geosat, ERS-1, and Topex/Poseidon satellites (Koblinsky *et al.*, 1993). Satellite altimetry was an attractive alternative for recording periodic water level variations. These data were used with in situ observations to derive river discharge (Birkett *et al.*, 2002; Kouraev *et al.*, 2004). Similar studies have continued to expand and refine the capabilities of radar altimetry in flow estimation (Nathanson *et al.*, 2012; Paris *et al.*, 2016; Pavelsky *et al.*, 2014; Schneider *et al.*, 2017). Nevertheless, radar satellite altimeters were confirmed to provide inaccurate results for rivers less than 1 kilometer wide, due to the signal contamination by river banks (Birkett *et al.*, 2002). In addition, their swath widths are quite narrow (~5 km) which means that many areas are not covered, limiting their applications (Biancamaria *et al.*, 2009). The temporal frequency of satellite altimeters, which varies between 10 and 35 days, is also a limiting factor. This is an important limitation in water level monitoring and flow estimation, especially since the latter requires continuous and regular data.

Several researchers have been directed toward optical sensors to estimate water flow (Ashmore et Sauks, 2006; Dongmei Feng *et al.*, 2019; Gleason *et al.*, 2015; Gleason et Smith, 2014; Huang *et al.*, 2018; King *et al.*, 2018; Young *et al.*, 2015). The idea is conceptually similar to the altimeter method described above, except that with optical sensors, the measurement is derived by the satellite and the variable chosen to calibrate the rating curve is the flow width. On the other hand, the signal from optical sensors cannot pass through clouds which strongly affect the quality of a large amount of acquired data. In addition, optical imagery does not allow the detection of water under dense vegetation cover. In doing so, some information is not processed (Huang *et al.*, 2018).

To overcome these problems, several studies have used radar data to estimate water flow (Birkinshaw *et al.*, 2014; Durand *et al.*, 2016; Fulton *et al.*, 2020; Pavelsky *et al.*, 2014; Sichangi *et al.*, 2016; Smith, 1997; Smith *et al.*, 1996). Indeed, due to the sensitivity of the radar signal to the presence of water and due to its ability to penetrate clouds and operate independently of weather conditions, these sensors have shown interesting potential for detecting river water surfaces and also for monitoring water flow (Chokmani *et al.*, 2014b; Smith et Pavelsky, 2008b). However, despite the potential of radar data to estimate water flow, they have a recurrent problem related to the level of noise that degrades the image quality and makes direct interpretation of hydrological information difficult. Indeed, all works use a proxy for the estimation of the flow, in

particular the width. With the noise of the radar imagery the estimation of this proxy is erroneous and therefore the flow as well.

In the last two decades, artificial intelligence (AI) -based approaches have been widely used for river flow estimation. For more details, Yaseen *et al.*, (2015) prepared a complete review of the application of AI for river flow estimation. In this regard, artificial neural networks (ANNs), one of the most frequently applied models (Chang *et al.*, 2014; Üneş *et al.*, 2020) due to its strong nonlinear adjustment capability. However, ANNs cannot give an effective representation of the spatial variability of the input data or extract their characteristics, so their applications are limited in practice (Nourani, 2017).

Recently, deep learning techniques have been successfully applied in hydrology and water resources analysis (Shen, 2018; Zhang *et al.*, 2016). Several researchers have shown the potential of deep learning application in water resource management (Fang *et al.*, 2019; Zhou *et al.*, 2019), water quality parameter monitoring and estimation (Barzegar *et al.*, 2020; Liu *et al.*, 2019), groundwater loading estimation (Damavandi *et al.*, 2019; Kratzert *et al.*, 2019), and in water level forecasting (Hrnjica et Bonacci, 2019; Ling *et al.*, 2019). Furthermore, convolutional neural network (CNN) is one of the deep learning algorithms that have been successfully applied to solve water management and hydrological parameter estimation problems (Shen, 2018).

Indeed, the CNN appears to be a promising and appropriate solution to the problems of flow estimation with satellite sensors. The water dynamics and the mechanism that manages the flow are complex and highly non-linear. The flow estimation depends on several parameters such as longitudinal and lateral slope, water level, landform, etc. The approach with CNN, due to its many stacked layers, is capable of representing complex and high-dimensional spatial data features (Krizhevsky *et al.*, 2012), as well as deducing robust and scalable information while minimizing the manual intervention rate (Goodfellow *et al.*, 2016; Saba *et al.*, 2019; Sengupta *et al.*, 2020). In addition, CNNs can represent nonlinear features and handle raw and noisy data (Sahiner *et al.*, 2019).

Thus, to overcome the complexity and difficulty of using traditional approaches and to minimize the inconvenience of using remote sensing data and ANN approaches, a CNN network was deployed to develop a river flow estimation model. The main objective of this work is to estimate river flow over the boreal ecozone in Canada based on the use of a CNN model and SAR data.

## 2. Study area

In this work, the study area is located in the Boreal Shield ecozone Eastern of Canada (Figure 4.1). The Boreal Shield is the largest terrestrial ecozone in Canada. it covers 1.8 million square kilometers (Eamer, 2015) and in terms of land area, it occupies 18.2% of the country (Eamer, 2015). There are several lakes and rivers within the area that represent 22% of Canada's freshwater area. The area encompasses five provinces: Alberta, Saskatchewan, Manitoba, Ontario, Quebec and Labrador (Eamer, 2015). Approximately 30% of the forested area in the eastern part of the Boreal Shield ecozone is covered by dense coniferous forests. 13% of this area is covered by mixed coniferous and deciduous forest. 35% is covered by open forest (Colombo *et al.*, 1998).The boreal ecozone is characterized by rocky headlands and rocky lakes. The precipitation in this zone is relatively abundant, 1000 mm in the eastern part (Lowe *et al.*, 1996).The average daily temperature is -10°C to -20°C in January and 15°C to 18°C in July (Lowe *et al.*, 1996). The most common soil type in the Boreal Shield is leached soil such as bare rock and very swampy areas (« The Boreal Shield », s.d.). The hydrological regime in the eastern part of the Boreal Shield is driven by rainfall and snowmelt (Les gouvernements fédéral, 2010).

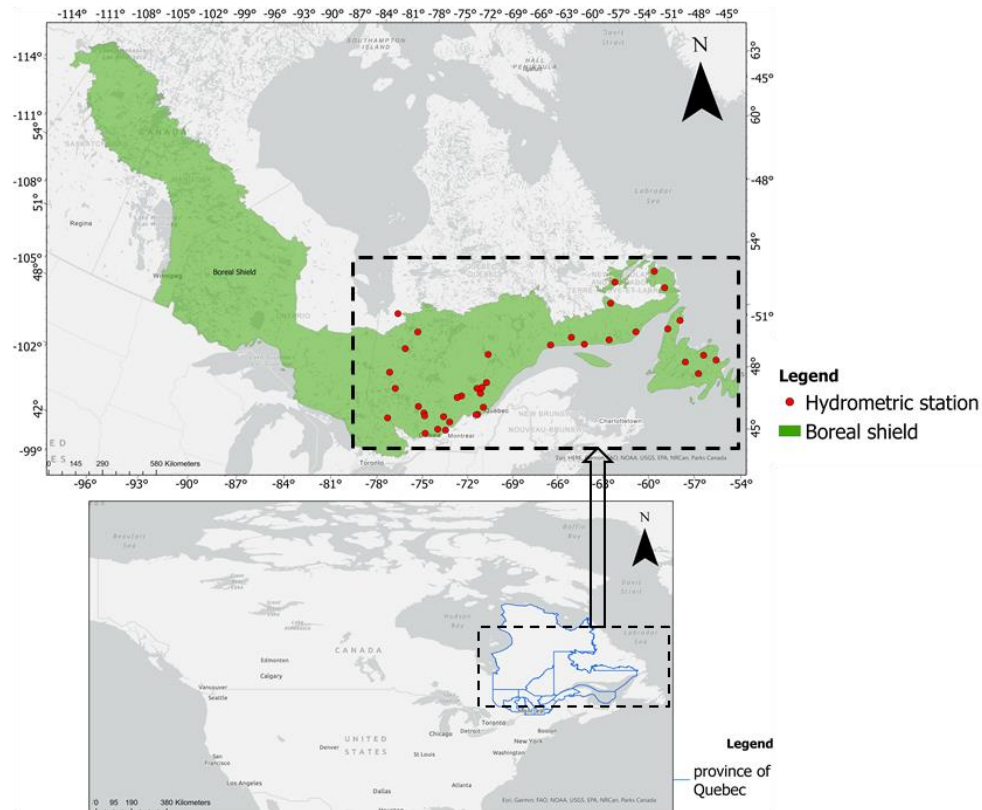


Figure 4.1 Overview of the study area and location of the hydrometric stations

### **3. Material and Methods**

#### **3.1 In-situ and satellite dataset**

##### **3.1.1. Satellite data**

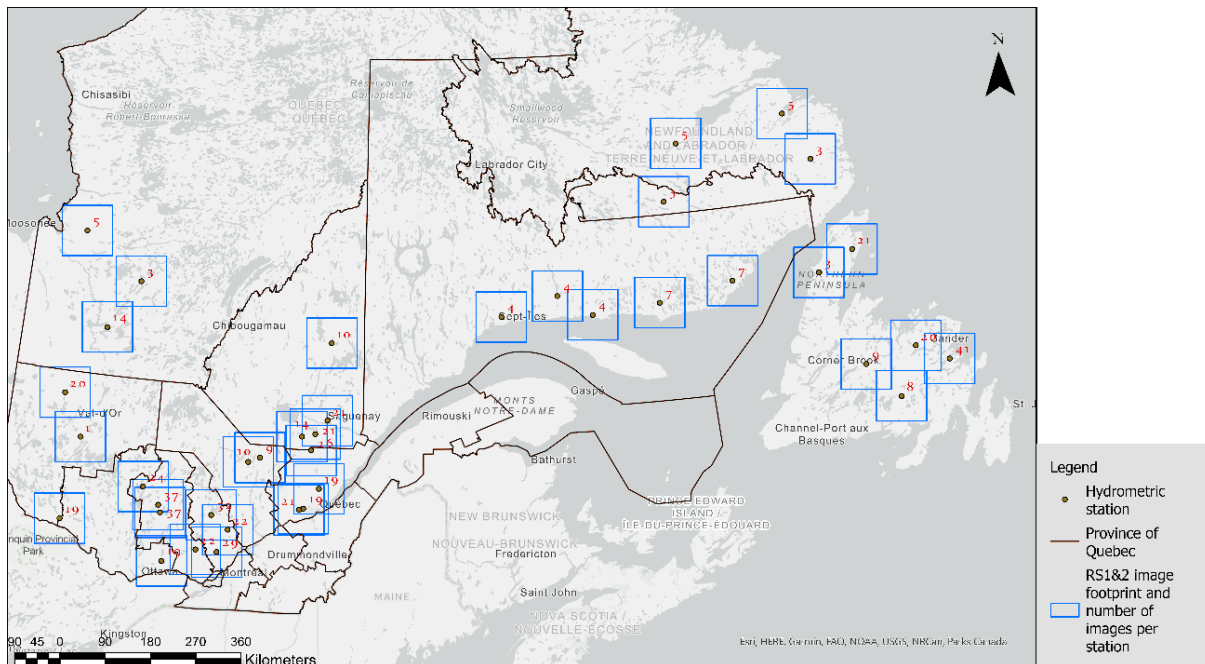
RADARSAT-1 and RADARSAT-2 (RS-1&RS-2) data in ascending and descending orbits were used in this study. These sensors emit C-band microwaves (5.6 cm wavelength), at a frequency of 5.3 GHz (RS-1) and 5.4 GHz (RS-2). They follow a helio-synchronous orbit and are equipped with a synthetic aperture radar (SAR) allowing to obtain high spatial resolution of images. The radar backscatter coefficients, noted  $\sigma_0$  of SAR images in this study, are calculated from the Ground-Range-Detected products.

The selected images are in fine acquisition mode, with a spatial resolution of 8 m and HH polarization. A total of 600 images (between 1997 and 2013) were collected and pre-processed for calibration and validation of the CNN model for river flow estimation. These images cover all hydrometric stations in the study area (Figure 4.2).

The data were acquired during summer periods between May and August, with conditions of open water flow without ice or snow.

Before data analysis, an initial pre-processing of the SAR data is necessary. Indeed, all RS-1&RS-2 images are orthorectified using orbital parameters and a digital terrain model (Geomatica Ortho Engine). A radiometric calibration was also applied. This last one allowed to obtain an image whose pixel value really corresponds to the radar backscatter coefficient of the scene. Then, the Lee SAR Speckle Filter (FLE) of size 3x3 was applied on the correctly orthorectified images in order to reduce speckle. This filter is available in PCI-Geomatica. The final  $\sigma_0$  image values are subsequently converted to logarithmic decibels (dB).





**Figure 4.2 Radar image coverage in the study zone**

### 3.1.2. In-situ gauging dataset

39 hydrometric stations (Figure 4.1, Table 4-1 Statistical information on hydrometric stations used for river flow , Appendix 1) that cover the study area were used. Instantaneous flow measurements for the period 1997 to 2013 were used for this study. These data correspond to the same dates and times of RS-1&RS-2 images (with an interval of +/- 1.5h). They are available for all the selected hydrometric stations. At these stations, river widths vary between 16 and 431 meters and water flows vary between 0.1 m<sup>3</sup>/s and 750 m<sup>3</sup>/s.

For the province of Quebec, the data were acquired from the direction de l'expertise hydrique (DEH). The stations located in Labrador were obtained from the Environment Canada website.

**Table 4-1 Statistical information on hydrometric stations used for river flow estimation.**

range of values in the stations	Width at station (in m)	Catchment area (KM <sup>2</sup> )	Flow river m <sup>3</sup> /s
minimum	16	99	0.1
mean	112	7456	80
maximum	431	92500	750
Standard deviation	105	15497	120

### 3.1.3 Topographic dataset

In this study, topographic data that are extracted from SRTM digital terrain models (DEMs) at 30 × 30 m spatial resolution were also used. The data covering the study area, were downloaded from the NASA web server (<https://earthexplorer.usgs.gov>) and processed to compute the Height Above Nearest Drainage (HAND).

In the HAND model, each point in the watershed is calculated as a function of the elevation above the nearest stream. The HAND normalizes the topography to the local relative heights found on the drainage network (Huang *et al.*, 2017; Nobre *et al.*, 2011). The HAND is calculated from the DEM using two steps:

- Modeling the water flow: To do this, the DEM is first adapted to this objective. Indeed, the DEM contains pits that pose a problem when determining the direction of flow (the pits are generally cells surrounded by higher cells). The objective is therefore to fill these pits. A hydrologically coherent DEM is thus obtained which allows to define the direction of the flow and to generate an accumulated surface grid. The latter is used to define the drainage network.
- The nearest drainage map is generated by combining data from the local flow direction and drainage network: Each pixel in the map corresponds to a DEM pixel draining to that pixel. DEM pixels are calculated by calculating their elevation difference from the nearest drainage pixel. The unit of the HAND value is the meter (Figure 4.3).

For this work, the HAND product was resampled from 30m to the same spatial resolution as the radar images (8m). To perform these steps, the pc raster tool (a library in Qgis) and jupyter notebook were used to automate the processing.

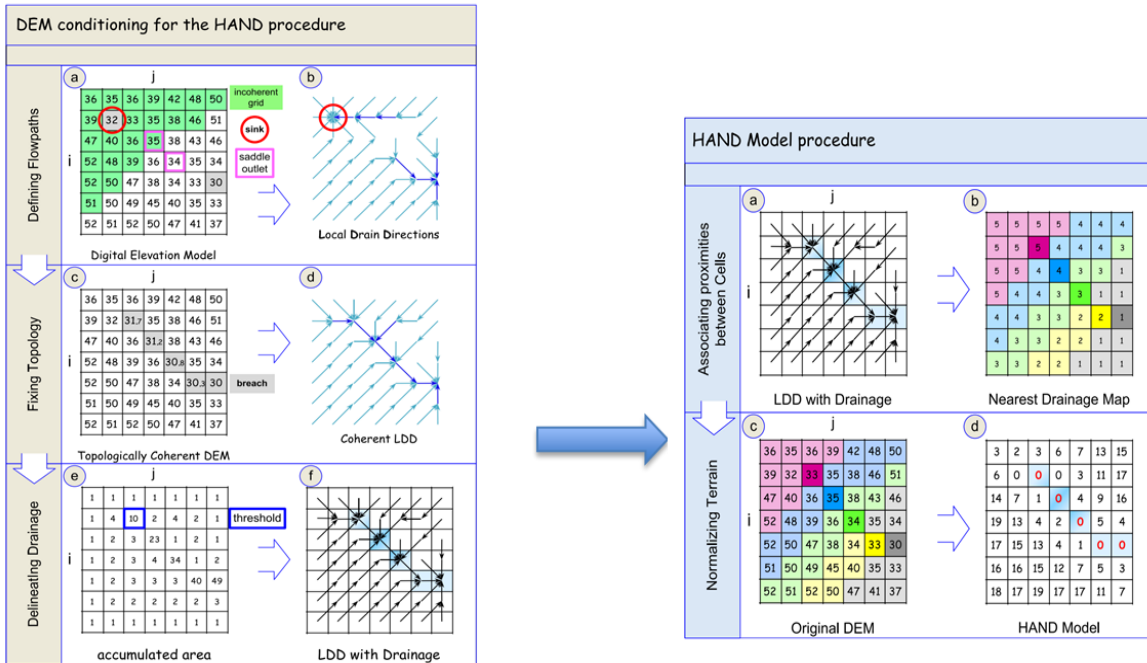


Figure 4.3 HAND computation workflow, source : (Nobre et al., 2011)

### 3.2 CNN approach for river flow estimation

The methodological approach involves two main components (Figure 4.4). The first part consists in presenting the database that served as input data for the CNN model and the second part focuses on the training and evaluation of the deep learning model.

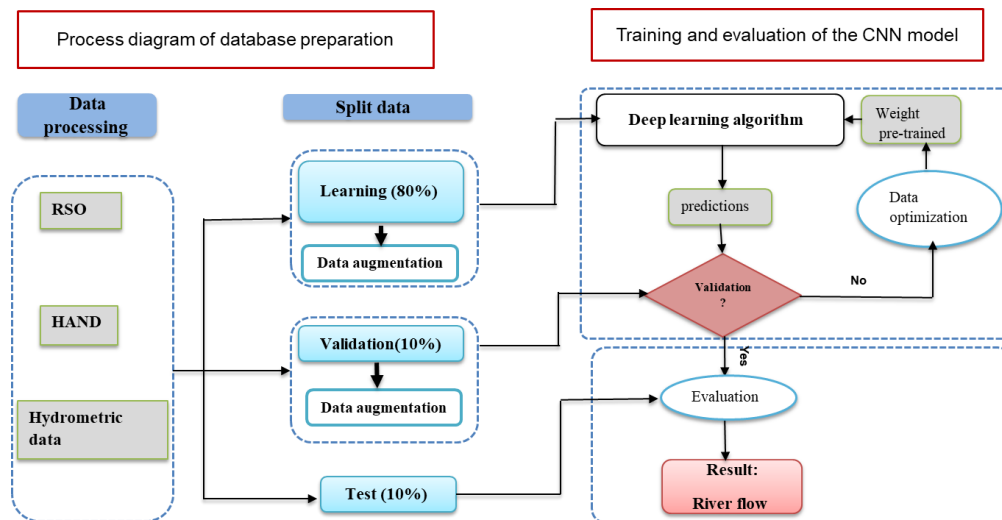


Figure 4.4 General flowchart of the river flow estimation approach

### 3.2.1 Input data pre-processing

In order to train a CNN algorithm, it is necessary to separate the data into different sets: training dataset used for calibration and adjusting weights, validation datasets used to evaluate the model at each calibration iteration and test dataset which is completely independent, that is used to measure the model's accuracy and evaluate its performance. The optimal ratio that was used to separate the data is as follows: 80% for training, 10% for validation and 10% for testing. The choice of this ratio was based on several trials. Different ratios were tested to find the best balance between optimal use of data for model training and reliable evaluation of model performance. This ratio showed the best performance in the tests.

The input data considered for this approach are the SAR data and the HAND. A series of preprocessing was performed on these data. First, a clipping step was applied on these data (Figure 4.5). In a 5km buffer zone around the hydrometric stations, the radar images were clipped into small patches along the 1km wide rivers. The same flow value was assigned to all the patches. This increased the number of thumbnails for the model training step. The size of the database was then increased from 600 images to 10000 images.

Once the clipping of the images is done, the normalization of the patches (eq. 1) is done. This step is important because it allows to standardize the input data to the model as well as the unit variance. It therefore reduces the variability between the data and ensures a more stable training.

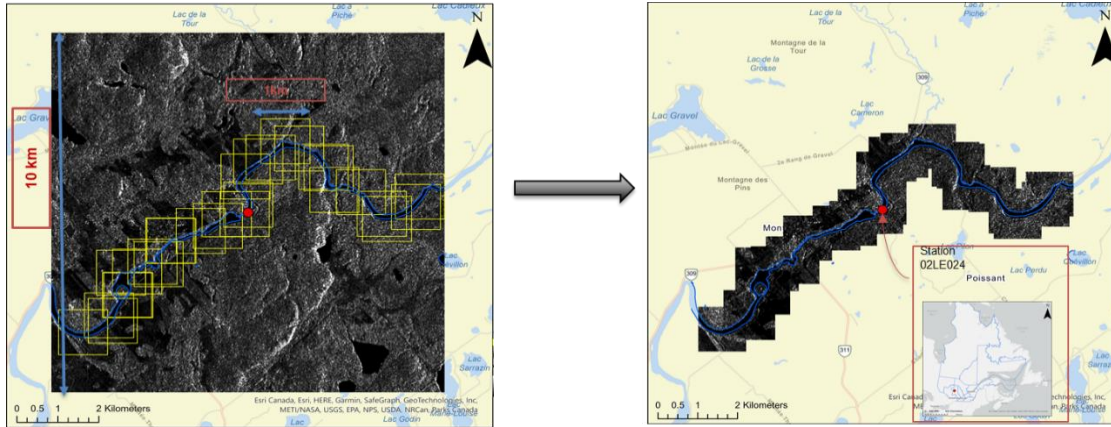
Normalization is calculated as follows:

$$z = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

with:

$z$  : Normalized pixel values,  $x$  : Pixel value before normalization

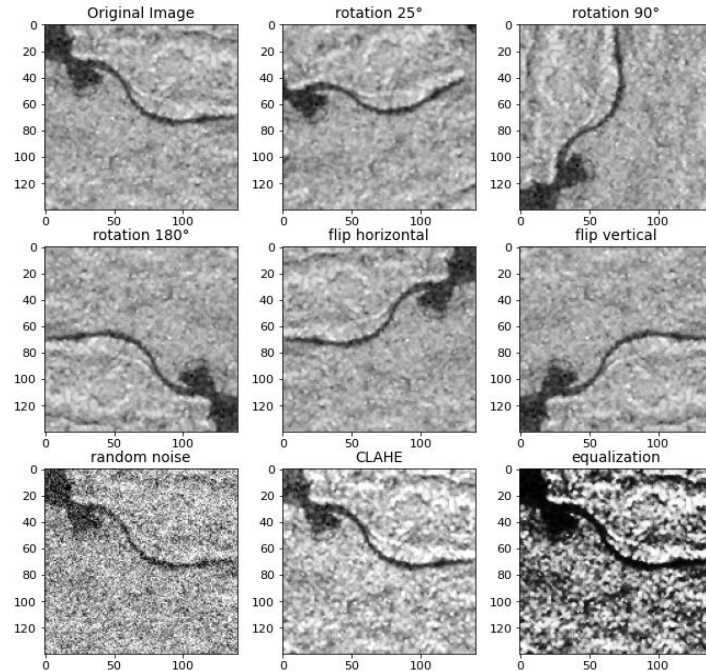
$x_{max}, x_{min}$ : Maximum and minimum pixel value.



**Figure 4.5 Clipping data**

To handle the problems of overfitting, a data augmentation was applied. Overfitting occurs when the model fits perfectly with the training data, but it estimates poorly with new data. This means that the model trains well on the training data, but is not generic enough with other data.

Data augmentation is a simple but effective method that reduces overfitting and contributes to improving the performance of the model (Simard *et al.*, 2003). This step allows to increase the amount of data by modifying the data already available. In this work each thumbnail has received a series of geometric and pixel transformations (Figure 4.6). For the geometric transformations, different kinds of rotations {25°, 90°, 180°} were applied. This helped the model to learn the invariant orientation of a channel. Vertical and horizontal flips were also applied. For the pixel transformations, a noise layer was added. The Gaussian noise was generated using a separate layer called GaussianNoise. Two adaptive filters Contrast-limited adaptive histogram equalization (CLAHE) and equalization were also applied on each thumbnail. The two latest ones increase the contrast and improve the visibility of the edges. They also allow to change the gray level on the image while preserving the details. These different transformations were done with the Keras library. After these transformations, the total number of patches in the database increased from 10,000 to 82,000.



**Figure 4.6 Data augmentation**

### **3.2.2 Training and evaluation of the deep learning model**

In this second section of the methodological approach, the CNN model and the loss function will be explained. In fact, for the CNN model development and through its different layers, the data are transformed and will have a new presentation. The last layer of the CNN will be used to estimate the water flow value. Due to a loss function or cost function (in this work it is the mean square error (MSE) as shown in eq 2)) that measures the performance of the algorithm, this estimated value will then be challenged with the validation data (the real data value). This will allow computing a loss score which indicates the error difference between the estimated and the true values. If the error difference is large, an optimization function will be used to adjust the weights and minimize the error difference. Back-propagation is the most commonly used technique (Rumelhart *et al.*, 1985) for calculating the gradient of the error function relative to one weight at a time. For updating weights and minimizing error variance, stochastic gradient descent algorithms or their variations, such as AdaGrad, RMSprop, and Adam, are commonly used. The Adam algorithm was chosen in this work.

The input data will then be exposed again to the CNN with adjusted weights. The above steps will be repeated until a minimum error difference between the predicted and true values is obtained

(Figure 4.4). It may be that the best accuracy reached after training is not considered sufficient. In this case, modifications of the developed architecture are necessary. This means modifying, especially the parameters, or adding/removing layers in order to find an architecture that correctly models the dataset in question.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Q_i^{obs} - Q_i^{estim})^2 \quad (2)$$

With :

$n$  : sample size

$Q_i^{obs}$  ,  $Q_i^{estim}$  : The observed and estimated flows respectively

### 3.2.3 The proposed CNN architecture

The typical structure of a CNN consists mainly of an **input layer**, a **hidden layers**, and an **output layer** (Anwar *et al.*, 2017). The hidden layers of a CNN are composed of convolutional layers, non-linear layer (also called activation function (Krizhevsky *et al.*, 2012), pooling layers (Nagi *et al.*, 2011) and fully connected layers (Wu, 2016).

*The input layer* allows to easily manipulate the input data of the whole CNN without destroying its internal structure. It generally represents the pixel matrix of the image.

*The hidden layers* are composed of convolutional layers that extract features from the input images of the previous layer. The results are transmitted to the next layer. This is done by using a filter that scans the entire input image and calculates the convolution product between the weights of the convolutional neuron (filter) and each portion of the scanned image. The convolution step is usually followed by a nonlinear transfer operation. A nonlinearity layer also called an activation function. This is an essential element in the neural network, it introduces a non-linearity in the model. ReLU (Rectified Linear Activation), sigmoid function, and hyperbolic tangent function are some of the most commonly used activation functions in deep learning. In this work, the ReLU function is used. The last one

allows to avoid negative values at the output of the neuron. It sets all negative values to zero while the positive values are unchanged. ReLU layers have been shown to work much better than other functions (Brown *et al.*, 2017). The advantage of ReLU over the others is that it allows for better generalization. It also allows the network to train much faster. It also helps to avoid the problem of gradient vanishing. Through the use of the pooling layer, the calculation is reduced and simplified by keeping only the most important features present in the input images. In the fully connected layer, the output of the convolution and pooling layers represents the input. This layer connects to all the features of the previous layer to obtain the final output (river flow estimation). In this work Global Average Pooling Layers (GAP) were used instead of Fully connected layers. GAP layers are a pooling operation, designed to replace fully connected layers (Al-Sabaawi *et al.*, 2021). The idea is to reduce the dimensionality of a tensor that will have a dimension of  $1 \times 1 \times d$  instead of size  $h \times w \times d$  (with h: height, w: width and d: depth). GAP layers therefore allow each hw feature map to be reduced to a single number while taking the average of all hw values. Among the advantages of using GAP layers is that there are no parameters to optimize during this step that helps, avoiding the overfitting problem. For more information on the design of CNN structures, see (Khan *et al.*, 2018; Sewak *et al.*, 2018).

*In the output layer*, a model's ability to make predictions depends on the type of activation function used. Indeed, the linear function has often been used in regression models. Since this work is aimed at estimating river flow, the activation function at the last output layer of this CNN model was defined as a linear function.

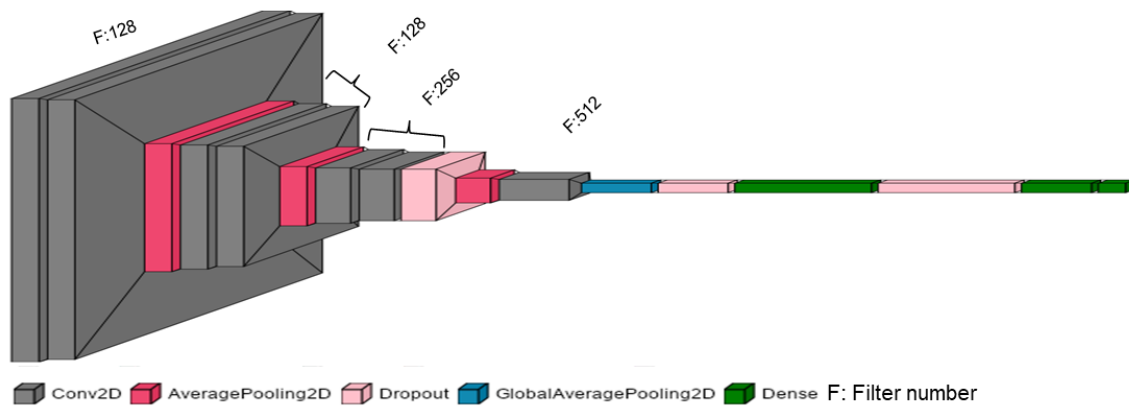
To improve the training generalization and avoid overfitting, the dropout layers technique (Srivastava *et al.*, 2014) was used. The idea of this technique amounts to disabling several nodes at each iteration during the learning process. This forces the network to be redundant and having alternative paths to correctly predict the output. This helps to reduce the overfitting problem. This method seemed the most suitable for this work, especially since it is extremely light in terms of computational requirements.

The learning rate is a very important parameter that influences the performance of the model. If the learning rate is very low, overfitting can occur (Smith, 2018). A high learning rate can lead to oscillations around global minima without reaching it. Therefore, different learning rates were tried in the context of this work. It was found that the CNN model worked best at a learning rate of  $10^{-04}$ .



The most difficult, but also the most important step is to identify the best performing architecture to achieve the best flow estimates. Indeed, hyperparameters need to be adjusted and there is no single rule to apply on the dataset. Several adjustments should be made in order to find the appropriate hyperparameters, and thus maximize the performance of the model.

Figure 4.7, Table 2 and 3 (Appendix 1) show the best design in terms of the number of layers used and the values of the hyperparameters, which are appropriate for the input data. This architecture is the result of many trials and tests. In total, 3,536,617 parameters were trained by the final version of this model.



**Figure 4.7 An overview of a convolutional neural network (CNN) architecture**

### 3.2.3 Model libraries

The criterion for choosing a deep learning library is based on its easy handling, its flexibility (i.e., the type of architecture supported and the possible operations), and also its free and open source. The Keras library (Gulli *et al.*, 2019) verified all three criteria hence its chosen for this work. It is well suited for developing one's own models due to its simplified coding. Other libraries that are used for data import, preprocessing, and visualization are Pandas, NumPy, Scikit-Learn, and PyLab. The Spyder package from Anaconda is used as an interface.

The used model training platform is a deep Learning EC2 instance of Amazon Web Services (AWS) featuring up to eight Gaudi accelerators, 256 GB of high-bandwidth memory and 768 GB of system memory.

### 3.2.4 Evaluation of model performance

After training the model, the CNN parameters were used to test the performance and generalization of the model with a test dataset, which has never been used before. The statistical performance indices chosen were: coefficient of determination ( $R^2$ ), root mean square error (RMSE), and Nash-Sutcliffe efficiency (NSE). The  $R^2$  is a statistical measure that assesses the degree to which a model explains variances in measured flows. The RMSE reflects the standard deviation of the residuals (prediction errors). A model's predictive power is evaluated using the NSE. It allows comparison between situations with different orders of magnitude of flows (Nash et Sutcliffe, 1970).

$$R^2 = \left( \frac{\sum_{i=1}^n (Q_i^{obs} - \mu_{obs})(Q_i^{estim} - \mu_{estim})}{\sqrt{\sum_{i=1}^n (Q_i^{obs} - \mu_{obs})^2} \sqrt{\sum_{i=1}^n (Q_i^{estim} - \mu_{estim})^2}} \right)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_i^{obs} - Q_i^{estim})^2} \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_i^{obs} - Q_i^{estim})^2}{\sum_{i=1}^n (Q_i^{obs} - \mu_{obs})^2} \quad (5)$$

$n$  : sample size

$Q_i^{obs}$  ,  $Q_i^{estim}$  : are the observed and estimated flows respectively

$\mu_{obs}$  : averages of observed flows

$\mu_{estim}$  : averages of estimated flows

## 4. RESULTS AND DISCUSSION

### 4.1 Analysis based on input data

First, the CNN model was trained using the SAR images only. The HAND data were added to the SAR images to train another CNN model for comparison purpose. Figure 4.8 summarizes the CNN training results obtained using the SAR data alone and the SAR data with HAND. This figure

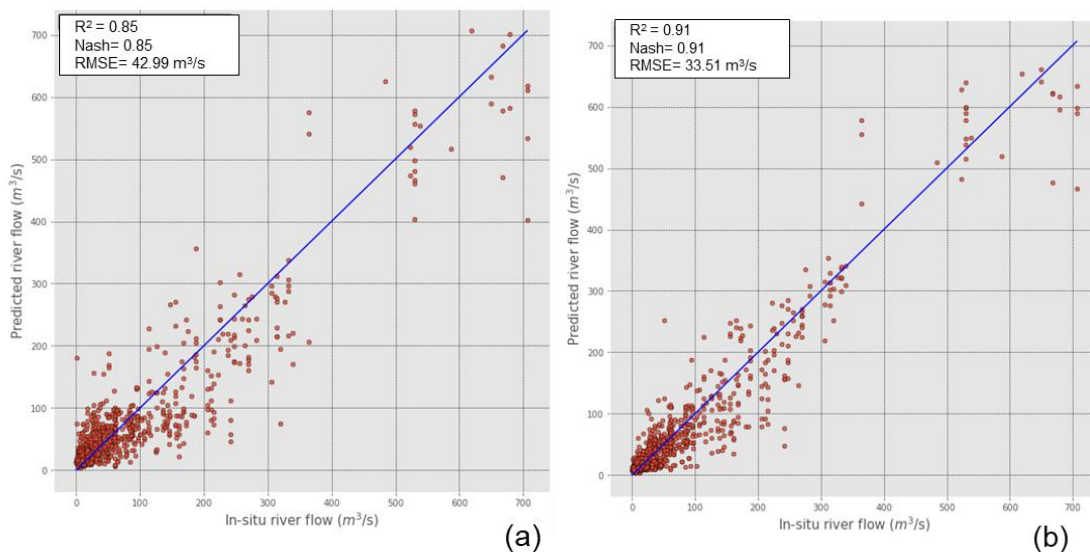
illustrates the relationship between the in situ flow measurements and the predicted flow measurements.

SAR images with HAND as input data led to a better training of the CNN for flow estimation with  $R^2 = \text{NSE} = 0.91$  and  $\text{RMSE} = 33.51 \text{ m}^3/\text{s}$ . The flow estimation using SAR data only as input data was less successful with  $R^2 = \text{NSE} = 0.85$  and  $\text{RMSE} = 42.99 \text{ m}^3/\text{s}$ . Based on the current results, it can be concluded that the CNN model trained on SAR images alone performed well, but these results were improved with the addition of HAND. Indeed, adding the topographic aspect, via the HAND data, to the SAR data has given the model a better ability to estimate in-river flow. The improvement is most visible in the observed flow values below  $300 \text{ m}^3/\text{s}$  where the point clouds are less dispersed with respect to the 1:1 line.

On the other hand, SAR data have been shown in several studies to be very useful for mapping the water surface of rivers and tracking change in its width (Bjerklie *et al.*, 2003b; Chokmani *et al.*, 2014b; Smith *et al.*, 1995). On the other hand, these data have several inconveniences. Among them is the land effect, which makes it difficult to map water because of shadow regions in the land that have similar values to those over water bodies. The use of elevation data acquired directly by the DEM does not always allow for an improvement in the mapping of surface water. On the one hand, because all DEMs contain errors, on the other hand, the simple thresholding in elevation to delimit shadows and water bodies becomes unreliable when the shadows are in low areas. For this reason, the use of another land variable that has a strong link with the presence of water and that can overcome the drawbacks of DEM is necessary, such as HAND. One of the key features of HAND is its ability to determine local drainage potential and calculate the elevation of each point in the watershed above the stream closest to its drainage direction (Nobre *et al.*, 2011). With HAND data, we are interested in relative spatial changes in elevation and not in height in an absolute value, making the elevation error on the DTM less significant. The use of HAND allowed for better delineation of water bodies and therefore improved flow estimation.

Indeed, several studies have shown a strong correlation between river width and flow (Moody et Troutman, 2002; Schubert *et al.*, 2015). River width information can provide the hydraulic geometry of the channel in which an increase or decrease in water extent results in a significantly higher or lower flow. Thus, proper delineation of water bodies is a key step in estimating river flow. Therefore, the use of SAR images with HAND data, which improved water delineation, allowed the CNN model to perform better training for river flow estimation.

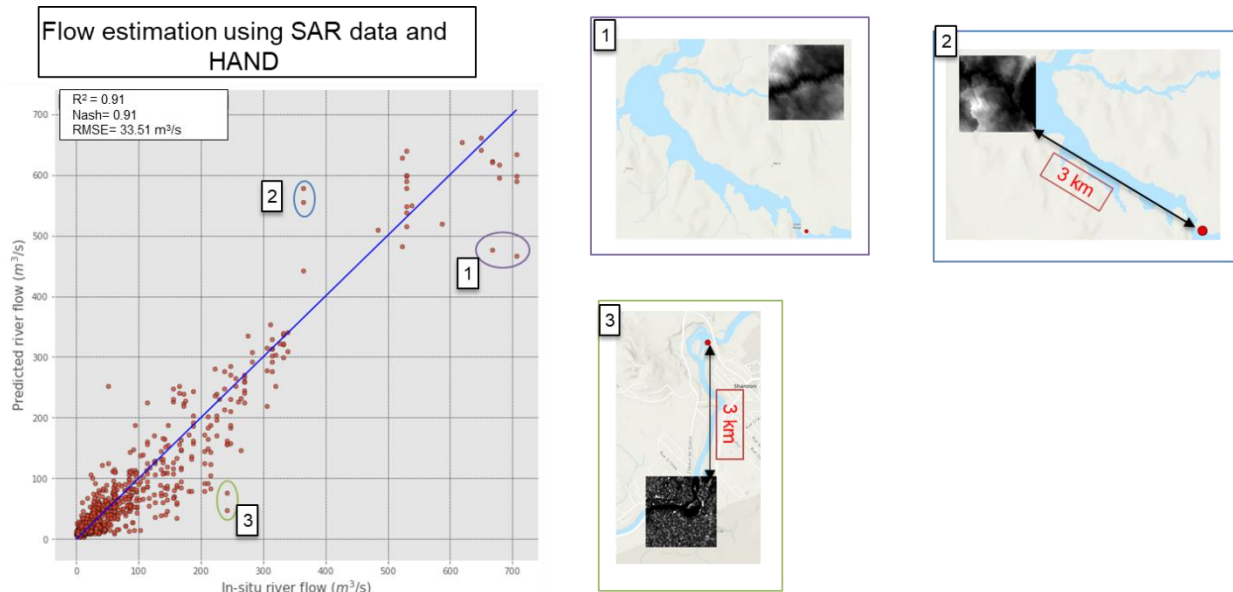
***PS: In the rest of the analyses, the focus will be only on the training results obtained using the SAR images with HAND.***



**Figure 4.8 An illustration of the effect of adding HAND on the accuracy of river flow estimation**  
**(a) flow estimation using only SAR data, (b) flow estimation using SAR data and HAND**

Figure 4.9 illustrates some examples of flow estimates that are over or underestimated. Their corresponding images were selected and visually analyzed. It was found that for the three selected examples, the distance between the image and the hydrometric station is large (about 3 km). For the image (1), the HAND was calculated on a tributary different from the one where the station is located, it is also noticed that the shape of the channel has changed which possibly induces a change in the flow rate as well. From these findings, it can be concluded that the distance from the hydrometric station as well as the change of the channel shape can significantly influence the flow estimation via the proposed method.

Various sources of uncertainty for river flow measurement have been highlighted in numerous studies (Hersch, 2002; Léonard *et al.*, 2000; Pelletier, 1988; Peña-Arancibia *et al.*, 2015; Tomkins, 2014). These uncertainties include physical factors used for flow calculation (current velocity, instrument exposure time, and location of the gauged section). A low level of uncertainty can also be observed in the extraction of remote sensing input parameters for use in the model. However, in this study the errors may be related to physical factors, not to physical parameters.



**Figure 4.9** Illustration of some examples of overestimated or underestimated flow predictions

#### 4.2 Analysis according to the morphometric characteristics of the estimation site

In this section, the number of test data was reduced. Only images within a 2 km buffer from the hydrometric station were considered. The goal is to highlight the effect of changing distance from the hydrometric station on the flow estimate. Figure 4.10 shows the relationship between the observed flow and the estimated flow following the reduction of the distance to the hydrometric station. Compared to the previous results, it is noticed that the use of images that are at a 2 km buffer zone from the hydrometric station reached a better estimation of the river flow. The values of  $R^2$ , RMSE, NSE have increased to 0.95, 22.55  $m^3/s$ , 0.95, respectively. The scatter plot became increasingly less dispersed with respect to 1:1 line.

The same test was reproduced, except that this time only the images that pass through the station were considered. Similarly, the flow estimation results improved again as shown in Figure 4.11. The  $R^2$  and NSE values increased from 0.91 with all test data to 0.98. The RMSE was also reduced from 33.51 to 17.28  $m^3/s$ . From the above, it is clear that the proximity is a key factor to accurate flow estimation. The closer the hydrometric station, flow prediction becomes.

These results can be justified by two reasons. Firstly, by the fact that the morphology of the stream changes along the river which induces a change in the flow as well. Thus, the value of the flow at the hydrometric station can be significantly changed along the channel section. This would

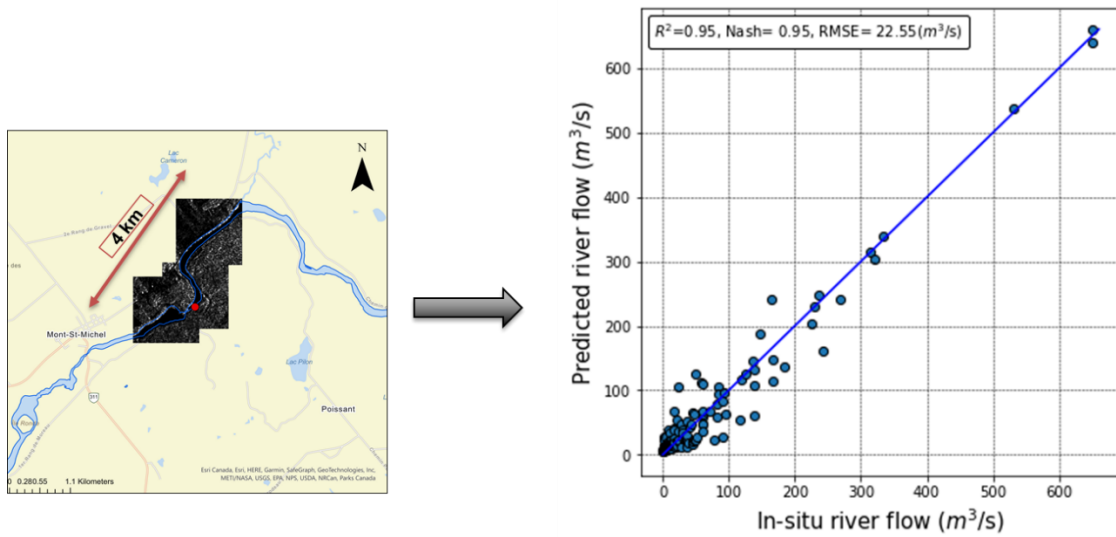
produce a significant error that would disturb and affect the quality of the flow estimation. Therefore, the more the model is trained with images acquired at the flow measurement stations, the more accurate the estimation result becomes. The second reason is that the accuracy of the deep learning algorithm depends on the quality of the input data. It has been shown in several studies that there is a positive correlation between the number of features and the accuracy of the results. Using the relevant features in the model can only increase its accuracy (Pristyanto *et al.*, 2019). On the other hand, the use of redundant and inappropriate features can lead to confusion in the trained model (Wang *et al.*, 2017). Removing these redundant features will help improve the learning performance of the model and avoid its degradation (Li *et al.*, 2017).

It is important to note that to date, to our knowledge, no work has successfully created a direct link between the radar image signal and the river flow estimation. The work that has been done in this direction often relates the flow to a derivative of the SAR images (river width, water surface, etc.). As an example, Ahmad et Kim, (2019) proposed an approach to streamflow estimation in the Han River basin in Korea using sentinel-1 SAR images. The working method is based on the use of the rating curve that relates the extracted water surface with SAR data and in situ flow measurements. Their results showed that the flow rate can be estimated accurately (mean R value = 0.80) when the water width is greater than 40 m as well as when the channel cross section has gentle side slopes. Mungen *et al.*, (2020) used also sentinel-1 SAR data to estimate river discharge. To do so, they used the satellite-estimated river width and the in situ measured flow and via a thresholding method were able to develop a model for the estimation of flow for a given river section. This thresholding method of flow estimation showed a good potential for flow rate estimation with a relative root mean square error (RRMSE) value of 38.5% and 34.5%, respectively.

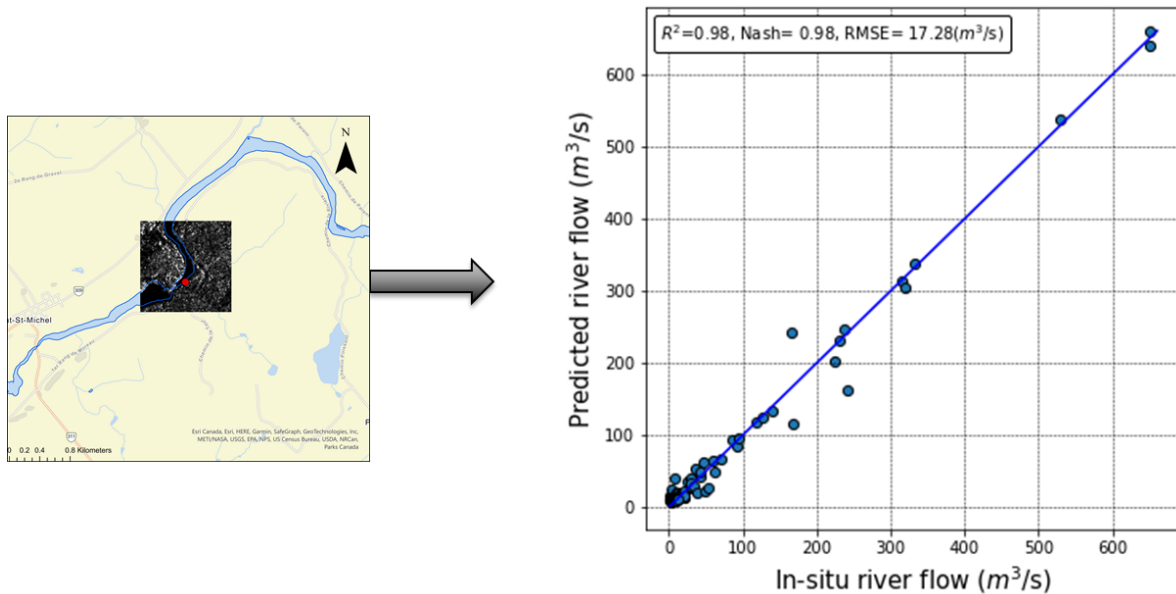
Despite the potential of the above works, they have several limitations. First, because of their empirical nature, rating curves are limited to specific sections and cannot be applied in other rivers (Bjerklie *et al.*, 2003b). Secondly, the quality of the results given by the rating curve is highly dependent on the quality of the calibration data, which depends on the availability and representativity of the satellite-derived observation data used to calibrate the model. Also, the satellite-derived measurements are not always available and of good quality. Thus, the uncertainty associated with the satellite-derived variables, physical parameters and rating curve equations results in many errors in the river flow estimate.

The present study demonstrated that a direct causal relationship between the SAR data signal and river discharge can be established. This relationship has several advantages over previous

studies, including its potential for large-scale application (gauged and ungauged sites) without the need for recalibration. On the other hand, in addition to the local aspect of the models presented in the literature, flow has always been modeled from proxies derived from SAR images. These derived variables are issued from an error-ridden modeling process. The least error in the estimation of this derived variable, by error propagation process, can easily lead to a distorted final result of the flow modeling. Finding a direct relationship between the SAR signal and flow inhibited a significant source of error, which likely led to the development of a robust model ( $R^2 = \text{Nash} = 0.98$ ) via a CNN.



**Figure 4.10** Flow prediction using data that is within a two km buffer from the hydrometric station



**Figure 4.11** Flow estimation using data passing through the hydrometric station

### 4.3 Analysis by individual station of the flow estimate

This part aims to analyse how the model estimates the flow at different dates with different flow conditions within the same hydrometric station. To do so, among the test data that pass through the hydrometric station those that have flow observations at different dates were taken. A station by station analysis was done. In total, 15 hydrometric stations were analyzed.

At first sight, two groups of performance are quite distinct. A first group of flow estimation with very good performance and a second group with less or mediocre performance (Figure 4.12). After analysis, it turned out that rivers with widths of 40 m and more belong to the first group and vice versa. Thus, for the stations belonging to the first group of rivers the  $R^2$  and NSE vary, respectively, between [0.79 to 1.00] and [0.76 to 0.97] and the RMSE varies between 2.49 and 42.65 m<sup>3</sup>/s (Figure 4.13 and Figure 4.14). For the stations in the second group of rivers, the  $R^2$  and NSE range, respectively, from [0.06 to 0.85] and [- 4.85 to 0.37] and the RMSE ranges from 1.22 to 8.39 m<sup>3</sup>/s. This shows that the model is capable of estimating the flow at the station itself with different flow conditions when the width of the river is greater than 40 meters. When the width is less than this threshold, it is noticed that the modeling quality is very low and the model tends to overestimate the flow of these rivers. This inaccurate modeling can be mainly related to the spatial resolution of the HAND data used (30m).

Indeed, it was shown during this study that the flow estimation was significantly improved by integrating the HAND data into the CNN model (Figure 4.8). This improvement is explained by the integration of the topographic aspect at the time of modeling. HAND allows for better delineation of the spatial extent of the river water surface. However, the accuracy of river delineation deteriorates with coarser spatial resolution data (Tate *et al.*, 2002), which is the case in this study (30 m for HAND versus 8 m for SAR data). With this difference in resolutions, radar signals returned from water bodies, within the 30 m HAND pixel (even when resampled to 8 m spatial resolution), are more likely to be contaminated by non-water surfaces, which can degrade the quality of the measurement. The work of de Bjerklie *et al.*, (2003) reinforces this idea. They showed that even if it is possible to distinguish the river from the surrounding landscape, the smaller channels would have greater uncertainty in estimating their widths due to the size of the satellite footprint which limits the good estimation of relative width. This, in turn, explains the less accurate performance of the CNN model for flow estimation for narrower rivers (< 40m).

This analysis illustrates that the developed model is mostly able to offer good flow estimates with different flow conditions and at different dates, especially for rivers with a width exceeding 40 m. This model can be and potentially transferred to other locations with appropriate engineering and



optimization of the characteristics to make an accurate estimation of the flow of rivers, especially the ungauged ones. However, caution should be exercised when estimating the flows of smaller rivers due to the associated uncertainty. Nevertheless, it is important to note that this source of uncertainty can be controlled by re-training the model with a HAND from the lidar at 1 m. With this spatial resolution, it would be possible to produce a high-precision delineation of river water surfaces and consequently a better estimation of flows, especially for smaller rivers.

Another important point to emphasize is that the initial approach adopted in this work for dividing training, validation and test data is based on successive trials, which enabled the ratio to be progressively adjusted until the best results were obtained. However, a more promising alternative is the use of block cross-validation. Block cross-validation is a variant of traditional cross-validation that considers the dependency structures present in the data, such as temporal, spatial or hierarchical structure (Roberts *et al.*, 2017). It aims to maintain the integrity of dependency structures when dividing data. The use of this method could open new perspectives for future studies on river flow estimation. This approach would make it possible to take better account of spatial and temporal variations in flow, as well as potential dependencies between observations. In this way, it could enable a more accurate assessment of the performance of deep learning models, while avoiding the pitfalls of overfitting or under learning.

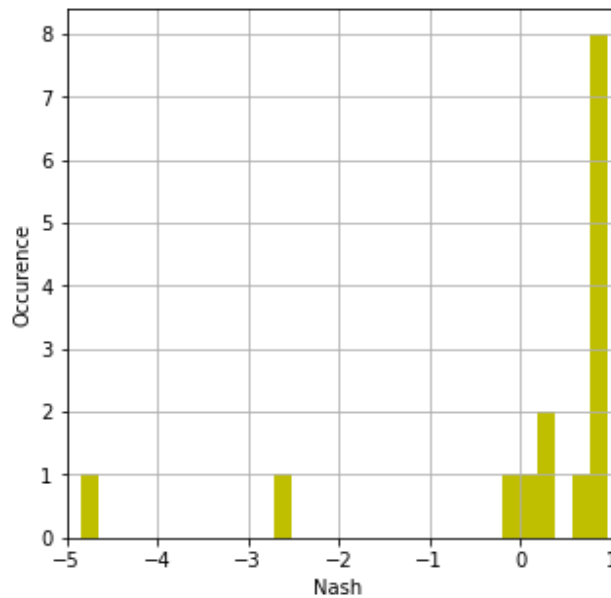
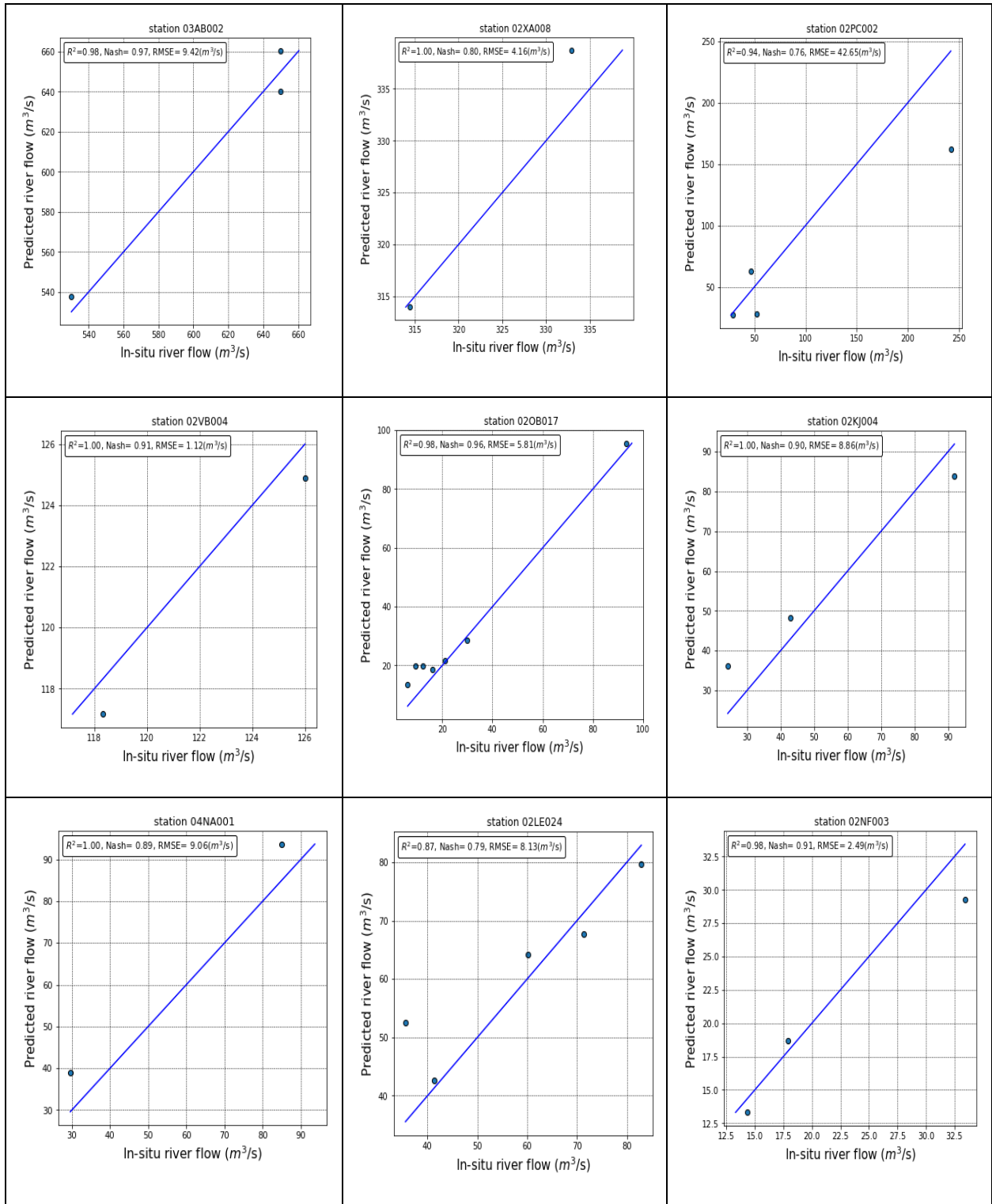
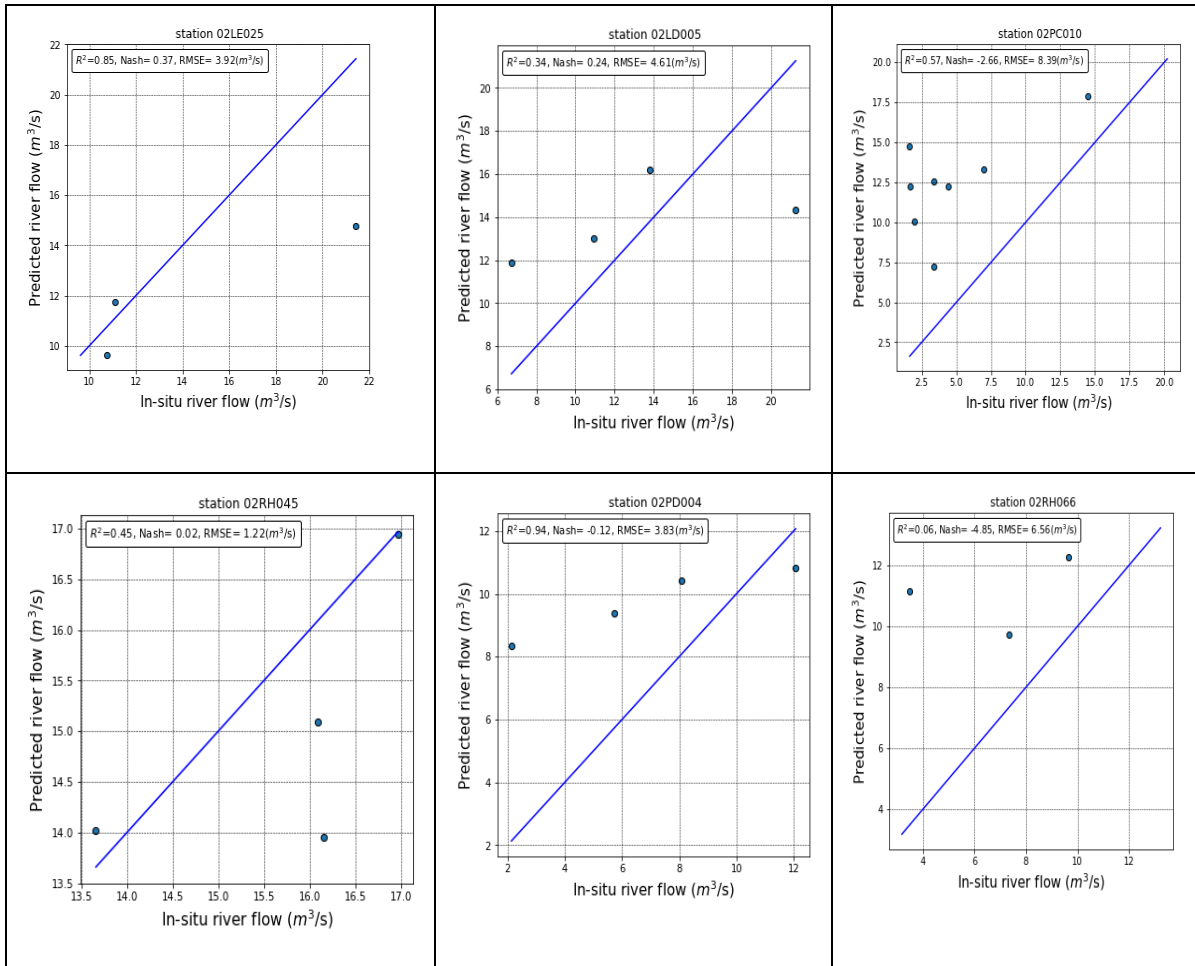


Figure 4.12 Occurrence of NSE in the 15 hydrometric stations



**Figure 4.13 Analysis of flow estimation station by station at different dates and different flow conditions. River flow varies from 33 to 650 m<sup>3</sup>/s**



**Figure 4.14 Analysis of flow estimation station by station at different dates and different flow conditions. River flow varies from 12 to 22 m<sup>3</sup>/s**

## 5. Conclusion

River flow estimation is an essential aspect of water resources management and represents a very important parameter in flood risk management and hydrological modeling. In this work, an innovative river flow estimation approach has been developed. This method proposes the use of a deep learning model (CNN) to estimate river discharge using SAR images and HAND. The study was conducted using a total of 39 hydrometric stations located in the eastern part of the boreal zone in Canada.

The results show that SAR imagery is a valuable tool that detects the change in flow along a river section. Regarding the CNN structure, the results indicate that selecting the appropriate number of features and quality of those features can effectively reduce redundant information. As a result, network learning is improved.

A performance evaluation of the CNN model also revealed that it was highly reliable and promising. An excellent correlation is obtained between the in-situ flow and the estimated flow. The integration of HAND to the SAR data to the input data of the CNN model contributed greatly to improve the modeling with a NSE value of 0.91, RMSE of 33.51 m<sup>3</sup>/s and R<sup>2</sup> of 0.91. It was also shown that the proximity of the SAR images to the hydrometric stations is a key factor for a good modeling of the river flow; the closest the images are to the stations, the better the statistical evaluation indexes are (R<sup>2</sup> = NSE = 0.91 and RMSE = 33.51 m<sup>3</sup>/s for images located 5 km away from the stations to R<sup>2</sup>= NSE = 0.98 and RMSE = 17.28 m<sup>3</sup>/s for images acquired on the stations) It was also shown that the change in stream morphology along the river induces a change in flow as well.

The results of the station-by-station flow estimation, at different dates and under different flow conditions, showed a good performance of the model at rivers that are greater than 40 m width. For small rivers, the results showed that the flow estimation by the CNN model has a tendency of overestimation. This is likely related to the HAND spatial resolution data used in this work (30m). However, the quality of the modeling of these small rivers can be improved by re-training the model with a finer HAND spatial resolution; to the meter for example generated from Lidar data.

In this sense, the contribution of this work has made it possible to develop a methodology for estimating flow in rivers with different flow conditions. The use of this approach is particularly important for ungauged rivers. In addition, a systematic application of this method to other ecozones could be considered. The availability of high spatial resolution satellite data may also

allow the successful application of this method on small rivers. Such improvements will be beneficial for automatic flow monitoring on different river morphologies.

**Credit author statement**

Conceptualization, SZ and KC; Methodology, SZ and KC; Formal analysis and interpretation, SZ and KC; Development of the code, SZ; Writing the paper, SZ; Supervision, KC; Technical support for the development of part of the code, CC; Reviewing and editing the paper, AE, KC, CC

All authors approved the published version of the manuscript

## 5 DISCUSSION GÉNÉRALE ET CONCLUSION

---

L'estimation du débit en rivière est une étape essentielle pour de la gestion des ressources en eau, qui est un paramètre très important pour la gestion des risques d'inondation et la modélisation hydrologique. Dans ce travail, une méthode innovante pour l'estimation du débit en rivière a été développée. Cette méthode propose l'utilisation d'un modèle d'apprentissage profond (CNN) pour estimer le débit en rivière en utilisant les images RSO et la hauteur par rapport au drainage le plus proche (HAND). Un total de 39 stations hydrométriques qui se trouvent dans la partie orientale de la zone boréale au Canada ont fait l'objet de cette étude. Les principales conclusions de la recherche sont résumées ci-dessous.

En premier lieu, le modèle CNN a été entraîné en utilisant les images RSO seulement. Les données HAND ont été rajoutées avec les images RSO pour entraîner un autre modèle CNN à des fins de fin de comparaison.

Les résultats montrent que l'imagerie RSO peut fournir des informations importantes sur la cartographie des eaux de surface ainsi que la détection du changement dans le débit le long d'un tronçon de la rivière. L'évaluation de performances du modèle CNN utilisé pour estimer le débit en rivière était également prometteuse et très fiable. Une forte corrélation est obtenue entre le débit in situ et le débit estimé par le modèle. En revanche, les données RSO présentent plusieurs inconvénients. Parmi eux, l'effet de terrain, qui rend difficile la cartographie de l'eau à cause des régions d'ombre du terrain qui ont des valeurs semblables à celles au-dessus des plans d'eau. L'utilisation de données d'élévation acquise directement par le MNT ne permet pas toujours une amélioration dans la cartographie des eaux de surface. D'une part, parce que tous les MNT contiennent des erreurs, d'autre part, le simple seuillage en élévation pour délimiter les ombres et les plans d'eau devient non fiable lorsque les zones d'ombre se trouvent dans de basses zones. C'est pour cette raison l'utilisation d'une autre variable de terrain qui a un lien fort avec la présence de l'eau et qui peut pallier aux inconvénients du MNT est nécessaire, tel que le HAND.

L'ajout du HAND aux données RSO dans le modèle CNN a contribué grandement à l'amélioration de la modélisation avec une valeur de NSE de 0,91, RMSE de 33,51 m<sup>3</sup>/s et R<sup>2</sup> de 0,91. Le HAND permet une meilleure délimitation de l'étendue spatiale de la surface de l'eau des rivières.

Il a été également démontré que la proximité des images RSO aux stations hydrométriques est un facteur clé pour une bonne modélisation du débit des rivières ; plus les images sont proches des stations plus les indices d'évaluation statistiques sont meilleurs (R<sup>2</sup> = NSE = 0,91 et RMSE

= 33,51 m<sup>3</sup>/s pour des images se trouvant à 5 km de part et d'autre des stations à  $R^2 = NSE = 0,98$  et  $RMSE = 17,28$  m<sup>3</sup>/s pour des images acquises sur les stations). Il a été démontré aussi que le changement de la morphologie de cours d'eau tout au long de la rivière induit à un changement de débit aussi.

Ces résultats peuvent être justifiés par deux raisons. Tout d'abord, par le fait que la morphologie du cours d'eau change tout au long de la rivière ce qui induit un changement du débit aussi. Donc, la valeur du débit au niveau de la station hydrométrique peut se voir sensiblement changé le long de la section du canal. Ceci produirait une erreur non négligeable qui perturberait et affecterait la qualité de l'estimation du débit. Par conséquent, plus le modèle est entraîné avec des images acquises au niveau des stations de mesures du débit, plus le résultat d'estimation devient précis. Ensuite, la deuxième raison est que la précision de l'algorithme d'apprentissage automatique dépend de la qualité des données d'entrée. Il a été démontré dans plusieurs études qu'il existe une corrélation positive entre le nombre de caractéristiques et la précision des résultats. L'emploi des caractéristiques pertinentes dans le modèle ne peut qu'augmenter sa précision (Pristyanto *et al.*, 2019). En revanche, l'utilisation des caractéristiques redondantes et non appropriées peut engendrer une confusion dans le modèle entraîné (Wang *et al.*, 2017). Enlever ces caractéristiques redondantes va aider à améliorer la performance d'apprentissage du modèle et éviter sa dégradation (Li *et al.*, 2017).

Une analyse par station individuelle de l'estimation de débit a été effectuée. Cette analyse vise à déterminer comment le modèle estime le débit à des dates différentes avec des conditions d'écoulement différentes au niveau de la station hydrométrique. Pour ce faire, parmi les données de test qui passent par la station hydrométrique celles qui ont des observations des débits à des dates différentes ont été prises. Une analyse station par station a été faite. Au total, 15 stations hydrométriques sont analysées. Les résultats ont montré une bonne performance du modèle au niveau des rivières qui ont une largeur plus grande que 40 m. Ceci montre que le modèle est capable d'estimer le débit au niveau de la station elle-même avec des conditions d'écoulement différentes lorsque la largeur de la rivière est supérieure à 40 mètres.

Pour les petites rivières et lorsque la largeur est inférieure à 40 m, les résultats ont montré que l'estimation du débit par le modèle CNN a une tendance de surestimation. Ceci est probablement lié à la résolution spatiale des données HAND utilisé dans ce travail (30 m). En effet, la précision de la délimitation des cours d'eau se détériore avec données à résolution spatiale plus grossière (Tate *et al.*, 2002), ce qui est le cas dans cette étude (30 m pour le HAND versus 8 m pour les données RSO). Avec cette différence de résolutions, les signaux radars renvoyés par les plans

d'eau, au sein du pixel de 30 m du HAND (même à 8 m de résolution spatiale), sont plus susceptibles d'être contaminés par des surfaces non aquatiques, ce qui peut dégrader la qualité de la mesure. Cependant, la qualité de la modélisation de ces petites rivières peut être améliorée en réentraînant le modèle avec un HAND à résolution spatiale plus fine ; au mètre par exemple issu des données Lidar.

Une autre remarque importante à souligner est que l'approche initiale adoptée dans ce travail pour diviser les données d'entraînement, de validation et de test est basée sur des essais successifs, ce qui a permis d'ajuster progressivement le ratio jusqu'à obtenir les meilleurs résultats. Cependant, une alternative plus prometteuse émerge c'est l'utilisation de la validation croisée par bloc. En effet, cette dernière est une variante de la validation croisée traditionnelle qui prend en compte les structures de dépendance présentes dans les données, telles que la structure temporelle, spatiale ou hiérarchique. Elle vise à maintenir l'intégrité des structures de dépendance lors de la division des données.

En incorporant la validation croisée par bloc dans des études ultérieures sur l'estimation du débit, il serait possible d'obtenir des évaluations plus rigoureuses des performances des modèles tout en évitant les pièges de surapprentissage ou de sous-apprentissage. Cette approche permettrait de mieux tenir compte des variations spatiales et temporelles du débit, ainsi que des dépendances potentielles entre les observations.

La contribution de ce travail a donc ouvert la porte au développement d'une méthodologie d'estimation du débit dans des rivières avec différentes conditions d'écoulement. L'utilisation de cette approche est particulièrement importante pour les rivières non jaugées. En outre, une application systématique de cette méthode sur d'autres écozones pourrait être envisagée. La présence de données satellitaire à haute résolution spatiale pourrait également permettre d'appliquer avec succès cette méthode sur de petites rivières. De telles améliorations seront bénéfiques pour le suivi de débit sur différentes morphologies des rivières.









## 6 BIBLIOGRAPHIE

---

- Ahmad, W. et Kim, D. (2019) Estimation of flow in various sizes of streams using the Sentinel-1 Synthetic Aperture Radar (SAR) data in Han River Basin, Korea. *International Journal of Applied Earth Observation and Geoinformation*, vol. 83, p. 101930.
- Al-Sabaawi, A., Ibrahim, H. M., Arkah, Z. M., Al-Amidie, M. et Alzubaidi, L. (2021) Amended convolutional neural network with global average pooling for image classification. *International Conference on Intelligent Systems Design and Applications*, p. 171-180. Springer.
- Alsdorf, D. E. (2003) Water storage of the central Amazon floodplain measured with GIS and remote sensing imagery. *Annals of the Association of American Geographers*, vol. 93, n°1, p. 55-66.
- Alsdorf, D. E. et Lettenmaier, D. P. (2003) Tracking fresh water from space. *Science*, vol. 301, n°5639.
- Alsdorf, D. E., Rodríguez, E. et Lettenmaier, D. P. (2007a) Measuring surface water from space. *Reviews of Geophysics*, vol. 45, n°2, p. 1-24.
- Alsdorf, D. E., Rodríguez, E. et Lettenmaier, D. P. (2007b) Measuring surface water from space. *Reviews of Geophysics*, vol. 45, n°2.
- Anwar, S., Hwang, K. et Sung, W. (2017) Structured pruning of deep convolutional neural networks. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 13, n°3, p. 1-18.
- Ashmore, P. et Sauks, E. (2006) Prediction of discharge from water surface width in a braided river with implications for at-a-station hydraulic geometry. *Water Resources Research*, vol. 42, n°3, p. 1-11.
- Bai, Y., Chen, Z., Xie, J. et Li, C. (2016) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology*, vol. 532, p. 193-206.

- Baldassarre, G. D., Processes, S. U.-H. et 2012, undefined (2012) Is the current flood of data enough? A treatise on research needs for the improvement of flood modelling. *academia.edu*, vol. 26, p. 153-158.
- Banerjee, A., Chakrabarty, M., Rakshit, N., Bhowmick, A. R. et Ray, S. (2019) Environmental factors as indicators of dissolved oxygen concentration and zooplankton abundance: Deep learning versus traditional regression approach. *Ecological Indicators*, vol. 100, p. 99-117.
- Barbetta, S., Franchini, M., Melone, F. et Moramarco, T. (2012a) Enhancement and comprehensive evaluation of the Rating Curve Model for different river sites. *Journal of Hydrology*, vol. 464-465, p. 376-387.
- Barbetta, S., Franchini, M., Melone, F. et Moramarco, T. (2012b) Enhancement and comprehensive evaluation of the Rating Curve Model for different river sites. *Journal of Hydrology*, vol. 464-465, p. 376-387.
- Barzegar, R., Aalami, M. T. et Adamowski, J. (2020) Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, p. 1-19.
- Bhola, P. K., Nair, B. B., Leandro, J., Rao, S. N. et Disse, M. (2019) Flood inundation forecasts using validation data generated with the assistance of computer vision. *Journal of Hydroinformatics*, vol. 21, n°2, p. 240-256.
- Biancamaria, S., Andreadis, K. M., Durand, M., Clark, E. A., Rodriguez, E., Mognard, N. M., Alsdorf, D. E., Lettenmaier, D. P. et Oudin, Y. (2009) Preliminary characterization of SWOT hydrology error budget and global capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, n°1, p. 6-19.

- Birkett, C. M., Mertes, L. A. K., Dunne, T., Costa, M. H. et Jasinski, M. J. (2002) Surface water dynamics in the Amazon Basin: Application of satellite radar altimetry. *Journal of Geophysical Research: Atmospheres*, vol. 107, n°D20, p. LBA-26.
- Birkinshaw, S. J., Moore, P., Kilsby, C. G., O'Donnell, G. M., Hardy, A. J. et Berry, P. A. M. (2014) Daily discharge estimation at ungauged river sites using remote sensing. *Hydrological Processes*, vol. 28, n°3, p. 1043-1054.
- Birkinshaw, S. J., O'donnell, G. M., Moore, P., Kilsby, C. G., Fowler, H. J. et Berry, P. A. M. (2010) Using satellite altimetry data to augment flow estimation techniques on the Mekong River. *Hydrological Processes*, vol. 24, n°26, p. 3811-3825.
- Bjerklie, D. M., Dingman, S. L., Vorosmarty, C. J., Bolster, C. H. et Congalton, R. G. (2003a) Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, vol. 278, n°1-4, p. 17-38.
- Bjerklie, D. M., Dingman, S. L., Vorosmarty, C. J., Bolster, C. H. et Congalton, R. G. (2003b) Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, vol. 278, n°1-4, p. 17-38.
- Bjerklie, D. M., Moller, D., Smith, L. C. et Dingman, S. L. (2005) Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of hydrology*, vol. 309, n°1-4, p. 191-209.
- Brakenridge, G. R., Nghiem, S. V., Anderson, E. et Mic, R. (2007) Orbital microwave measurement of river discharge and ice status. *Water Resources Research*, vol. 43, n°4.
- Brown, M. J., Hutchinson, L. A., Rainbow, M. J., Deluzio, K. J. et De Asha, A. R. (2017) A comparison of self-selected walking speeds and walking speed variability when data are collected during repeated discrete trials and during continuous walking. *Journal of Applied Biomechanics*, vol. 33, n°5, p. 384-387.

- Calmant, S., Seyler, F. et Cretaux, J. F. (2008) Monitoring continental surface waters by satellite altimetry. *Surveys in geophysics*, vol. 29, n°4, p. 247-269.
- Chang, F.-J., Chen, P.-A., Lu, Y.-R., Huang, E. et Chang, K.-Y. (2014) Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, vol. 517, p. 836-846.
- Chokmani, K., Perreault, S., Jacome, A., Bernier, M., Poulin, J. et Gauthier, Y. (2014a) Estimation du débit en rivière à l' aide de l' imagerie RADARSAT : concept de la station hydrométrique virtuelle.
- Chokmani, K., Perreault, S., Jacome, A., Bernier, M., Poulin, J. et Gauthier, Y. (2014b) Estimation du débit en rivière à l' aide de l' imagerie RADARSAT : concept de la station hydrométrique virtuelle.
- Chollet, F. (2018) *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG.
- Coe, M. T. et Birkett, C. M. (2004) Calculation of river discharge and prediction of lake height from satellite radar altimetry: Example for the Lake Chad basin. *Water Resources Research*, vol. 40, n°10.
- Colombo, S. J., Cherry, M. L., Graham, C., Greifenhagen, S., McAlpine, R. S., Papadopol, C. S., Parker, W. C., Scarr, T., Ter-Mikaelian, M. T. et Flannigan, M. D. (1998) The impacts of climate change on Ontarios forests.
- Dai, A. et Trenberth, K. E. (2002) Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *Journal of hydrometeorology*, vol. 3, n°6, p. 660-687.
- Damavandi, H. G., Shah, R., Stampoulis, D., Wei, Y., Bosovic, D. et Sabo, J. (2019) Accurate prediction of streamflow using long short-term memory network: a case study in the Brazos

- River Basin in Texas. *International Journal of Environmental Science and Development*, vol. 10, n°10, p. 294-300.
- Despax, A. (2017) Incertitude des mesures de débit des cours d'eau au courantomètre . Amélioration des méthodes analytiques To cite this version : HAL Id : tel-01496704 Incertitude des mesures de débit des cours d'eau au courantomètre . Amélioration des méthodes analyt.
- Di Baldassarre, G. et Uhlenbrook, S. (2012) Is the current flood of data enough? A treatise on research needs for the improvement of flood modelling. *Hydrological Processes*, vol. 26, n°1, p. 153-158.
- Dottori, F., Martina, M. L. V. et Todini, E. (2009a) A dynamic rating curve approach to indirect discharge measurement. *Hydrology and Earth System Sciences*, vol. 13, n°6, p. 847-863.
- Dottori, F., Martina, M. L. V. et Todini, E. (2009b) A dynamic rating curve approach to indirect discharge measurement. *Hydrology and Earth System Sciences*, vol. 13, n°6, p. 847-863.
- Dottori, F, Martina, M., System, E. T.-H. and E. et 2009, undefined (2009) A dynamic rating curve approach to indirect discharge measurement. *hess.copernicus.org*, vol. 13, p. 847-863.
- Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., Rodriguez, E., Bates, P. D., Pavelsky, T. M., Monnier, J., Chen, X., Di Baldassarre, G., Fiset, J. M., Flipo, N., Frasson, R. P. D. M., Fulton, J., Goutal, N., Hossain, F., Humphries, E., Minear, J. T., Mukolwe, M. M., Neal, J. C., Ricci, S., Sanders, B. F., Schumann, G., Schubert, J. E. et Vilmin, L. (2016) An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, vol. 52, n°6, p. 4527-4549.
- Eamer, J. (2015) Boreal Shield and Newfoundland Boreal ecozones+ evidence for key findings summary.
- Fang, K. et Shen, C. (2017) Full-flow-regime storage-streamflow correlation patterns provide insights into hydrologic functioning over the continental US. *Water Resources Research*, vol. 53, n°9, p. 8064-8083.



- Fang, Q., Zhang, J., Xie, C. et Yang, Y. (2019) Detection of multiple leakage points in water distribution networks based on convolutional neural networks. *Water Supply*, vol. 19, n°8, p. 2231-2239.
- Feng, D, Gleason, C., ... X. Y.-W. R. et 2019, undefined (2019) Comparing discharge estimates made via the BAM algorithm in high-order Arctic rivers derived solely from optical CubeSat, Landsat, and Sentinel-2 data. *Wiley Online Library*, vol. 55, n°9, p. 7753-7771.
- Feng, Dongmei, Gleason, C. J., Yang, X. et Pavelsky, T. M. (2019) Comparing discharge estimates made via the BAM algorithm in high-order Arctic rivers derived solely from optical CubeSat, Landsat, and Sentinel-2 data. *Water Resources Research*, vol. 55, n°9, p. 7753-7771.
- Fleischmann, A. S., Paiva, R. C., Collischonn, W., Sorribas, M. V. et Pontes, P. R. (2016) On river-floodplain interaction and hydrograph skewness. *Water Resources Research*, vol. 52, n°10, p. 7615-7630.
- Fulton, J. W., Anderson, I. E., Chiu, C. L., Sommer, W., Adams, J. D., Moramarco, T., Bjerklie, D. M., Fulford, J. M., Sloan, J. L., Best, H. R., Conaway, J. S., Kang, M. J., Kohn, M. S., Nicotra, M. J. et Pulli, J. J. (2020) QCam: SUAS-based doppler radar for measuring river discharge. *Remote Sensing*, vol. 12, n°20, p. 1-23.
- Gleason, C. J. et Smith, L. C. (2014) Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, n°13, p. 4788-4791.
- Gleason, C. J., Smith, L. C., Finnegan, D. C., Lewinter, A. L., Pitcher, L. H. et Chu, V. W. (2015) Semi-automated effective width extraction from time-lapse RGB imagery of a remote, braided Greenlandic river. *hess.copernicus.org*, vol. 19, p. 2963-2969.
- Gleason, C. J., Wada, Y. et Wang, J. (2018) A hybrid of optical remote sensing and hydrological modeling improves water balance estimation. *Journal of Advances in Modeling Earth Systems*, vol. 10, n°1, p. 2-17.

- Gleason, C., Wada, Y., in, J. W.-J. of A. et 2018, undefined (2018) A hybrid of optical remote sensing and hydrological modeling improves water balance estimation. Wiley Online Library, vol. 10, n°1, p. 2-17.
- Goodfellow, I., Bengio, Y., Courville, A. et Bengio, Y. (2016) Deep learning (Vol. 1). MIT press Cambridge.
- Guerrero, J. L., Westerberg, I. K., Halldin, S., Xu, C. Y. et Lundin, L. C. (2012a) Temporal variability in stage-discharge relationships. Journal of Hydrology, vol. 446-447, p. 90-102.
- Guerrero, J. L., Westerberg, I. K., Halldin, S., Xu, C. Y. et Lundin, L. C. (2012b) Temporal variability in stage-discharge relationships. Journal of Hydrology, vol. 446-447, p. 90-102.
- Gulli, A., Kapoor, A. et Pal, S. (2019) Deep learning with TensorFlow 2 and Keras: regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API. Packt Publishing Ltd.
- Hahmann, T., Roth, A., Martinis, S., Twele, A. et Gruber, A. (2008) Automatic extraction of water bodies from TerraSAR-X data. IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium, vol. 3, p. III-103. IEEE.
- Herschy, R. W. (2002) The uncertainty in a current meter measurement. Flow measurement and instrumentation, vol. 13, n°5-6, p. 281-284.
- Hinton, G. E., Osindero, S. et Teh, Y.-W. (2006) A fast learning algorithm for deep belief nets. Neural computation, vol. 18, n°7, p. 1527-1554.
- Horritt, M. S., Mason, D. C., Cobby, D. M., Davenport, I. J. et Bates, P. D. (2003) Waterline mapping in flooded vegetation from airborne SAR imagery. Remote Sensing of Environment, vol. 85, n°3, p. 271-281.

- Hossain, F., ... A. S.-E.-A.-I. J. et 2013, undefined (s.d.) Proof of concept of an altimeter-based river forecasting system for transboundary flow inside Bangladesh. [ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Hossain, F., Siddique-E-Akbor, A. H., Mazumder, L. C., ShahNewaz, S. M., Biancamaria, S., Lee, H. et Shum, C. K. (2013) Proof of concept of an altimeter-based river forecasting system for transboundary flow inside Bangladesh. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, n°2, p. 587-601.
- Hrnjica, B. et Bonacci, O. (2019) Lake level prediction using feed forward and recurrent neural networks. *Water Resources Management*, vol. 33, n°7, p. 2471-2484.
- Huang, C., Nguyen, B. D., Zhang, S., Cao, S. et Wagner, W. (2017) A comparison of terrain indices toward their ability in assisting surface water mapping from sentinel-1 data. *ISPRS International Journal of Geo-Information*, vol. 6, n°5.
- Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., Hou, A. et Hong, Y. (2018) Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River. *Remote Sensing of Environment*, vol. 219, p. 115-134.
- Jeong, J. et Park, E. (2019) Comparative applications of data-driven models representing water table fluctuations. *Journal of Hydrology*, vol. 572, p. 261-273.
- Jeong, J., Park, E., Chen, H., Kim, K.-Y., Han, W. S. et Suk, H. (2020) Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. *Journal of Hydrology*, vol. 582, p. 124512.
- Junqueira, A. M., Mao, F., Mendes, T. S. G., Simões, S. J. C., Balestieri, J. A. P. et Hannah, D. M. (2021) Estimation of river flow using CubeSats remote sensing. *Science of the Total Environment*, vol. 788, p. 147762.

- Kapoor, A., Gulli, A., Pal, S. et Chollet, F. (2022) Deep Learning with TensorFlow and Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd.
- Kebede, M. G., Wang, L., Li, X. et Hu, Z. (2020) Remote sensing-based river discharge estimation for a small river flowing over the high mountain regions of the Tibetan Plateau. *International Journal of Remote Sensing*, vol. 41, n°9, p. 3322-3345.
- Khan, S., Rahmani, H., Shah, S. A. A. et Bennamoun, M. (2018) A guide to convolutional neural networks for computer vision. *Synthesis lectures on computer vision*, vol. 8, n°1, p. 1-207.
- King, T. V., Neilson, B. T. et Rasmussen, M. T. (2018) Estimating Discharge in Low-Order Rivers With High-Resolution Aerial Imagery. *Water Resources Research*, vol. 54, n°2, p. 863-878.
- Koblinsky, C. J., Clarke, R. T., Brenner, A. C. et Frey, H. (1993) Measurement of river level variations with satellite altimetry. *Wiley Online Library*.
- Kouraev, A. V., Zakharova, E. A., Samain, O., Mognard, N. M. et Cazenave, A. (2004) Ob'river discharge from TOPEX/Poseidon satellite altimetry (1992–2002). *Remote sensing of environment*, vol. 93, n°1-2, p. 238-245.
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S. et Nearing, G. S. (2019) Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, vol. 55, n°12, p. 11344-11354.
- Krizhevsky, A., Sutskever, I. et Hinton, G. E. (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, vol. 25, p. 1097-1105.
- Kumar, D., Singh, A., Samui, P. et Jha, R. K. (2019) Forecasting monthly precipitation using sequential modelling. *Hydrological sciences journal*, vol. 64, n°6, p. 690-700.

- Le Coz, J., Blanquart, B., Pobanz, K., Dramais, G., Pierrefeu, G., Hauet, A. et Despax, A. (2016) Estimating the uncertainty of streamgauging techniques using in situ collaborative interlaboratory experiments. *Journal of Hydraulic Engineering*, vol. 142, n°7, p. 04016011.
- Le Coz, J., Pierrefeu, G., SAYSSET, G., Brochot, J.-F. et Marchand, P. (2008) Mesures hydrologiques par profileur Doppler. Editions Quae.
- LeFavour, G. et Alsdorf, D. (2005) Water slope and discharge in the Amazon River estimated using the shuttle radar topography mission digital elevation model. *Geophysical Research Letters*, vol. 32, n°17.
- LeFavour, G., Letters, D. A.-G. R. et 2005, undefined (2005) Water slope and discharge in the Amazon River estimated using the shuttle radar topography mission digital elevation model. *Wiley Online Library*, vol. 32, n°17, p. 1-5.
- Léonard, J., Mietton, M., Najib, H. et Gourbesville, P. (2000) Rating curve modelling with Manning's equation to manage instability and improve extrapolation. *Hydrological sciences journal*, vol. 45, n°5, p. 739-750.
- Les gouvernements fédéral, provinciaux et territoriaux du C. (2010) Biodiversité canadienne : état et tendances des écosystèmes en 2010.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J. et Liu, H. (2017) Feature selection: A data perspective. *ACM computing surveys (CSUR)*, vol. 50, n°6, p. 1-45.
- Ling, F., Boyd, D., Ge, Y., Foody, G. M., Li, Xiaodong, Wang, L., Zhang, Y., Shi, L., Shang, C. et Li, Xinyan (2019) Measuring river wetted width from remotely sensed imagery at the subpixel scale with a deep convolutional neural network. *Water Resources Research*, vol. 55, n°7, p. 5631-5649.
- Liu, P., Wang, J., Sangaiah, A. K., Xie, Y. et Yin, X. (2019) Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability*, vol. 11, n°7, p. 2058.

- Lowe, J. J., Power, K. et Marsan, M. W. (1996) Inventaire des forêts du Canada 1991: sommaire par écorégion et par écozone terrestres (Vol. 364).
- Martin, Y. et Ham, D. (2005) Testing bedload transport formulae using morphologic transport estimates and field data: lower Fraser River, British Columbia. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, vol. 30, n°10, p. 1265-1282.
- McCulloch, W. S. et Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, vol. 5, n°4, p. 115-133.
- Mengen, D., Ottinger, M., Leinenkugel, P. et Ribbe, L. (2020) Modeling river discharge using automated river width measurements derived from sentinel-1 time series. *Remote Sensing*, vol. 12, n°19, p. 3236.
- Michailovsky, C. I., McEnnis, S., Berry, P. A. M., Smith, R. et Bauer-Gottwein, P. (2012) River monitoring from satellite radar altimetry in the Zambezi River basin. *Hydrology and Earth System Sciences*, vol. 16, n°7, p. 2181-2192.
- Mo, S., Zhu, Y., Zabarás, N., Shi, X. et Wu, J. (2019) Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resources Research*, vol. 55, n°1, p. 703-728.
- Moody, J. A. et Troutman, B. M. (2002) Characterization of the spatial variability of channel morphology. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, vol. 27, n°12, p. 1251-1266.
- Morlot, T. (2014) La gestion dynamique des relations hauteur-débit des stations d'hydrométrie et le calcul des incertitudes associées: un indicateur de gestion, de qualité et de suivi des points de mesure (PhD Thesis). Thèse de doctorat, Grenoble.
- Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J. et Gambardella, L. M. (2011) Max-pooling convolutional neural networks for vision-based hand

- gesture recognition. 2011 IEEE international conference on signal and image processing applications (ICSIPA), p. 342-347. IEEE.
- Nash, J. E. et Sutcliffe, J. V. (1970) River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, vol. 10, n°3, p. 282-290.
- Nathanson, M., Kean, J. W., Grabs, T. J., Seibert, J., Laudon, H. et Lyon, S. W. (2012) Modelling rating curves using remotely sensed LiDAR data. *Hydrological Processes*, vol. 26, n°9, p. 1427-1434.
- Negrel, J., Kosuth, P. et Bercher, N. (2011) Estimating river discharge from earth observation measurements of river surface hydraulic variables. *Hydrology and Earth System Sciences*, vol. 15, n°6, p. 2049-2058.
- Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M. et Saleska, S. (2011) Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *Journal of Hydrology*, vol. 404, n°1-2, p. 13-29.
- Nourani, V. (2017) An emotional ANN (EANN) approach to modeling rainfall-runoff process. *Journal of Hydrology*, vol. 544, p. 267-277.
- Pan, F. et Nichols, J. (2013) Remote sensing of river stage using the cross-sectional inundation area-river stage relationship (IARSR) constructed from digital elevation model data. *Hydrological Processes*, vol. 27, n°25, p. 3596-3606.
- Pan, F., Wang, C. et Xi, X. (2016) Constructing river stage-discharge rating curves using remotely sensed river cross-sectional inundation areas and river bathymetry. *Journal of Hydrology*, vol. 540, p. 670-687.
- Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P.-A., Collischonn, W., Bonnet, M.-P. et Seyler, F. (2016) Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon basin. *Water Resources Research*, vol. 52, n°5, p. 3787-3814.

- Pavelsky, T. M. (2014) Using width-based rating curves from spatially discontinuous satellite imagery to monitor river discharge. *Hydrological Processes*, vol. 28, n°6, p. 3035-3040.
- Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C., Allen, G. H. et Miller, Z. F. (2014) Assessing the potential global extent of SWOT river discharge observations. *Journal of Hydrology*, vol. 519, p. 1516-1525.
- Pelletier, P. M. (1988) Uncertainties in the single determination of river discharge: a literature review. *Canadian Journal of Civil Engineering*, vol. 15, n°5, p. 834-850.
- Peña-Arancibia, J. L., Zhang, Y., Pagendam, D. E., Viney, N. R., Lerat, J., van Dijk, A. I., Vaze, J. et Frost, A. J. (2015) Streamflow rating uncertainty: Characterisation and impacts on model calibration and performance. *Environmental Modelling & Software*, vol. 63, p. 32-44.
- Pristyanto, Y., Adi, S. et Sunyoto, A. (2019) The effect of feature selection on classification algorithms in credit approval. 2019 International Conference on Information and Communications Technology (ICOIACT), p. 451-456. IEEE.
- Qin, J., Liang, J., Chen, T., Lei, X. et Kang, A. (2019) Simulating and Predicting of Hydrological Time Series Based on TensorFlow Deep Learning. *Polish Journal of Environmental Studies*, vol. 28, n°2.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B. et Thuiller, W. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, vol. 40, n°8, p. 913-929.
- Rumelhart, D. E., Hinton, G. E. et Williams, R. J. (1985) Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science.
- Saba, L., Biswas, M., Kuppili, V., Godia, E. C., Suri, H. S., Edla, D. R., Omerzu, T., Laird, J. R., Khanna, N. N. et Mavrogeni, S. (2019) The present and future of deep learning in radiology. *European journal of radiology*, vol. 114, p. 14-24.



- Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., Summers, R. M. et Giger, M. L. (2019) Deep learning in medical imaging and radiation therapy. *Medical physics*, vol. 46, n°1, p. e1-e36.
- Sanyal, J. et Lu, X. X. (2004) Application of remote sensing in flood management with special reference to monsoon Asia: a review. *Natural Hazards*, vol. 33, n°2, p. 283-301.
- Schmidt, A. R. (2003) Analysis of stage-discharge relations for open-channel flows and their associated uncertainties.
- Schneider, R., Godiksen, P. N., Villadsen, H., Madsen, H. et Bauer-Gottwein, P. (2017) Application of CryoSat-2 altimetry data for river analysis and modelling. *Hydrology and Earth System Sciences*, vol. 21, n°2, p. 751-764.
- Schubert, J. E., Monsen, W. W. et Sanders, B. F. (2015) Metric-resolution 2D river modeling at the macroscale: Computational methods and applications in a Braided River. *Frontiers in Earth Science*, vol. 3, p. 74.
- Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., Ravi, V. et Peters, A. (2020) A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, vol. 194, p. 105596.
- Sewak, M., Karim, M. R. et Pujari, P. (2018) *Practical convolutional neural networks: implement advanced deep learning models using Python*. Packt Publishing Ltd.
- Shen, C. (2018) A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, vol. 54, n°11, p. 8558-8593.
- Sichangi, A. W., Wang, L. et Hu, Z. (2018) Estimation of river discharge solely from remote-sensing derived data: an initial study over the Yangtze river. *Remote Sensing*, vol. 10, n°9, p. 1385.

- Sichangi, A. W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., Zhou, J., Liu, W. et Kuria, D. (2016) Estimating continental river basin discharges using multiple remote sensing data sets. *Remote Sensing of Environment*, vol. 179, p. 36-53.
- Simard, P. Y., Steinkraus, D. et Platt, J. C. (2003) Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 2003-Janua, p. 958-963.
- Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y. et Demir, I. (2020) A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*.
- Smith, L. C. (1997) Satellite remote sensing of river inundation area, stage, and discharge: A review. *Hydrological Processes*, vol. 11, n°10, p. 1427-1439.
- Smith, L. C., Isacks, B. L., Bloom, A. L. et Murray, A. B. (1996) Estimation of discharge from three braided rivers using synthetic aperture radar satellite imagery: Potential application to ungaged basins. *Water Resources Research*, vol. 32, n°7, p. 2021-2034.
- Smith, L. C., Isacks, B. L., Forster, R. R., Bloom, A. L. et Preuss, I. (1995) Estimation of discharge from braided glacial rivers using ERS 1 synthetic aperture radar: First results. *Water Resources Research*, vol. 31, n°5, p. 1325-1329.
- Smith, L. C. et Pavelsky, T. M. (2008a) Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia. *Water Resources Research*, vol. 44, n°3, p. 1-11.
- Smith, L. C. et Pavelsky, T. M. (2008b) Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia. *Water Resources Research*, vol. 44, n°3.
- Smith, L. N. (2018) A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay, p. 1-21.

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. et Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, vol. 15, n°1, p. 1929-1958.
- Syed, T. H., Famiglietti, J. S., Chambers, D. P., Willis, J. K. et Hilburn, K. (2010) Satellite-based global-ocean mass balance estimates of interannual variability and emerging trends in continental freshwater discharge. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, n°42, p. 17916-17921.
- Tao, Y., Gao, X., Ihler, A., Sorooshian, S. et Hsu, K. (2017) Precipitation identification with bispectral satellite information using deep learning approaches. *Journal of Hydrometeorology*, vol. 18, n°5, p. 1271-1283.
- Tarpanelli, A., Amarnath, G., Brocca, L., Massari, C. et Moramarco, T. (2017) Discharge estimation and forecasting by MODIS and altimetry data in Niger-Benue River. *Remote sensing of environment*, vol. 195, p. 96-106.
- Tarpanelli, A., Barbetta, S., Brocca, L. et Moramarco, T. (2013) River discharge estimation by using altimetry data and simplified flood routing modeling. *Remote Sensing*, vol. 5, n°9, p. 4145-4162.
- Tate, E. C., Maidment, D. R., Olivera, F. et Anderson, D. J. (2002) Creating a terrain model for floodplain mapping. *Journal of Hydrologic Engineering*, vol. 7, n°2, p. 100-108.
- The Boreal Shield (s.d.)Canada's Biodiversity: Ecozones.
- Tomkins, K. M. (2014) Uncertainty in streamflow rating curves: methods, controls and consequences. *Hydrological processes*, vol. 28, n°3, p. 464-481.
- Tourian, M. J., Schwatke, C. et Sneeuw, N. (2017) River discharge estimation at daily resolution from satellite altimetry over an entire river basin. *Journal of Hydrology*, vol. 546, p. 230-247.

- Tourian, M., Schwatke, C., Hydrology, N. S.-J. of et 2017, undefined (s.d.) River discharge estimation at daily resolution from satellite altimetry over an entire river basin. Elsevier.
- Tramblay, Y., Ouarda, T. B. M. J., St-Hilaire, A. et Poulin, J. (2010a) Regional estimation of extreme suspended sediment concentrations using watershed characteristics. *Journal of Hydrology*, vol. 380, n°3-4, p. 305-317.
- Tramblay, Y., Ouarda, T. B. M. J., St-Hilaire, A. et Poulin, J. (2010b) Regional estimation of extreme suspended sediment concentrations using watershed characteristics. *Journal of Hydrology*, vol. 380, n°3-4, p. 305-317.
- Üneş, F., Demirci, M., Zelenakova, M., Çalışıcı, M., Taşar, B., Vranay, F. et Ziya Kaya, Y. (2020) River flow estimation using artificial intelligence and fuzzy techniques. *Water (Switzerland)*, vol. 12, n°9.
- Wang, S., Tang, J. et Liu, H. (2017) Feature Selection.
- Wu, J.-N. (2016) Compression of fully-connected layer in neural network by kronecker product. 2016 Eighth International Conference on Advanced Computational Intelligence (ICACI), p. 173-179. IEEE.
- Xiaodong, S., Ganlin, Z., Feng, L. I. U., Decheng, L. I., Yuguo, Z. et Jinling, Y. (2016) Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model.
- Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A. et Sayl, K. N. (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, vol. 530, p. 829-844.
- Young, D. S., Hart, J. K. et Martinez, K. (2015) Image analysis techniques to estimate river discharge using time-lapse cameras in remote locations. *Computers & Geosciences*, vol. 76, p. 1-10.

- Yu, Z., Yang, K., Luo, Y. et Shang, C. (2020) Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network. *Journal of Hydrology*, vol. 582, p. 124488.
- Zhang, D., Lindholm, G. et Ratnaweera, H. (2017) Use Long Short-Term Memory to Enhance Internet of Things for Combined Sewer Overflow Monitoring. *Journal of Hydrology*.
- Zhang, Liangpei, Zhang, Lefei et Du, B. (2016) Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, n°2, p. 22-40.
- Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K. et Fu, G. (2019) Deep learning identifies accurate burst locations in water distribution networks. *Water research*, vol. 166, p. 115058.
- Zhou, Z., Shi, L. et Zha, Y. (2020) Seeing macro-dispersivity from hydraulic conductivity field with convolutional neural network. *Advances in Water Resources*, vol. 138, p. 103545.





## 7 ANNEXE

---

**Table1: Detailed measurements of the hydrometric station and satellite images used in the period of analysis for river flow estimates.**

Station Name	Latitude Degree	Longitude Degree	Width at station (in m)	Catchment area (KM <sup>2</sup> )	No. of RS-1&2 Images
02JB009	47.843861	-77.5487	93.282	10300	1
02KJ004	46.346389	-77.8157	60.1223	3760	19
02LC021	46.046419	-74.2525	15.8505	311	22
02LD005	45.791283	-75.0911	18.7614	1330	19
02LE024	46.785028	-75.3116	82.2435	4530	37
02LE025	46.650306	-75.247	96.7865	883	37
02LG005	47.08325	-75.7535	127.824	6840	24
02ND003	47.676972	-73.0408	82.06	2640	10
02NE011	47.7685	-72.7349	76.2652	1570	9
02NF003	46.683578	-73.9136	70	1390	32
02OB017	46.030694	-73.7049	42.4822	1270	29
02OC021	46.441667	-73.4619	33.6825	186	22
02PC002	46.8925	-71.5261	62.3041	2010	19
02PC010	46.8675	-71.6372	20.46	213	21
02PD004	47.260028	-71.1372	22.4434	269	19
02RB004	49.881426	-70.9261	69.4586	1955	10
02RH027	47.941861	-71.3822	36.5656	495	26
02RH035	48.182694	-71.6448	84.8771	1110	14
02RH045	48.487944	-70.9722	37.8548	746	21
02RH066	48.235944	-71.2885	31.7567	355	21
02UC002	50.3525	-66.1867	117.467	19000	4
02VB004	50.685556	-64.5786	346.92	7230	4
02VC001	50.307778	-63.6186	125	13000	4
02WB003	50.4275	-61.7122	431.483	15600	7
02XA003	52.22981	-61.31694	122.143	4540	5
02XA008	50.680833	-59.6019	139.106	19200	7
02YC001	50.60747	-57.15161	21.29	624	3
02YD002	50.92442	-56.11169	34.1091	200	21
02YO011	48.84439	-56.26967	200	6300	9
02YQ001	49.01628	-54.85067	109.013	4450	20
02YS005	48.66275	-54.01525	77.3786	2000	41
02ZE004	48.16875	-55.48281	31.6417	99.5	8



03AB002	49.8575	-77.1872	85	31291	14
03BD002	50.745806	-76.3872	387.963	9684	3
03BF001	51.533583	-78.0966	194.305	6020	5
03OE001	53.24831	-60.78511	318.658	92500	5
03QC001	53.53428	-57.49386	287.417	10900	5
03QC002	52.64861	-56.87122	82.297	2310	3
04NA001	48.59775	-78.1102	94	3680	20

Table 2. Detail of the CNN model architecture.

Layers	Output shape
conv2d_input	(140,140,2)
conv2d	(138,138,128)
conv2d_1	(136,136,128)
average_pooling2d	(67,67,128)
conv2d_2	(65,65,128)
conv2d_3	(63,63,128)
average_pooling2d_1	(31,31,128)
conv2d_4	(29,29,256)
conv2d_5	(27,27,256)
Dropout	(27,27,256)
average_pooling2d_2	(13,13,256)
conv2d_6	(11,11,512)
global_average_pooling2d	(512)
dropout_1	(512)
Dense	(512)
dropout_2	(512)
Dense_1	(1)

Table 3. Hyperparameters used to train the CNN model.

Hyperparameter name	Hyperparameter value
Learning rate	1e-04
Optimizer	Adam
Loss function	MSE
Batch size	16
Epochs	50
Size of filter	3